6-24-2021

Human population history at the crossroads of East and Southeast Asia since 11,000 years ago

Tianyi Wang

Wei Wang

Guangmao Xie

Zhen Li

Xuechun Fan

See next page for additional authors

Recommended Citation

Tianyi Wang, Wei Wang, Guangmao Xie, Zhen Li, Xuechun Fan, Qingping Yang, Xichao Wu, Peng Cao, Yichen Liu, Ruowei Yang, Feng Liu, Qingyan Dai, Xiaotian Feng, Xiaohong Wu, Ling Qin, Fajun Li, Wanjing Ping, Lizhao Zhang, Ming Zhang, Yalin Liu, Xiaoshan Chen, Dongju Zhang, Zhenyu Zhou, Yun Wu, Hassan Shafiey, Xing Gao, Darren Curnoe, Xiaowei Mao, E. Andrew Bennett, Xueping Ji, Melinda A. Yang, Qiaomei Fu. "Human population history at the crossroads of East and Southeast Asia since 11,000 years ago," *Cell*, Volume 184, Issue 14, 2021, 3829-3841.e21, https://doi.org/10.1016/j.cell.2021.05.018.

Human population history at the crossroads of East and Southeast Asia since 11,000 years ago

Tianyi Wang[1,2,3]*, Wei Wang[4]*, Guangmao Xie[5,6]*, Zhen Li[4]*, Xuechun Fan[7,8]*, Qingping Yang[5], Xichao Wu[9], Peng Cao[1], Yichen Liu[1,3], Ruowei Yang[1], Feng Liu[1], Qingyan Dai[1], Xiaotian Feng[1], Xiaohong Wu[9], Ling Qin[10], Fajun Li[11], Wanjing Ping[1], Lizhao Zhang[1], Ming Zhang[1], Yalin Liu[1,12], Xiaoshan Chen[13], Dongju Zhang[13], Zhenyu Zhou[14], Yun Wu[15], Hassan Shafiey[1], Xing Gao[1,12], Darren Curnoe[16], Xiaowei Mao[1,3], E. Andrew Bennett[1], Xueping Ji[15,17]§, Melinda A. Yang[1,18]§, Qiaomei Fu[1,3,12,19]§

[1]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, China;
[2]Northwest University, Xi'an 710069, China;
[3]Shanghai Qi Zhi Institute, Shanghai, 200232, China
[4]Institute of Cultural Heritage, Shandong University, Qingdao 266237, China;
[5]Guangxi Institute of Cultural Relic Protection and Archaeology, Nanning 530022, China;
[6]College of History, Culture and Tourism, Guangxi Normal University, Guilin, 541001, China
[7]International Research Center for Austronesian Archaeology, Pingtan 350000, China;
[8]Fujian Museum, Fuzhou 350001, China;
[9]Fujian Longyan Museum, Longyan 364000, China;
[10]School of Archaeology and Museology, Peking University, Beijing 100871, China;
[11]Department of Anthropology, School of Sociology and Anthropology, Sun Yat-Sen University, Guangzhou 510275, China;
[12]University of the Chinese Academy of Sciences, Beijing 100049, China;
[13]Key Laboratory of Western China's Environmental Systems (Ministry of Education), College of Earth and Environmental Sciences, Lanzhou University, Lanzhou 730000, China;
[14] Institute of Archaeology, Chinese Academy of Social Sciences, Beijing 100710, China;
[15]Yunnan Institute of Cultural Relics and Archaeology, Kunming 650118, China;
[16]Australian Museum Research Institute, Australian Museum, 1 William Street, Sydney, NSW, 2010, Australia;
[17]Yunnan Key Laboratory of Earth System Science, Yunnan University, Kunming 650500, China;
[18]Department of Biology, University of Richmond, Richmond VA 23173, USA;
[19]Lead Contact
*These authors contributed equally to this work.
§Corresponding authors
Correspondence to: fuqiaomei@ivpp.ac.cn; myang@richmond.edu; jxpchina@foxmail.com

SUMMARY

Past human genetic diversity and migration between southern China and Southeast Asia has not been well-characterized, in part due to poor preservation of ancient DNA in hot and humid regions. We investigated 31 newly sequenced ancient genomes from southern China (Guangxi and Fujian), including two ~12,000-10,000-year-old individuals representing the oldest humans sequenced from southern China. We discovered a novel and deeply diverged East Asian ancestry in the Guangxi region that persisted until at least 6,000 years ago. We found ~9,000-6,000-year-old Guangxi populations were a mixture of local ancestry, southern ancestry previously sampled in Fujian, and deep Asian ancestry related to Southeast Asian Hòabìnhian hunter-gatherers, showing broad admixture in the region predating the appearance of farming. Historical Guangxi populations dating to ~1,500 to 500 years ago are closely related to Tai-Kadai and Hmong-Mien speakers. Our results show heavy interactions between three distinct ancestries at the crossroads of East and Southeast Asia.

INTRODUCTION

Modern humans have a long history of occupation in East and Southeast Asia. Recent studies sampling ancient human DNA have revealed distinct demographic patterns in Southeast Asia and southern China (Lipson et al., 2018; McColl et al., 2018; Ning et al., 2020; Yang et al., 2020). In Southeast Asia, ~8,000-4,000-year-old Southeast Asian Hòabìnhian hunter-gatherers possessed deeply diverged Asian ancestry (denoted **Hòabìnhian ancestry** since it was first detected in Hòabìnhian-related samples, see Box 1) (Lipson et al., 2018; McColl et al., 2018), whereas the first Southeast Asian farmers beginning ~4,000 years ago show a mixture of ancestry associated with present-day southern Chinese populations and deeply diverged Hòabìnhian ancestry. In southern China, ~9,000-4,000-year-old individuals from Fujian province show ancestry distinct from that found in northern China, but not as deeply diverged as Hòabìnhian ancestry. This ancestry (denoted **Fujian ancestry** since it was first detected in Fujian, see Box 1) is found in partial amounts in present-day southern Chinese populations, but is closely associated with ancestry found in today's Austronesians, a seafaring population that migrated away from mainland Asia several thousand years ago (Yang et al., 2020). These findings show that using ancient DNA techniques to examine ancestral populations and early population dynamics (especially before the transition to farming) is key for a better understanding of past population history.

Anthropological and archeological evidence also highlight demographic complexity in East and Southeast Asia. Surveys of material culture indicate that the culture associated with Hòabìnhian ancestry may have been found in southern China (Hung et al., 2017; Institute of Archaeology Chinese Academy of Social Sciences, 2003; Ji et al., 2016). Comparisons of skeletal morphology from prehistoric humans along the border of southern China and Southeast Asia show patterns suggestive of deep ancestry unlike that observed in present-day East and Southeast Asians (Matsumura et al., 2019). One comparative archaeological study (Zhang and Hung, 2008) suggested that in southern China there were two different cultural traditions: one predominantly in coastal southern China and nearby islands, and another in the region bordering Vietnam – mirroring the two distinct genetic patterns observed in ancient individuals from Southeast Asia and southern China (Lipson et al., 2018; McColl et al., 2018; Ning et al., 2020; Yang et al., 2020).

Despite more clarity on East and Southeast Asian history, regions like Guangxi, a province in southern China bordering Vietnam, show that population history across southern China and Southeast Asia is still not well-established. In Guangxi, an individual from a >10,000-year-old cave site (Longlin) was found to possess cranial morphology with a mix of archaic and modern features (Curnoe et al., 2012), which suggested a possible ancestry similar to or deeper than Hòabìnhian ancestry – a pattern not observed in ancient East and Southeast Asians to date. Though Hòabìnhian-related material culture can be found in other parts of southern China (Hung et al., 2017; Institute of Archaeology Chinese Academy of Social Sciences, 2003; Ji et al., 2016), Hòabìnhian ancestry has yet to be found in any ancient human outside Southeast Asia. Populations today in Guangxi are Tai-Kadai and Hmong-Mien speakers (Wang et al., 2021), who possess a mix of Fujian and northern Chinese ancestry (Yang et al., 2020). Despite Guangxi's central location bridging southern China and Southeast Asia, ancient DNA (aDNA) techniques have not been applied to ancient humans in this region, largely due to the difficulties presented by low preservation of aDNA in hot and humid regions. Despite this sampling challenge, we surveyed ancient humans in the Guangxi region over the last 11,000 years to investigate (1) what role deeply divergent ancestries played in the region, particularly with regard

to the Longlin specimen; (2) whether Hòabìnhian and Fujian ancestries extended to this region and if so, how they interacted with each other; and (3) how past humans in this region contributed to present-day populations.

RESULTS

To address these questions, we screened 170 specimens from 30 sites in Guangxi (Table S1). Despite the difficulty of retrieving ancient DNA in southern regions, we successfully obtained genomic material from 30 individuals from Guangxi with radiocarbon dates ranging from 10,686 to 294 calibrated years before present (cal BP, BP is before 1950 AD, Table 1, Table S1, Figure 1A-1B), including from a specimen excavated at Longlin Cave who possessed both archaic and modern cranial features (Curnoe et al., 2015; Curnoe et al., 2012). We also obtained genomic data from an additional individual (Qihe3; 11,747-11,356 cal BP) from the Qihe cave site in Fujian, China (Wu et al., 2014; Yang et al., 2020). Longlin and Qihe3 date to ~12,000-10,000 years ago, allowing an unprecedented look into the diversity of East Asia at the Pleistocene-Holocene transition. Collectively, we find that the aDNA sampled from the Guangxi region reveals a genetic history unlike that observed in other regions, including Southeast Asia and Fujian in southern China.

We used large-scale nuclear aDNA capture techniques (Haak et al., 2015) to enrich for endogenous DNA at 1.2 million single nucleotide polymorphisms (SNPs) (Fu et al., 2015). To ensure SNPs were correctly called for each individual, we first identified characteristic aDNA damage signatures suggesting the presence of endogenous DNA (Briggs et al., 2007). Then, we estimated modern human contamination rates for each sample, using all fragments from samples with ≤3% contamination. Three samples showed modern human mtDNA contamination levels above 3% (Table 1). One of these, identified as male, showed negligible levels of contamination for nuclear DNA (2.9%), so we used all fragments for subsequent analyses. The other two, identified as female, could not be assessed for nuclear DNA contamination, so we restricted our downstream analyses to only DNA fragments that possessed a characteristic aDNA damage signature (Fu et al., 2013a; Korneliussen et al., 2014). In total, we obtained genetic information from 31 individuals sequenced to between 0.01 to 4.06 x fold coverage for the 1.2 million targeted SNPs (Table 1).

We first performed a kinship analysis to test whether any samples were related to each other. Of the 30 Guangxi individuals sequenced, seven sets of close familial relationships (1st and 2nd degree) were found. For each of these sets we retained the individual with the highest SNP count (Table S1), resulting in 23 unrelated individuals for population genetic analysis. For these 23 unrelated Guangxi individuals, we used the results from principal component (Patterson et al., 2006) (PCA, Figure 1C), outgroup-f_3 (Patterson et al., 2012) (Figure S1A), f_4-statistic (Patterson et al., 2012), and ADMIXTURE (Alexander et al., 2009) analyses to separate them into nine groups (Figure 1A-1B, see STAR Methods). Qihe3 from Fujian clusters genetically with a previously published individual from the same site (Figure S1D, see STAR Methods).

Novel East Asian ancestry found 11,000 years ago in Guangxi

The oldest individual sampled in this study, Longlin (10,686-10,439 cal BP, Laomaocao Cave, Guangxi, China), possesses a cranial morphology with a mix of archaic and early modern human features. Longlin's genetic profile, however, falls well within the genetic diversity found in modern human populations from Asia, and with similar levels of archaic ancestry as that observed in East Asians (Table S2).

Comparisons to 9,000-4,000-year-old individuals sampled in China (Yang et al., 2020) show that Longlin is not closely related to presently sampled East Asians. In an outgroup-f_3 analysis, Longlin shares little genetic similarity with ancient humans who are closely related to present-day East Asians (Figure S1D), namely, 9,500-7,500-year-old Shandong populations in northern China (**Shandong ancestry**, Box 1) and 9,000-7,500-year-old Fujian populations from southern China (**Fujian ancestry**) (Yang et al., 2020). The Shandong (*EN_SD*) and Fujian (*EN_FJ*) populations in fact share a closer relationship to each other than to Longlin, i.e. f_4*(Mbuti, EN_SD/EN_FJ; Longlin, EN_SD/EN_FJ)>0* ($3.2<Z<14.4$, Table S2), and neither population shares excess affinity with Longlin, i.e. f_4*(Mbuti, Longlin; EN_SD, EN_FJ)~0* ($-2.1<Z<2.3$, Table S2). This suggests that the lineage to which Longlin belongs branched prior to the separation of Shandong and Fujian ancestries in the north and south, respectively. After modeling the phylogenetic relationship between Longlin and Neolithic East Asians with both Admixture Graph and Treemix analyses (Figure 2A-2B, see STAR

Methods), we find further support for scenarios whereby Longlin is an outgroup to northern and southern East Asian ancestries represented by the Shandong and Fujian populations.

To explore how deeply diverged Longlin is from East Asians, we compared Longlin and Neolithic East Asians to other individuals with deeply diverged Asian ancestries ('Deep Asians', see Box 1), including the ~40,000-year-old Tianyuan (Fu et al., 2013a; Yang et al., 2017), present-day Papuans from Papua New Guinea (Mallick et al., 2016), Onge from the Andamanese Islands (Mallick et al., 2016), and an ~8,000-year-old Hòabìnhian from Southeast Asia (McColl et al., 2018). We find that Longlin is more closely related to the Shandong and Fujian populations than to any of the deeply diverging Asian ancestries, i.e. f_4*(Mbuti, Longlin; EN_SD/EN_FJ, Deep Asians)<0* (-12.8<Z<-2.4, Table S2) and f_4*(Mbuti, EN_SD/EN_FJ; Longlin, Deep Asians)<0* (-13.1<Z<-2.7, Table S2). Our genetic analyses show Longlin to be an offshoot of the East Asian branch of modern humans, with no close relationship to more deeply diverged Asian ancestries.

An ~2,700-year-old individual from Japan associated with the Jōmon culture, Ikawazu, shows a similar pattern to East Asians and deeply diverged Asians as observed for Longlin (McColl et al., 2018). Ikawazu and Longlin share a closer relationship to each other than either share with deeply diverged Asians, i.e. f_4*(Mbuti, Longlin/Ikawazu; Ikawazu/Longlin, Deep Asians)<0* (-7.3<Z<-3.4, Table S2). To assess who is more closely related to East Asians, we compared Ikawazu and Longlin to Shandong and Fujian populations (see STAR Methods). In an f_4-analysis, we find that Shandong and Fujian populations are similarly related to Longlin and Ikawazu, i.e. f_4*(Mbuti, EN_SD/EN_FJ; Longlin, Ikawazu)~0* (0.5<Z<2.2, Table S2), and both have connections to Shandong and Fujian populations not found in the other individual, i.e. f_4*(Mbuti, Longlin; Ikawazu, EN_SD/EN_FJ)>0* (2.4<Z<5.2, Table S2) and f_4*(Mbuti, Ikawazu; Longlin, EN_SD/EN_FJ)>0* (3.1<Z<6.4, Table S2). These patterns suggest that Longlin, Ikawazu, and Neolithic East Asians likely separated from each other at about the same time.

We thus find that Longlin's ancestry (hereafter referred to as **Guangxi ancestry**) is unlike both Fujian and Hòabìnhian ancestries, the two previous ancestries observed in the region encompassed by southern China and Southeast Asia. Similar to the Jōmon ancestry found in Ikawazu in Japan, Guangxi ancestry is more closely related to East Asian ancestry (e.g. Fujian and Shandong) than deeply diverging Asian ancestry (e.g. Hòabìnhian). However, unlike Ikawazu, Longlin was not geographically isolated from other mainland East Asians. These patterns indicate that the genetic diversity in Asia 11,000 years ago was higher than in more recent periods of human history.

Admixture in southern China by 9,000-6,400 years ago

We observed Guangxi ancestry in an ~11,000-year-old human, so we next examined whether younger populations from the region also carried Guangxi ancestry. We recovered genome-wide data from two individuals (Dushan, Baojianshan) from Guangxi dating to ~9,000 - 6,400 BP. If Dushan, a male individual directly dating to 8,974 - 8,593 cal BP, is a descendent of a population more closely related to northern and southern East Asians than Longlin, we would expect that f_4*(Mbuti, Longlin; Dushan, East Asians)~0*. Instead, we observed that relative to some East Asians from Siberia (Sikora et al., 2019) and Fujian (Yang et al., 2020), f_4*(Mbuti, Longlin; Dushan, DevilsCave_N/Qihe)<0* (Z=-3.7/-3.3, Figure 2C, Table S2), indicating the presence of a genetic connection between Longlin and Dushan. When we compared Longlin to ancient East and Southeast Asians in an outgroup f_3-analysis, i.e. f_3*(Mbuti; Longlin, X),* the highest value observed was for Dushan (Figure S1B, Figure S1D), demonstrating that Longlin shares the most genetic similarity with Dushan. These patterns suggest that Guangxi ancestry is present in Dushan.

However, rather than Dushan possessing solely Guangxi ancestry, outgroup f_3-analysis with Dushan shows high genetic similarity to Fujian populations and Southeast Asian farmers, a pattern not observed for Longlin (Figure S1D). Phylogenetic analyses allowing migration events Figure 2A-2B) consistently model Dushan as a mixture of two sources - one related to Longlin (17%) and one related to a Fujian population (Qihe, 83%). f_4-analysis supports that Dushan shares a connection with a population of Fujian ancestry relative to Siberian-related northern East Asians and Shandong populations (Box 1), i.e. f_4*(Mbuti, Dushan; EN_FJ, Siberian-related northern East Asians)*<0 (-5.8<Z<-2.1) and f_4*(Mbuti, Dushan; Liangdao1, EN_SD)*<0 (-3.1<Z<-1.7, Table S2). The genetic patterns observed for Dushan suggest that by around 9,000 BP, gene flow between populations carrying Guangxi ancestry and Fujian ancestry was occurring, resulting in admixed populations possessing a mixed Guangxi-Fujian ancestry.

Given the increased allele sharing between Dushan and populations carrying Fujian ancestry (Figure 2B), we next examined whether the admixed Guangxi-Fujian ancestry impacted the Fujian region. We found that Dushan shows more affinity to Late Fujian populations (grouped set of Xitoucun, Tanshishan) relative to Early Fujian populations (grouped set of Qihe, Qihe3, Liangdao1/2), i.e. we observed that f_4*(Mbuti, Dushan, LN_FJ, EN_FJ)<0* (Z = -5.1). This pattern persisted for transversions only (Z=-3.2). In an expanded analysis keeping individuals from different archaeological sites separate, we assessed affinity to Dushan relative to the 12,000-year-old Qihe3 individual, i.e. f_4*(Mbuti, Dushan; X, Qihe3)* (see Figure 2D, Figure S2, see STAR Methods). The affinity to Dushan persisted not only for the Late Fujian populations dating to 4,100-2,000 years ago (Xitoucun and Tanshishan, Z=-4.5/-3.7), but also 1,900-1,100-year-old Taiwan islanders (Taiwan_Hanben, Z=-3.9, Figure 2D, Figure S2A-S2B), a pattern that remained consistent after we applied a post-hoc Benjamini-Hochberg correction (-3.8<Z<-3.3, see STAR Methods). We also observed the same Dushan affinity in 4,100-2,000-year-old Southeast Asian populations from Vietnam (Man_Bac and Nui_Nap, Z=-3.1), and 1,500-year-old populations from Guangxi (BaBanQinCen, Z=-4.2, Figure 2D, Figure S2A-S2B). With the Benjamini-Hochberg correction, the ancient Southeast Asian populations no longer showed a significant affinity, but BaBanQinCen did. Lastly, we found that for transversions only, the Dushan affinity only persisted for Xitoucun. Overall, these patterns indicate that ancestry related to Dushan, perhaps an admixed Guangxi-Fujian ancestry, played a prominent role in southern China's prehistory.

An admixed Guangxi-Fujian ancestry seems to persist for a couple thousand years in Guangxi, based on genetic patterns found for a female individual from Baojianshan who was found in an archaeological layer dated to between 8,300 – 6,400 years ago (see STAR Methods). Like Dushan, Baojianshan shares the highest genetic similarity with Fujian populations and Southeast Asian farmers (Figure S1D). Baojianshan also shares more alleles with Dushan relative to both Shandong and Fujian populations, i.e. f_4*(Mbuti, Baojianshan; Dushan, northern East Asians)<0* (-6.7<Z<-2.8) and f_4*(Mbuti, Baojianshan; Dushan, Qihe/Qihe3)<0* (-3.2<Z<-2.4, Table S2).

While Baojianshan shares ancestry with Dushan, unlike Dushan and other prehistoric Guangxi individuals, Baojianshan also shares alleles with the deeply diverged Hòabìnhian hunter-gatherers of Southeast Asia (McColl et al., 2018). In an f_4-analysis, the Hòabìnhian hunter-gatherers show a connection to Baojianshan relative to northern East Asians that is not observed for Longlin and Dushan (Figure 2E), i.e. f_4*(Mbuti, Hòabìnhian; Baojianshan, DevilsCave_N)<0* (Z=-3.2), while f_4*(Mbuti, Hòabìnhian; Longlin/Dushan, DevilsCave_N)~0* (-1.7<Z<0.4, Table S2). When we estimated admixture proportions with qpAdm (see STAR Methods), we observed that Baojianshan can be modeled as a mixture of 72.3% Dushan-related ancestry and 27.7% Hòabìnhian-related ancestry (Table S3), with similar proportions estimated using qpGraph analysis (Figure 2B). In a Treemix analysis allowing migration events (see STAR Methods), Baojianshan clusters with Dushan, sharing a migration event from the Longlin branch, and additionally experiences migration from Hòabìnhians (Figure 2A). Thus, Fujian and Hòabìnhian ancestry are both found in the Guangxi region by 8,300-6,400 years ago, and collectively, all three southern ancestries can be found in admixed form in the Guangxi region through Baojianshan.

From ~9,000-6,400 years ago, admixture played a prominent role in prehistoric populations along the border of southern China and Southeast Asia. Dushan belonged to a population that possessed a mixture of Guangxi and Fujian ancestry, while Baojianshan is similar to Dushan, but additionally shares Hòabìnhian ancestry. These patterns support that Hòabìnhian ancestry extended into southern China, as has been suggested from study of material culture at some southern Chinese archaeological sites (Ji et al., 2016). However, these patterns highlight that neither Hòabìnhian nor Fujian ancestry is sufficient to describe the populations that existed along the border of southern China and Southeast Asia. Guangxi ancestry persisted in partial amounts until at least 6,400 years ago, and ancestry associated with Dushan likely influenced prehistoric populations outside of the Guangxi region as well. Our findings show that the prehistoric period from 9,000-6,400 years ago is replete with admixed populations containing different levels of each of the southern ancestries. The timing and archaeological associations of these admixed populations suggest that admixture profoundly influenced the human landscape in southern China and Southeast Asia well before the advance of farming cultures such as those that were sampled in Southeast Asia ~4,000 years ago (McColl et al., 2018). The pattern in Guangxi contrasts greatly with the pattern observed in Fujian (Yang et al., 2020) around the same time period, where Fujian ancestry persisted for several millennia.

Changes in historical populations of Guangxi

With sampling in Guangxi from 1,500 – 500 years ago, we lastly assessed what role, if any, the three southern ancestries played in the historical period. We found that historical Guangxi populations do not cluster with prehistoric populations in a PCA (Figure 1C). Instead, the majority of historical individuals dating to ~1,500 years ago share a similar genetic profile, forming a tight cluster and overlapping with Tai-Kadai speakers (Figure 1C, Figure S1D). However, the ~500-year-old GaoHuaHua population is distinct from the ~1,500-year-old cluster, falling near Hmong-Mien speakers both in PCA (Figure 1C) and in outgroup-f_3 analyses (Figure S3A). To directly compare their relationships with present-day populations, we calculated f_4*(Mbuti, present-day East Asians; 1500BP Guangxi, 500BP Guangxi)* and showed that Hmong-Mien speakers always show a significant affinity to the ~500-year-old GaoHuaHua population (Figure 3A). All historical Guangxi populations were sampled from Cave Burial sites (see STAR Methods). Based on the inscription and coffin typology, cave burials in Guangxi were believed to belong to ancestors of the Zhuang (Tai-Kadai speakers) (Guangxi Museum and Tiandong County Museum, 1991). However, cave burials where the ~500-year-old GaoHuaHua are sampled have been hypothesized to be connected to Miao-Yao populations (Hmong-Mien speakers) (Zhou, 1991). Our genetic analyses suggest that populations in Guangxi at these two periods are indeed genetically very distinct and belong to different populations, as suggested previously (Peng, 2001). Thus, the genetic structure of present-day Guangxi populations belonging to Tai-Kadai and Hmong-Mien groups was present by at least 500 years ago.

We further explored the genetic structure of historical Guangxi populations using qpAdm to model mixture proportions from different source ancestries. We found historical Guangxi populations can be modeled as a mixture of 58.2%-90.6% Dushan-related (or Qihe3-related) ancestry, with 9.4%-41.8% northern East Asian-related ancestry (see STAR Methods). For all populations but BaBanQinCen, we do not observe any significant signal of deep ancestry associated with Dushan (Figure 2D), which suggests that the southern ancestry found in these historical Guangxi populations is closely related to Fujian ancestry.

Similar to present-day southern East Asians (Yang et al., 2020), historical Guangxi populations (~1,500 years ago) also show admixture from northern East Asians. We further compared previously published ancient populations from different areas in northern East Asia to test which ancestries had the strongest influence on historical Guangxi populations (Ning et al., 2020; Sikora et al., 2019; Wang et al., 2021; Yang et al., 2020). In an outgroup f_3-analysis, historical Guangxi populations show the closest genetic affinity with ancient populations found near the Lower Yellow River, e.g. Shandong populations dating to 9,500-7,900 BP (Yang et al., 2020) (Bianbian, Xiaojingshan) and Central Plains populations dating to 4,225-2,000 BP (Ning et al., 2020) (YR_LN, YR_LBIA, Figure 3B). The genetic affinity between Guangxi and the ~7,900 BP Xiaojingshan persisted from the earliest historical Guangxi group BaBanQinCen (max. date 1,688 BP), i.e. f_4*(Mbuti, BaBanQinCen; large panel of northern East Asians, Xiaojingshan)>0* ($2.1<Z<10.6$, Table S2), to the youngest Guangxi population sampled, the GaoHuaHua (max. date 513 BP, $2.5<Z<9.2$, Table S2). Thus, the northern influence found in historical Guangxi populations from 1,500-500 years ago was most closely related to Shandong ancestry dating to 9,500-7,900 years ago.

DISCUSSION

Our analysis of individuals spanning ~11,000-6,000 years ago from the Guangxi region of southern China reveals a previously unsampled genetic lineage that is deeply diverged from East Asians. This lineage, best represented by the ~11,000-year-old Longlin individual, acts as an outgroup to the northern and southern East Asian ancestries present in Shandong and Fujian (Yang et al., 2020), revealing that deep branching in East Asian lineages is found not only in isolated regions such as the Japanese archipelago (Kanzawa-Kiriyama et al., 2019; McColl et al., 2018) but also in mainland East Asia. Another ~12,000-year-old individual was sampled from the Fujian region along China's southern coast who, unlike Longlin, shows Fujian ancestry (Yang et al., 2020). Together, these two individuals show that ~12,000-10,000 years ago, southern China was characterized by at least two highly diverse human populations. However, while Fujian-related ancestry (represented by Qihe3) existed in the Fujian region from ~12,000 – 4,000 years ago, this pattern did not extend to the Guangxi region.

More recent sampling shows that population continuity was not a feature of the Guangxi region, and gene flow played a formative role ~9,000-6,400 years ago. The ~9,000-year-old Dushan is best characterized as

a mixture of Fujian and Guangxi ancestry, and ancestry related to Dushan appears later in a Fujian population (Xitoucun) dating to ~4,000 years ago (Yang et al., 2020). In contrast, Baojianshan, who dates between 8,300-6,400 years ago, is a mixture of those two ancestries and additionally Hòabìnhian ancestry, a deeply diverged Asian ancestry that was widespread in Southeast Asia prior to 4,000 years ago (McColl et al., 2018). The presence of Hòabìnhian ancestry in Baojianshan suggests that the range for Hòabìnhian ancestry extended from Southeast Asia into southern China. However, its presence in a population composed of a mixture of Fujian, Guangxi, and Hòabìnhian ancestry shows that the Guangxi region on the border of Southeast Asia and southern China cannot be simply characterized by ancestry related to a single population. Mixture between these three diverse ancestries in southern China and Southeast Asia from 9,000-6,400 years ago shows that admixture had a marked influence on prehistoric populations prior to the introduction of farming in Guangxi and Southeast Asia.

Previous studies have suggested that the cranial morphology of prehistoric populations in Japan and Guangxi share similarities with Australo-Papuans, similar to Hòabìnhians from Southeast Asia (Hung, 2019; Hung et al., 2017; Matsumura et al., 2019). A model (see Box 1) has been proposed whereby two layers of ancestry are present in East and Southeast Asia, a first layer represented by an early ancestry associated with prehistoric populations closely related to Australo-Papuans and a second layer that originated from northern East Asia from populations which gradually replaced the first layer with the expansion of farming (Matsumura et al., 2019). However, similar cranial features across specimens from southern China, Southeast Asia, and Japan that have been grouped as a first layer do not show similar groupings genetically in this study and in others (Kanzawa-Kiriyama et al., 2019; McColl et al., 2018; Yang et al., 2020). This suggests that the studied cranial features may not be capturing the diversity across these pre-farming populations accurately. Lineages of deep Asian ancestry, e.g. Hòabìnhian ancestry (McColl et al., 2018), existed, but humans sampled from the last 11,000 years across East Asia, including Guangxi, Fujian, and the Japanese archipelago, share more common ancestry with each other, revealing many offshoots of an East Asian lineage.

In historical Guangxi populations dating from ~1,500-500 years ago, Shandong ancestry related to northern East Asians along the Yellow River is prominent, a pattern observed across southern China and Southeast Asia (McColl et al., 2018; Yang et al., 2020) . We do not observe northern ancestry in Guangxi individuals dating from 11,000-6,400 years ago, which suggests that movement of populations carrying Shandong ancestry occurred sometime between 6,400-1,500 years ago. Historical Guangxi populations, unlike Austronesians, show heavy influence from populations carrying northern East Asian ancestry, similar to present-day East Asian populations. The absence of detectable Guangxi ancestry suggests that this early East Asian lineage had vanished from southern China by this time, with no substantial contribution to the genetic diversity found in this region today. Sampling of historical Guangxi populations resolves some debate related to the recent population history of the Guangxi region (Guangxi Museum and Tiandong County Museum, 1991; Peng, 2013b; Zhang et al., 1986; Zhou, 1991). Two major language groups are found in Guangxi today – one associated with Tai-Kadai speakers and the other with Hmong-Mien speakers. The historical Guangxi populations in our current data show that ancestry related to Tai-Kadai speakers can be found by at least ~1,500 years ago, while ancestry related to Hmong-Mien speakers is found in individuals dating to ~500 years ago. Thus, these two populations have lived continuously in Guangxi for at least 500 years.

By 11,000 years ago, the Guangxi region shows a deeply diverged ancestry of no relation to Hòabìnhian or Fujian ancestry, that gave way to highly admixed populations by 9,000 – 6,400 years ago. Unlike in the Fujian region, the existence of highly admixed populations in Guangxi suggests that this region was an interaction zone between indigenous populations from Guangxi, populations from the Fujian region, and populations related to Hòabìnhians of Southeast Asia. Unlike in Southeast Asia, we find that gene flow well before farming played an important role in forming the pre-agricultural populations in these regions. These prehistoric individuals do not share a close relationship to present-day populations of Guangxi, but we have found ancestry associated with present-day Tai-Kadai and Hmong-Mien speakers in the historical period since 1,500 years ago. Sampling in regions near the Yangtze River and southwest China may clarify what genetic shifts occurred between 6,000 and 1,500 years ago that gave rise to the genetic composition we see today in southern China, and further clarify the remarkably diverse genetic prehistory of humans across southeastern Asia.

Box 1: Definition of terms used in this study.

Geographic definition

Southern China: The geographic region representing the southern regions of China. Here we primarily examine regions represented by two provinces of China (Guangxi and Fujian). Guangxi, along the border of southern China and Southeast Asia, is where we sampled more ancient humans in this study, while the Fujian region was primarily sampled previously (Yang et al., 2020).

Southeast Asia: The geographic region consists of mainland Southeast Asia and Maritime Southeast Asia, where many ancient humans were sampled across multiple countries (Lipson et al., 2018; McColl et al., 2018). We focus on ancient humans from mainland Southeast Asia, particularly in Laos and Vietnam.

Archaeological background

Two-layer hypothesis: A model proposed based on cranial morphometrics and dental characteristics, which is widely used to explain human migration and interaction across Southeast and East Asia (Matsumura et al., 2019). This model proposes that Asia was occupied by a first wave of humans (first-layer) who were hunter-gatherers associated with flexed burials and shell midden sites and may have contributed to Australo-Papuans today. Those assigned to the Hòabìnhian culture show cranial features associated with Australo-Papuans (Matsumura, 2006; Matsumura et al., 2017; Matsumura et al., 2011). This first-layer was largely replaced by populations with cranial morphology associated with East Asians today (second-layer). Second-layer populations show an association with agriculture, extended position burials, and materials related to Neolithic culture. In this hypothesis, second-layer populations originated in the earliest agricultural regions along the Yellow River, expanding southwards to replace first-layer populations (Matsumura and Oxenham, 2014). Genetic sampling shows Hòabìnhian ancestry diverged deeply along the Asian lineage (McColl et al., 2018), which supports that they may have belonged to the first-layer population. Hòabìnhian ancestry in Southeast Asia became diminished with the rise of southern Chinese ancestry in farming-related populations, further lending strength to the two-layer hypothesis. However, genetic sampling in Japan and southern China of populations associated craniometrically with the first-layer show that they are more closely related genetically to second-layer East Asian populations, indicating that the two-layer model is not sufficient to describe the population movement, replacement, and mixture in prehistoric Asia.

Hòabìnhian industry: This culture was defined from material recovered from the caves in Hòa Bình Province and neighboring provinces in northern Vietnam (Colani, 1927). Later, it was re-described as an industry represented by different stone artifact assemblages containing flaked and cobble artifacts across Southeast Asia, existing from the Late Pleistocene to the Holocene c.50,000 to 5,000 BP (Solheim, 1970). In China, a Hòabìnhian lithic assemblage was reported from Xiaodong rockshelter in Yunnan Province (Ji et al., 2016), but none of the archaeological sites from southern China described in this study show evidence of Hòabìnhian culture.

Genetic populations

Deep Asians: Those distantly related to present-day East Asians but genetically more closely related to Asians than non-Asians. Tianyuan (Fu et al., 2013a; Yang et al., 2017), the ~40,000-year-old Early Asian from Beijing, China represents one branch of deep ancestry. Present-day Papuan and Onge (Mallick et al., 2016), and the ~7,950-7,795 years ago Southeast Asia Hòabìnhian (McColl et al., 2018) hunter-gatherers, represent a separate branch of Deep Asian ancestry.

Guangxi ancestry: First defined in this study, this ancestry refers to ancestry found in the ~11,000-year-old Longlin, the oldest human sampled from Guangxi province. This ancestry persists in admixed form from 9,000 – 6,000 years ago, and it is not observed in present-day populations.

Hòabìnhian ancestry:This ancestry was first defined by (McColl et al., 2018), specifically referring to ancient hunter-gatherers from Laos and Malaysia associated with Hòabìnhian material culture for whom genetic data was sampled. Materials associated with Hòabìnhian industry have been found in an extended region of Southeast Asia and southern China, but genetic evidence has only been described in Southeast Asia.

Fujian ancestry: Early Neolithic southern East Asians from Fujian and surrounding areas: Qihe3 (this study); Qihe(Qihe2), Liangdao1, and Liangdao2 (Yang et al., 2020). In the main text, *EN_FJ* is the abbreviation for Early Neolithic Fujian ancestries in f_4-statistics, while *LN_FJ* is the abbreviation for Late Neolithic Fujian ancestries in f_4-statistics.

Shandong ancestry: Early Neolithic northern East Asians from Shandong, China: Bianbian, Boshan, Xiaojingshan, Xiaogao (Yang et al., 2020). In the main text, *EN_SD* is the abbreviation for Early Neolithic Shandong ancestries in f_4-statistics.

Siberian-related northern East Asian: Yumin (Yang et al., 2020) an early Neolithic individual from Inner Mongolia, China. Two Neolithic northeast Asians from coastal Siberia, DevilsCave_N (Sikora et al., 2019) and Boisman_MN (Wang et al., 2021).

STAR★METHODS

Detailed methods are provided in the online version of this paper
and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Sites and specimen description
- METHOD DETAILS
  - Ancient DNA extraction, sequencing, and data processing
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Present-day Datasets
  - Relatedness analysis
  - Principal components analysis
  - *f*-statistics
  - ADMIXTURE analysis
  - Inferring admixture and estimating mixture proportions
  - Admixture Graph modeling
  - Estimating a maximum likelihood phylogeny with migration events
  - Northern East Asian influence on historical Guangxi samples
  - Archaic ancestry estimation
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at XXXX

ACKNOWLEDGMENTS

We thank Shiyu Qiao, Vikas Kumar for comments and Lihua Wu for help in sampling. We also thank our anonymous peer reviewers for insightful comments that have greatly helped us enhance the key findings. This work was supported by the Chinese Academy of Sciences (CAS, XDB26000000) to Q.F. and X.J., National Natural Science Foundation of China (41925009,41630102), National Key R&D Program of China (2016YFE0203700), CAS (XDA1905010, QYZDB-SSW-DQC003), "Research on the roots of Chinese civilization" of Zhengzhou University (XKZDJC202006), Tencent Foundation through the XPLORER PRIZE, and the Howard Hughes Medical Institute (grant no. 55008731).

AUTHOR CONTRIBUTIONS

Q.F. conceived the idea for this study. W.W., G.X., Z.L., Q.Y., X.F., X.W., X.W., L.Q., F.L., L.Z., M.Z., X.C., D.Z., Z.Z., Y.W., X.G., D.C., X.J. and Q.F. assembled archaeological materials and performed dating. Q.F., P.C., R.Y., F.L., Q.D, X.F. and W.P performed or supervised wet laboratory work. M.A.Y. and H.S. did preliminary data analysis at IVPP. T.W and Q.F. did the primary data analysis for this manuscript. T.W, M.A.Y. and Q.F. wrote the manuscript with critical input from all authors. T.W., M.A.Y., W.W., X.J., Y.L.,G.X., Z.L., X.F., Q.L., F.L., X.Z. and Q.F. wrote and edited the supplement. X.J., W.W., G.X., Z.L, L.Q., D.C. Y.L, E.A.B. and X.M. helped to revise the manuscript and supplement.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome research *19*, 1655-1664.

Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M., and Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nature genetics *23*, 147.

B. I. G. Data Center Members (2018). Database Resources of the BIG Data Center in 2018. Nucleic acids research *46*, D14-D20.

Bai, F., Zhang, X., Ji, X., Cao, P., Feng, X., Yang, R., Peng, M., Pei, S., and Fu, Q. (2020). Paleolithic genetic link between Southern China and Mainland Southeast Asia revealed by ancient mitochondrial genomes. Journal of Human Genetics *65*, 1125-1128.

Baum, B.R. (1989). PHYLIP: Phylogeny Inference Package. Version 3.2. Joel Felsenstein. The Quarterly Review of Biology *64*, 539-541.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological) *57*, 289-300.

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. Proceedings of the National Academy of Sciences *104*, 14616-14621.

Colani, M. (1927). L',ge de la pierre dans la province de Hoa-Binh (Tonkin) (HanoÔ: impr. d'ExtrÍme-Orient).

Curnoe, D., Ji, X., Taçon, P.S.C., and Yaozheng, G. (2015). Possible Signatures of Hominin Hybridization from the Early Holocene of Southwest China. Scientific Reports *5*, 12408.

Curnoe, D., Xueping, J., Herries, A.I., Kanning, B., Tacon, P.S., Zhende, B., Fink, D., Yunsheng, Z., Hellstrom, J., Yun, L.*, et al.* (2012). Human remains from the Pleistocene-Holocene transition of southwest China suggest a complex evolutionary history for East Asians. PLoS One *7*, e31918.

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.T., Weihmann, A., Nickel, B., Valdiosera, C., Garcia, N., Paabo, S., Arsuaga, J.L.*, et al.* (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proceedings of the National Academy of Sciences of the United States of America *110*, 15758-15763.

de Barros Damgaard, P., Marchi, N., Rasmussen, S., Peyrot, M., Renaud, G., Korneliussen, T., Moreno-Mayar, J.V., Pedersen, M.W., Goldberg, A., Usmanova, E.*, et al.* (2018a). 137 ancient human genomes from across the Eurasian steppes. Nature *557*, 369-374.

de Barros Damgaard, P., Martiniano, R., Kamm, J., Moreno-Mayar, J.V., Kroonen, G., Peyrot, M., Barjamovic, G., Rasmussen, S., Zacho, C., Baimukhanov, N.*, et al.* (2018b). The first horse herders and the impact of early Bronze Age steppe expansions into Asia. Science *360*, eaar7711.

Durand, E.Y., Patterson, N., Reich, D., and Slatkin, M. (2011). Testing for ancient admixture between closely related populations. Molecular Biology and Evolution *28*, 2239-2252.

Forestier, H. (2000). De quelques chaînes opératoires lithiques en Asie du Sud-Est au pléistocène supérieur final et au début de l'holocène. L'Anthropologie *104*, 531-548.

Fu, Q., Hajdinjak, M., Moldovan, O.T., Constantin, S., Mallick, S., Skoglund, P., Patterson, N., Rohland, N., Lazaridis, I., Nickel, B.*, et al.* (2015). An early modern human from Romania with a recent Neanderthal ancestor. Nature *524*, 216-219.

Fu, Q., Li, H., Moorjani, P., Jay, F., Slepchenko, S.M., Bondarev, A.A., Johnson, P.L.F., Aximu-Petri, A., Prüfer, K., de Filippo, C.*, et al.* (2014). Genome sequence of a 45,000-year-old modern human from western Siberia. Nature *514*, 445-449.

Fu, Q., Meyer, M., Gao, X., Stenzel, U., Burbano, H.A., Kelso, J., and Paabo, S. (2013a). DNA analysis of an early modern human from Tianyuan Cave, China. Proceedings of the National Academy of Sciences *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. Current Biology *23*, 553-559.

Gansauge, M.T., and Meyer, M. (2013). Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. Nat Protoc *8*, 737-748.

Gorman, C.F. (1970). Excavations at Spirit Cave, North Thailand: SOME INTERIM INTERPRETATIONS. Asian Perspectives *13*, 79-107.

Guangxi Museum, and Tiandong County Museum (1991). Survey and Research on Cliff Burials of the Left and Right River Valleys in Guangxi (in Chinese). Jianghan Archaeology *3*.

Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K.*, et al.* (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. Nature *522*, 207-211.

Hung, H.-c. (2019). Prosperity and complexity without farming: the South China Coast, c. 5000–3000 BC. Antiquity *93*, 325-341.
Hung, H.-c., Chi, Z., Matsumura, H., and Zhen, L. (2017). Neolithic Transition in Guangxi: A Long Development of Hunting-Gathering Society in Southern China. In Bio-anthropological studies of early Holocene hunter-gatherer sites at Huiyaotian and Liyupo in Guangxi, China (Tokyo National Museum of Nature and Science, November 2017.), pp. 205-228.
Institute of Archaeology Chinese Academy of Social Sciences (2003). Guilin Zengpiyan (in Chinese) (Cultural Relics Publishing House).
Jeong, C., Ozga, A.T., Witonsky, D.B., Malmstrom, H., Edlund, H., Hofman, C.A., Hagan, R.W., Jakobsson, M., Lewis, C.M., Aldenderfer, M.S.*, et al.* (2016). Long-term genetic stability and a high-altitude East Asian origin for the peoples of the high valleys of the Himalayan arc. Proceedings of the National Academy of Sciences of the United States of America *113*, 7485-7490.
Ji, X., Kuman, K., Clarke, R.J., Forestier, H., Li, Y., Ma, J., Qiu, K., Li, H., and Wu, Y. (2016). The oldest Hoabinhian technocomplex in Asia (43.5 ka) at Xiaodong rockshelter, Yunnan Province, southwest China. Quaternary International *400*, 166-174.
Ji, X., Wu, X., Wu, Y., and Liu, W. (2014). The temporal bony labyrinthine morphology of Homo Longlin 1 from the Pleistocene–Holocene transition of South China (in Chinese). Chinese Science Bulletin.
Kanzawa-Kiriyama, H., Jinam, T.A., Kawai, Y., Sato, T., Hosomichi, K., Tajima, A., Adachi, N., Matsumura, H., Kryukov, K., Saitou, N.*, et al.* (2019). Late Jomon male and female genome sequences from the Funadomari site in Hokkaido, Japan. Anthropological Science *127*, 83-108.
Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics *15*, 356.
Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D.C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., Sirak, K.*, et al.* (2016). Genomic insights into the origin of farming in the ancient Near East. Nature *536*, 419-424.
Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754-1760.
Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., Cann, H.M., Barsh, G.S., Feldman, M., Cavalli-Sforza, L.L.*, et al.* (2008). Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation. Science *319*, 1100.
Liao, W., Xing, S., Li, D., Martinón-Torres, M., Wu, X., Soligo, C., Bermúdez de Castro, J.M., Wang, W., and Liu, W. (2019). Mosaic dental morphology in a terminal Pleistocene hominin from Dushan Cave in southern China. Scientific reports *9*, 2347-2347.
Lipson, M., Cheronet, O., Mallick, S., Rohland, N., Oxenham, M., Pietrusewsky, M., Pryce, T.O., Willis, A., Matsumura, H., Buckley, H.*, et al.* (2018). Ancient genomes document multiple waves of migration in Southeast Asian prehistory. Science *361*, 92-95.
Liu, D., Duong, N.T., Ton, N.D., Van Phong, N., Pakendorf, B., Van Hai, N., and Stoneking, M. (2020). Extensive Ethnolinguistic Diversity in Vietnam Reflects Multiple Sources of Genetic Diversity. Molecular Biology and Evolution *37*, 2503-2519.
Lu, D., Lou, H., Yuan, K., Wang, X., Wang, Y., Zhang, C., Lu, Y., Yang, X., Deng, L., Zhou, Y.*, et al.* (2016). Ancestral Origins and Genetic History of Tibetan Highlanders. American journal of human genetics *99*, 580-594.
Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., Tandon, A.*, et al.* (2016). The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature *538*, 201-206.
Marwick, B. (2008). What attributes are important for the measurement of assemblage reduction intensity? Results from an experimental stone artefact assemblage with relevance to the Hoabinhian of mainland Southeast Asia. Journal of Archaeological Science *35*, 1189-1200.
Matsumura, H. (2006). The population history of Southeast Asia viewed from morphometric analyses of human skeletal and dental remains. In Bioarchaeology of Southeast Asia, M. Oxenham, and N. Tayles, eds. (Cambridge: Cambridge University Press), pp. 33-58.
Matsumura, H., Hung, H.-c., Cuong, N., Zhao, Y.f., He, G., and Chi, Z. (2017). Mid-Holocene Hunter-Gatherers 'Gaomiao' in Hunan, China: The First of the Two-layer Model in the Population History of East/Southeast Asia. In, pp. 61-78.
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. Scientific Reports *9*, 1451.
Matsumura, H., and Oxenham, M.F. (2014). Demographic transitions and migration in prehistoric East/Southeast Asia through the lens of nonmetric dental traits. American Journal of Physical Anthropology *155*, 45-65.
Matsumura, H., Oxenham, M.F., and Kt, N. (2011). Population history of mainland Southeast Asia: the Two Layer model in the context of Northern Vietnam. In, pp. 153-178.

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*, 88-92.
Meacham, W. (1996). Defining the Hundred Yue. Indo-Pacific Prehistory Association Bulletin *15*, 93.
Meyer, M., Kircher, M., Gansauge, M.T., Li, H., Racimo, F., Mallick, S., Schraiber, J.G., Jay, F., Prufer, K., de Filippo, C.*, et al.* (2012). A high-coverage genome sequence from an archaic Denisovan individual. Science *338*, 222-226.
Monroy Kuhn, J.M., Jakobsson, M., and Günther, T. (2018). Estimating genetic kin relationships in prehistoric populations. PLoS One *13*, e0195491.
Moreno-Mayar, J.V., Potter, B.A., Vinner, L., Steinrücken, M., Rasmussen, S., Terhorst, J., Kamm, J.A., Albrechtsen, A., Malaspinas, A.-S., Sikora, M.*, et al.* (2018). Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. Nature *553*, 203-207.
Narasimhan, V.M., Patterson, N., Moorjani, P., Rohland, N., Bernardos, R., Mallick, S., Lazaridis, I., Nakatsuka, N., Olalde, I., Lipson, M.*, et al.* (2019). The formation of human populations in South and Central Asia. Science *365*, eaat7487.
Ning, C., Li, T., Wang, K., Zhang, F., Li, T., Wu, X., Gao, S., Zhang, Q., Zhang, H., Hudson, M.J.*, et al.* (2020). Ancient genomes from northern China suggest links between subsistence changes and human migration. Nat Commun *11*, 2700.
Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. Genetics *192*, 1065-1093.
Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLoS genetics *2*, e190.
Peng, C. (2001). Preliminary study on early cave burials in Guangxi (in Chinese). Guangxi Ethnic Studies, 84-90.
Peng, S. (2009). Archaeological Discovery of Tooth Extraction Custom of Ancient Residents in Lingnan (in Chinese). Cultural Relics of Southern China, 80-88.
Peng, S. (2013a). Guangxi ancient cave burials (in Chinese) (Nanning: Guangxi Science & Technology Publishing House).
Peng, S. (2013b). On the Form of the Wooden Coffin Burial in Ancient Cliff Caves in Guangxi (in Chinese). In The 16th Annual Meeting of China Baiyue Ethnic History Research Association, C.B.E.H.R. Association, G.I.o.C.R.a. Archaeology, G.I.o.C.R.a. Archaeology, and S.Y.-s.U. Department of Anthropology, eds. (Guangzhou,Guangdong,China), pp. 315-327.
Peter, B.M. (2020). 100,000 years of gene flow between Neandertals and Denisovans in the Altai mountains. bioRxiv, 2020.2003.2013.990523.
Pickrell, J.K., and Pritchard, J.K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. PLoS genetics *8*, e1002967.
Pinhasi, R., Fernandes, D., Sirak, K., Novak, M., Connell, S., Alpaslan-Roodenberg, S., Gerritsen, F., Moiseyev, V., Gromov, A., Raczky, P.*, et al.* (2015). Optimal Ancient DNA Yields from the Inner Ear Part of the Human Petrous Bone. PLoS One *10*, e0129102.
Pinhasi, R., Fernandes, D.M., Sirak, K., and Cheronet, O. (2019). Isolating the human cochlea to generate bone powder for ancient DNA analysis. Nat Protoc *14*, 1194-1205.
Posth, C., Nakatsuka, N., Lazaridis, I., Skoglund, P., Mallick, S., Lamnidis, T.C., Rohland, N., Nagele, K., Adamski, N., Bertolini, E.*, et al.* (2018). Reconstructing the Deep Population History of Central and South America. Cell *175*, 1185-1197 e1122.
Prüfer, K., de Filippo, C., Grote, S., Mafessoni, F., Korlević, P., Hajdinjak, M., Vernot, B., Skov, L., Hsieh, P., Peyrégne, S.*, et al.* (2017). A high-coverage Neandertal genome from Vindija Cave in Croatia. Science *358*, 655-658.
Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., Heinze, A., Renaud, G., Sudmant, P.H., de Filippo, C.*, et al.* (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. Nature *505*, 43-49.
Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., De Bakker, P.I., and Daly, M.J. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics *81*, 559-575.
Raghavan, M., Skoglund, P., Graf, K.E., Metspalu, M., Albrechtsen, A., Moltke, I., Rasmussen, S., Stafford, T.W., Jr., Orlando, L., Metspalu, E.*, et al.* (2014). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature *505*, 87-91.
Reimer, P.J., Austin, W.E.N., Bard, E., Bayliss, A., Blackwell, P.G., Bronk Ramsey, C., Butzin, M., Cheng, H., Edwards, R.L., Friedrich, M.*, et al.* (2020). The IntCal20 Northern Hemisphere Radiocarbon Age Calibration Curve (0–55 cal kBP). Radiocarbon *62*, 725-757.
Renaud, G., Stenzel, U., and Kelso, J. (2014). leeHom: adaptor trimming and merging for Illumina sequencing reads. Nucleic acids research *42*, e141.
Sawyer, S., Krause, J., Guschanski, K., Savolainen, V., and Pääbo, S. (2012). Temporal Patterns of Nucleotide

Misincorporations and DNA Fragmentation in Ancient DNA. PLOS ONE *7*, e34131.
Seguin-Orlando, A., Korneliussen, T.S., Sikora, M., Malaspinas, A.S., Manica, A., Moltke, I., Albrechtsen, A., Ko, A., Margaryan, A., Moiseyev, V.*, et al.* (2014). Paleogenomics. Genomic structure in Europeans dating back at least 36,200 years. Science *346*, 1113-1118.
Shi, C. (1995). Five Migration Waves in the History of the Miao (in Chinese). Guizhou Ethnic Studies *01*, 120-128.
Shinde, V., Narasimhan, V.M., Rohland, N., Mallick, S., Mah, M., Lipson, M., Nakatsuka, N., Adamski, N., Broomandkhoshbacht, N., Ferry, M.*, et al.* (2019). An Ancient Harappan Genome Lacks Ancestry from Steppe Pastoralists or Iranian Farmers. Cell *179*, 729-735.e710.
Sikora, M., Pitulko, V.V., Sousa, V.C., Allentoft, M.E., Vinner, L., Rasmussen, S., Margaryan, A., de Barros Damgaard, P., de la Fuente, C., Renaud, G.*, et al.* (2019). The population history of northeastern Siberia since the Pleistocene. Nature *570*, 182-188.
Skoglund, P., Northoff, B.H., Shunkov, M.V., Derevianko, A.P., Pääbo, S., Krause, J., and Jakobsson, M. (2014). Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. Proceedings of the National Academy of Sciences *111*, 2229.
Solheim, W.G. (1970). Northern Thailand, Southeast Asia, and World Prehistory. Asian Perspectives *13*, 145-162.
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.
Wang, Y., Song, F., Zhu, J., Zhang, S., Yang, Y., Chen, T., Tang, B., Dong, L., Ding, N., Zhang, Q.*, et al.* (2017). GSA: Genome Sequence Archive. Genomics Proteomics Bioinformatics *15*, 14-18.
Wu, X., Fan, X., Li, S., Gao, X., Zhang, Y., and Fang, Y. (2014). The early Neolithic human skull from Qihe Cave, Zhangping, Fujian (in Chinese). Acta Anthropol Sinica *33*, 448-459.
Xu, J., and Wei, X. (2008). Review on Studies of Origins of ethnic groups in the Lingnan (in Chinese with English abstract). Guangxi Ethnic Studies *93*, 115-124.
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, 282-288.
Yang, M.A., Gao, X., Theunert, C., Tong, H., Aximu Petri, A., Nickel, B., Slatkin, M., Meyer, M., Pääbo, S., Kelso, J.*, et al.* (2017). 40,000-year-old individual from Asia provides insight into early population structure in Eurasia. Current Biology *27*, 3202-3208 e3209.
Zhang, C., and Hung, H.-c. (2008). The hunter-gatherer groups in southern China and its adjacent regions during the Neolithic. Kaoguxue Yanjiu [Archaeology Studies], 415-434 (in Chinese).
Zhang, C., and Hung, H.-c. (2012). Later hunter-gatherers in southern China, 18 000–3000 BC. Antiquity *86*, 11-29.
Zhang, S., Peng, S., and Zhou, S. (1986). Investigation report of Lihu cave burial in Nandan County, Guangxi (in Chinese). Cultural Relics, 67-77+107.
Zhang, X., Pei, S., and Wu, X. (2017). Qingshuiyuan Dadong: A newly discovered Late Paleolithic site in Guizhou province, China. Archaeological Research in Asia *9*.
Zhou, J. (1991). Discussion on the Cave Burial in Guangxi (in Chinese). Guangxi Wenwu *2*.
Zhou, J., and Tian, F. (1991). Investigation and Study on Cliff Burial in the Left and Right River Basins of Guangxi (in Chinese). Jianghan Archaeology, 30-38.

Figure Legends

Figure 1. Geographic, temporal and genetic information for newly sequenced individuals.
(A) Geographic locations of newly sampled individuals, the map also shows published individuals from East and Southeast Asia. Associated information is provided in Table S1.
(B) The calibrated radiocarbon dates of newly sampled individuals.
(C) Principal component analysis (PCA) of ancient individuals projected onto present-day East and Southeast Asians. The color of the present-day population indicates their language affiliation: Austronesian-speakers (Gray), Austroasiatic-speakers (Green), Hmong-Mien-speakers (Blue), Tai-Kadai-speakers (Teal), Sino-Tibetan speakers (Orange).

Figure 2. Genetic structure and admixture of prehistoric Guangxi individuals.
(A) Treemix phylogeny allowing three migration events.
(B) Admixture graph fitting early Asians and East Asians. The vertical timeline shows the radiocarbon date of the individual, but does not accurately reflect population split times. The estimated genetic drift on each branch is given, and the admixture events with the estimated mixture proportions are shown in dashed lines.
(C) Populations across geographically southern East Asians and Southeast Asians (X) who share more alleles with Longlin than northern East Asians (red, Z<-3) in f_4*(Mbuti, Longlin; X, DevilsCave_N)*, where DevilsCave_N is a northern East Asian from coastal Siberia (~7,700 BP).
(D) Populations from X who share more alleles with Dushan than coastal southern East Asians (red, Z<-3) in f_4*(Mbuti, Dushan; X, Qihe3)*, where Qihe3 is a southern East Asian (~12,000 BP).
(E) Populations from X who share more alleles with a Hòabìnhian (La368) than northern East Asians (red, Z<-3) in f_4*(Mbuti,* Hòabìnhian*; X, DevilsCave_N)*.

Figure 3. Genetic relationships of historical Guangxi populations.
(A) Plot of f_4*(Mbuti, X; 1500BP GX, GaoHuaHua)*, where X are present-day populations with different language affiliations, and 1,500BP GX are historical Guangxi populations dated to ~1,500 BP. Hmong-Mien speakers show a significantly closer relationship with the ~500 BP GaoHuaHua from Guangxi.
(B) Outgroup f_3-statistics of f_3*(Mbuti; X, Y)* where X are historical Guangxi populations, and Y are various ancient northern East Asians from different regions. All historical Guangxi populations share the most genetic drift with northern East Asians from Shandong and the Central Plain. Population information of northern East Asians can be found in STAR Methods.

741 **STAR METHOD**

STAR METHODS

1 LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Qiaomei Fu (fuqiaomei@ivpp.ac.cn).

2 EXPERIMENTAL MODEL AND SUBJECT DETAILS

2.1 Sites and specimen description

In this study, we sampled the remains of 170 ancient humans from Guangxi Zhuang Autonomous Region, China (Table S1). 30 individuals were successfully sequenced from 16 sites with radiocarbon dates ranging from 10,686 to 294 calibrated years before present (cal BP, Table 1, Table S1). In addition, we also sequenced an additional individual (Qihe3) from Qihe cave, Fujian, China, where one individual was previously sequenced (Yang et al., 2020).

All samples but the one from Baojianshan Cave were directly dated using radiocarbon (^{14}C) dating techniques through accelerator mass spectrometry (AMS), which were then calibrated using the Int Cal 20 calibration curve (Reimer et al., 2020) (Table 1, Table S1). All ages are reported as cal BP, where BP means years before present (present is AD 1950).

These samples were collected from the related archaeological institutes and research universities, with their appropriate permissions. A review board at the Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Science (IVPP-CAS) surveyed the samples from which we successfully retrieved ancient DNA for this project and approved their use for this project (review no. 202005160005).

2.1.1 Prehistoric Caves

Longlin - The Longlin Laomocao Cave site is located in Longlin Autonomous County, Baise City, Guangxi Zhuang Autonomous Region, China. Human fossils – including one incomplete skull, one mandible, more than ten vertebrae, and ribs – were recovered from the cave in 1979 (Ji et al., 2014). No associated artifacts were collected and no excavation was carried out thereafter. The skull morphology of Longlin_1 exhibits unusual characteristics for a modern human, with a mixture of both archaic and modern human features (Curnoe et al., 2015; Curnoe et al., 2012). A contemporaneous early human, Maludong, found in Mengzi (Yunnan, Southwest China) shows similar characteristics as Longlin, possibly indicative of similar ancestry (Curnoe et al., 2012). Such unusual cranial morphological features are not seen among Pleistocene or present-day populations of modern human (Curnoe et al., 2012). Three possible hypotheses have been proposed for the presence of unusual cranial features in Longlin and Maludong: First, they represent a late-surviving archaic population (Curnoe et al., 2012). Second, they resulted from the retention of a large number of ancestral polymorphisms in a population of *H. sapiens* (Curnoe et al., 2012). Third, Longlin may have descended from a modern population that interbred with one or more archaic groups (Curnoe et al., 2015).

We successfully obtained genome-wide data from the temporal bone of Longlin_1, who was directly radiocarbon dated to 10,686-10,439 cal BP.

Dushan - The Dushan cave site is located in Linfeng Town, Tiandong County, Guangxi Zhuang Autonomous Region, China. This cave was found in 2010 and excavated in 2011 by the Natural History Museum of Guangxi. It sits at a low isolated hill of Paleozoic limestone in a small valley 8 meters above the valley floor, surrounding by typical karst peak clusters. The cave extends about 15 meters in length from southwest to northeast, with an average width of 4 meters. In aerial view, the interior floor is an oblique triangle, with a marked wide entrance and narrow terminal. The sediment is almost undisturbed, mainly concentrated at the entrance and becoming thinner from southwest to northeast. One archaeological test pit was excavated with an exposed area of six square meters (2 by 3 meters). The test pit was excavated in intervals of 10 cm, down to the deepest horizontal layer (layer 19). Four stratigraphic units were identified (from top to bottom) based on varying characteristics of deposits. Unit I consists of thin cemented yellow silty clay containing a small number of stone artifacts and a few mammal teeth and vertebrae fragments; Unit II consists of thick grey

yellow silty clay with limestone breccia, yielding flaked stone artifacts, ground stone tools, and hominin fossils; Unit III is a thick pale yellow silty clay with breccia, producing flaked stone artifacts and hominin remains, however, ground stone implements are absent in this unit; Unit IV contains cemented clay and is restricted to the northeast part of the test pit, without any cultural or animal remains (Liao et al., 2019). This excavation recovered more than one thousand stone artifacts and about two hundred hominin remains.

Radio carbon dating (AMS) results indicate that the age of Dushan sedimentation covers a range from the terminal Pleistocene to the early Holocene, roughly from 15,000 to 7,000 BP (7753 ± 49 cal BP to layer 17: 14995 ± 369 cal BP). Interestingly, the lithic assemblage in this site can be clearly categorized as either Neolithic or Paleolithic in good correspondence with the sedimentary strata. In Unit II (7,000-12,000 BP), the lithic assemblage includes choppers, scrapers, utilized flakes and ground stone tools including adzes and grinders. The appearance of the ground stone tools seems to imply a threshold for the Neolithic period in this region. In Unit III (12,000-15,000 BP), the dominant stone tools are small flake-based tools, including well retouched scrapers and utilized flakes, and a small number of cores and choppers. All of these stone artifacts are produced from medium sized fluvial cobbles that were transported from the ancient Youjiang River, more than 10 km north of the cave.

Dushan cave is close to northern Vietnam (~100 km to the border) where the Hòabìnhian technocomplex was first characterized by large, flat and long, largely unifacial, cobble tools. However, the typical Hòabìnhian-like tools, shaped on cobbles with a plano-convex cross-section, or the "sumatralith" (flaking usually around the circumference of a unifacial tool) (Forestier, 2000; Gorman, 1970; Ji et al., 2016; Marwick, 2008) do not occur at this site. We have noted that the choppers, chopping-tools and small flake tools are also common in the Hòabìnhian technocomplex; nevertheless, the Dushan lithic assemblage is more like that of traditional south or central China. To date, archeological evidence indicates that the Hòabìnhian technocomplex has a broad distribution in Southeast Asia and beyond, yet the appearance of this complex in southwest China is rare, except for a recent finding at Xiaodong in Yunnan, southwest China (Ji et al., 2016). Therefore, we tend to consider that there is no distinct relationship of lithic technology between the Dushan assemblage and Hòabìnhian complex. Yet because the existence of technocomplex diversity in this area is a significant issue, a more precise categorization awaits future research.

The human samples analyzed in this paper are from the fourth horizontal layer in Unit II (40 cm beneath the surface) of Dushan Cave. In this layer, rich stone artifacts were found to accompany the human bones, mainly including stone grinders and scrapers. No evidence associated with early agriculture has been found in Dushan Cave.

We successfully obtained genome-wide data from the temporal bone of a human excavated from Layer 4 in Unit II of Dushan Cave, Dushan4_1, who was directly radiocarbon dated to 8,974-8,593 cal BP.

A separate 15,000 BP individual (Dushan1, not sampled in this study) from the same Dushan Cave, shows morphological features that are rare in modern humans but more commonly found in Middle Pleistocene archaic humans (Liao et al., 2019). Like Longlin, plausible explanations are that Dushan1 represents a late surviving individual representing some of the earliest modern humans or the ancestors of Dushan1 admixed with late-surviving archaic humans (Liao et al., 2019).

Baojianshan - The Baojianshan Cave site (Baojianshan Cave A) is located in Longzhou County, Chongzuo City, Guangxi Zhuang Autonomous Region, China. This cave sits at a western cliff of the Zuojiang River, 10 m above water surface and 115 m above sea level. The cave is relatively spacious, with an area of 120 square meters. In 2013, it was excavated by the Guangxi Institute of Cultural Relic Protection and Archaeology. Two test pits (5 by 5 m and 2 by 2 m) were dug to an average depth of about 1.5 m. The strata can be divided into nine layers from top to bottom, preliminary radiocarbon dated to about 3,000 to 8,400 years BP. From Layer 1 to Layer 3, the sediment mainly consists of silt clay, heavily disturbed by late human activities. Some human and animal bones, pottery fragments, implements made of stone and shell are found scattered at these layers and estimated to be from 3,000 to 4,000 years BP. Layer 4 and Layer 6 consist of shell middens, with an average thickness of 20 – 30 cm respectively, containing human and animal bones, as well as shell and stone artifacts. Layer 5 and Layer 7 to Layer 9 consist of silt clay, containing some animal bones, stone and shell artifacts, and a few pottery fragments. The thickness of these layers varies from 5 to 40 cm. Importantly, two human individual skeletons (M1 and M2) were found under Layer 5 and

another one (M3) was in Layer 7.

This excavation resulted in a discovery of 1292 cultural remains in total, including 34 flaked stone artifacts and 32 ground stone tools, and more than 1100 pottery fragments. The flaked stone artifacts consist of stone anvils, cores, flakes, choppers and scrapers. The ground stone artifacts contain stone axes, stone adzes, and stone grinders. A bone sword and some implements made of shell were also unearthed during this excavation. The characteristics of the lithic assemblage of this site is similar to that of typical Neolithic sites broadly distributed along riversides in south China.

Although the Baojianshan site is close to north Vietnam and Hòabìnhian sites, the elements of the Hòabìnhian technocomplex does not occur at this site. In addition, no evidence of agriculture has been found in this site, such as domesticated animal or cultivated rice.

We sequenced two individuals from M1 and M2 of Baojianshan Cave under Layer 5. Many bones from the human skeleton from M1 were very fragmented, and most of the limb bones were broken. Based on the bone placement, the individual in M1 showed a supine with legs flexed burial. The human skeleton from M2 was identified to be a juvenile. This child was badly preserved, with most bones fragmented. The child was found in flexed burial. Both individuals were placed on and surrounded by a large number of shells. We attempted direct radiocarbon dating from these human skeletons several times, but all attempts failed. We instead radiocarbon-dated charcoal from Layer 4, the layer above Layer 5, for which we determined a calibrated date ranging from 6,400-6,290 cal BP (2σ 95.4%). We additionally radiocarbon-dated an animal skeleton excavated from Layer 7, for which we determined a calibrated date ranging from 8,415-8,335 cal BP (2σ 95.4%). The two human specimens from Baojianshan Cave sampled in this study were both discovered under Layer 5, between Layer 7 and Layer 4. Thus, these individuals are likely older than 6,400 cal BP but younger than 8,335 cal BP. For this study, we used the date range of 8,335 - 6,400 BP for the Baojianshan individuals.

We found that these two individuals, Baojianshan5_M1 (786,870 SNPs) and Baojianshan5_M2 (37,557 SNPs), had a familial relationship, with kinship to the second degree (Table S1). For population genetic analyses, we used the higher coverage Baojianshan5_M1.

Qihe - The Qihe cave is located in Zhangping, Fujian, China. Three human skulls were excavated from the same cultural phase in Qihe cave. The Qihe1 specimen is a small fragment of a child's skull, which was not well preserved due to severe damage (Wu et al., 2014), and we failed to retrieve genome-wide data . Genome-wide data for a second specimen, Qihe2 (8,428-8,359 cal BP), was retrieved in a previous study, where Qihe2's genetic ancestry was closely related to that of Austronesians, suggesting that they either were or contributed to early ancestors of Austronesians (Yang et al., 2020). In this study, we sequenced a newly excavated individual, Qihe3, who is located earlier in the strata (Wu et al., 2014). The Qihe2 individual is buried with stone tools, sand tempered pottery pieces and animal bones. The Qihe3 individual was not found during the site excavation, but was discovered beneath Qihe2 during processing of the sediment block in the morphology lab at IVPP-CAS. Qihe3 was accompanied by a small amount of stone flakes and red burnt soil. Study of Qihe3's skull morphology shows Qihe3 has a long head, large cranial capacity, high narrow face, broad and low nasal shape, consistent with other late *H. sapiens* (Wu et al., 2014). Qihe3 was directly radiocarbon dated to 11,747-11,356 cal BP.

2.1.2 Historical Cave burials

Cave Burial (Yandongzang) is a burial custom where the dead are placed in natural caves. This custom is distinct from both hanging coffins (wooden coffins placed on beams secured to a cliff) found in the Yangtze River region and cliff burials (excavated artificial caves on a cliff) found in Sichuan, China (Peng, 2013a). Guangxi is the oldest region of China where cave burials have been found. They have been found from the end of the Late Neolithic up to the Ming and Qing dynasties, lasting for more than 4,000 years. Thus, Guangxi is believed to be an important birthplace of the cave burial. Cave burials are densely distributed in Guangxi, as this region is filled with developed karst features containing many natural caves. Most of the cave burial sites were chosen at the foot of mountains and mountainsides, while some were chosen on cliffs and near the top of mountains, in natural caves or rock buildings that were hidden and not accessible to people; some of the cave entrances were artificially blocked. Clan burials, in which many people are buried together, is dominant. Both primary and secondary burials are observed (Peng, 2013a).

Based on the inscription and coffin typology, cave burials in Guangxi were believed to belong to ancestors of the Zhuang (Tai-Kadai speakers) (Guangxi Museum and Tiandong County Museum, 1991). However, cave burials from later periods in Lihu Yaozu Town, Nandan County, Hechi City, Guangxi have been hypothesized to be connected to Miao-Yao populations (Hmong-Mien speakers) (Zhou, 1991). It is argued that the Zhuang-Dong (Tai-Kadai speakers) in Guangxi are the original populations in the Lingnan region (Guangxi, Guangdong, Hainan, Hong Kong, Macao), closely related to the "Baiyue" populations, various ethnicities who inhabited southern China during the 1st millennium BC to the 1st millennium AD (Meacham, 1996; Xu and Wei, 2008). Although the origin and migration of the Miao-Yao speakers (Hmong-Mien speakers) is not completely settled (Shi, 1995; Xu and Wei, 2008), most scholars believe that the Miao-Yao group living in Guangxi today moved into this region in a later period spanning from the Yuan Dynasty (1271-1368 AD) to the Qing Dynasty (1636-1912 AD), before later migrating from Yunnan and Guangxi to Vietnam, Laos and Thailand (Zhang et al., 1986).

We sequenced 26 individuals from 12 cave burials with coffins in Guangxi. Individuals from nine sites date to 1,688-1,278 cal BP, but three sites (Gaofeng, Huaqiao, Huatuyan) date to 513-294 years ago and are located in Lihu Yaozu Town, Nandan County, Guangxi, where the Baikuyao, a subbranch of the Yao population, primarily live today. Thus, our historical individuals range from 1,688-294 years ago.

Banda - The Banda cave site is located in Dahua Yao Autonomous County, Hechi City, Guangxi Zhuang Autonomous Region, China. Banda shows characteristics of the late period of Cave Burials in Guangxi. The coffins in Banda have heads and tails in the style of horns (Peng, 2013b). The human skulls found in this cave show the custom of tooth ablation (Peng, 2009). We successfully sequenced two individuals from Banda cave, BandaKD11 and BandaKD15. BandaKD15 was directly radiocarbon dated to 1,517-1,353 cal BP, and BandaKD11 was directly radiocarbon dated to 1,467-1,307 cal BP.

Layi - The Layi (Laba) cave site is located on the bank of Hongshui River, Baida Village, Beijing Town, Dahua Yao Autonomous County, Hechi City, Guangxi Zhuang Autonomous Region, China. There are four caves associated with this site, where Caves 1 and 2 are on the left bank of the river while Caves 3 and 4 are on the right. Human bones, pottery fragments, stone tools, bone tools and coffins were collected from the caves (Peng, 2001). The coffins were made of intact wood with the head and tail decorated in the shape of horns and swallowtails (Peng, 2013a). The human skulls found in this cave show the custom of tooth ablation (Peng, 2009). We sequenced LayiKD01 from KD01, Cave 1. We directly radiocarbon dated LayiKD01 to 1,532-1,403 cal BP.

Qinchang - The Qinchang cave site, near the Hongshui River, is located on Nongshi Hillside, Qinchang, Yantan Town, Dahua Yao Autonomous County, Hechi City, Guangxi Zhuang Autonomous Region, China. Coffins with torch-shaped heads were found in Qinchang Cave (Peng, 2013b). The human skulls found in this cave have the custom of tooth ablation (Peng, 2009). We sequenced genome-wide data for two individuals (KD13 in Grave M1:1 and KD14 in Grave 6) at Qinchang Cave using bone samples from their teeth. QinchangKD13 was directly radiocarbon dated to 1,520-1,363 cal BP, and QinchangKD14 was directly radiocarbon dated to 1,545-1,407 cal BP.

Balong - The Balong cave site is located in Beijing Town, Dahua Yao Autonomous County, Hechi City, Guangxi Zhuang Autonomous Region, China. Four complete human skulls were found in coffins that had horn-shaped tails and heads (Peng, 2013a). We sequenced four individuals from Balong. From our kinship analysis, we found that two of the sampled individuals share kinship with another individual: BalongKD06 shows first-degree kinship with BalongKD10, and BalongKD08 shows second-degree kinship with BalongKD10 (Table S1). We thus excluded BalongKD06 and BalongKD08 from population genetic analyses, keeping only the unrelated BalongKD07 and BalongKD10 for further analysis. We directly radiocarbon dated BalongKD07 to 1,688-1,414 cal BP and BalongKD10 to 1,568-1,409 cal BP.

Lada - The Lada cave site is located in Jinchengjiang District, Hechi City, Guangxi Zhuang Autonomous Region, China. The human skulls found in this cave show the custom of tooth ablation (Peng, 2009). We sequenced genome-wide data for the temporal bone of the specimen LadaKH01, whom we directly radiocarbon dated to 1,467-1,307 cal BP.

Yiyang - The Yiyang (Bayang) cave site is located in Pingguo County, Baise City, Guangxi Zhuang Autonomous Region, China. There are 21 coffins in this site. One individual YiyangKP17 was sequenced, and directly radiocarbon dated to 1,467-1,307 cal BP.

Shenxian - The Shenxian cave site is located in Pingguo County, Baise City, Guangxi Zhuang Autonomous Region, China. The human skulls found in this cave show the custom of tooth ablation for eight of twelve individuals (Peng, 2009). We sequenced ShenxianKP09 from the temporal bone and directly radiocarbon dated the specimen to 1,350-1,278 cal BP.

Cenxun - The Cenxun cave site (Cenxundong) is located on Cenxun Mountain, Taiping Town, Pingguo County, Baise City, Guangxi Zhuang Autonomous Region, China. Cave burials of this site possibly extended from the Sui and Tang Dynasties to the early Ming Dynasty based on the archaeological evidence (Zhou and Tian, 1991). A secondary burial style was observed (Zhou and Tian, 1991), and three of six human skulls found in this cave show the custom of tooth ablation (Peng, 2013a). Our study sequenced three individuals and directly radiocarbon dated them: CenxunKP05 (1,467-1,307 cal BP), CenxunKP07 (1,366-1,293 cal BP) and CenxunKP13 (1,511-1,310 cal BP).

Gaofeng - The Gaofeng site is a cave located on Gaofeng Mountain, about 0.5 kilometers east of Huatu Village, Lihu Yaozu Town, Nandan County, Hechi City, Guangxi Zhuang Autonomous Region, China. There was a well-preserved coffin with traces of remaining lime daub. The buried individual was found lying in an extended side position and was an original burial (Zhang et al., 1986). The tooth ablation custom is found at very low frequency in Lihu Yaozu Town, Nandan County, suggesting that it was not practiced in Nandan County (Peng, 2013a). The specimen GaofengNL23 was sequenced and directly radiocarbon dated to 421 cal BP. Kinship analysis revealed that this individual share kinship to the second degree with an individual from the Huatayan cave site, HuatuyanNL04.

Huaqiao - The Huaqiao site, which has three caves, is located on Baitai Mountain, in Huaqiao Village, Lihu Yaozu Town, Nandan County, Hechi City, Guangxi Zhuang Autonomous Region, China. There are four coffins in Cave 1, 20 coffins in Cave 2, and seven coffins in Cave 3 (Zhang et al., 1986). We obtained genome-wide data from a tooth and a temporal bone belonging to the individual HuaqiaoNL26, found in Grave 4, Cave 2. We directly radiocarbon dated this individual to 514-428 cal BP.

Huatuyan - The Huatuyan site is a cave located on the hillside southeast of Huatu Village, Lihu Yaozu Town, Nandan County, Hechi City, Guangxi Zhuang Autonomous Region, China. There are 28 coffins with wooden frames. Each coffin contained two to four human individuals lying in a straight-limbed position (Zhang et al., 1986). We sequenced eight individuals in this study, and excluded three individuals for high kinship patterns with other sequenced individuals. HuatuyanNL04 in Grave 3 shares kinship to the second degree with HuatuyanNL17 and GaofengNL23. HuatuyanNL06 in Grave 5 shares kinship to the second degree with HuatuyanNL21 and HuatuyanNL17. HuatuyanNL18 in Grave 16 shares kinship to the second degree with HuatuyanNL21. The five individuals we used in our population genetic analyses are HuatuyanNL02 (NL02, Grave 2, 466-306 cal BP), HuatuyanNL11 (NL11, Grave 7, 477-312 cal BP), HuatuyanNL17 (NL17, Grave 15, 509-320 cal BP), HuatuyanNL19 (NL19, Grave 16, 455-294 cal BP), HuatuyanNL21 (NL21, Grave 18, 495-315 cal BP), all of whom were directly radiocarbon dated.

Yinwang - The Yinwang cave site (Yinwangdong) is located on Nian Mountain, in Liming Village, Liming Township, Pingguo County, Baise City, Guangxi Zhuang Autonomous Region, China. This individual shows a second-degree kinship with HuatuyanNL02, suggesting that this individual was possibly contemporaneous to Huatuyan individuals around 500 years ago. Only 12,700 SNPs were successfully sequenced from this individual, so this kinship is also possibly due to the low number of SNPs available for analyses. We excluded Yinwang from further population genetic analysis.

3 METHOD DETAILS

3.1 Ancient DNA extraction, sequencing, and data processing

The ancient DNA work was carried out in dedicated ancient DNA clean-room facilities at the Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, IVPP-CAS. For each of 170 ancient human remains from Guangxi, China (Table S1), we drilled powder either from the petrous portion

of the temporal bone (Pinhasi et al., 2015; Pinhasi et al., 2019) or from a tooth. Using the bone or tooth powder, we extracted DNA following a previously published protocol (Dabney et al., 2013). A single-stranded protocol ("SS") (Dabney et al., 2013; Gansauge and Meyer, 2013; Meyer et al., 2012) was used to prepare the libraries for all samples. We treated eight libraries with uracil-DNA-glycosylase (UDG) from *E. coli* and endonuclease (Endo VIII) ("SS UDG") to remove deaminated cytosine residues (Briggs et al., 2007) (Table 1). Library amplifications were performed using the AccuPrimepfx DNA enzyme, for 35 cycles (Yang et al., 2020).

To capture DNA in solution, we used oligonucleotide probes synthesized by Agilent Technologies (California, USA). Mitochondrial DNA (mtDNA) was captured using oligonucleotide probes synthesized from a complete human mitochondrial genome (Fu et al., 2013a). The nuclear genome was enriched for approximately 1.2 million SNPs (Fu et al., 2015).

The enriched mitochondrial DNA libraries were sequenced on Illumina Miseq instruments with 2×76 base pairs (bp) paired-end reads, and the enriched nuclear DNA libraries were sequenced on the Illumina Hiseq4000 instruments with 2×100 bp and 2×150 bp paired-end reads. We then utilized *leeHom* (Renaud et al., 2014) (https://github.com/grenaud/leeHom) to trim adaptors and merge paired reads into a single sequence (overlap > 11 base pairs). Merged reads at least 30 bp in length were then mapped to the revised Cambridge Reference Sequence (rCRS) (Andrews et al., 1999) (for mtDNA), and to the human reference genome hg19 (for nuclear DNA) with the Burrows-Wheeler Aligner (BWA, version 0.6.1) (Li and Durbin, 2009) using the *samse* command (-n 0.01 and -l 16500). For duplicated reads with the same orientation, we kept the highest quality sequence for analysis and removed the duplicates, along with reads with mapping quality scores less than 30.

To ensure the authenticity of ancient DNA, we calculated the C-to-T deamination proportion (Sawyer et al., 2012) (Table 1). We estimated the contamination rates based on mtDNA and X chromosome contamination rates. The mtDNA contamination rate was determined by ContamMix (Fu et al., 2013b). For males, we tested contamination for the X-chromosome (Korneliussen et al., 2014). For both methods, if the contamination was >3%, the library was treated as contaminated (Table 1). For libraries with low contamination of the nuclear DNA as determined by the X-chromosome, but slightly higher contamination of the mtDNA, we used all fragments for further analysis.

For those libraries with substantial contamination (>3% nuclear DNA), we restricted our analyses to only the fragments having characteristics typical of ancient DNA damage in order 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→T substitution in the first three positions at the 5'-end and the last three positions at the 3'-end by using *pmdtools0.60* (Skoglund et al., 2014) with the --customterminus parameter. These libraries were referred to as damage-restricted libraries in Table 1.

We ignored the first and last five positions of each fragment and generated pseudo-haploid genotype calls by randomly sampling one fragment per position to determine an allele for that individual (Fu et al., 2015).

4 QUANTIFICATION AND STATISTICAL ANALYSIS

4.1 Present-day Datasets

We used two panels of present-day datasets. For PCA, ADMIXTURE, and f_3-statistics, we took populations from the Human Origin (HO) SNP Panel (Patterson et al., 2012), Tibetan and Han populations from Lu et al. (Lu et al., 2016), Southeast Asian populations from Liu et al. (Liu et al., 2020), and populations in southern China from Wang et al. (Wang et al., 2021). For f_4-statistics and qpAdm analysis, we assembled the panel of 1240k capture SNPs from the Simons Genome Diversity Panel (SGDP) (Mallick et al., 2016), the Human Genome Diversity Project (HGDP)-shotgun data (Li et al., 2008), and Tibetan populations from Lu et al. (Lu et al., 2016).

4.2 Relatedness analysis

The degrees of kinship among newly sampled individuals were estimated using the software READ (Monroy Kuhn et al., 2018), which was developed specifically to handle pseudo-haploid genotypes for

prehistoric populations. We kept all the unrelated individuals for subsequent analyses. For each set of individuals sharing kinship, we determined the number of SNPs that were successfully sequenced and kept the individual from that kinship set with the highest number of SNPs available for analysis. We ultimately excluded seven individuals (Table S1) using this kinship criteria, leaving 23 Guangxi individuals which we used for subsequent population genetic analyses.

4.3 Principal components analysis

Principal components analysis (PCA) was performed with the smartpca program of the EIGENSOFT package (Patterson et al., 2006) using default options except lsqproject: YES, numoutlieriter: 0 and shrinkmode: YES for all present-day East Asians. All the newly sampled ancient individuals and the previously published ancient Asians were projected onto the PCA determined for present-day East Asians (Figure S1C). To increase resolution, we visualized estimated principal components for southern East Asians and Southeast Asians (Figure 1C).

4.4 *f*-statistics

We used the software qp3Pop (version 412) and qpDstat (version 712) in AdmixTools (Patterson et al., 2012) to calculate the f_3- and f_4-statistics, respectively. For qpDstat, we used "f4mode: YES". If the number of individuals in a group was greater than one, we used frequency data to calculate f-statistics (Table 1); otherwise, we used a 0/1 count (Durand et al., 2011). Outgroup f_3-analysis (Raghavan et al., 2014) had the form $f_3(Mbuti; X, Y)$, and f_4-statistics had the form $f_4(Mbuti, X; Y, Z)$, where the present-day Central African Mbuti are used to represent an outgroup.

4.4.1 Genetic clustering among new samples

Our methodology to determine which individual samples could be grouped together used a combination of outgroup f_3, PCA, and f_4 comparisons. We computed outgroup-f_3 statistics of the form f_3(X, Y; Mbuti) to measure the shared genetic drift between newly sampled Guangxi individuals (Figure S1A). To differentiate each individual, we used the "Individual ID" in Table 1. We found that the three prehistoric individuals could not be clustered, and the historical individuals formed three clusters, with three individuals that could not fit into any cluster.

Using these results, we re-categorized our individuals into one of nine new IDs, as follows:

- "Longlin": Longlin_1
- "Dushan": Dushan4_1
- "Baojianshan": Baojianshan5_M1
- "LaCen": LadaKH01, CenxunKP07, CenxunKP13
- "BaBanQinCen": BalongKD10, BalongKD07, BandaKD11, BandaKD15, QinchangKD13, QinchangKD14, CenxunKP05
- "GaoHuaHua": GaofengNL23, HuaqiaoNL26, HuatuyanNL02, HuatuyanNL17, HuatuyanNL11, HuatuyanNL19
- "Shenxian": ShenxianKP09
- "Yiyang": YiyangKP17
- "Layi": LayiKD01

The three clusters of historical individuals were labeled BaBanQinCen, LaCen, and GaoHuaHua. The eight BaBanQinCen individuals and three LaCen individuals date to ~1,500 years ago, while the six GaoHuaHua individuals date to ~500 years ago. GaoHuaHua individuals dating to about 500 years ago cluster with each other and differentiate from other historical individuals who date to about 1,500 BP, showing a population shift occurred between 1,500 years ago to 500 years ago. This grouping is consistent with the results of a PCA (Figure 1C), where the GaoHuaHua cluster differentiates from other historical individuals. Other historical individuals are located near each other in the PCA but are slightly differentiated.

To further confirm our choice of clustering, we checked the pairwise f_4-statistics in $f_4(Mbuti, various\ populations; Individual\_X, Individual\_X)$ in Table S2. In all cases, we can separate historical individuals into two major clusters based on time (~500-year-old and ~1500-year-old). To avoid genetic structure across individuals influencing a cluster, we used clear separations in pairwise outgroup-f_3 statistics (Figure S1A) to

form several subgroups within the temporal clusters. Those with low genetic similarity to each other were not grouped together, and in some cases, single individuals were kept separate in analyses.

4.4.2 Prehistoric populations in this study

4.4.2.1 Population relationships with Deep Asians

In the PCA, Longlin is the most centrally located of the newly sampled individuals and does not cluster with any particular present-day East Asians (Figure S1C). In an outgroup-f_3 analysis (Figure S1D), we found that both Longlin and Baojianshan do not share high genetic similarity with East Asians sampled to date.

To test how deeply the lineages of prehistoric samples diverged among sampled Asian populations, we compared them to a set of "Deep Asians", who diverged deeply in the Eastern Eurasian lineage. The "Deep Asians" set includes the 40,000-year old Asian Tianyuan (Yang et al., 2017), the present-day Andaman islander Onge, the New Guinea indigenous Papuan, an ~8,000 year-old Southeast Asian Hòabìnhian hunter-gatherer (La368) (McColl et al., 2018), and ~3,000-year-old prehistoric individuals from Japan (Ikawazu (McColl et al., 2018) and Japan_Jōmon (Wang et al., 2021)).

We calculated two sets of f_4-statistics, *f_4(Mbuti, aGX; X, Deep Asians)* (Table S2), and *f_4(Mbuti, X; aGX, Deep Asians)* (Table S2) where aGX are the three prehistoric Guangxi individuals Longlin, Baojianshan, and Dushan, and X includes various ancient East Asians representing different ancestries found across East Asia. Using these two statistics, we could determine whether our new samples share more alleles with ancient East Asians or with one or more of the Deep Asians. We found that Longlin, Dushan, and Baojianshan all cluster with Early Neolithic northern and southern East Asians who mainly contributed to present-day East Asians (Yang et al., 2020) rather than with Papuan, Onge, Hòabìnhian, and Tianyuan who possess a deep Asian lineage. This suggests that these three Guangxi individuals are not as deeply diverged from East Asians as these four Deep Asians.

We then compared prehistoric Guangxi populations with the Jōmon from Japan. Comparing to deep Asians (Tianyuan, Papuan, Onge), Longlin and Ikawazu are genetically closer to each other, i.e. *f_4(Mbuti, Longlin/Ikawazu; Ikawazu/Longlin, Tianyuan/Papuan/Onge)*<0 (-7.3<Z<-3.4, Table S2). Relative to the Jōmon, Dushan significantly clusters with Early Neolithic Shandong and Fujian East Asians, but both Longlin and Baojianshan show equal amounts of genetic similarity with these East Asians as with the Jōmon. Early Neolithic Shandong and Fujian East Asians are also similarly related to Longlin and Ikawazu, i.e. *f_4(Mbuti, EN_SD/EN_FJ; Longlin, Ikawazu)~0* (0.5<Z<2.2, Table S2). Both the Jōmon and Longlin have connections to the Early Neolithic Shandong and Fujian East Asians not found in the other population, i.e. *f_4(Mbuti, Longlin; Ikawazu, EN_SD/EN_FJ)>0* and *f_4(Mbuti, Ikawazu; Longlin, EN_SD/EN_FJ)>0*. We also observed a similar pattern for Baojianshan as found for Longlin, indicating that they are as deeply diverged from Early Neolithic Shandong and Fujian East Asians as the Jōmon, i.e. *f_4(Mbuti, Baojianshan; Ikawazu, EN_SD/EN_FJ)>0* (4.8<Z<7.3, Table S2), *f_4(Mbuti, Ikawazu; Baojianshan, EN_SD/EN_FJ)>0* (4.3<Z<6.3, Table S2), and *f_4(Mbuti, EN_SD/EN_FJ; Baojianshan, Ikawazu)~0* (-1.3<Z<0.5, Table S2). Thus, the separation of Longlin-related and Baojianshan-related ancestries from Early Neolithic Shandong and Fujian East Asians occurred more recently than that of Tianyuan-, Onge-, Papuan-, and Hòabìnhian-related ancestries. Longlin, Early Neolithic Shandong and Fujian East Asians, and the Jōmon, however, are similarly related.

To explore whether there is any shared affinity with Deep Asians contributing to the genetic distance between Longlin and Baojianshan from Neolithic East Asians, we computed *f_4(Mbuti, Deep Asian; Longlin/Baojianshan, Early Neolithic East Asian)*, where Early Neolithic East Asian includes some Northern-related ancestries from Siberia, Far East and Mongolia (E_N_northern-related: Shamanka_EN (de Barros Damgaard et al., 2018b), Lokomotiv_EN (de Barros Damgaard et al., 2018b), DevilsCave_N (Sikora et al., 2019), Boisman_MN (Wang et al., 2021), Mongolia_N_North (Wang et al., 2021) and Yumin (Yang et al., 2020)), and Early Neolithic Shandong East Asians (EN_SD: Bianbian (Yang et al., 2020), Boshan (Yang et al., 2020), Xiaojingshan (Yang et al., 2020), Xiaogao (Yang et al., 2020)) and Early Neolithic Fujian East Asians (EN_FJ: Qihe (Yang et al., 2020), Liangdao1 (Yang et al., 2020), Liangdao2 (Yang et al., 2020)). Interestingly, Longlin shows no affinity with any of the Deep Asians, as *f_4(Mbuti, Tianyuan/Papuan/Onge/Hòabìnhian; Longlin, Early Neolithic East Asian)~0* (-2.7<Z<0.3, Table S2). However, Baojianshan shows significant Hòabìnhian-related affinity in *f_4(Mbuti, Hòabìnhian; Baojianshan, E_N_northern-related East Asians)<0* (-4.1<Z<-3.2, Table S2). We further calculated this statistic using

transversions only, and we find that this pattern remains significant (-4.8<Z<-2.5, Table S2). This connection is consistent with results from qpAdm, qpGraph and Treemix in the later method sections, giving robust support to a genetic connection between Baojianshan and Hòabìnhian.

Several major Asian lineages have been described to date – one related to the geographically northern East Eurasian 40,000-year-old Tianyuan (Fu et al., 2013a; Yang et al., 2017) in northern East Eurasia, one related to the Onge and Hòabìnhians (McColl et al., 2018) in southern East Eurasia, and one that gave rise to the Early Neolithic Shandong and Fujian populations that have broadly contributed to present-day East Asians (Yang et al., 2020). The deep divergence of Tianyuan- and Hòabìnhian-related lineages from East Asian ancestry found today in mainland East and Southeast Asia shows that many diverse human groups were found in Asia. Here, we tested the relationship between Longlin and the geographically southern and northern East Eurasian deep lineages represented by Onge/Hòabìnhians and Tianyuan. We do not observe excess similarity of Longlin to either Onge or Tianyuan, i.e. f_4*(Mbuti, Longlin; Onge, Tianyuan)~0* (Z=0.3, Table S2). When we modeled the phylogenetic relationship between these populations, we found that Longlin has no Onge-related affiliation, and instead Longlin represents another deep lineage in southern East Eurasia (Figure 2A-2B). Substituting the present-day Onge with the ancient Hòabìnhian La368, we saw connections between Longlin and Hòabìnhians, i.e. f_4*(Mbuti, Longlin; La368, Tianyuan)~0* (Z=-2.7), f_4*(Mbuti, La368; Longlin, Tianyuan)<0* (Z=-3.7), and f_4*(Mbuti, Tianyuan; Longlin, La368)~0* (Z=-0.9). This pattern is not found using transversions only, i.e. f_4*(Mbuti, Longlin; La368, Tianyuan)~0*. (Z=-0.3), f_4*(Mbuti, La368; Longlin, Tianyuan)~0* (Z=-1.4), and f_4*(Mbuti, Tianyuan; Longlin, La368)~0* (Z=-1.1). Thus, the relationship between Tianyuan, Hòabìnhians, and Longlin is still unclear. However, combining the statistics above and different phylogenetic tools (Figure 2A-2B), we find that Hòabìnhians and Longlin definitively do not share the same ancestry. Thus, hunter-gatherers in southeastern Asia dating to the last 11,000 years are composed of at least two lineages – one related to Longlin and another related to the Onge and Hòabìnhians (Figure 2B).

4.4.2.2 Population relationships with Early East Eurasians

We next tested how these three prehistoric Guangxi populations compared to a set of "Early East Eurasians", namely Neolithic and Bronze Age populations of East Eurasia (Table S2), including those carrying northern East Asian-related (northern East Asians from Shandong (Yang et al., 2020), denoted as *EN_SD*), southern East Asian-related (southern East Asians from Fujian (Yang et al., 2020), denoted as *EN_FJ*), Tibetan-related (e.g. Chokhopani (Jeong et al., 2016)) , and Siberian-related ancestry (e.g. Kolyma (Sikora et al., 2019), Shamanka_EN (de Barros Damgaard et al., 2018b)). Due to the close relationship Native American ancestry shares with East Asian ancestry, we also compared against present-day populations and ancient individuals carrying Native American ancestry (Moreno-Mayar et al., 2018). The genetic relationships between Guangxi populations and this larger panel of Early East Eurasians are all performed with f_4-statistics and presented as tables. To have a better understanding of the genetic affiliation with East Asians, we focused on Early Neolithic northern East Asians from Shandong and southern East Asians from Fujian and surrounding regions.

Using f_4-statistics comparing prehistoric Guangxi individuals to the above East Eurasians and Native Americans, we find that both Longlin and Baojianshan behave as an outgroup relative to Early Neolithic Shandong and Fujian East Asians. That is, f_4*(Mbuti, EN_FJ/EN_SD; Longlin/Baojianshan, EN_SD/EN_FJ)>0* (2.3<Z<19, Table S2). For Dushan, the f_4-analysis does not clearly place Dushan as an outgroup to northern and southern East Asians as found for Longlin and Baojianshan (-1.5<Z<3.5, Table S2), though we observe that in an outgroup f_3-analysis, Longlin shares the highest genetic similarity with Dushan (Figure S1B). We further observe that Dushan shows significant affinity to Early Neolithic Shandong and Fujian East Asians but not with Native Americans, Siberians, and Plateau populations, i.e. f_4*(Mbuti, Dushan; EN_SD/EN_FJ, Native American/ancient Siberian/Plateau)<0* (-11.1<Z<-3.2, Table S2). These results highlight that Dushan shares more alleles with Early Neolithic Shandong and Fujian East Asians than what is observed for Longlin and Baojianshan, though Dushan does show some patterns similar to an outgroup.

In an outgroup f_3-analysis, we observe that Dushan shares higher genetic drift with southern East Asians (Figure S1D). To assess whether Dushan shares a connection with Early Neolithic Fujian populations specifically, we tested f_4*(Mbuti, Dushan; EN_FJ, EN_SD)~0* (-3.1<Z<-0.2), which shows that Dushan shows a slight affinity to these southern East Asians relative to Early Neolithic Shandong populations, but similarly related to both in most configurations. We also observe that f_4*(Mbuti, EN_FJ; Dushan, EN_SD)~0* (-

1.5<Z<1.5) and f_4*(Mbuti, EN_SD; Dushan, EN_FJ)~0* (0.3<Z<3.5, Table S2), which suggests that there is no clear affinity to southern East Asians relative to northern East Asians.

The newly sequenced individual from Fujian, Qihe3, clusters generally with southern East Asians (Yang et al., 2020). In particular, Qihe3 clusters closely with the other Qihe (Yang et al., 2020) individual relative to northern East Asians, i.e. f_4*(Mbuti, Qihe3; Qihe, EN_SD)<0* (-6.2<Z<-5.1, Table S2), and f_4*(Mbuti, Qihe; Qihe3, EN_SD)<0* (-6.1<Z<-3.7, Table S2). But interestingly, Qihe3 shows a less close relationship to the other Qihe individual relative to the other southern East Asians, i.e. f_4*(Mbuti, Qihe3; Qihe, Liangdao1/Liangdao2)~0* (-2.2<Z<-2.1, Table S2), and f_4*(Mbuti, Qihe; Qihe3, Liangdao1/Liangdao2)<0* (-2.1<Z<-2.0, Table S2).

4.4.2.3 Population relationships with southern East Asians and Southeast Asians since the Late Neolithic

Prehistoric Guangxi populations, especially Dushan, share the most genetic drift with southern East Asians, Southeast Asians, and the historical Guangxi populations (Figure 1C). Geographically, we see shared ancestry in both southern Chinese provinces of Fujian and Guangxi during the Neolithic. To determine if younger southern East Asians and Southeast Asians share more ancestry with Neolithic Guangxi or Fujian southern East Asians, we compared f_4*(Mbuti, younger populations; Bianbian, Qihe3)* to f_4*(Mbuti, younger populations; Bianbian, Dushan)*, where Qihe3 is from Fujian (~12 kBP), Dushan is from Guangxi, and Bianbian is an ancient northern East Asian (~9.5k BP), younger populations are southern East Asians and Southeast Asians dated since Late Neolithic and historical Guangxi individuals (Figure S2A). Among more recent populations, we find that coastal southern East Asians, Oceania Vanuatu, and island Austronesian populations from Southeast Asia (Group 6 (McColl et al., 2018)) are closer to Qihe3 than to Dushan. In contrast, ancient mainland Southeast Asians and historical Guangxi populations are closer to Dushan. Using f_4*(Mbuti, prehistoric GX; X, Qihe3)*, we confirmed directly that prehistoric Guangxi populations are closer to younger southern East Asians and Southeast Asians than Qihe3 (Figure S2B-S2C).

When comparing Late Neolithic populations to prehistoric Guangxi individuals, the Late Neolithic Southeast Asian farmer population Man_Bac (4,100-year-old individuals from Vietnam) shows the highest affinity with Dushan (Figure S2A). The strong affiliation between Man_Bac and Dushan is further supported in that they share significantly more alleles with each other than to Qihe3, i.e. f_4*(Mbuti, Man_Bac/Dushan; Dushan/Man_Bac, Qihe3)<0* (Z=-3.9 and -3.1, Figure S2B-S2C). Furthermore, both Dushan and Baojianshan show significantly more alleles with Man_Bac than with Qihe3, i.e. f_4*(Mbuti, Dushan/Baojianshan; Man_Bac, Qihe3)<0* (Z<-3.1, Figure S2B). However, while there is a similar pattern using transversions only, the comparison is no longer significant (Figure S2C). Despite the lack of concrete results from the f_4-analysis, we find other supporting evidence of a genetic affinity between Dushan and Man_Bac, through clustering in an outgroup-f_3 analysis (Figure S1D), and shared ancestry in qpAdm and qpGraph analyses.

In a previous study, Man_Bac was shown to possess a mixture of ancestry belonging to deeply diverged East Eurasians and East Asians (Lipson et al., 2018). In another study, Southeast Asian hunter-gatherer Hòabìnhian-related ancestry (McColl et al., 2018) was shown to be a deeply diverged East Eurasian lineage, which suggested that Man_Bac's deep ancestry was likely related to Hòabìnhian ancestry. Our results suggest that the deep lineage associated with Man_Bac is not related to Hòabìnhians, but rather to the deep Longlin-lineage found in Dushan in qpAdm and qpGraph section. We assessed affinity to Hòabìnhians relative to Early Neolithic East Asians, which includes Early Neolithic ancestries dating to before 7,000 years ago. These Early Neolithic populations include those with far northern-related ancestries (E_N northern-related ancestries: Shamanka_EN (de Barros Damgaard et al., 2018b), Lokomotiv_EN (de Barros Damgaard et al., 2018b), DevilsCave_N (Sikora et al., 2019), Boisman_MN (Wang et al., 2021), Mongolia_N_North (Wang et al., 2021) and Yumin (Yang et al., 2020)), northern-related ancestry from Shandong (Yang et al., 2020) (EN_SD), and southern-related ancestry from Fujian (Yang et al., 2020)(EN_FJ). We find that Man_Bac does not share excess ancestry with Hòabìnhians, i.e. f_4*(Mbuti, Hòabìnhian; Man_Bac, Early Neolithic East Asian)~0* (-1.6<Z<0.7, Table S2). Using qpAdm, we found that Man_Bac can be modeled as mixture of 65.8% Dushan-related ancestry and 34.2% of Longlin-related ancestry.

We also observe that populations contemporaneous with Man_Bac dating to 4,600-4,200 BP from Fujian (Xitoucun and Tanshishan) also show a significant affinity to Dushan and Baojianshan relative to Early

Neolithic southern East Asians, i.e. f_4*(Mbuti, Dushan/Baojianshan; Xitoucun/Tanshishan, Qihe3)<0* (Z<-3, Figure S2B). In a qpGraph analysis, Xitoucun could be modeled as a mixture of ancestry related to Longlin and Qihe (Figure S4E). We then used qpAdm to estimate the ancestry proportions in Xitoucun and Tanshishan, and we found that both are best modeled as a mixture of Dushan-related (34.8%-54.1%), Qihe3-related (8.2%-17%) and northern East Asian-related (34.4%-44.2%) ancestries, as well as a small amount of deep ancestry represented by IndusPeriphery populations (Narasimhan et al., 2019) (3.4%-3.9%, Table S2). In populations younger than 4,000 BP, the 2,000-year-old Nui_Nap in northern Vietnam and the 1,500-year-old BaBanQinCen in Guangxi (-4.2<Z<-3.1, Figure S2B) show similar patterns indicating affinity to Dushan and Baojianshan. Nui_Nap and BaBanQinCen cluster together in the PCA (Figure 1C) and f_3-statistics (Figure S1D), together with other historical populations dated to around 1,500 years ago in Guangxi. Using qpAdm, both populations can be modeled as a mixture of Dushan-related (~65%) and northern East Asian-related (~35%) ancestry.

None of the 4,000-year-old and younger southern East Asian and Southeast Asian populations described above show a significant connection to Hòabìnhians, i.e. f_4*(Mbuti, Hòabìnhian; Man_Bac/Xitoucun/Tanshishan/Nui_Nap/BaBanQinCen, Early Neolithic East Asian)~0* (-2.7<Z<2.1, Table S2), but they tend to share a connection with Dushan, who possesses deep ancestry related to Longlin. Longlin shows patterns consistent with some affinity to these southern East Asian and Southeast Asian populations (Figure S2B), although the patterns are not significant. Using transversions only, however, the connection between Dushan/Baojianshan and Xitoucun is the only connection that remains significant, though we still observe a consistent pattern for other populations (Figure S2C). Meanwhile, some younger Southeast Asians do possess deep ancestry related to Hòabìnhians, like Vt_G2, G3, and Vt778_G4_1, i.e. f_4*(Mbuti, Hòabìnhian; Vt_G2/G3/Vt778_G4_1, E_N_northern-related East Asians)<0* (-5.4<Z<-2.4, Table S2). This suggests that deep ancestry in younger Southeast Asians is diverse and complex, associated with either Longlin- or Hòabìnhian-related ancestry.

We further tested the connections between the younger southern East Asian and Southeast Asian populations and Dushan in f_4(Mbuti, Dushan; X, Qihe3). To correct for multiple comparisons and minimize the probability of type I errors, we carried out the Benjamini-Hochberg correction (Benjamini and Hochberg, 1995), using the function p.adjust in R. After the correction, we then used a p-value of 0.001 (correspond to Z-score of -3.09) as the significance threshold. We then converted adjusted p-values to Z-scores as a direct comparison (Table S2). After correction, Xitoucun (Z=-3.8), Tanshishan (Z=-3.3), TaiwanHanben (Z=-3.4), and BaBanQinCen (Z=-3.6) still shared significantly more alleles with Dushan.

To test whether the Guangxi ancestry shared an affinity with Late Neolithic Fujian populations, we performed similar analyses as above, but grouped Qihe, Qihe3, Liangdao1, and Liangdao2 as EN_FJ, and Tanshishan, Xitoucun as LN_FJ. We found a closer genetic affinity between Dushan and LN_FJ relative to EN_FJ in f_4*(Mbuti, Dushan; LN_FJ, EN_FJ)<0*, Z=-5.1. The result is still significant for transversions only (Z = -3.2). We found that in this case, f_4*(Mbuti, Longlin; LN_FJ, DevilsCave_N/Boisman_MN)<0*, Z=-3.1/-3.3 and f_4*(Mbuti, Longlin; EN_FJ, DevilsCave_N/Boisman_MN)~0*, Z=-0.9/-0.7. This shows that populations carrying Guangxi-related ancestry share connections to Late Neolithic Fujian populations that are not shared with Early Neolithic Fujian populations. However, we did not observe a significantly negative result for the direct f4-analysis, f_4*(Mbuti, Longlin; LN_FJ, EN_FJ)<0* (Z= -2.3). One possibility is that Longlin might have some admixture with EN_FJ that offsets any Longlin-LN_FJ connection. Another explanation is f_4*(Mbuti, Dushan; LN_FJ, EN_FJ)<0* (Z= -5.1)may be due to ancestry unrelated to Longlin, i.e. a third ancestry in Dushan that could not be observed directly with the analyses currently available.

4.4.2.4 Relationship with present-day populations

We projected ancient Guangxi populations onto the East Asian PCA, where we observe that the three prehistoric populations from Guangxi cluster near Austro-Asiatic speakers (Figure 1C). A similar connection was observed in a previous study (Zhang et al., 2017), where the mitochondrial (mtDNA) haplogroup of Longlin and another 11,201-11,079-year-old early human from Qingshuiyuan Dadong (Guizhou, China) were named as a new subhaplogroup, M71d (Bai et al., 2020). Longlin was located at the basal position on the lineage leading to M71d, sharing a maternal genetic connection with present-day populations from mainland Southeast Asia (Bai et al., 2020).

To have a better understanding of the affinities shown in present-day populations, we compared f_4*(Mbuti, X;*

Bianbian, Qihe3) to f_4*(Mbuti, X; Bianbian, Dushan)*. We found that present-day Austronesians are closer to Qihe3, while present-day Austro-Asiatic groups share more alleles with Dushan. Present-day populations belonging to other language groups have a moderate allele-sharing rate between Qihe3 and Dushan (Figure S2D). However, unlike ancient populations in Fujian, which show a close relationship to present-day Austronesians in f_4-statistics analyses suggesting shared ancestry (Yang et al., 2020), prehistoric populations in Guangxi do not share a significant affiliation with any language speakers, i.e. f_4*(Mbuti, prehistoric GX; present-day populations, Qihe3)~0* ($-2.6<Z<1.7$, Figure S2E).

4.4.3 Historical populations in this study

Above, we observed the connection between Dushan and the historical Guangxi populations (Figure S2A). Among historical populations, BaBanQinCen shows the strongest affiliation with both Dushan and Baojianshan ($Z<-3$), and others also show a closer relationship to Dushan than to Qihe3 (Figure S2B). Meanwhile, they also show high genetic similarity with southern and northern East Asians in outgroup-f_3 statistics (Figure S1D). Historical Guangxi populations fall within the genetic variation observed in southern East Asians and Southeast Asians. Looking more closely, they cluster with each other, and share high genetic drift with the southern East Asian Xitoucun, Tanshishan, and the Southeast Asian Nui_Nap (Figure S1D). In fact, all historical Guangxi individuals share more alleles with northern East Asians than prehistoric Guangxi individuals (Figure S1D).

4.4.3.1 Relationship between historical Guangxi samples and present-day populations

Among historical populations, we see individuals dating to 1,500 years ago cluster with each other and overlap with Tai-Kadai groups (Figure 1C). In contrast, the 500-year-old GaoHuaHua individuals cluster separately from the 1,500-year-old cluster but cluster closely with Hmong-Mien speakers in the PCA (Figure 1C). To further define the genetic relationships between historical Guangxi populations and present-day populations, we performed the outgroup-f_3 statistics using a panel of present-day populations representing many different language groups. The results show consistently that the ~500-year-old Guangxi populations cluster with the Hmong-Mien speakers (Figure S3A).

4.5 ADMIXTURE analysis

We pruned the HO dataset to account for linkage disequilibrium using ADMIXTURE (Alexander et al., 2009) and PLINK (Purcell et al., 2007) (v1.90b3.40) with parameters "--indep-pairwise 200 25 0.4". A model-based maximum likelihood (ML) clustering algorithm was implemented to estimate individual ancestries and determine population structure with cross-validation. We re-ran the software 100 times using different seeds for each value of K, and we presented K=4 to K=7 results in Figure S3B-S3C; the lowest CV is when K=4.

We observed that Austronesians all share a component (pink), which is also observed at high proportions in ancient southern East Asians and Vanuatu. Northern East Asians share a component (yellow), which is found in ancient northern East Asians and widely found in present-day East Asians. The deep lineages, such as G1 (Hòabìnhian), the Indus Valley ancestry Harappan, and Juang (an Austroasiatic-speaking group from India), all share a component (orange) – however, a close relationship is not observed between these populations (Shinde et al., 2019). We thus do not have high confidence that the orange component reflects shared ancestry. The Southeast Asian Mlabri have a separate component (blue), that can be found in some ancient and present-day Southeast Asians.

For K=4, we found that Longlin primarily contained deep ancestry (orange), similar to other deep ancestries, e.g. Hòabìnhian, Ikawazu. Longlin and Ikawazu harbor both northern (yellow) and southern (pink) East Asian-related ancestry components, which is consistent with their genetic relationship with East Asians mentioned above. Dushan and Baojianshan show deep ancestry (orange) mixed with Austronesian-related southern East Asian (pink) ancestry. The historical Guangxi populations show a similar genetic structure, i.e. primarily a southern East Asian (pink) ancestry with some deep ancestry (orange) and a small amount of northern East Asian ancestry (yellow). The more recent GaoHuaHua shows more northern East Asian components than other historical Guangxi populations (Figure S3C).

4.6 Inferring admixture and estimating mixture proportions

We applied qpWave (Meyer et al., 2012) and qpAdm (Haak et al., 2015) to infer ancestral sources and estimate admixture proportions for admixed populations. In all analyses, we used all SNPs (allsnps: YES). The strategy we adopted is as follows: (1) We considered all new samples and previously published ancient and present-day Southeast and southern East Asians as potential target populations by running one-way, two-way, three-way, four-way models to fit their ancestry. (2) We began with an outgroup set of distantly related populations to these potential targets, denoted "Fixed rightgroups". (3) We assigned potential source populations, with some as a fixed source ("Fixed leftgroups"), and some that rotated through as a possible source ("Rotating populations"). We used these to run combinations of one-, two-, three-, and four-way models. When populations were treated as a target population, we did not include this population in the potential set of sources ("Fixed leftgroups") for those analyses. (4) For the "Rotating populations", if a population could not fit as a potential source or target population, we systematically added that population to the outgroup set. Individuals from the "Fixed leftgroups" set were never included into the outgroup set.

Rotating potential sources into the outgroups ("Fixed rightgroups") increases the ability to identify optimal admixture models, and avoid the effects of more recent gene flow. "Rotating populations" are those that share different degrees of relationship with the "Fixed rightgroups". Individuals in "Fixed Leftgroups" are either:

(1) those that are genetically related to another population in the "Rotating populations" but more recently dated (e.g. Boshan, who belongs to "Fixed Leftgroup" is genetically close to Bianbian, a "Rotating population", and Liangdao2, who belongs to "Fixed Leftgroup", is genetically close to Qihe3, also a "Rotating population") or
(2) populations that have a deep lineage but date from a fairly recent period (e.g. Jōmon).

Fixed rightgroups: Mbuti, UstIshim (Fu et al., 2014), Kostenki14 (Lazaridis et al., 2016; Seguin-Orlando et al., 2014), Iran_N (Lazaridis et al., 2016), Yana (Sikora et al., 2019), Papuan (Mallick et al., 2016), Onge (Mallick et al., 2016), Tianyuan (Yang et al., 2017), Clovis (Posth et al., 2018), Shamanka_EN (de Barros Damgaard et al., 2018a), Yumin (Yang et al., 2020)

Rotating populations: Longlin, Dushan, Qihe3, DevilsCave_N (Sikora et al., 2019), Kolyma (Sikora et al., 2019), Bianbian (Yang et al., 2020), IndusPeriphery (merged Gonur2_BA and Shahr_I_Sokhta_BA2 from (Narasimhan et al., 2019)), Hòabìnhian (the 7,950-7,795 cal BP individual La368 from (McColl et al., 2018))

Fixed leftgroups: Boshan (Yang et al., 2020), Liangdao2 (Yang et al., 2020), Jōmon (merged Ikawazu and Jōmon (McColl et al., 2018; Wang et al., 2021))

Applying the strategy described above, we start with one-way modeling and then proceed to higher ranks, up to four-way modeling. Using 'n' to refer to number of source populations, we considered $p>0.05$ to indicate that the n-source model is possible. A second p-value (p-nest) was determined by comparing the n-source model with the n-1-source model with the highest p-value. A $p\text{-nest}<0.05$ indicates that the higher ranking n-source model is significantly better than the n-1-source model, so the best fitting model is one that includes n-sources (Yang et al., 2020). Below, we highlight the highest-ranking n-source model where $p>0.05$ and $p\text{nest}<0.05$, and we do not report models that do not show this fit to the data.

4.6.1 Prehistoric Fujian populations

Qihe3

In a previous study (Yang et al., 2020), Liangdao1 was shown to have more northern East Asian ancestry than Liangdao2 and Qihe. To better understand differences among Early Neolithic Fujian populations, we considered Qihe3, Qihe (Qihe2) (Yang et al., 2020), and Liangdao2 (Yang et al., 2020) each as a potential target population. We also allowed Liangdao2 and Qihe3 to be a source population when not used as a target population. We did not use Qihe as a potential source population as Qihe possesses a lower number of SNPs (328,913) than Qihe3 (616,335).

In one-way modeling, Liangdao2 and Qihe3 can be modeled as the source for each other, and Qihe can be modeled with Qihe3 as the source (Table S3). We then tested 2-way modeling, and we found that Liangdao2 is best modeled as a mixture of northern East Asian ancestry (e.g. Bianbian, Boshan, DevilsCave_N, 10-18%) and Qihe3-related ancestry (82-90%, $p>0.05$, $p\text{nest}<0.05$, Table S3), indicating Liangdao2 has more

northern East Asian influence than Qihe3. Qihe3 can be modeled as a mixture of ancestry related to East Asians (e.g. Boshan, Liangdao2) and a population of deeper ancestry (e.g. Longlin, IndusPeriphery), possibly indicating that Qihe3 contains a deep lineage that Liangdao2 does not share or is diluted below the sensitivity of these tests. The 2-way models for Qihe are not a better fit than the 1-way model in Table S3.

4.6.2 Prehistoric Guangxi populations

Longlin

The oldest prehistoric Guangxi individual Longlin shares little genetic similarity with ancient and present-day East Asians in outgroup-f_3 statistics and f_4-statistics. Also, there is no evidence to support that any deeply diverged Asian ancestry previously sampled shares affinity with Longlin relative to other East Asians. Because we found no supporting evidence of admixture in Longlin through other analyses, we did not test Longlin as a potential target.

Dushan

In outgroup-f_3 statistics and f_4-statistics we found Dushan has connections to Longlin, but shares more alleles with southern East Asians. To estimate admixture proportions for Dushan, we treated Dushan as a potential target. We find that Dushan cannot be modeled using a single source. In a two-way model, Dushan can be modeled as a mixture of ancestry related to Longlin and Liangdao2 (p=0.47, Table S3), consistent with previous analyses showing an affiliation with Longlin and southern East Asians.

Baojianshan

Baojianshan can be modeled in a one-way model with ancestry related to Longlin (p= 0.41, Table S3). However, a two-way model is a significantly better fit than a one-way model (pnest<0.05). In this two-way model, Baojianshan is significantly better modeled as a mixture of Dushan-related (72%) and Hòabìnhian-related (28%) ancestries (Table S3). This is consistent with f_4-statistics where Baojianshan has an affiliation with both Dushan and the Hòabìnhian.

4.6.3 Historical Guangxi populations

Layi

In a one-way model, Layi shares ancestry with Liangdao2 (p = 0.11, Table S3). In a two-way model, however, Layi is significantly better modeled as a mixture of Boshan-related ancestry (22-27%) and either Longlin-related (78%) or Dushan-related (73%) ancestry (Table S3). Thus, Layi possesses ancestry found in prehistoric Guangxi individuals, with an additional 22% to 27% northern East Asian ancestry.

Shenxian

Shenxian can only be modeled with two sources. Shenxian is best modeled as a mixture of northern East Asian-related ancestry (9%-22%) and southern East Asian-related ancestry (78-91%, Table S3), suggesting Shenxian also has a northern East Asian component. Unlike Layi, a prehistoric Guangxi population is not needed to model Shenxian's ancestry.

Yiyang

Yiyang can be modeled using a single source when using Liangdao2 (p=0.10, Table S3). Like Shenxian, Yiyang can be modeled as a mixture of northern East Asian (18%-42%), and southern East Asian (Liangdao2, 58%-83%) ancestry. However, Yiyang can also be modeled as a mixture of northern East Asian-related ancestry (27%-42%) and Dushan-related ancestry (58%-73%, Table S3).

BaBanQinCen

Only three-way models show feasible combinations for BaBanQinCen, where BaBanQinCen can be described as a mixture of ancestry related to Dushan (5%-64%), northern East Asians (19%-40%) and southern East Asians (5%-72%, Table S3). Like for other historical Guangxi populations, this is consistent with an affiliation to the admixed Dushan (southern East Asian and Longlin-related ancestries) and northern East Asians observed in other analyses. The proportions of Dushan- and southern East Asian-related (e.g. Qihe3, Liangdao2) ancestry vary, possibly because qpAdm cannot easily differentiate the southern East Asian ancestry found in Dushan from that found in coastal southern East Asians.

LaCen

LaCen, like Layi, can only be modeled using a 2-way approach, where LaCen is best modeled as a mixture of northern East Asian ancestry (22%-30%) and Dushan-related ancestry (70%-78%, Table S3).

GaoHuaHua

GaoHuaHua, like LaCen and Layi, is also best modeled through a 2-way approach. The best model is one where GaoHuaHua is a mixture of northern East Asian ancestry (Boshan, 34%) and Dushan-related ancestry (66%, Table S3).

4.6.4 Ancient southern East Asians and Southeast Asians

We targeted previously published ancient southern East Asians and Southeast Asians, to estimate the genetic contribution of prehistoric Guangxi populations in the neighboring region. Since the data quality would influence the power of the modeling, here we reported results for populations with greater than 100,000 SNPs.

We found that applying a two-source model to ancient southern East Asians and Southeast Asians led to feasible admixture models that were significantly better than one-way models with either Longlin- or Hòabìnhian-related ancestry. In the mixture models, G4, La_G2, Ma912_G2, Oakaie1, Vt_G2, Vt778_G4_1 are best modeled as a mixture of Longlin-related ancestry (57%-94%) and northern East Asian-related ancestry (6%-43%). In addition, Oakaie1, Vt778_G4_1, Nui_Nap, Chuanyun, and G3 can be modeled as mixture of Dushan-related ancestry and northern East Asian ancestry. Finally, Ma912_G2 and Vt_G2 can also be modeled as a mixture of Longlin-related ancestry and southern East Asian-related ancestry (Qihe3, Liangdao2).

Interestingly, Man_Bac, who we found in f_4-analyses to have a connection to prehistoric Guangxi individuals, is best modeled as a mix of Dushan-related ancestry (65.8%) and Longlin-related ancestry (34.2%, Table S3), with no ancestry specific to southern or northern East Asians. Man_Bac can also be fit as primarily Dushan-related ancestry with some IndusPeriphery-related ancestry (3.5%, Table S3), which suggests that Dushan-related ancestry is the primary contributor to Man_Bac.

Vt_G2 also can be modeled as a mixture of Dushan-related ancestry (19%) and Longlin-related ancestry (81%, Table S3). However, like the connection with Hòabìnhian we see in f_4-statistics, Vt_G2 also can be modeled as containing 9%-19% Hòabìnhian-related ancestry admixed with 81%-91% southern East Asian-related ancestry (Liangdao2, Qihe3, Table S3). Similarly, G5 is best modeled as a mix of southern East Asian-related ancestry (Qihe3, 67%) and Hòabìnhian-related (34%, Table S3), which is consistent with the contribution of Hòabìnhians mentioned in a previous study (McColl et al., 2018).

We did not observe a plausible three-way model for any ancient southern East Asians or Southeast Asians with Longlin or Dushan as a potential source, but some four-way models were possible and significantly better than lower-ordered models. Both Late Neolithic Fujian populations Xitoucun and Tanshishan are best modeled as a mixture of Dushan-related ancestry (35/54%), northern East Asian ancestry (44/34%), Qihe3-related ancestry (17/8%) and IndusPeriphery-related ancestry (4/3%, Table S3).

4.6.5 Present-day East Asians and Southeast Asians

Of present-day populations, only Mlabri, Cambodian, Thai, and Burmese can be modeled as having ancestry related to prehistoric Guangxi individuals. Cambodian, Mlabri, and Thai can be modeled as a three-way mixture of Longlin-related, Liangdao2-related, and DevilsCave_N-related ancestries (Table S3). Mlabri can also be modeled as a mixture of Dushan-related, DevilsCave_N-related, and Hòabìnhian-related ancestries. For the Burmese, we observe that the best model uses four sources – Dushan or Longlin, DevilsCave_N, Hòabìnhian, and IndusPeriphery or Jōmon (Table S3).

Based on the qpAdm results, we can see that prehistoric Guangxi populations profoundly influenced later populations, though this contribution is considerably less in later periods. In fact, several previously published Southeast Asians primarily associated with the deep lineage related to Hòabìnhians (McColl et al., 2018) can be better described as a mixture of northern East Asian, southern East Asian, and prehistoric

Guangxi ancestry. Furthermore, some previously published Southeast Asians with deep ancestry (Lipson et al., 2018), such as Man_Bac, can be better modeled with prehistoric Guangxi-related ancestry as the source, rather than Hòabìnhian-related ancestry. For present-day populations, the contributions of both prehistoric Guangxi-related and Hòabìnhian-related ancestries are limited. In summary, the qpAdm analyses here reveals the diverse and complicated genetic picture in southern China and Southeast Asia, and the important role Guangxi ancestry played within this region.

4.7 Admixture Graph modeling

We modeled the relationship between populations using qpGraph in ADMIXTOOLS (Patterson et al., 2012), with allsnps:YES. To build an Admixture Graph Model, we added samples chronologically, where each sample's best fitting node or set of two nodes (admixture) are cataloged. Then, the set of best fitting models including that sample is used as the base graph for adding the next set of samples. We began with a basic model that included the Central African Mbuti, the early European Kostenki14 (Seguin-Orlando et al., 2014), the early Asian Tianyuan (Yang et al., 2017), and the 7,950-7,795 year old Hòabìnhian Hunter-gatherer La368 (McColl et al., 2018) (denoted as G1, Figure S4A).

(1) Adding the Paleolithic East Asian: Longlin (10,686-10,439 Cal BP)

We first added the Late Paleolithic individual Longlin, and found the only feasible model positioned Longlin on the East Eurasian lineage, with Tianyuan. Even though the tree showed limited shared ancestry between Hòabìnhian and Longlin (Figure S4B), it is also likely the relationship is defined by a polytomy as suggested in the f_4-analysis using transversions only.

(2) Adding Neolithic East Asians

We then added different Neolithic East Asians in turn to the graph in Figure S4C. Following chronologically, we first added those samples dating to before 8,000 years ago: the northern East Asian Boshan (~8,300 BP (Yang et al., 2020)), the southern East Asian Qihe (~8,400 BP (Yang et al., 2020)), and Dushan (8,974-8,593 cal BP), presented in this study. We show all feasible models (maximum |Z|<3) in Figure S4C.

- Dushan can be modeled in two ways: a mixture of a lineage related to Tianyuan and a lineage related to Longlin (Figure S4C) or a mixture of Longlin-related ancestry and southern East Asian Qihe-related ancestry (Figure S4C).
- Boshan can be modeled in two ways. One is clustering with Qihe, where their common ancestry was derived from Longlin-related and Tianyuan-related ancestry (Figure S4C). In the second, Boshan received ancestry from a Dushan-related or Qihe-related lineage, and from a population that is deeply diverged (Figure S4C).
- Qihe in most cases clusters with Boshan (Figure S4C), but in some cases Qihe can be modeled as a mixture of a Longlin-related lineage and a Dushan-related lineage (Figure S4C). In Figure S4C, Qihe can be modeled as a mix of a Tianyuan-related ancestry and Longlin-related ancestry. Similarly, Qihe can be modeled as a mixture of Dushan-related ancestry with a Tianyuan-related ancestry.

Then, we added two later ancient East Asians dating to between 8,000 to 5,000 BP: Liangdao2 (~7,600 BP) (Yang et al., 2020) and Baojianshan (8,335-6,400 BP). We found eight models that fit the observed patterns (maximum |Z|<3, Figure S4D).

- In most models (Figure S4D), Liangdao2 forms a clade with Qihe, consistent with previous findings (Yang et al., 2020). Liangdao2 can also be modeled as admixture of Qihe- and Boshan-related ancestry (Figure S4D), indicating Liangdao2 received more northern East Asian influence than Qihe, which is consistent with the qpAdm analysis.
- Baojianshan can be modeled in three ways. First, Baojianshan can be fit as a mixture of ancestry related to Dushan and the Hòabìnhian G1 (Figure S4D). Second, Baojianshan can be described as sharing common ancestry with both northern and southern East Asians, but separating prior to the northern and southern divergence (Figure S4D). Third, Baojianshan can be modeled as a mixture of Longlin-related ancestry and the shared northern and southern East Asian ancestry (Figure S4D).

To these graphs, we next added Late Neolithic individuals who date to around 4,000 BP, i.e. Man_Bac (~4,100 BP (Lipson et al., 2018)) and Xitoucun (~4,600 BP (Yang et al., 2020)). We found two models that

fit the data well with maximum |Z|<3 (Figure S4E).

- Xitoucun can be modeled as a mixture of Longlin-related and Qihe-related ancestry in both models (Figure S4E).
- Man_Bac can be modeled as a mixture of Dushan-related ancestry and a southern East Asian-related ancestry. In some cases, Man_Bac is modeled directly as receiving ancestry from a southern East Asian-related population (Xitoucun, Figure S4E). In other cases, Man_Bac is modeled as receiving ancestry from an admixed lineage related to Dushan, where Dushan always has a connection to southern East Asians (Figure S4E).

We provide possible models here without explicitly supporting a given model as the most accurate. Using these models, we summarize patterns regarding the complicated ancient genetic history of East Asia.

- Longlin fits as a separate lineage sharing limited ancestry with the Hòabìnhian (Figure S4B). However, it is more likely the relationship was a polytomy based on an f_4-analysis using transversions only. Longlin shares a closer relationship with later East Asian populations than with deep Asians (Tianyuan and Hòabìnhians, Figure S5).
- Dushan can be predominantly modeled as a mixture of Longlin-related and southern East Asian-related ancestry (Figure S4C).
- Baojianshan fits as a mixture of Dushan-related and G1-related ancestry, consistent with results from f_4-analysis (Figure S4D).
- Liangdao2 mostly clusters with Qihe. However, in some cases Liangdao2 can be modeled as mixture of northern East Asian-related ancestry and Qihe-related ancestry, indicating a difference between Qihe and Liangdao2 (Figure S4D).
- Man_Bac can be modeled as a mixture of Dushan-related and Qihe-related ancestry, consistent with the results described in f_4-analysis. Furthermore, Man_Bac does not show evidence of Hòabìnhian-related ancestry (Figure S4E).

4.8 Estimating a maximum likelihood phylogeny with migration events

The phylogenetic relationships were determined by Treemix v1.13 (Pickrell and Pritchard, 2012). We rooted the tree by the Central African Mbuti, made blocks of 500 SNPs and used global rearrangements, i.e. the parameters "-root Mbuti –k 500 -global" were used, allowing 0, 1, 2, or 3 migration events (m). We ran 1,000 replicates for each tree, adding the options "-bootstrap -q". And the bootstrap trees were assessed in phylip with the command "consense" (Baum, 1989). Results are shown in Figure S5A-S5D for m=0 to m=3, and m=3 is shown in Figure 2A with a visualization of the residuals in Figure S5H. Here, the Hòabìnhian La368 is represented by the label G1.

When $m = 0$, Longlin is an outgroup to southern East Asians, northern East Asians, Dushan, and Baojianshan. Longlin clusters with them relative to Tianyuan and the Onge/G1 clade (Hòabìnhian-related clade). Baojianshan clusters with Dushan. After allowing $m = 1$, a migration event occurred between Tianyuan and the northern East Asians DevilsCave_N and Yumin. When $m = 2$, both Dushan and Baojianshan received gene flow from a Longlin-related population. When $m = 3$, Baojianshan received gene flow from the Hòabìnhian-related clade.

4.9 Northern East Asian influence on historical Guangxi samples

In qpAdm analyses, we found historical Guangxi populations possess partial northern East Asian ancestry, with the mixture proportion estimated to ~20% in qpAdm. To understand which northern East Asians sampled thus far best represent the source population(s), we compared historical Guangxi populations with previously published ancient northern East Asians: Early Neolithic Shandong individuals (Yang et al., 2020), Neolithic Mongolians (Wang et al., 2021), Primorye populations (Sikora et al., 2019), Amur River populations (Ning et al., 2020), West Liao River populations (Ning et al., 2020), and Central Plain populations (Ning et al., 2020).

We first calculated the outgroup f_3-statistic, $f_3(X, Y; Mbuti)$ to measure the shared drift between historical Guangxi populations and ancient northern East Asians listed above. We found that of all ancient northern East Asians tested, northern populations from Early Neolithic Shandong and Central Plain populations share the most genetic drift with historical Guangxi populations. These Shandong populations date to ~9,500-

8,000 years ago, representing the oldest samples among northern East Asians who share high genetic drift with historical Guangxi populations (Figure 3B).

To confirm the connection between historical Guangxi (GX) populations with ancient Shandong populations, we performed f_4*(Mbuti, historical GX; X, Y)* where Y are ancient Shandong populations and X are all other ancient northern East Asians (Table S2). This comparison allows us to assess in which specific instances historical Guangxi populations share more alleles with the ancient Shandong populations than other ancient northern East Asians. Our results show that the historical Guangxi populations share a closer relationship to ancient Shandong populations than ancient northern East Asians from the Amur River region, West Liao River region, Coastal Siberia, and Mongolia, as most f_4*(Mbuti, historical GX; X, ancient Shandong)*>0 ($-1.1<Z<10.6$, Table S2).

In particular, the strongest affiliation is with the 7,900-year-old Xiaojingshan population from Shandong, who shows a connection to the historical Guangxi individuals even relative to other ancient Shandong populations, i.e. f_4*(Mbuti, historical GX; Xiaojingshan, other ancient Shandong)* tends negative, and is significantly negative when the historical Guangxi population is the 500-year-old Gaohuahua (Table S2). When compared with ancient populations from the Central Plain region, we found that most f_4*(Mbuti, historical GX; ancient Central Plain, ancient Shandong)*~ 0 (Table S2), suggesting that ancient Central Plain populations and ancient Shandong populations are similarly related to historical Guangxi populations. The Central Plain populations are younger than the Shandong populations, and they also show evidence of southern East Asian-related ancestry (Ning et al., 2020). Thus, the northern East Asian ancestry most associated with historical Guangxi populations is that related to Early Neolithic Shandong individuals and ancient populations from the Central Plain.

4.10 Archaic ancestry estimation

To estimate introgressed archaic fragments in ancient Guangxi populations, we used *admixfrog* (Peter, 2020) (version 0.5.6, https://github.com/BenjaminPeter/admixfrog/). Admixfrog is capable of inferring introgressed segments from highly degraded and contaminated data (Peter, 2020). Using this software, we modeled target individuals as a mixture of three different sources: two high-coverage Neanderthal genomes (the high coverage Altai (Prüfer et al., 2014) and and Vindija (Prüfer et al., 2017) Neanderthal genomes, NEA), one high-coverage Denisova genome (Denisova 3, DEN, (Meyer et al., 2012)), and 44 genomes of present-day Sub-Saharan Africans from the Simons Genome Diversity Panel (Mallick et al., 2016) (AFR). We used the "1240k" SNP panel to infer the archaic introgressed fragments in the given target genome.

We first converted the target individuals from the BAM file format to the input file format for *admixfrog*, with the command '--length-bin-size 35 –minmapq 25 --deam-cutoff 3'. These parameters filter BAM files for fragments of at least 35 base pairs (bp), mapping quality greater than 25, and remove the deamination of C→T substitution at the first three and/or the last three bases. Then, using these input files, we ran the analysis to infer introgressed archaic fragments. The potential sources were set to be Africans, Neanderthals, or Denisovans (--states AFR NEA DEN). The chimpanzee (*panTro4*) reference genome was used to infer the ancestral state of each allele (--ancestral PAN). The bin size for every individual was set to 5,000 bp (--bin-size 5000). Other parameters were configured using default options (Peter, 2020).

Individuals with less than 200,000 SNPs (marked in gray) gave the highest and lowest archaic proportions, likely because their low number of SNPs skews the estimates for these samples. Other ancient Guangxi individuals with greater than 200,000 SNPs gave estimates ranging from 0.5-2.5% for Denisovan introgressed segments and 1.9-5.2% for Neanderthal introgressed segments (Table S2). None of the results indicated archaic ancestry above that which has been shown for similarly dated individuals from this region.

Although the cranial morphology of Longlin shows a mixture of archaic-related features, she does not have extra archaic ancestry proportions greater than that found in similarly dated samples with more typically modern human morphological features. Our analysis suggests that Longlin does not show high archaic-related ancestry. Thus, these features are possibly retained within the variation found among early modern humans, without necessarily implicating a direct link to archaic humans. One possible explanation for the different morphological and genetic results is that the archaic ancestor contributing to Longlin is currently unknown and not closely related to Neanderthal or Denisovan ancestries. As the analysis depends on reference populations to use as admixture sources, we may not be able to detect unknown archaic ancestry.

5 DATA AND CODE AVAILABILITY

BAM files and genotype calls for the newly sequenced individuals are available at the Genome Sequence Archive (Wang et al., 2017) in BIG Data Center (B. I. G. Data Center Members, 2018) (https://bigd.big.ac.cn/gsa-human; accession number: PRJCA003870). All newly generated code is available upon request from the Lead contact. All software used are freely available online and are referenced in Key Resources Table (Wang et al., 2017).

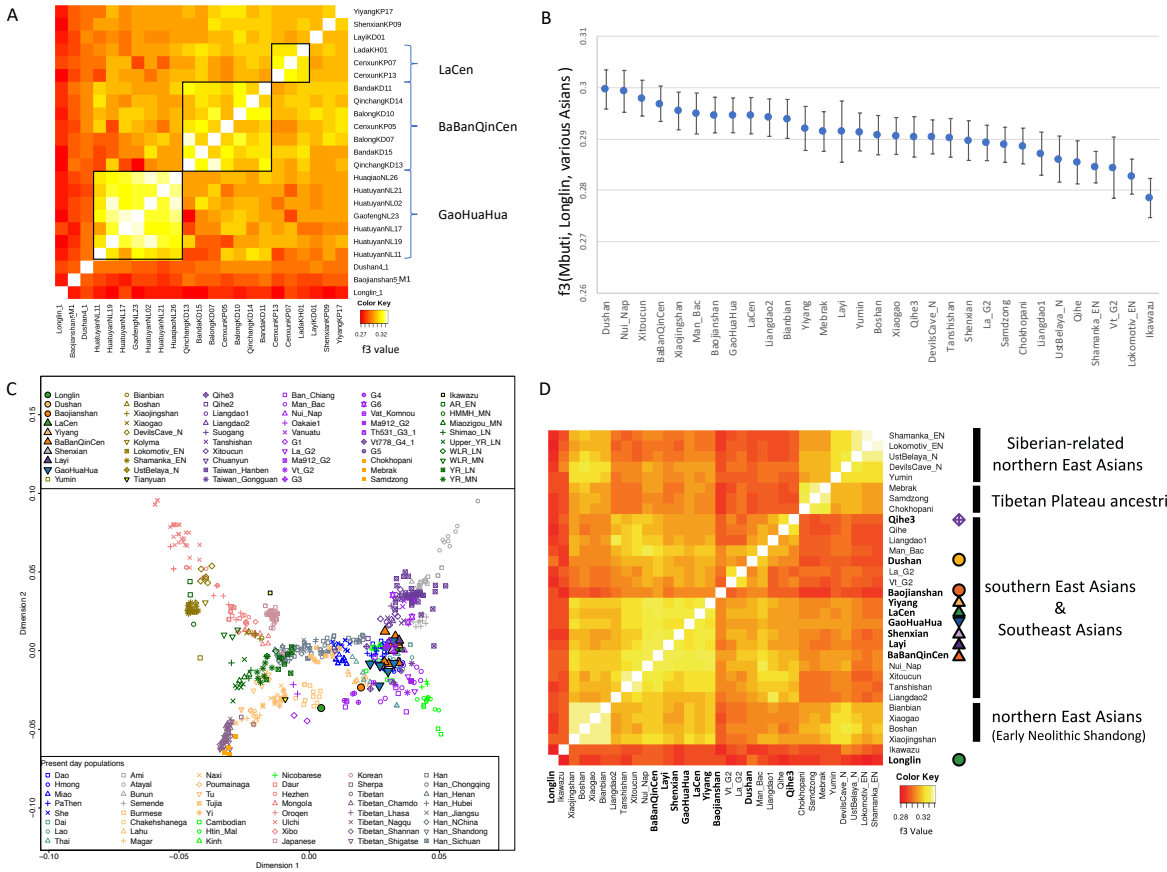

Figure S1. Genetic structure of new individuals. Related to Figure 1

(A) Pairwise outgroup-f_3 analysis of newly sampled individuals, for *f_3(Mbuti; newly sampled individuals, newly sampled individuals)*. The Mbuti are a central African population that acts as an outgroup to the Asian populations belonging to the newly sampled individuals. Based on their clustering pattern, we grouped several historical individuals into one of three major clusters, LaCen, BaBanQinCen, and GaoHuaHua. Related to Figure 1.

(B) Outgroup-f_3 statistics of *f_3(Mbuti; Longlin, X)*. Related to Figure 1.

(C) PCA projecting ancient Asians onto diverse present-day Asians. Ancient populations are listed in the key at the top. Newly sampled ancient individuals are symbols with a black outline and different fill colors (first column at top). Ancient northern East Asians are in dark tan and green, while ancient southern East Asian and Southeast Asians are in dark purple and light purple. Present-day populations are listed in the key at the bottom, with coloring based on their associated language group. Related to Figure 1.

(D) Pairwise outgroup f_3-statistics in the form of *f_3(Mbuti; X, Y)* to measure the shared genetic drift among ancient East Asians and Southeast Asians, where yellow indicates higher genetic similarity between pairs. Mbuti represents a central African population and is used as an outgroup to Asian populations.

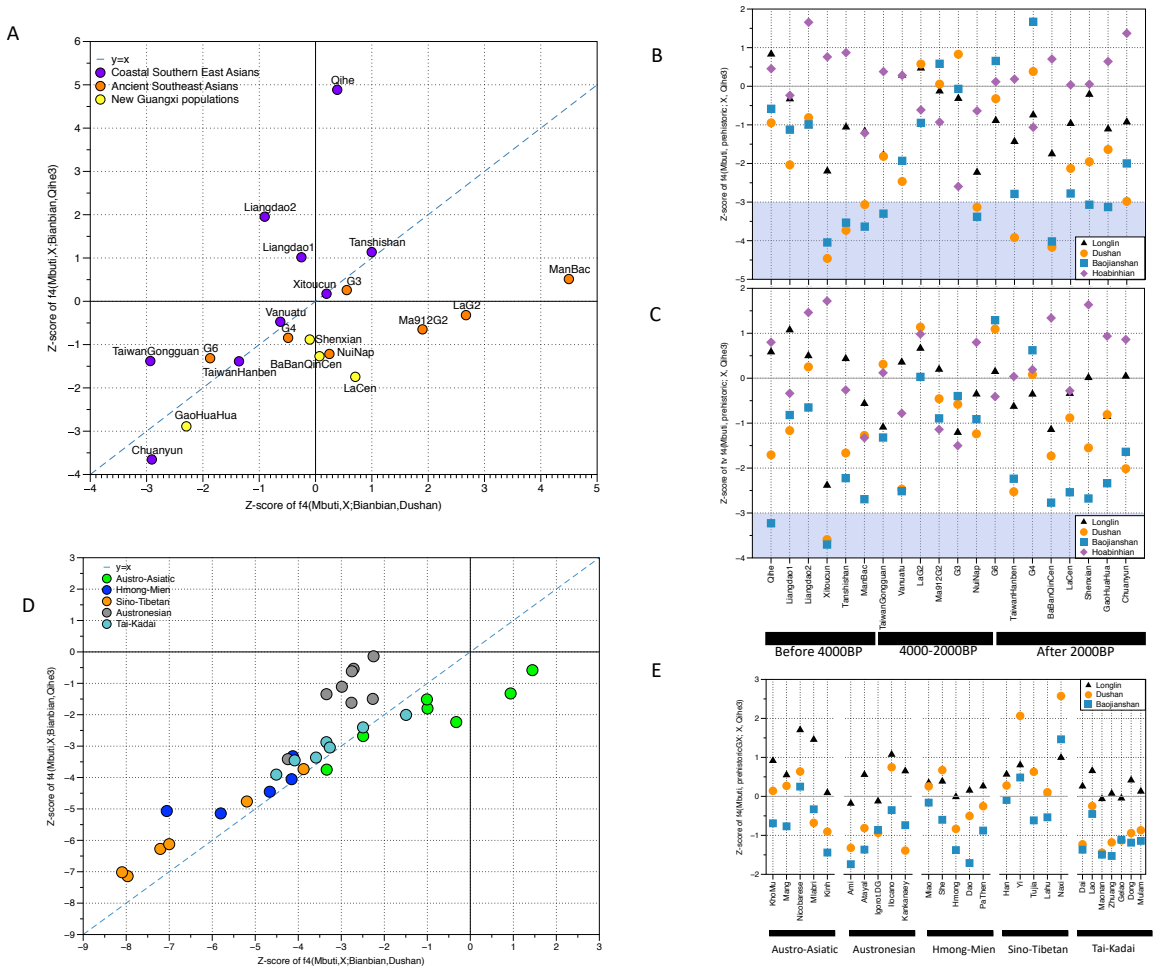

Figure S2. Genetic affiliation with younger populations. Related to Figure 2

(A-C) The genetic affiliation with younger southern East Asians and Southeast Asians. (A) f_4-statistics of allele sharing with Dushan or with Qihe3 relative to Bianbian. We compared *f_4(Mbuti, X; Bianbian, Dushan)* and *f_4(Mbuti, X; Bianbian, Qihe3)* to determine whether Dushan contributed to younger populations, and to distinguish from shared southern East Asian ancestry. Relative to Bianbian, populations below the diagonal are closer to Dushan than Qihe3, while populations above the diagonal are closer to Qihe3 than Dushan. Those who are closer to the third quadrant share more northern East Asian alleles. (B)Z-scores for *f_4(Mbuti, prehistoric; X, Qihe3)* confirm younger populations share more alleles with prehistoric Guangxi individuals than the southern East Asian Qihe3. Prehistoric populations include Longlin, Dushan, Baojianshan, and Hòabìnhian. "X" populations are historical Guangxi populations and previously published ancient southern East Asians/Southeast Asians. For "X" populations, we keep only those greater than 300,000 SNPs to decrease biases due to low data quality. (C) Z score for *f_4(Mbuti, prehistoric; X, Qihe3)* using transversions only. In b and c, the blue highlighted region indicates where there is a significant affinity between a prehistoric population and X relative to Qihe3. Related to Figure 2.

(D-E) The genetic affiliation with present-day Asians. (D) f_4-statistics of allele sharing with the Neolithic Guangxi Dushan or the Neolithic coastal southern East Asian Qihe3 relative to a coastal northern East Asian, Bianbian. We compared *f_4(Mbuti, X; Bianbian, Dushan)* and *f_4(Mbuti, X; Bianbian, Qihe3)* to determine whether Dushan contributed to present-day populations, and to distinguish from shared southern East Asian ancestry. (E) Z scores for *f_4(Mbuti, prehistoricGX; X, Qihe3)* confirming whether present-day populations share more alleles with prehistoric Guangxi individuals than the southern East Asian Qihe3. prehistoricGX includes Longlin, Dushan, Baojianshan, and "X" populations are present-day Austro-Asiatic, Austronesian, Hmong-Mien, Sino-Tibetan, Tai-Kadai speakers. Related to Figure 2.

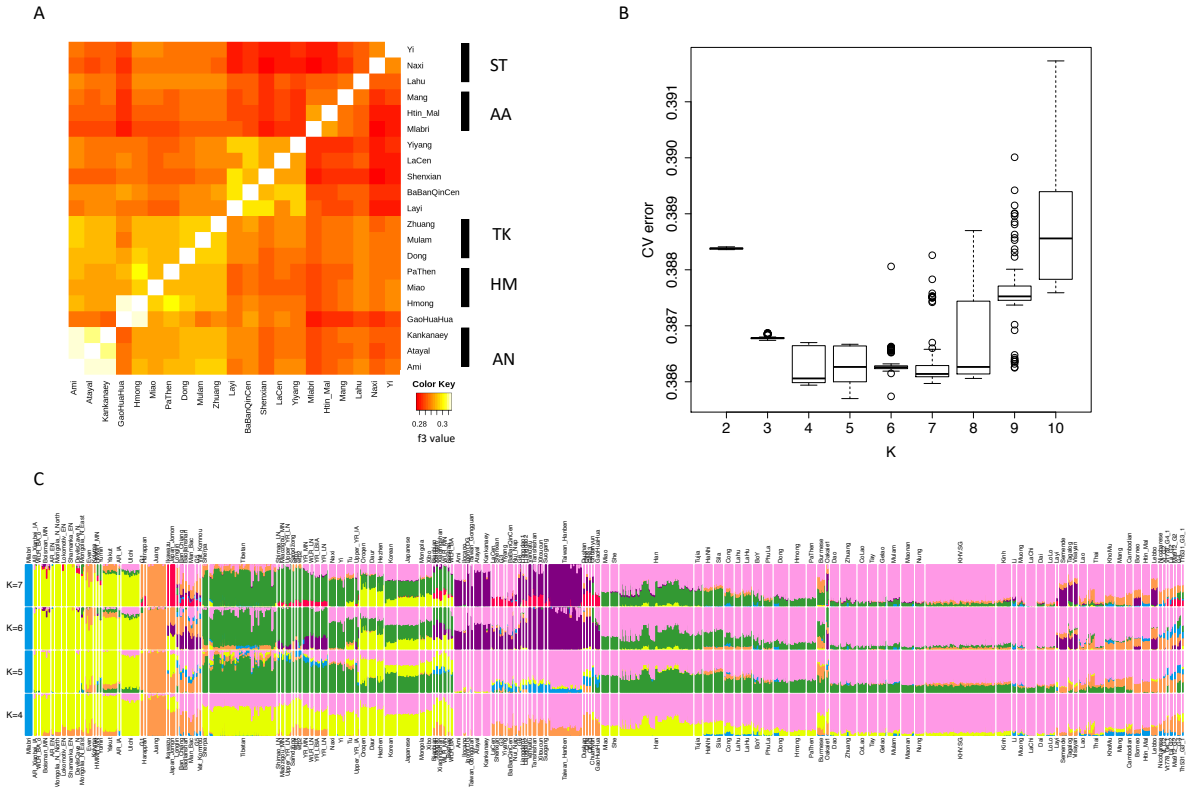

Figure S3. Genetic component for historical Guangxi populations and present-day East Asians. Related to Figure 3

(A) Pairwise outgroup-f_3 of historical Guangxi populations and present-day populations. "AA" represent Austro-Asiatic speakers, "HM" is Hmong-Mien speakers, "TK" is Tai-Kadai speakers, "ST" is Sino-Tibetan speakers, "AN" is Austronesian speakers. Related to Figure 3.

(B) Cross-validation results for different K values. The cross validation (CV) is lowest when K=4, the lowest CV error often correlates to the 'best' K.

(C) ADMIXTURE results for K=4 to K=7. We include previously published ancient and present-day populations. The genetic components of ancient southern East Asians and Vanuatu represented by pink; Northern East Asians share a component in yellow; the deep lineages show in orange; and the Southeast Asian Mlabri have a separate component in blue. Related to Figure 1.

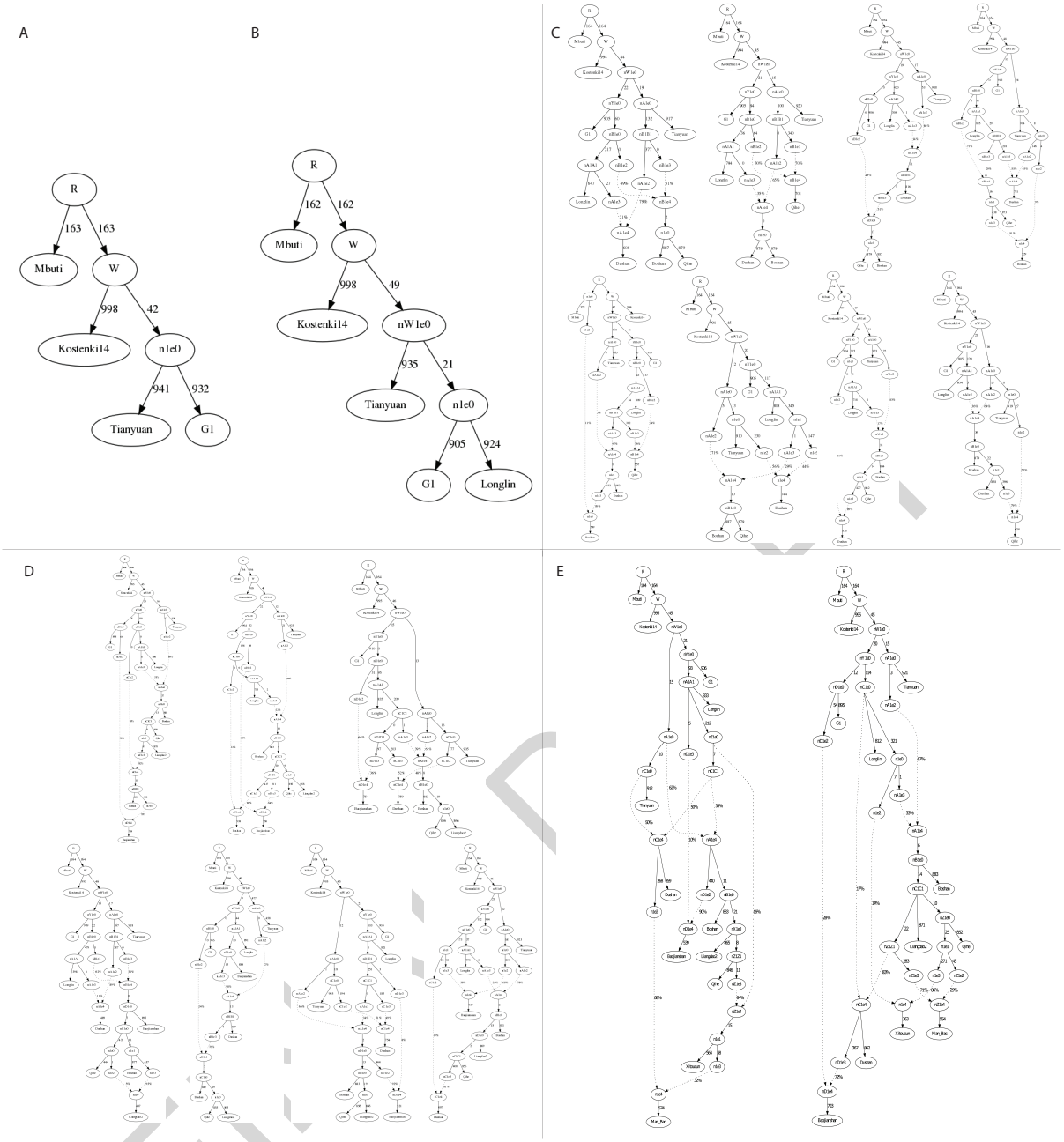

Figure S4. Admixture Graph. Related to Figure 2 and STAR Methods

(A) Admixture Graph of the basic model. |maxZ|= -0.640. Related to Figure 2.

(B) Admixture graph models adding Longlin to the basic model. |maxZ|= 1.598. Related to Figure 2.

(C) Admixture graph models adding Dushan, Qihe, Boshan. Corresponding order from left to right and top to bottom: |maxZ|= -2.530; |maxZ|=-2.675; |maxZ|=-2.305; |maxZ|=-2.699; |maxZ|=2.911; |maxZ|=-2.440; |maxZ|=-2.454; |maxZ|=-2.757. Related to Figure 2.

(D)Admixture graph models adding Baojianshan, Liangdao2. Corresponding order from left to right and top to bottom: |maxZ|= 2.855; |maxZ|= 2.855; |maxZ|= 2.855; |maxZ|= -2.606; |maxZ|= 2.884; |maxZ|= 2.855; |maxZ|= 2.855. Related to Figure 2.

(E) Admixture graph models adding the Late Neolithic Fujian Xitoucun and the Southeast Asian Man_Bac. Corresponding order from left to right: |maxZ|= 2.855; |maxZ|= 2.898. Related to Figure 2.

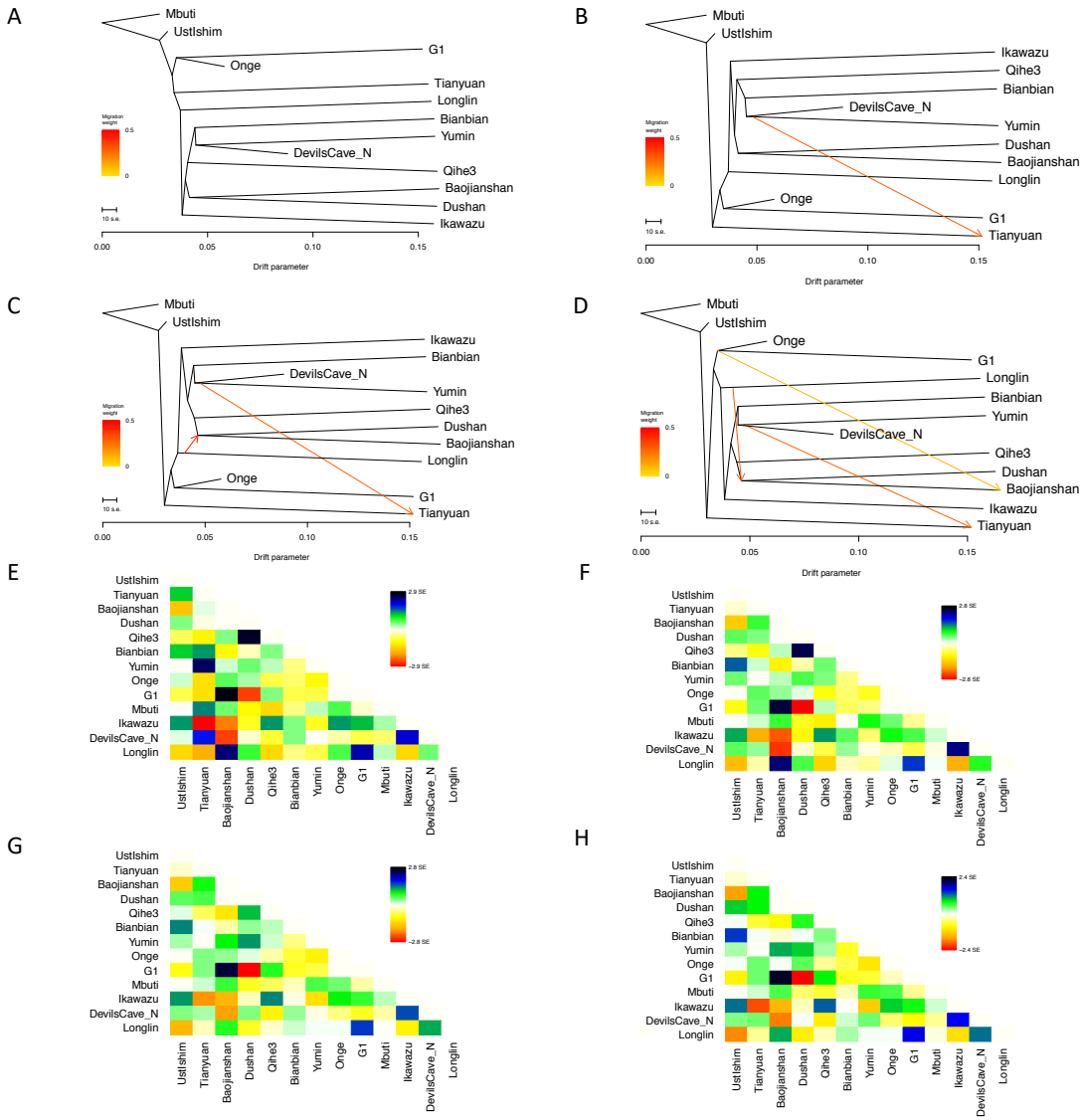

Figure S5. *Treemix* results and Pairwise residuals. Related to Figure 2 and STAR Methods

(A-D) *Treemix* results for zero to three migration events.

(E-H) Pairwise residuals for the phylogenies for 0, 1, 2, and 3 migration events. Related to Figure 2.