Prehistoric genomes from Yunnan reveal ancestry related to Tibetans and Austroasiatic speakers

Tianyi Wang, Melinda A Yang, Zhonghua Zhu, Minmin Ma, Han Shi, Leo Speidel, Rui Min, Haibing Yuan, Zhilong Jiang, Changcheng Hu, et al.

RESEARCH ARTICLE SUMMARY

ANCIENT DNA

Prehistoric genomes from Yunnan reveal ancestry related to Tibetans and Austroasiatic speakers

Tianyi Wang *et al.*

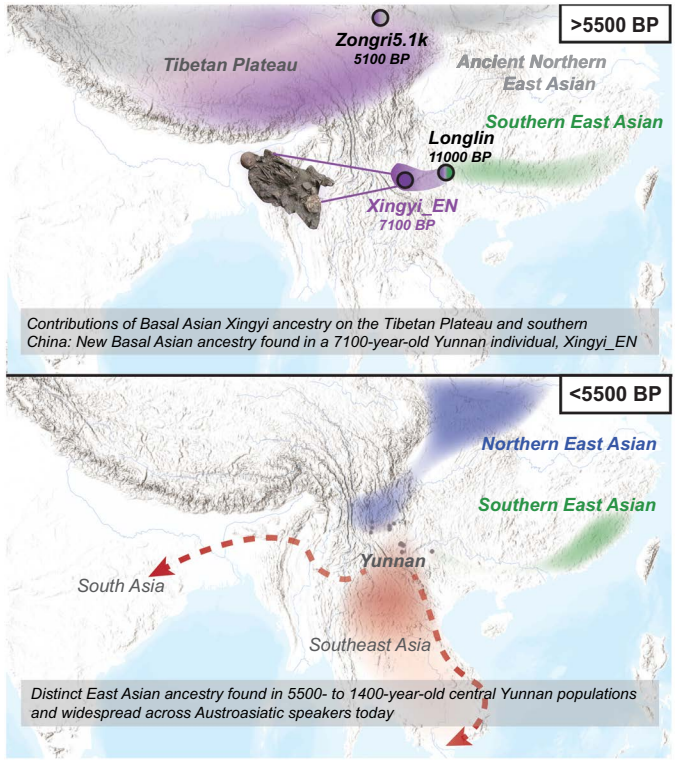
Full article and list of author affiliations: https://doi.org/10.1126/science.adq9792

INTRODUCTION: In East and Southeast Asia, a long history of population movement, replacement, and mixture has heavily influenced the genetic composition of humans up to the present day. For example, early populations found across this region that diversified before 40,000 years ago carrying Basal Asian ancestries have largely been replaced by a single lineage that diversified into the ancestries found in most present-day East and Southeast Asians. However, these ancestries sampled to date have yet to explain the genetic diversity observed across all present-day East and Southeast Asians, with ancestries still uncharacterized in many populations, including in Tibetans and Austroasiatic speakers.

RATIONALE: Sequencing of 127 ancient human genomes dating to 7100 to 1500 years ago from Yunnan province in southwestern China may help to further clarify human population history in East and Southeast Asia. Yunnan is at the intersection of the Tibetan Plateau, Southeast Asia, and southern China and is home to the highest ethnic and linguistic diversity in China today. Thus, genomic sampling of ancient populations from this region provides an opportunity to capture the migration, interaction, and change over time in prehistoric populations of East and Southeast Asia.

RESULTS: Analysis of a 7100-year-old individual from the Xingyi archaeological site in central Yunnan revealed a previously unsampled Basal Asian ancestry that is deeply diverged from East Asian ancestries and persisted in southern East Asia into the mid-Holocene. This Basal Asian Xingyi ancestry is also related to deeply diverged ghost ancestry found in Tibetan Plateau populations, providing new information for studying the origin of Tibetan Plateau populations. The 5500- to 1500-year-old populations from central Yunnan do not show Basal Asian Xingyi ancestry, but carry an East Asian ancestry distinct from northern and southern East Asian ancestries previously characterized, denoted here as Central Yunnan ancestry. This distinct East Asian ancestry can be found across present-day Austroasiatic speakers, indicating that these ancient populations in central Yunnan were likely a proto-Austroasiatic population. By contrast, 3800- to 1700-year-old populations from western Yunnan primarily have northern East Asian ancestry, whereas a 3400-year-old population from eastern Yunnan shows a mixture of southern East Asian and Central Yunnan ancestries. Collectively, these patterns indicate that multiple diverse East Asian ancestries coexisted in Bronze Age Yunnan populations.

CONCLUSION: Sampling ancient humans from Yunnan province, we discovered a new Basal Asian ancestry that is closely related to a "ghost" ancestry that influenced populations who lived on the Tibetan Plateau. In addition, we found a distinct East Asian ancestry that sheds light on the origin of Austroasiatic populations. Multiple East Asian ancestries in Yunnan during the Bronze Age reveal high human genetic diversity and dynamic population movements across East and Southeast Asia. □



Population migration, replacement, and the preservation of deeply diverged ancestry in southern East Asia. A 7100-year-old individual from Yunnan shows a Basal Asian ancestry that is related to a deeply diverged ghost ancestry contributing to Tibetan Plateau populations. After 5500 years before present (BP), populations in Yunnan exhibited diverse ancestries, including contributions from northern East Asia in western Yunnan, coastal southern East Asia in eastern Yunnan, and a newly identified East Asian lineage in central Yunnan that later contributed to present-day Austroasiatic speakers.

Corresponding author: Qiaomei Fu (fuqiaomei@ivpp.ac.cn) Cite this article as T. Wang *et al.*, *Science* **388**, eadq9792 (2025). DOI: 10.1126/science.adq9792

ANCIENT DNA

Prehistoric genomes from Yunnan reveal ancestry related to Tibetans and Austroasiatic speakers

Tianyi Wang[1,2]†, Melinda A. Yang[3]†, Zhonghua Zhu[4,5]†, Minmin Ma[6]†, Han Shi[1,2], Leo Speidel[7,8], Rui Min[5], Haibing Yuan[4,9], Zhilong Jiang[5], Changcheng Hu[5], Xiaorui Li[5], Dongyue Zhao[10], Fan Bai[1,2], Peng Cao[1], Feng Liu[1], Qingyan Dai[1], Xiaotian Feng[1], Ruowei Yang[1], Xiaohong Wu[11], Xu Liu[12], Ming Zhang[10], Wanjing Ping[1], Yichen Liu[1], Yang Wan[5], Fan Yang[5], Ranchao Zhou[5], Lihong Kang[5], Guanghui Dong[6,13], Mark Stoneking[14], Qiaomei Fu[1,2]*†

The human landscape in East and Southeast Asia is vastly complex, and successful retrieval of genome-wide data from prehistoric humans of southern East Asia is sparse. We successfully sampled 127 ancient human genomes from southwestern China. A 7100-year-old female individual from central Yunnan shows a previously unsampled Basal Asian ancestry related to a ghost population that contributed to Tibetan Plateau populations. Central Yunnan populations dating to 5500 to 1400 years before present show an East Asian ancestry distinct from northern or southern East Asian ancestries that contributed to present-day East and Southeast Asians, particularly Austroasiatic speakers, and emphasizes the importance of the Red River valley for proto-Austroasiatic population history. Diverse Asian ancestries are represented in humans sampled from Yunnan, clarifying past population dynamics related to both Tibetan and Austroasiatic origins.

The ancestry of present-day humans in East and Southeast Asia can be traced to ancient DNA from humans of the Late Pleistocene through the Holocene dating as early as 19,000 years ago (*1*). Prehistoric population splits and mixtures reveal the presence of distinct northern and southern East Asian lineages. In northern East Asia, a shared ancestry is found across populations spanning the Tibetan Plateau (*2*–*4*) to the Yellow River region (*5*–*7*) and as far up as the Amur River region (*1*). In southern East Asia, populations from the Fujian region of coastal China were shown to be closely related to present-day Austronesian speakers (*5*, *8*) and to have contributed ancestry partially to humans further inland in Guangxi, as well as Southeast Asia (*8*–*10*). During the Late Neolithic to Bronze Age periods ~4000 to 3000 years ago, more humans in southern China (*5*) and Southeast Asia (*9*, *10*) transitioned to agricultural subsistence strategies and experienced higher gene flow from populations carrying northern East Asian ancestry.

Gaps remain in the genetic record of humans in East and Southeast Asia, leaving many questions unanswered, such as the genetic origins of many present-day East Asian populations, including Tibetan Plateau populations and Austroasiatic speakers. It is well established that populations on the Tibetan Plateau carry northern East Asian–related ancestry, but studies have shown that they also carry a deeply diverged ghost ancestry that is not well characterized (*2*, *3*, *11*). Some have suggested an archaic origin (*11*) due to high frequencies of a Denisovan-related *EPAS1* haplotype in Tibetan populations (*12*) and the physical presence of Denisovans on the plateau (*13*, *14*). Others have proposed a modern origin related to Paleolithic Eurasians (*2*, *3*, *15*), including an early Asian lineage (*7*) given the many deeply diverged Asian individuals that have been sampled to date (*1*, *10*, *16*). The lack of a representative source population for the deeply diverged ghost ancestry has stymied efforts to resolve the origins of Tibetans.

Population dynamics during the transition to agriculture in East and Southeast Asia are still unclear in many regions, particularly in southern China and Southeast Asia. Most proposed transmission routes to Southeast Asia have been along waterways such as the Mekong or Red River watersheds or the eastern coast of Asia in the South China Sea (*17*–*19*). In mainland Southeast Asia, agricultural expansion, particularly of rice, has also been associated with the ancestors of Austroasiatic speakers (*18*, *20*, *21*), who today are found in northeastern India, southern China, and Southeast Asia. Multiple hypotheses on Austroasiatic origins tied to these regions have been developed, but little consensus has been found to date (*20*–*26*). Studies of genome-wide data from 4000- to 3000-year-old humans in Vietnam and Laos indicate the presence of ancestry related to Austroasiatic speakers in the earliest Southeast Asian farmers (*9*, *10*). These farmers carried a mixture of ancestry related to hunter-gatherers, e.g., Hòabìnhians (*10*), and early farmers from China. However, genetic sampling in southern China has largely been limited to ancient populations in Fujian and Guangxi, which are not associated with Austroasiatic speakers (*5*, *8*). Without more genetic material from ancient humans associated with Austroasiatic speakers, it is difficult to resolve how proto-Austroasiatic speakers expanded across southern China and Southeast Asia.

Nestled between the Tibetan Plateau, Southeast Asia, and southern China, the region known today as the Chinese province of Yunnan is home to the highest ethnic and linguistic diversity in China today. Ancient humans that lived in this region may be key to addressing several remaining questions on the prehistoric populations of East and Southeast Asia. The role of prehistoric humans in Yunnan in the expansion of proto-Austroasiatic populations in Southeast Asia is much debated, with some arguing that prehistoric populations from Yunnan had little to no impact on Southeast Asian prehistory, and others insisting that Yunnan populations, particularly those in the Sanjiang region of western Yunnan, played an important role in the expansion of proto-Austroasiatic populations from or into Southeast Asia (*17*, *20*, *27*). Yunnan is also home to the oldest archaeological site associated with the Hòabìnhian culture (*28*, *29*), the 43,500-year-old Xiaodong rock shelter (*28*, *30*, *31*), indicating a long record of human occupation that is potentially associated with deeply diverged Asian ancestry (*10*).

To address this gap, we obtained genome-wide data from 147 individuals from Yunnan dating from 7100 years ago to the modern era (Fig. 1, A and B), more than doubling the number of ancient humans with genome-wide data from southern East Asia and providing insights into human prehistory in a critical region of southeastern Asia during the transition to agriculture.

[1]Key Laboratory of Vertebrate Evolution and Human Origins, Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, Beijing, China. [2]University of the Chinese Academy of Sciences, Beijing, China. [3]Department of Biology, University of Richmond, Richmond, VA, USA. [4]School of Archaeology and Museology, Sichuan University, Chengdu, China. [5]Yunnan Institute of Cultural Relics and Archaeology, Kunming, China. [6]Key Laboratory of Western China's Environmental Systems (Ministry of Education), College of Earth and Environmental Sciences, Lanzhou University, Lanzhou, China. [7]Genetics Institute, University College London, London, UK. [8]Ancient Genomics Laboratory, Francis Crick Institute, London, UK. [9]Center for Archaeological Science, Sichuan University, Chengdu, China. [10]China-Central Asia "The Belt and Road" Joint Laboratory on Human and Environment Research, Key Laboratory of Cultural Heritage Research and Conservation, School of Culture Heritage, Northwest University, Xi'an, China. [11]School of Archaeology and Museology, Peking University, Beijing, China. [12]Yunnan Museum, Kunming, China. [13]State Key Laboratory of Loess Science, Institute of Earth Environment, Chinese Academy of Sciences, Xi'an, Shaanxi, China. [14]Laboratoire de Biométrie et Biologie Evolutive, UMR 5558, Université Lyon 1, CNRS, Villeurbanne, France. ***Corresponding author. Email: fuqiaomei@ivpp.ac.cn** †These authors contributed equally to this work.

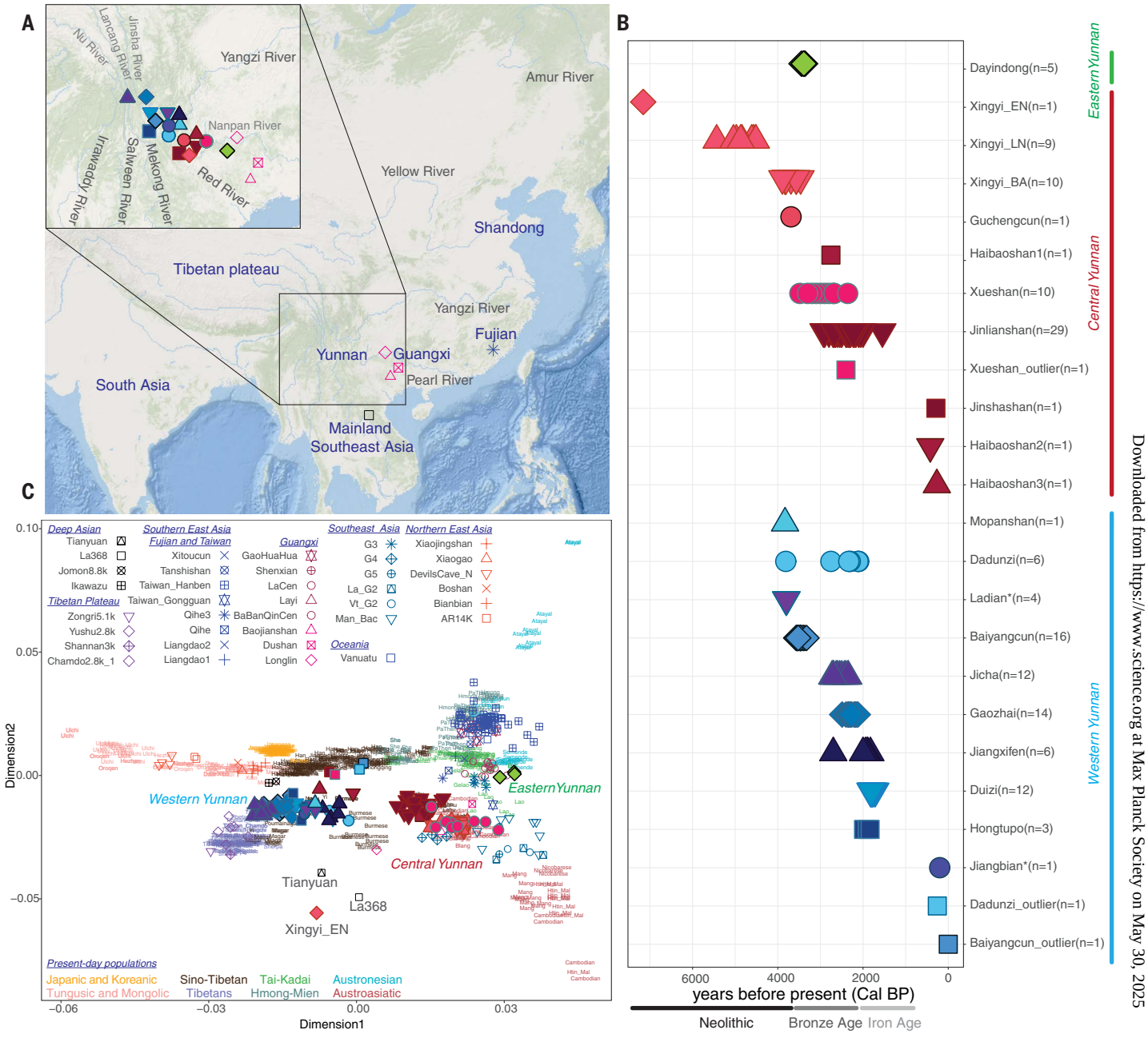


Fig. 1. Chronological, geographic, and genetic distribution of newly sampled ancient individuals. (**A**) Location of newly sampled archaeological sites on a map of East and Southeast Asia. For the Xingyi site, only Xingyi_EN's symbol is shown. Five sites for representative previously published individuals dating to the Early Neolithic or Paleolithic from southern China or Southeast Asia are also shown (unshaded symbols). (**B**) Chronological data for 147 newly sampled individuals based primarily on radiocarbon dating (asterisk indicates the site was not successfully dated using direct radiocarbon dating), with the total number of individuals sampled from each site indicated on the *y* axis. (**C**) PCA of present-day East and Southeast Asians with projection of ancient individuals. A total of 127 ancient individuals from Yunnan were included, because low-coverage individuals with <20,000 SNPs were omitted. To gain a clearer understanding of the genetic patterns of the Tibetan Plateau populations, we have specifically labeled "Tibetans" separately.

Results

We generated genome-wide data from 147 individuals. For 102 individuals, we obtained DNA enriched for 1.2 million single nucleotide polymorphisms (SNPs) using a large-scale ancient nuclear DNA capture technique (*32*). For seven of the 102 individuals, and an additional 45 individuals who had high endogenous DNA, we applied shotgun sequencing (data S1a). Direct radiocarbon dating for 58 individuals showed that 53 of them ranged from 7150 to 1400 years ago, and five individuals spanned the 15th century to 1950 CE (Fig. 1B and data S1a).

For samples with contamination rates >5%, we restricted our analysis to only damaged fragments (see the materials and methods). We obtained genetic information from 127 individuals, after excluding 20 individuals with <20,000 SNPs available. Overall, average coverage across the 1.2 million targeted SNPs ranged from 0.3× to 6.37×, covering 22,158 to 956,334 SNPs. Familial and uniparental haplogroup analyses were additionally performed, and 21 individuals from joint tombs in both central [Jinlianshan tomb M15, ~2000 to 1400 years before present (BP)] and western (Gaozhai tomb M4, ~2500 to 2100 BP) Yunnan

allowed determination of multigenerational family pedigrees spanning from three to potentially more than 10 generations (supplementary text section S3).

Deeply diverged Asian lineage found in a 7100-year-old Yunnan individual

We characterized ancestry in Yunnan dating to the Early Neolithic, where we were able to sequence an individual from the Xingyi archaeological site in central Yunnan dating to 7158 to 6888 calibrated (cal) BP (Xingyi_EN, female; Fig. 1, A and B). Radiocarbon dating showed that this individual lived 1500 years earlier than other individuals excavated at higher layers (supplementary text section S1). We compared the ancestry carried by this individual with a diverse panel of ancient humans from Europe, West Asia, and Siberia, and found that Xingyi_EN is most closely related to ancient and present-day humans from East and Southeast Asia in both f4 analyses (data S2c) and a maximum likelihood phylogeny (bootstrap = 100%; Fig. 2A). Further comparison with ancient and present-day East and Southeast Asians revealed the presence of a deeply diverged Asian ancestry in Xingyi_EN. In an outgroup f3 analysis comparing relative levels of genetic similarity, Xingyi_EN showed low similarity to 19,000- to 4000-year-old East Asians from northern (Yellow River, Amur River, and Shandong) and southern (Fujian) China, who carry ancestry related to that found in present-day East and Southeast Asians. A similar pattern is observed for deeply diverged ancient Asians such as the ~40,000-year-old Tianyuan (*16*, *33*) and the ~8000-year-old Hòabìnhian from Southeast Asia [La368 (*10*); Fig. 2, B and C]. Further exploration of the genetic structure within Asian populations using f4 analyses shows that the three ancestries represented by Xingyi_EN, Tianyuan, and La368 are distinct and equidistant from each other (data S2c and supplementary text section S4), indicating that East and Southeast Asia were home to a highly diverse set of Basal Asian populations.

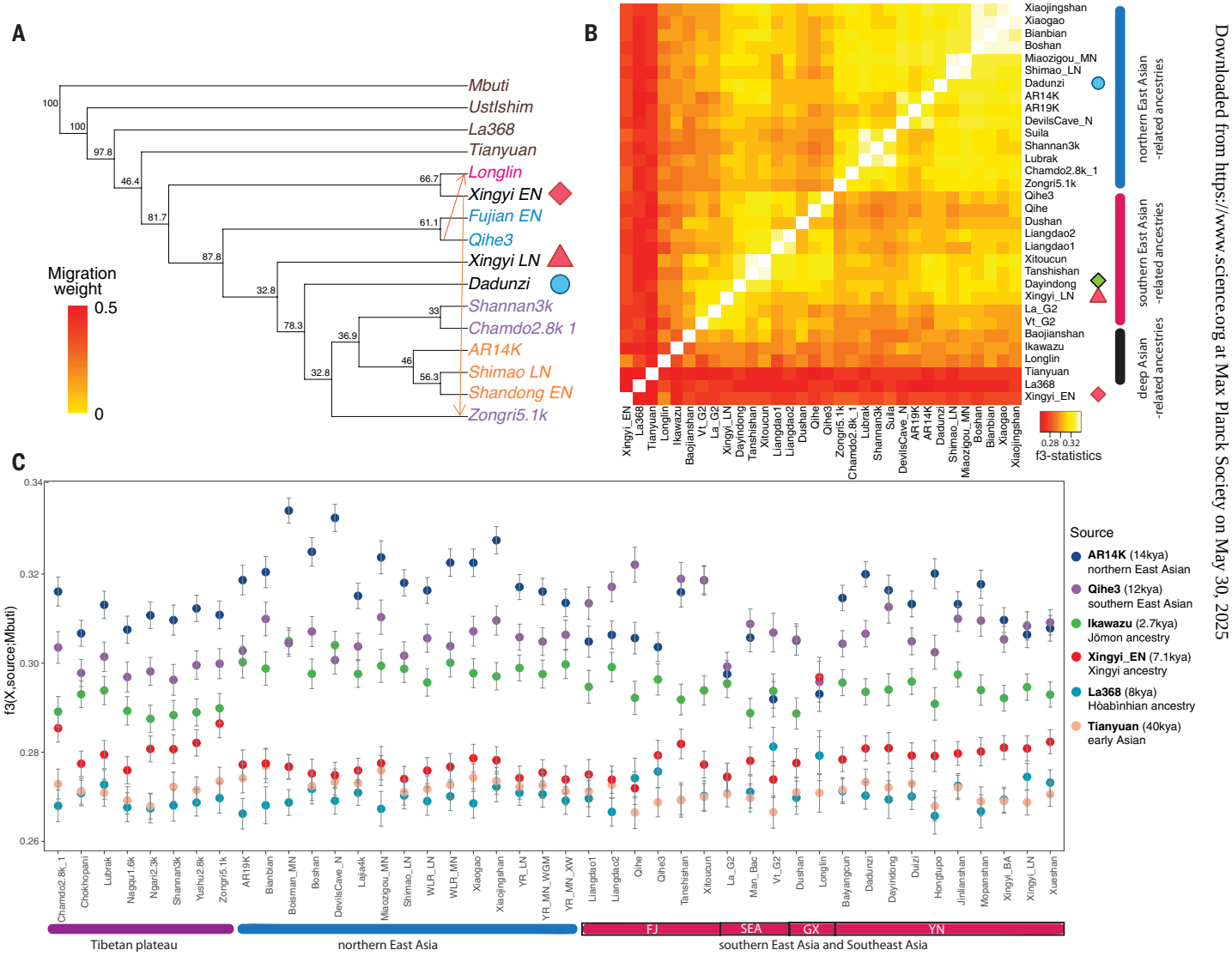


Fig. 2. Genetic relationships of ancient Yunnan individuals to other East Asians. (**A**) Maximum likelihood tree with two migration events showing the genetic relationship of ancient humans from East Asia. The bootstrap percentage for 1000 replicates is added by each node. (**B**) Pairwise outgroup f3 analysis using f3(Mbuti; X, Y) where X and Y are ancient East and Southeast Asians. For (A) and (B), icons next to the label indicate a newly sampled individual, and the icon matches the key in Fig. 1. (**C**) Outgroup f3 results comparing ancient populations from East and Southeast Asia against source populations, where a higher f3 value indicates higher genetic similarity to the source population. Error bars indicate ±1 SE. Source populations include the northern East Asian AR14K, southern East Asian Qihe3, deep East Asian Jōmon Ikawazu, and Basal Asians such as the newly sequenced Xingyi_EN, Tianyuan, and the Hòabìnhian La368. FJ, Fujian; SEA, Southeast Asia; GX, Guangxi; and YN, Yunnan.

Estimates of archaic ancestry in Xingyi_EN show similar proportions to that found in today's East Asians (data S1e), indicating that the signal for deeply diverged ancestry found in Xingyi_EN is not due to the presence of extra Denisovan or Neanderthal ancestry. The above findings show that Xingyi_EN represents a previously unsampled deeply diverged Asian ancestry. We thus refer to Xingyi_EN– and Xingyi_EN–related ancestries (i.e., ancestry from unsampled populations more related to Xingyi_EN than other humans sampled to date) as Basal Asian Xingyi ancestry. This ancestry is equally distant from the lineages that contributed to Hòabìnhian- and Tianyuan-related populations, as well as the ancestral population that contributed ancestry to mainland East Asians. Thus, Xingyi_EN carries a Basal Asian ancestry that diverged at least 40,000 years ago, the age of the Tianyuan individual (*16*, *33*). Together with the 8000- to 6000-year-old individuals from Southeast Asia carrying Hòabìnhian ancestry, Xingyi_EN shows the persistence of deeply diverged prehistoric humans in southern East Asia until the mid-Holocene.

Contributions of deeply diverged Basal Asian Xingyi ancestry on the Tibetan Plateau and southern China

To date, the impact of other Basal Asian ancestries (e.g., Hòabìnhian and Tianyuan) on East and Southeast Asian populations who contributed to present-day human genetic diversity has been minimal. To determine whether Basal Asian Xingyi ancestry follows similar patterns, we examined the impact of this ancestry on ancient individuals of East Asian ancestry by using f-statistic, phylogenetic, and haplotype-based mixture modeling approaches. First, we found that an 11,000-year-old individual from Guangxi, Longlin (*8*), who was previously found to be a deeply diverged East Asian individual, shows partial ancestry related to Xingyi_EN (Fig. 2, A and B). Longlin and Xingyi_EN share excess genetic similarity apart from ancient southern East Asians from Fujian (sEA) and ancient northern East Asians spanning across the Yellow to the Amur River regions [nEA (*1*, *5*, *6*)], i.e., f4(Mbuti, Xingyi_EN; Longlin, nEA/sEA) < 0 ($-6.2 < Z < -3.2$; Fig. 3A and data S2d). Testing for admixture using qpAdm and qpGraph, we found that Longlin is a mixture of ancestry related to Xingyi_EN (48.8 to 71.5%) and an East Asian ancestry (Fig. 4B and data S3c). Additional genetic affinity to ancient southern East Asians (e.g., Qihe3; Fig. 2, A and B) indicates that Longlin is likely a mixture of southern East Asian ancestry and Basal Asian Xingyi ancestry. By 8000 to 6000 years ago in Guangxi, however, the influence of Basal Asian Xingyi ancestry is much decreased (data S2d), and southern East Asian ancestry more heavily affects the region.

On the Tibetan Plateau northwest of Yunnan, ancient and present-day Tibetans were previously shown to derive from a mixture of northern

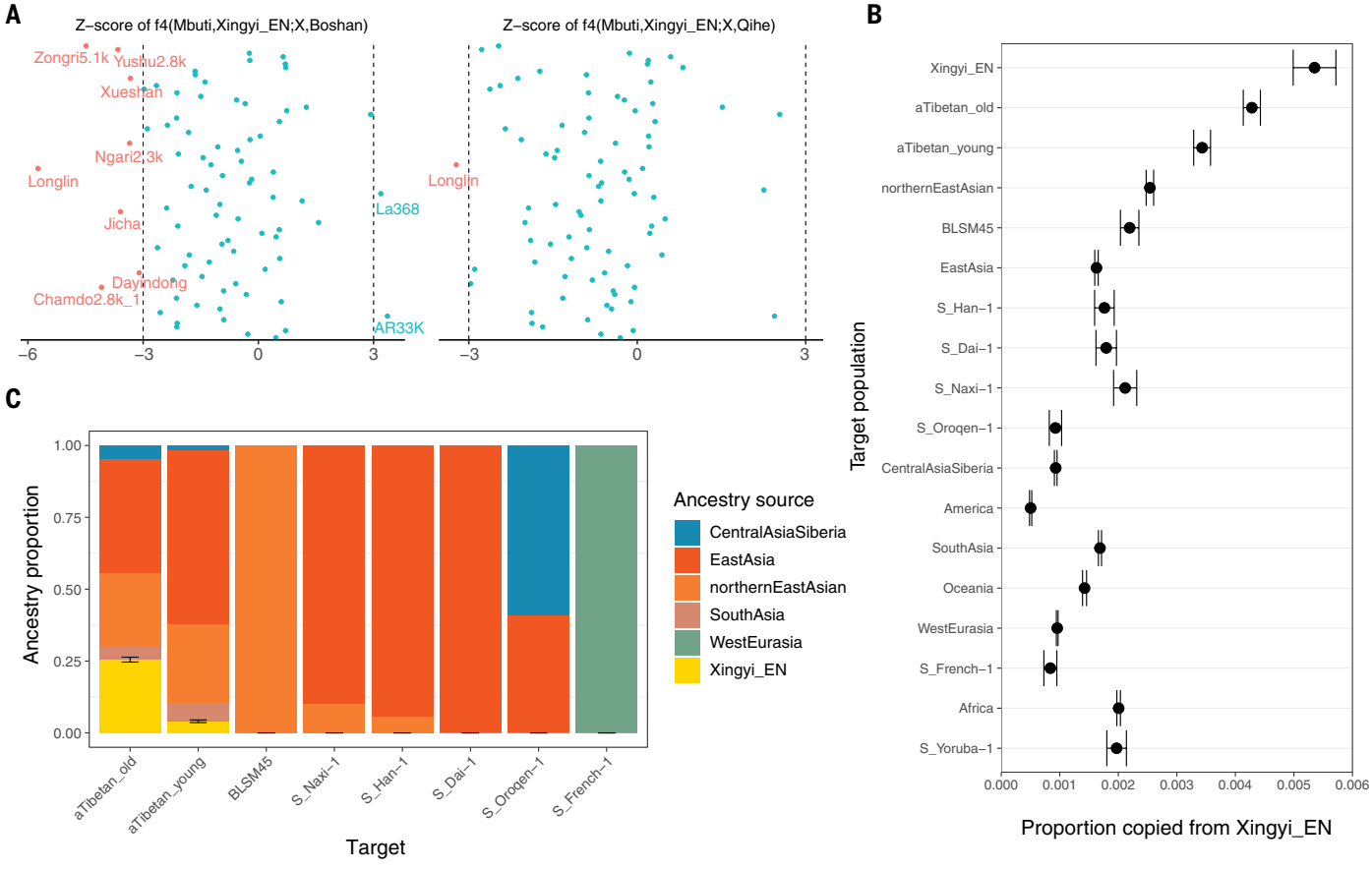


Fig. 3. Basal Asian Xingyi ancestry found in Tibetans and an 11,000-year-old individual from Guangxi. (**A**) Z scores for f4 analyses showing the relationship between ancient East Asians with respect to Xingyi_EN. Red points show $Z < -3$, indicating a significant genetic connection with Xingyi_EN relative to the northern East Asian Boshan (left) or the southern East Asian Qihe (right). (**B**) Proportion of the genome copied from Xingyi_EN, when the reference population set includes present-day groups, ancient northern East Asians, and the Xingyi_EN individual. We also included several present-day individuals (S_Yoruba-1, S_French-1, S_Naxi-1, S_Han-1, S_Dai-1, and S_Oroqen-1), as well as an ancient northern East Asian individual (BLSM45) who was not included in the reference population set. Error bars indicate 1 SE inferred from 1000 block bootstrap samples. (**C**) Ancestry proportions inferred by fitting a non-negative least-squares model that uses the target painting profiles obtained in (B) as a mixture of putative source painting profiles. Error bars correspond to the 95% confidence interval of Xingyi_EN-related ancestry obtained from 1000 block bootstrap samples. aTibetan_old includes ancient Tibetans older than 2500 BP (Zongri5.1k, Yushu2.8k, Nagqu2.5k, and Rhirhi), and aTibetan_young includes ancient Tibetans younger than 2500 BP (Nagqu1.1k, Shigatse0.7k, Kyang, and Mebrak).

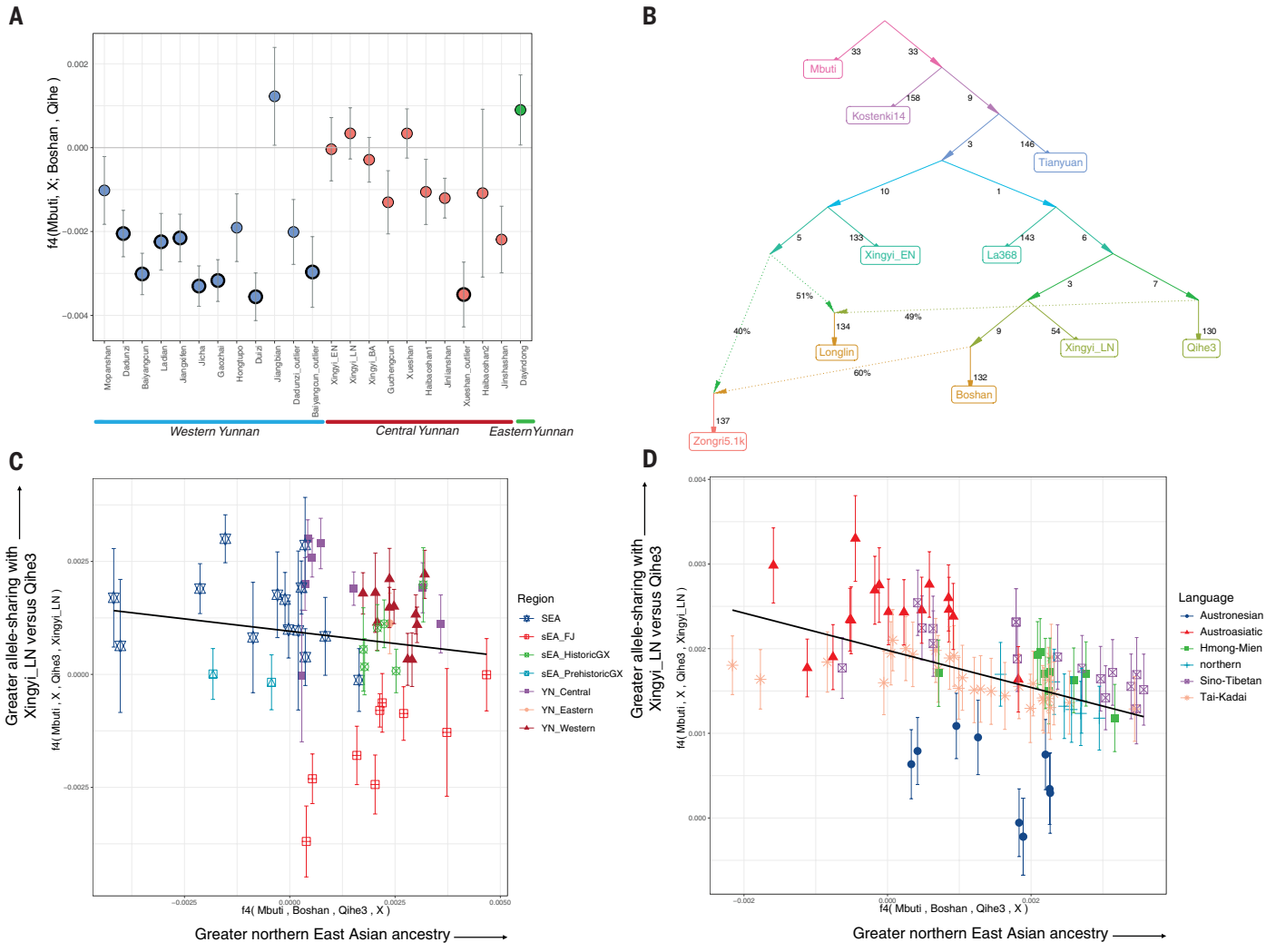


Fig. 4. Impact of Central Yunnan ancestry across ancient and present-day East Asians. (**A**) f4 analysis comparing newly sampled individuals with northern East Asians (Boshan) and southern East Asians (Qihe). The black circle indicate that |Z| > 3. The dashed line indicates where f4 = 0; points below the gray line indicate affinity to the northern East Asian Boshan, and points higher than the line indicate affinity to the southern East Asian Qihe. (**B**) qpGraph admixture graph showing a demographic model including ancient East and Southeast Asians that depicts observed genetic relationships for newly sampled Yunnan and Guangxi individuals. (**C**) Regression symmetry f4 analysis examining whether ancient southern East Asians (X) share a greater genetic affinity to the coastal Qihe3 versus Xingyi_LN. SEA, Southeast Asians; sEA_FJ, southern East Asians from Fujian; sEA_HistoricGX, historical Guangxi populations; sEA_Prehistoric_GX, prehistoric Guangxi populations; YN, Yunnan. (**D**) Regression symmetry f4 analysis examining whether present-day East Asians (X) share a greater genetic affinity to Qihe3 or Xingyi_LN. Data S1f includes population names for associated language categories. northern, Mongolic, and Tungusic populations. For (C) and (D), the regression line was calculated using all individuals in the plot. The line indicates a symmetric relationship to both Qihe3 and Xingyi_LN, points above the line indicate greater affinity to Xingyi_LN, and points below the line indicate greater affinity to Qihe3.

East Asian ancestry and a deeply diverged, unsampled "ghost" ancestry (*2*, *3*). We found that ancestry related to Xingyi_EN affected plateau populations, because we consistently observed excess shared genetic similarity between Xingyi_EN and plateau populations using PCA, outgroup f3, and f4 analyses (Figs. 1C; 2, B and C; and 3A and data S2d). Testing for admixture, plateau populations could only be modeled as a mixture of ancestry related to Xingyi_EN and northern East Asians in a qpAdm analysis (data S3a and supplementary text section S5), and the addition of a migration event from the Xingyi_EN branch to a Tibetan-related branch greatly improves the fit of the phylogenetic tree (Fig. 2A, fig. S25, and supplementary text section S6). We next inferred whole-genome genealogies using Relate (*34*) for Xingyi_EN and individuals from the Tibetan Plateau to estimate the genomic proportions in Tibetans that matched Xingyi_EN and found that ancient Tibetans, especially those dated earlier than 2500 BP, derived a substantially higher proportion of their genomes from Xingyi_EN than any other human population tested in our analysis (25.5%, 95% confidence interval: 24.7 to 26.4%; Fig. 3, B and C, and supplementary text section S8). These findings suggest that the deeply diverged ghost ancestry in ancient Tibetans (*2*, *3*, *11*) derives from a population that was closely related to Xingyi_EN. Admixture graph modeling using qpGraph also consistently shows partial ancestry related to Xingyi_EN in ancient Tibetan populations spanning the plateau (figs. S31 to S33 and supplementary text section S7), although a model with additional ghost ancestry from a population unrelated to Xingyi_EN was also feasible (fig. S32). An adaptive *EPAS1* haplotype found at high frequency in plateau populations associated with Denisovan introgression (*2*, *3*, *12*) was not observed in Xingyi_EN or Longlin (table S1).

One possibility is that the population represented by Xingyi_EN was polymorphic for *EPAS1*, i.e., multiple *EPAS1* haplotypes were present in the population to which Xingyi_EN belonged and we happened to sample an individual without the adaptive haplotype. However, another possibility is that unsampled populations not related to Xingyi_EN also contributed to the ghost ancestry observed in Tibetans, and one of those populations contributed the adaptive *EPAS1* haplotype.

A new East Asian ancestry in central Yunnan distinct from northern or southern East Asian ancestries

In addition to the 7100-year-old Xingyi_EN individual, we sequenced and analyzed genome-wide data from 146 individuals spanning 5500 years of Yunnan history, which were associated with the Late Neolithic, Bronze Age, and Iron Age, including those associated with the prominent Dian culture known for their bronze metalworks (*35*). Nineteen individuals also derive from the Xingyi site in central Yunnan, and 127 are distributed across 16 sites spanning western, central, and eastern Yunnan. Populations younger than 5000 years in Yunnan, including those at the Xingyi archaeological site and from previous studies (*36*), are not closely related to Xingyi_EN and instead share high levels of ancestry with ancient northern and southern East Asians (data S2c).

Projection of ancient Yunnan individuals onto a principal components analysis (PCA) shows three main clusters that primarily assort by geography: central, eastern, and western Yunnan (Fig. 1C). Focusing first on central Yunnan, on the Yungui Plateau near Lake Dian, 40 unrelated individuals with sufficient coverage from five sites were securely dated to 5500 to 1500 years ago, Xingyi_LN (5435 to 4296 cal BP), Xingyi_BA (3830 to 3385 cal BP), and populations associated with the Dian culture (Xueshan, Jinlianshan, Haibaoshan_1; 2753 to 1413 cal BP). These individuals cluster together near the southern East Asian cline but are shifted toward the Tibetan cline (Fig. 1C).

To investigate the relationship of central Yunnan populations to previously sampled East Asians, we compared Late Neolithic Xingyi individuals, Xingyi_LN, with ancient northern and southern East Asians (nEA/sEA) and found that Xingyi_LN cannot be placed within the context of known East Asian ancestries. In a phylogenetic analysis, Xingyi_LN's placement is uncertain, with 32.8% bootstrap replicates grouping Xingyi_LN with ancient northern East Asians and 18.5% replicates grouping Xingyi_LN with ancient southern East Asians (Fig. 2A and supplementary text section S6). Assessing population genetic relationships using f4 analyses, we found that Xingyi_LN is similarly related to northern East Asians from the Yellow River (*6*), Amur River (*1*), and Shandong (*5*), and southern East Asians from Fujian (*5*, *8*), i.e., f4(Mbuti, Xingyi_LN; nEA, sEA) ~ 0 ($-1.8 < Z < 2.0$; data S2e). Modeling ancestral relationships using qpGraph, Xingyi_LN falls within East Asian genetic diversity, showing a closer relationship with the northern East Asian Boshan and the southern East Asian Qihe3 than with individuals carrying Basal Asian ancestries (Fig. 4B). Collectively, our results show that Xingyi_LN does not have a Basal Asian ancestry, but rather an East Asian ancestry that is unlike previously described northern or southern East Asian ancestries (*1*, *5*, *8*).

Ancestry related to the 5500-year-old Xingyi_LN persists in more recent central Yunnan populations dating from 4000 and 1500 years ago. Comparing the relative affinity of more recent individuals from central Yunnan to Xingyi_LN and the southern East Asian Qihe3 (*5*, *8*) using a regression symmetry f4 analysis, greater genetic affinity is shared with Xingyi_LN (Fig. 4C). Several f4 analyses confirm that more recent central Yunnan populations are more related to Xingyi_LN than northern or southern East Asians, i.e., f4(Mbuti, Xingyi_LN; cYN, nEA/sEA) < 0 ($-11.5 < Z < -2.0$; data S2e). Modeling genetic admixture in Bronze Age central Yunnan populations using qpAdm analyses, we found that they are predominantly composed of ancestry related to Xingyi_LN with some northern East Asian-related gene flow (data S3d), a pattern also supported in f4 analyses (data S2e).

The ancestry observed across 5500 to 1500 years ago in central Yunnan, best represented by Xingyi_LN, is unlike that found in northern China and the Fujian region of southern China, indicating that this ancestry is distinct from those previously sampled in East Asia. Given its prevalence in central Yunnan over several millennia, we denote ancestry from Xingyi_LN and related populations as Central Yunnan ancestry.

With 12 individuals from the Xingyi site spanning 7100 to 3300 years, we also investigated how cultural and genetic patterns changed over time using a single archaeological site. The 7100-year-old individual at this site (Early Neolithic, Xingyi_EN) was found in an ash pit below all other cultural layers excavated at the Xingyi site under three large stones, with no associated pottery shards or other artifacts (supplementary text section S1). The discontinuity in the archaeological record between the ash pit and higher cultural layers where 5500- to 3300-year-old individuals (Xingyi_LN and Xingyi_BA) were excavated makes it difficult to discern whether this site was occupied between 7100 and 5500 BP. Xingyi_LN and Xingyi_BA individuals showed a completely different ancestry from that found in the Xingyi_EN individual, which suggests that population changes occurred in this region over this 1500-year period. Stable isotope analyses at the Xingyi site (*37*) suggest that the 7100-year-old Xingyi_EN and the two oldest Xingyi_LN individuals sampled (~5500 to 5000 BP) were primarily foragers, with the first evidence of some millet consumption in younger Xingyi_LN individuals starting around 4900 BP (*37*). Individuals younger than 4900 BP at the Xingyi site had stable isotope patterns suggesting that the subsistence strategy in the Late Neolithic from 4900 to 4300 years ago was mixed, indicating the use of millet but the persistence of foraging practices. By the Bronze Age ~3800 to 3300 years ago, the subsistence strategy at the Xingyi site was highly reliant on agriculture (Table 1). Because the oldest Xingyi_LN individuals date to 5500 BP and show a foraging-based diet, the first appearance of agriculture at 4900 BP seems to have occurred after the establishment of Xingyi_LN's ancestry at this site. Thus, the shift to agricultural practices was not associated with a shift in ancestry. In fact, the difference in ancestry between the Early and Late Neolithic at the Xingyi site may indicate large ancestral shifts before the onset of agriculture. However, the lack of additional individuals spanning 7000 to 5500 years makes it difficult to clarify how long the ancestry related to the 7100-year-old Xingyi_EN individual persisted at this locality or when the ancestry related to the 5500-year-old Xingyi_LN individuals first appeared in this region. By contrast, the Late Neolithic to the Bronze Age at the Xingyi site reveals high population continuity at a time when the subsistence strategy was shifting from an economy that was foraging based to one that was fully reliant on agriculture.

To determine whether Central Yunnan ancestry extends to other parts of Yunnan, we examined the ancestry of two individuals dating to ~3400 BP from the Dayindong site in eastern Yunnan, near the Nanpan River Basin that feeds into the Pearl River. Unlike populations from central Yunnan, individuals from the Dayindong site are shifted toward Neolithic southern East Asians in a PCA (Fig. 1C) and share the most genetic similarity with Neolithic southern East Asians in an outgroup f3 analysis (Fig. 2B). In f4 analyses, the Dayindong individuals share ancestry with both ancient southern East Asians, e.g., Xitoucun and Tanshishan (*5*), and central Yunnan populations i.e., f4(Mbuti, Dayindong; Xitoucun/Tanshishan, central Yunnan) < 0 ($-5.6 < Z < -2.4$) and f4(Mbuti, central Yunnan; Dayindong, Xitoucun/Tanshishan) < 0 ($-4.9 < Z < 0$; data S2h). In a regression symmetry f4 analysis comparing affinity to southern East Asian ancestry (Qihe3) or Central Yunnan ancestry (Xingyi_LN), Dayindong is to the left of central Yunnan populations but near the regression line to Qihe3 (Fig. 4C). These patterns indicate that in eastern Yunnan 3500 years ago, populations share a genetic affinity to both coastal southern East Asians and central Yunnan populations.

To explore the extent of Central Yunnan ancestry in regions outside of Yunnan, we applied the same regression f4-symmetry analysis to

Table 1. Subsistence strategy and dating determined using stable isotope analysis for humans sampled from the Xingyi archaeological site from central Yunnan. Subsistence strategies are conclusions summarized from Ma *et al*. (*37*). "Forage" indicates complete foraging with no evidence of agriculture, "mixed forage/agriculture" indicates primarily foraging with agriculture as a supplement, and "agriculturally dependent" indicates high levels of agriculture. For the populations, radiocarbon dates, and kinship determined in this study, see the supplementary text sections S1 to S3 (data S1).

Burial	Population	Radiocarbon dates (cal BP)	Kinship	Subsistence
XingyiH4	Xingyi_EN	7158 to 6888		Forage
XingyiM11	Xingyi_LN	5435 to 5070		Forage
XingyiM20	Xingyi_LN	5040 to 4867		Forage
XingyiM15	Xingyi_LN	4973 to 4857		Mixed forage/agriculture
XingyiM9	LowCov†	4870 to 4655		Mixed forage/agriculture
XingyiM13	Xingyi_LN	4843 to 4650		Mixed forage/agriculture
XingyiM12	Xingyi_LN	4612 to 4418	Second kinship with M11 and M13	Mixed forage/agriculture
XingyiM6	Com*	4569 to 4421		Mixed forage/agriculture
XingyiM14	Xingyi_LN	4520 to 4411	First kinship with M11	Mixed forage/agriculture
XingyiM7	Xingyi_LN	4515 to 4296		Mixed forage/agriculture
XingyiW4	Xingyi_BA	3894 to 3722	Second kinship with M5	Agriculturally dependent
XingyiM18	Xingyi_BA	3830 to 3646		Agriculturally dependent
XingyiM2	Xingyi_BA	3828 to 3581	Second kinship with M1	Agriculturally dependent
XingyiM19	Xingyi_BA	3822 to 3571	First kinship with M18	Agriculturally dependent
XingyiM3	Xingyi_BA	3811 to 3573		Agriculturally dependent
XingyiM5	Xingyi_BA	3811 to 3573		Agriculturally dependent
XingyiM17	Xingyi_BA	3692 to 3491		Agriculturally dependent
XingyiW3	Xingyi_BA	3570 to 3467		Agriculturally dependent
XingyiM4	Xingyi_BA	3565 to 3463	Second kinship with M5 and M18	Agriculturally dependent
XingyiM1	Xingyi_BA	3465 to 3385		Agriculturally dependent

*Com: Individual with evidence of high contamination who upon restricting to only damaged fragments, had less than 20,000 SNPs available for analysis. This individual was not included in population genetic analyses.
†LowCov: Individual has more than 20,000 SNPs but fewer than 50,000. This individual was only included in the PCA and ADMIXTURE and was not included in other demographic analyses.

compare the affinities of other ancient individuals from southern East Asia and Southeast Asia with the southern East Asian (Qihe3) and Central Yunnan (Xingyi_LN) ancestries. This approach revealed genetic affinity to Xingyi_LN in many ancient Southeast Asians (Fig. 4C). In particular, 4000- to 2000-year-old Southeast Asians from Laos and Vietnam [La_G2, Vt_G2 (*10*)] are more genetically similar to Xingyi_LN than Qihe3 in both outgroup f3 and f4 analyses (Fig. 2B and data S2e). When testing for admixture using qpAdm, Vt_G2 shows partial ancestry related to Xingyi_LN (66.4 to 73.4%; data S3c). By contrast, the 9000-year-old Dushan and the 8300- to 6400-year-old Baojianshan individuals from Guangxi (*8*) share greater affinity with Qihe3 in the regression symmetry f4 analysis (Fig. 4C) and cannot be modeled with ancestry related to Xingyi_LN (data S3c), a pattern also observed for ancient individuals from coastal and island southern East Asia in the Fujian region (*5*, *8*) and Taiwan (*5*, *7*). Although populations in Guangxi and eastern Yunnan showed substantial levels of southern East Asian ancestry, populations in central Yunnan 5500 to 1500 years ago and northern regions of Southeast Asia 4000 to 2000 years ago were instead highly affected by Central Yunnan ancestry.

Multiple East Asian ancestries in Yunnan during the Bronze Age

In western Yunnan, in the Tibetan-Yi corridor at the foothills of the Tibetan Plateau, 55 unrelated individuals with sufficient coverage from 10 sites radiocarbon dated to ~3800 to 1700 BP were analyzed, along with six previously published ~3000-year-old individuals from the Haimenkou site (*36*). Most sites are from the Jinsha River Basin, which composes the upper reaches of the Yangzi River. Unlike central or eastern Yunnan individuals, western Yunnan individuals cluster on the Tibetan cline in a PCA (Fig. 1C). Using phylogenetic (Fig. 2A and supplementary text section S6) and f4 analyses (Fig. 4A and fig. S20) to assess the genetic relationship of western Yunnan individuals dating to 3800 to 1700 BP to diverse East Asians, we found that they are most closely related to northern East Asians. This pattern includes those from the two oldest sites in western Yunnan, Baiyangcun and Dadunzi, and holds relative to populations carrying Central Yunnan ancestry, e.g., f4(Mbuti, Baiyangcun/Dadunzi; nEA, Xingyi_LN) < 0 (–9.3 < Z < 0.4; data S2f to S2g).

In contrast to ancient Tibetans, western Yunnan individuals do not show substantial levels of ancestry related to the Basal Asian individual, Xingyi_EN. One exception is the 2700- to 2300-year-old Jicha group, who share excess alleles with Xingyi_EN relative to Boshan (Fig. 3A and data S2c), but this signal is not recovered in a qpAdm analysis. To discern which northern East Asian ancestry is more related to the northern East Asian ancestry observed in ancient Tibetans, we performed an f4-symmetry analysis assessing affinity to the northern East Asian Yellow River and Shandong ancestries (Shimao_LN, Boshan) versus western Yunnan ancestry (Dadunzi, Baiyangcun), and found that ancient Tibetans tend to share slightly more genetic affinity with those of Yellow River and Shandong ancestries (figs. S18 and S19).

Although western Yunnan individuals all carry northern East Asian–related ancestry, they also show gene flow related to Central Yunnan ancestry. Successful qpAdm mixture models describe western Yunnan individuals as a mixture of ancestry related to northern East Asians (39.9 to 69.8%; data S3d) and Xingyi_LN. In a regression f4-symmetry analysis assessing affinity to Xingyi_LN relative to the southern East Asian

Qihe3 or the Southeast Asian La_G2, western Yunnan individuals share more affinity with Xingyi_LN (fig. S15). Furthermore, they share excess alleles with central Yunnan individuals relative to Qihe3 or La_G2, i.e., f4(Mbuti, Dadunzi/Baiyangcun; central Yunnan, Qihe3/La_G2) < 0 ($-8.7 < Z < -3.3$; data S2f to S2g). These patterns indicate some gene flow from central Yunnan into western Yunnan. Reciprocal gene flow from western Yunnan to central Yunnan seems to have also occurred but at much lower levels and less consistently (data S3d). The observed patterns indicate that there was interaction between western and central Yunnan populations, but maintenance of high population structure persisted throughout the Bronze Age, with those in western Yunnan showing substantial levels of northern East Asian ancestry that did not strongly affect populations in central Yunnan.

Impact of Central Yunnan ancestry in present-day populations, particularly Austroasiatic speakers

Because Central Yunnan ancestry is genetically distinct from northern and southern East Asian ancestries, we next investigated the extent to which Central Yunnan ancestry contributed to present-day East and Southeast Asians (data S1f). Focusing first on a symmetry f4 analysis examining genetic affinity to Central Yunnan ancestry (e.g., Xingyi_LN) versus southern East Asian Fujian ancestry (e.g., Qihe3), most mainland southern East Asians share greater genetic affinity to Xingyi_LN than Qihe3, whereas most Austronesian speakers share greater genetic affinity to Qihe3 (Fig. 4D and fig. S16). We then used an ADMIXTURE analysis to visualize the genetic ancestry of present-day Austroasiatic speakers and found that they are similar to ancient central Yunnan populations, whereas Austronesian speakers are more similar to those found in ancient southern East Asians (fig. S8). All successful admixture models using qpAdm for present-day Austroasiatic speakers (Mlabri, Lawa_Western, Mang, Bateq, CheWong, Jehai, Kintaq, Mendriq, Jakun, Seletar, Nicobarese, and Palaung) required substantial ancestry associated with Xingyi_LN (29.8 to 88.3%; fig. S22 and data S3e). Not all South Asian Austroasiatic speakers could be successfully modeled using qpAdm, but the Ho and Korwa showed partial ancestry related to Xingyi_LN. Some Sino-Tibetan (Naxi, Yi, KarenSkaw, and KarenPwo) and Kra-Dai (Laotian and Saek) speakers also required partial ancestry related to Xingyi_LN (data S3e). These patterns show the wide influence of Central Yunnan ancestry in today's mainland southern East Asian and Southeast Asian populations, particularly Austroasiatic speakers.

Focusing specifically on ethnic groups that currently live in Yunnan (*38*), we modeled admixture proportions using qpAdm and found that most showed primarily a mixture of Central Yunnan ancestry and northern East Asian ancestry, with two major patterns. First, ancestral proportions are dependent on the affiliated language groups, with Sino-Tibetan speakers (Hani, Bai, and Pumi) having higher levels of northern East Asian ancestry (43.9 to 83.3%) and Austroasiatic speakers (Blang and Wa) having higher levels of Central Yunnan ancestry (60.1 to 88%; data S3e). Second, Sino-Tibetan speakers from the Jinsha River Basin in the northwest show lower Central Yunnan ancestry (0 to 41.3%, Pumi and Bai), whereas Sino-Tibetan speakers from more southern regions (Hani) show higher Central Yunnan ancestry (56.1%, Hani; data S3e). These patterns suggest that the population structure in Yunnan is associated with different populations belonging to distinct language groups converging in Yunnan, where the 3800- to 1700-year-old western Yunnan populations were likely more similar to proto–Sino-Tibetan speakers, whereas the 5500- to 1400-year-old central Yunnan populations were likely more similar to proto-Austroasiatic speakers. These populations coexisted in Yunnan as far back as 3800 years ago, and population structure observed in these Bronze Age Yunnan populations can still be observed today. Gene flow did occur, because nearly all populations carry both northern East Asian and Central Yunnan ancestry, with a higher signature of gene flow in more southern regions of Yunnan.

Discussion

Determining the genetic ancestry of ancient humans from southwest China has greatly influenced our understanding of East and Southeast Asian prehistory. At the Xingyi archaeological site in central Yunnan, China, a 7100-year-old individual, Xingyi_EN, derives from a population that is closely related to an unsampled population who contributed to Tibetan origins. Xingyi_EN–related ancestry (denoted here as Basal Asian Xingyi ancestry) falls on the Asian branch of human prehistory but is distinct from those previously discovered in 40,000- to 33,000-year-old individuals from northern China [Tianyuan (*1*, *16*, *33*)] and 8000- to 6000-year-old individuals from Southeast Asia [Hòabìnhians (*10*)], indicating that Basal Asian Xingyi ancestry diverged at least 40,000 years ago. The distant relationship between Tianyuan, Xingyi_EN, Hòabìnhians, and other East and Southeast Asians suggests that the diversification of these deeply diverged Asian lineages likely occurred over a short period of time. Despite morphological features suggesting more archaic ancestry in some deeply diverged Asians (*39*), to date, the sampled deeply diverged Asians (*10*, *16*), including Xingyi_EN, group completely with anatomically modern humans, with no elevated levels of archaic ancestry.

In southern China and Southeast Asia, the earliest presence of Basal Asian Xingyi and Hòabìnhian ancestries (*10*) is still unknown, but by the Early to Middle Holocene, these two deeply diverged ancestries coexisted with populations carrying southern East Asian ancestry (*5*, *8*) and Central Yunnan ancestry, indicating the presence of high genetic diversity in southern East Asia at this time. Others have proposed that southern East Asia may have had greater population continuity during the Last Glacial Maximum because of the warmer and more hospitable environment (*40*). This may explain the persistence of deeply diverged Basal Asian populations in southern China and Southeast Asia to the mid-Holocene. By contrast, further to the north, full-scale replacement of deeply diverged Asian populations, e.g., Tianyuan (*16*), occured during the Pleistocene to Holocene transition (*1*). Thus, diversification of multiple populations on the Asian lineage before 40,000 years ago shaped much of the prehistory of East and Southeast Asia, with negligible influence from non-Asian populations. This pattern differs greatly from the population history of modern humans in Europe, where Holocene migrants from the Near East and Central Asia had a substantial impact on human ancestry in Europe (*41*–*43*).

The persistence of deeply diverged Basal Asian individuals further south led to gene flow of these ancestries into nearby populations. Contact between populations carrying Basal Asian Xingyi ancestry and populations carrying southern East Asian ancestry occurred as early as 11,000 years ago given the age of an admixed individual from Longlin Cave, Guangxi. Others have shown that Hòabìnhian ancestry could be found in partial amounts in early Southeast Asian farmers and across some isolated ethnic groups today (*10*). However, the greatest impact of Basal Asian Xingyi ancestry may have been on the Tibetan Plateau, where plateau populations are a mixture of a northern East Asian ancestry and a deeply diverged ghost ancestry (*2*, *3*) that is related to Basal Asian Xingyi ancestry. Our findings strongly support that Xingyi_EN–related ancestry contributed to Tibetan populations, although with only a single representative of this ancestry, further characterization of the relationship between Xingyi_EN and ghost ancestry in Tibetans is difficult. Xingyi_EN–related ancestry in Tibetans does not preclude the possibility that other deeply diverged, currently unknown ancestries also contributed to Tibetans, but additional sampling is needed to address this issue. Both archaeological surveys (*44*–*47*) and our findings show that Tibetans are more associated with populations from the Yellow River and Shandong regions than with populations from western Yunnan. This indicates that the geographic origin of admixture was likely further north, emphasizing the need for further sampling in the northeastern fringes of the Tibetan Plateau.

Comparison of 5000- to 3500-year-old central Yunnan populations, including those at Dian-associated archaeological sites [~2700 to 2000 BP (*48*, *49*)], to other ancient East Asians reveal an ancestry that differs substantially from northern and southern East Asian ancestries previously sampled (*1*, *5*, *6*, *8*). This ancestry, denoted here as Central Yunnan ancestry, contributed to many present-day populations in southern East Asia and Southeast Asia, particularly to Austroasiatic speakers. With the oldest northern East Asian individual dated to 19,000 years (*1*), the separation of prehistoric East Asians into different populations carrying northern East Asian–, southern East Asian–, or Central Yunnan–related ancestries must have been before 19,000 years ago. Populations carrying Central Yunnan ancestry extended to Southeast Asia by 4000 to 3000 BP, but showed little impact along coastal regions, where populations carrying southern East Asian ancestry persisted. Populations carrying southern East Asian ancestry already contributed partially to populations in Guangxi by the Early Neolithic (*5*, *8*), as well as populations in eastern Yunnan by 3500 years ago, perhaps facilitated by the Nanpan and other tributaries that fed into the Pearl River system.

Various locations across southern East Asia and Southeast Asia have been proposed for the origin of the Austroasiatic language family (*22*, *50*, *51*), and many have associated the expansion of Austroasiatic speakers with the expansion of rice agriculture (*18*, *20*, *21*). The ancestry related to present-day Austroasiatic speakers in ancient humans from central Yunnan, dating to 5500 years ago, is the oldest Austroasiatic-associated ancestry in ancient humans observed to date. Stable isotope analysis at the Xingyi site shows that the first appearance of domesticated crop consumption occurred no earlier than 4900 BP (*37*), but our fine-scale genetic sampling of ancient central Yunnan populations shows shared ancestry spanning 5500 to 3300 years BP. This suggests that the first appearance of the Austroasiatic-associated Central Yunnan ancestry preceded the adoption of agriculture. If proto-Austroasiatic populations expanded into Yunnan, then our findings suggest that they must have done so before 5500 years ago in an expansion unrelated to the adoption of agriculture in central Yunnan. If rice agriculture motivated a single, initial Austroasiatic expansion, then the central Yunnan region may be a candidate homeland for Austroasiatic origins given the associated age and local transition to farming. Alternatively, the ancestry found in proto-Austroasiatic populations may have already been widely dispersed across East and Southeast Asia such that these central Yunnan populations were perhaps tangential to the Austroasiatic expansion proposed for Southeast Asia and northeast India (*20*). The dominance of northern East Asian ancestry in the Sanjiang region in western Yunnan, a pattern also noted by others (*36*), supports linguistic and archaeological evidence (*21*, *27*) that argues against western Yunnan as a potential source of an Austroasiatic expansion. The Austroasiatic association in prehistoric central Yunnan populations suggests that the Red River valley that connects central Yunnan to northern Vietnam may have played an important role in the movement and interaction of proto-Austroasiatic populations. Sampling of Late Neolithic and Bronze Age humans in Yunnan indicates that a more nuanced model of Austroasiatic prehistory is needed, because current models do not incorporate ancient populations from central Yunnan.

The presence of northern East Asian–related ancestry in the Jinsha River Basin of western Yunnan, Central Yunnan ancestry in the Lake Dian region of central Yunnan, and a mixture of Central Yunnan and the southern East Asian Fujian ancestries in the Nanpan River Basin of southeastern Yunnan around 3500 years ago indicates that human genetic diversity in southwestern China was incredibly high in the Bronze Age. Today, ethnic groups in Yunnan also show high genetic diversity (*38*, *52*), and the remnants of genetic structure found in Bronze Age Yunnan populations can still be observed in Sino-Tibetan and Austroasiatic speakers, who carry mostly northern East Asian or Central Yunnan ancestry, respectively. Gene flow did affect these populations, because all present-day ethnic groups in Yunnan associate with both ancestries, but here we see that diverse ethnic groups coexisted in Yunnan as far back as 3800 years ago.

Our findings on the genetic relationships of ancient humans from southwest China over the past 7100 years provide important new insights into many long-standing questions concerning East Asian prehistory. Analysis of ancient genome-wide data from the 7100-year-old Xingyi_EN shows the existence of a second deeply diverged Asian population that lived in southern latitudes during the mid-Holocene and reveals a Basal Asian ancestry that affected ancient and present-day Tibetan populations (*2*, *3*). Central Yunnan populations dating back to 5500 BP represent a distinct East Asian ancestry related to Austroasiatic speakers of East and Southeast Asia, whereas western Yunnan populations dating back to 3800 BP show ancestry related to northern East Asians. In sum, the rich genetic history represented in the prehistoric humans of Yunnan in southwest China casts light on the complex population dynamics across East and Southeast Asia.

Materials and methods

Ethics and inclusion statement

Permission to test for ancient DNA in the human specimens from this study was obtained through discussions with local archaeologists who excavated them, with final approval granted by the institute in Yunnan where they are managed and cared for, the Yunnan Institute of Cultural Relics and Archaeology. Additional oversight and approval were obtained from the Institutional Review Board at the Institute of Vertebrate Paleontology and Paleoanthropology of the Chinese Academy of Sciences to sample the genomes of the ancient humans included in this study (202201060011). Protocols used to sample the genomes follow the highest standards used in archaeogenomic research. The work was done in collaboration with several local archaeologists, who were included as co-authors for their contributions to collation of archaeological material, dating of specimens, and/or discussions that contributed to the connections made to archaeological research cited in this study.

Ancient DNA extraction, sequencing, and data processing

We sampled 283 human remains from Yunnan, China, and of these, 147 individuals were successfully sequenced and analyzed in this study (data S1a). All extraction and sequencing procedures on ancient human samples were performed in a dedicated ancient DNA laboratory at the Institute of Vertebrate Paleontology and Paleoanthropology (IVPP), Chinese Academy of Sciences. Following a previously published protocol (*53*), a drill was used to obtain <100 mg of sample material from the bones of these individuals. Single-stranded libraries ("SS") without uracil-DNA-glycosylase (UDG) treatment were prepared for all samples. Libraries were then amplified for 35 cycles using AccuPrimepfx DNA enzymes (*5*).

To increase the amount of endogenous ancient DNA and reduce contamination from environmental DNA, oligonucleotide probes were synthesized to capture the ancient DNA fragments in solution. Oligonucleotide probes for mitochondrial DNA were designed based on a complete human mitochondrial genome (*33*). For nuclear DNA, nearly 1.2 million SNPs (the '1240k' panel) were used in a capture analysis (*32*) for 147 individuals. An additional 45 individuals were shotgun sequenced. Both nuclear capture on the 1240k panel and shotgun sequencing were used for seven remaining individuals. For most population genetic analyses, we focused on the 1240k panel to maximize comparisons with previously published ancient humans. We performed f4-statistics for the individual Xingyi_EN using shotgun data and capture data in parallel to confirm associated findings were robust to sequencing techniques. In RELATE and DATES analyses, we focused on shotgun data only, comparing to previously published ancient humans with shotgun data available.

For the enriched mitochondrial and nuclear DNA libraries, Illumina Miseq with 2×76 base pairs (bp) paired-end reads, and Illumina Hiseq4000 and Illumina HiSeq X Ten with 2×100 bp and 2×150 bp paired-end reads were used for sequencing, respectively. Then *leeHom* (*54*) (https://github.com/grenaud/leeHom) was applied to trim adaptors and merge paired reads with more than 11 base pairs to overlap into a single sequence. Only merged reads with lengths longer than 30 bp were kept and mapped to reference genomes using the Burrows-Wheeler Aligner [BWA, version 0.5.10-evan.10-1-gfb1ff83 (*55*)] with the *samse* command (-n 0.01 and -l 16500). Mitochondrial DNA was mapped to the revised Cambridge Reference Sequence [rCRS (*56*)], and nuclear DNA was mapped to the human reference genome *hg19* (*57*). For duplicate reads in the same orientation and start and end positions, the sequence with the highest quality was retained and duplicates were discarded.

We relied on C to T misincorporations to authenticate ancient DNA sequences (*58*) (data S1a), as deamination of cytosines into uracils can occur post-mortem, upon which the cytosine would be sequenced incorrectly as a thymine. Contamination rates were estimated using three approaches. For all samples, mitochondrial DNA contamination rates were analyzed by ContamMix (*59*), which compares the sequenced mitochondrial genomes with present-day sequences. Libraries with a contamination level higher than 5% were treated as contaminated. For males, because they only have one copy of the X chromosome, we also applied the ANGSD software (*60*) to test contamination levels. If the contamination rate was higher than 5%, the library was regarded as contaminated (data S1a). For females, we compared data restricted to damaged fragments only (female_d) with data from all fragments (female) in the same female individual using f4(Mbuti, X; females, female_d) (fig. S4). We also assessed autosomal contamination in both females and males using ContamLD version 1.0 (*61*), relying on the decay of linkage disequilibrium. Default settings were used for ContamLD analyses. For samples with contamination rates > 5% in any of the three contamination estimate tests (ChrX, contamLD, or mtDNA), we restricted to only damaged fragments for genotyping and used the individuals with more than 20,000 SNPs for further analyses. Details of all contamination outcomes are documented in the supplementary materials (data S1a).

We restricted analyses to fragments with damage characteristic of ancient DNA for libraries with high contamination to preserve as many individuals as possible for study. Five individuals ("_d" added to LibraryID, data S1a) originally showed contamination higher than 5% - for those individuals, we restricted our analyses to damaged fragments only, resulting in lower contamination estimates ranging from 0.2 to 3.1%. The damaged fragments were obtained by using *pmdtools 0.60* (*62*) with the --customterminus parameter, after 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.

After merging multiple libraries obtained for one individual, DuiziM98_1, we then called pseudo-haploid genotypes for all individuals. To do this, we randomly sampled one base for each SNP to which reads mapped to determine the corresponding allele. In this way, we generated pseudo-haploid genotypes (*32*). All newly sequenced individuals were prepared with SS "No-UDG" libraries. Thus, we masked both ends (5 bp) of the read and used "pileupCaller --singleStrandMode" with default data quality filters (i.e., base and map quality larger than 30) to determine the pseudo-haploid genotypes by randomly sampling one fragment per position after discarding forward mapping reads at C/T SNPs and reverse mapping reads at G/A SNPs.

Sex determination

We calculated genetic sex based on the fraction of the number of sequences that mapped to the Y chromosome compared with the overall number of sex chromosome sequences (*63*), where the sequences were filtered for a mapping quality of no less than 37. Results of assigned sex are presented in data S1a.

Uniparental haplogroups assignment

Mitochondrial haplogroups were determined using HaploGrep2 (*64*) with PhyloTree Build 17 (*65*) for mitochondrial consensus sequences. Y chromosome haplogroups were assigned by determining the most derived allele upstream and the most ancestral allele downstream in the phylogenetic tree, using the International Society of Genetic Genealogy dataset version 10.01 (www.isogg.org/tree). Both sets of haplogroups are listed in data S1a, and the related sex-biased practices are further discussed in supplementary text section S3.

Data imputation

To better refine the genotype likelihoods for low-coverage ancient DNA (*66*, *67*), we applied imputation analyses for IBD inference and RELATE analyses. All samples underwent no UDG treatment, so five base pairs were masked from each end of the read in the BAM files, using trimBam in bamUtil to decrease the influence of aDNA degradation influence on non-UDG-treated samples. Imputation analyses were performed using the GLIMPSE2 software (*66*, *68*), with the 1000 Genomes Phase 3 dataset used as the reference panel (http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/) with default parameters (https://odelaneau.github.io/GLIMPSE/docs/tutorials/getting_started/). Using these criteria, we imputed 24 shotgun sequences in parallel for RELATE analyses, including Xingyi_EN and previously published ancient Tibetans and northern East Asian individuals (supplementary text section S8).

IBD inference

We selected 74 ancient Yunnan individuals with high data quality to determine the sharing of identity-by-descent (IBD) sequences. This includes 31 individuals captured using the 1240K Panel covering at least 500,000 1240k SNPs, and 39 individuals with whole genome sequencing data covering at least 200,000 1240k SNPs, and seven individuals with both shotgun and capture data (*69*, *70*). After base and map quality control (above 30) and imputation, phased and imputed VCF files were merged together using bcftools and were processed using ancIBD software (*69*) with all recommended default parameters (https://ancibd.readthedocs.io/en/latest/run_ancIBD.html). Information concerning inferred IBD segments is recorded in data S1c.

Kinship and family pedigree reconstruction

We first investigated the degree of kinship between samples from the same site using the software READ (Relationship Estimation from Ancient DNA) (*71*), which can infer up to second-degree kinships for ancient individuals with low coverage and pseudo-haploid genotype calls. Individuals with less than 50,000 SNPs were not included in the READ analysis, as estimated kinship calls are often incorrect with low SNP availability. On the basis of the kinship relationships in READ (data S1b), mitochondrial and Y haplogroup (data S1a), genetic sex, IBD sharing (data S1c), and radiocarbon dating results (data S1a), we reconstructed family pedigrees among 21 individuals in two joint tombs, the Jianlianshan site in central Yunnan and the Gaozhai site in western Yunnan (supplementary text section S3). For further demographic genetic analyses, we excluded 21 individuals with kinship relationships to individuals who had more SNP data available.

ROH estimation

Runs of homozygosity (ROH) segments were estimated using the software hapROH which has demonstrated robust capability in detecting ROH longer than 4 cMs in ancient individuals with a minimum coverage of 0.3× (*72*). We selected 64 ancient Yunnan individuals whose SNP numbers were higher than 400,000 and analyzed them in hapROH using default parameters for all chromosomes. ROHs are recorded in data S1d, and kinship relatedness and background relatedness in ancient Yunnan are further discussed in supplementary text section S3.

Present-day datasets

Two different panels of present-day populations were used in different analyses. For PCA and ADMIXTURE, we assembled populations from the Human Origin (HO) SNP Panel (*73*), Tibetan and Han populations from Lu et al. (*11*), Southeast Asian populations from Liu *et al.* (*74*) and Kutanan et al. (*75*), and populations from southern China from Wang et al. (*7*). For f4 and qpAdm analyses, we included a panel of 1240k captured SNPs from the Simons Genome Diversity Panel [SGDP (*76*)], shotgun data from the Human Genome Diversity Project [HGDP (*77*)], and Tibetan populations from Lu et al. (*11*). For present-day populations in East Asia, Southeast Asia, and South Asia, language categories were assigned based on the original study in which they were sampled, as described in data S1f.

For qpAdm analyses targeting present-day populations, we additionally included: 1) the HO SNP panel (*73*) (~600,000 SNPs) for Southeast Asian populations from Liu et al. (*74*), Kutanan *et al.* (*75*), and Wang *et al.* (*7*); 2) an overlapping dataset (~305,000 SNPs) for present-day populations from India and Malaysia (*78*, *79*); and 3) an overlapping dataset (~175,000 SNPs) for present-day populations from Yunnan (*38*) (Dali_Bai, Honghe_Hani, Lincang_Wa, Nujiang_Pumi, Puer_Lahu, Xishuangbanna_Blang, supplementary text section S5).

PCA

We applied the *smartpca* program from the EIGENSOFT package (*80*) to perform a PCA with default parameter settings, except lsqproject: YES, numoutlieriter: 0, and shrinkmode: YES. The principal components were calculated using present-day East and Southeast Asians, with all newly sequenced or previously published ancient individuals projected onto the calculated components (Fig. 1C).

f-statistics

We conducted an outgroup f_3-analysis using *qp3Pop* (version 412) in AdmixTools (*73*), to measure the genetic similarities shared between populations relative to an outgroup. We used the present-day Central African Mbuti population as the outgroup and compared newly sequenced and previously published ancient individuals of Asia (X, Y), where the f_3-statistic is represented by f_3(*Mbuti; X, Y*) (Fig. 2, B and C). An f_4-analysis was performed using *qpDstat* (version 712) in AdmixTools (*73*) with optional parameter: "f4mode: YES". F_4-statistics of the form f_4(*Mbuti, X; Y, Z*), with Mbuti as the outgroup, were used in analyses to determine genetic groupings by archaeological site (data S2a-2b, supplementary text section S2), as well as for population genetic analyses to determine genetic relationships of newly sequenced humans to a diverse array of present-day and ancient humans, primarily from Asia (data S2c-2h). Additionally, symmetry tests designed to examine whether individuals or populations shared greater affinity to one of two different ancestries (*8*, *81*) were used to examine the relationship of ancient and present-day populations to different northern and southern populations of East Asia, with details described further in supplementary text section S4. In both f_3- and f_4-analyses, frequency data were used when there is more than one individual in a group; otherwise, 0/1 counts are used for the analyses (*82*). The group clustering and more description of the f-statistics highlighted in this study can be found in supplementary text sections S2 and S4.

Inferring admixture and estimating mixture proportions

To model ancestry proportions for any target populations, we applied *qpAdm* (*41*) in AdmixTools (*73*) with the parameter "allsnps: YES". We developed a standard set of outgroups called the 'Right populations' and a potential set of source populations for the 'Left populations' as described for the qpAdm software. We selected representative populations of different lineages of modern humans worldwide, including Mota (*83*), Kostenki14 (*43*, *84*), Iran_N (*43*), IndusPeriphery (*85*), LBK_EN (*41*, *86*, *87*), Kotias (*88*), Karelia (*41*, *87*), Yana (*89*), AR33K (*1*), Kolyma (*89*), Onge (*76*), AR19K (*1*), Qihe3 (*8*), and Yumin (*5*) as the distal "Right populations" set. We used two different analyses to address specific questions: 1) To address the question of ghost ancestry previously found in ancient and present-day Tibetan Plateau populations (*2*–*4*, *11*), we tested a broad set of deeply diverged Asians as putative sources [UstIshim (*90*), Tianyuan (*16*, *33*), Xingyi_EN, La368 (*10*), Jomon8.8k (*91*)] for ancient and present-day Tibetans, fixing a second source to be an ancient population primarily of northern East Asian ancestry [fixing AR14K (*1*), Boshan (*5*), and Shimao_LN (*6*) in parallel] in data S3a. 2) To assess ancestries across a comprehensive set of ancient and present-day East and Southeast Asians, we developed a diverse set of representative Asian sources from this study and previous studies who were shown to contribute to ancient and present-day East and Southeast Asians from the past 11,000 years in data S3b-3e. The source set included representative lineages from East and Southeast Asia as potential sources, including Xingyi_EN, La368 (*10*), Jomon8.8k (*91*), AR14K (*1*), Boshan (*5*), Shimao_LN (*6*), Liangdao2 (*5*), and Xingyi_LN. For both qpAdm analyses, the potential sources were included using a rotating qpAdm algorithm, where sources not included in the immediate qpAdm analysis were rotated into the outgroup (*92*). A detailed description of how we estimated mixture proportions, along with descriptions of the standard outgroups and potential source populations can be found in supplementary text section S5.

Admixture Graph modeling

We applied *qpGraph* from ADMIXTOOLS (*73*) to model the relationship between populations by using the "*qpgraph*" function from the ADMIXTOOL2 package (*93*). We first ran fully automated graph explorations in *findGraphs* (*93*) allowing admixture from zero to eight in 100 algorithm iterations. Then we constrained deeply diverged Asians as non-admixed to further test the model in 100 algorithm iterations, allowing admixture events from zero to three. To construct the Admixture Graph, we included the Central African Mbuti, the ~36,200-year-old early European Kostenki14 (*84*), the ~40,000-year-old early Asian Tianyuan (*16*), the ~8000-year-old Hòabìnhian from Southeast Asia La368 (*10*), the ~11,000-year-old early East Asian Longlin (*8*), the ~12,000-year-old southern East Asian Qihe3 (*8*), the ~8300-year-old northern East Asian Boshan (*5*), the ~5100-year-old Tibetan Zongri5.1k (*2*), and the newly sequenced Xingyi_EN and Xingyi_LN. Details on Admixture Graph modeling can be found in supplementary text section S7.

RELATE

We used Relate v1.2.1 (*34*) to infer whole genome genealogies using a per base per generation mutation rate of 4e-9, and pre-inferred average coalescence rates for the Simons Genome Diversity Project [SGDP (*76*)]. We merged 24 imputed ancient shotgun genomes (coverages 0.3× to 6.8×) of Xingyi_EN, ancient Tibetan, and ancient northern East Asian ancestries, with present-day genomes available through the SGDP, and subsequently restricted our analysis to transversions and sites where all imputed ancient individuals had an imputation posterior exceeding 0.8. We applied the 'strict mask' setting available through the 1000 Genomes Project. Given these inferred genealogies, we inferred effective population sizes and cross coalescence rates. We constructed ancestry models using a genealogy-based chromosome painting approach to validate the ghost ancestry in ancient Tibetans detected using f-statistics. For more details, see supplementary text section S8.

Estimating a maximum likelihood phylogeny with migration events

We used TreeMix (*94*) (version 1.13) to infer the maximum likelihood phylogeny with migration events. The trees were rooted using the Central African Mbuti, and we set each block to contain 500 SNPs with the option "-root Mbuti –k 500 -global" (supplementary text section S6). The bootstrap values for each node in the tree (Fig. 2A) were obtained by running 1000 replicates with the option "-bootstrap -q".

Bootstrap trees were assessed in *phylip* with the command *consense* (*95*). Populations included in the Treemix analysis are described in supplementary text section S6.

ADMIXTURE analysis

We used ADMIXTURE (*96*) to elucidate the population structure within our recently sequenced individuals, encompassing a total of 938 present-day and 325 ancient individuals. Before the analysis, we pruned genotypes for high linkage disequilibrium ($r^2 > 0.4$), excluded SNPs with a missing rate exceeding 20%, and filtered out SNPs with minor allele frequencies below 5% using PLINK (*97*), using the parameters "--geno 0.2 --maf 0.05 --indep-pairwise 200 25 0.4". This process retained 234,719 SNPs. ADMIXTURE analysis was conducted across K values ranging from 2 to 10 (fig. S7). For each K value, we executed 100 runs with a randomized seed to pinpoint the K values associated with the lowest cross-validation (CV) errors (fig. S6). As K=5 yielded the minimum CV errors, we visualized these ADMIXTURE results in fig. S8.

Estimating the timing of admixture

We used DATES (*98*) (version 4010, https://github.com/MoorjaniLab/DATES_v4010) to estimate the timing of an admixture event between the ancestral population for ancient Tibetans and a population related to Xingyi_EN, using shotgun data for Xingyi_EN, ancient Tibetans, and northern East Asians. We started with a genetic distance of 0.45 cM with "lovafit: 0.45," and the maximum genetic distance was set to the maximum of one Morgan with "maxdis: 1" to make sure it is greater than the admixture LD blocks. We used the recommended optimal binsize of 0.001 Morgan with "binsize: 0.001." The standard error was estimated by performing a weighted block jackknife with the parameter "jackknife: YES". Results with NRMSD < 0.7 were taken into consideration. We assumed one generation corresponds to 29 years (*99*) to convert the generation results into years. Results primarily estimated times older than 200 generations (data S3f), indicating that DATES could not infer a time of admixture for Basal Asian Xingyi and northern East Asian ancestries in plateau populations. Because the oldest available ancient Tibetan genome is 5100 BP, this suggests that the timing of admixture from a Xingyi_EN-related population into a past northern East Asian population was before 11,000 years ago.

Archaic ancestry estimation

We applied admixfrog (*100*) (version 0.5.6, https://github.com/BenjaminPeter/admixfrog/) to estimate archaic introgressed fragments in the genomes of ancient populations from Yunnan. First, each target individual was converted from the BAM file format to the input file format of admixfrog. All BAM files were filtered for fragments of at least 35 base pairs (--length-bin-size 35), mapping quality greater than 25 (--minmapq 25), and deaminated C→T substitutions were removed at the first and last three bases (--deam-cutoff 3). Second, we inferred introgressed archaic fragments using generated input files. The potential sources were set to be Africans, Neanderthals, or Denisovans (--states AFR NEA DEN), described in supplementary text section S9. The bin size for each individual was set to 5000 bp (--bin-size 5000). The chimpanzee (panTro4) reference genome was used to infer the ancestral state of each allele (--ancestral PAN). Results are shown in data S1e.

REFERENCES AND NOTES

1. X. Mao *et al.*, The deep population history of northern East Asia from the Late Pleistocene to the Holocene. *Cell* **184**, 3256–3266.e13 (2021). doi: 10.1016/j.cell.2021.04.040; pmid: 34048699
2. H. Wang *et al.*, Human genetic history on the Tibetan Plateau in the past 5100 years. *Sci. Adv.* **9**, eadd5582 (2023). doi: 10.1126/sciadv.add5582; pmid: 36930720
3. C.-C. Liu *et al.*, Ancient genomes from the Himalayas illuminate the genetic history of Tibetans and their Tibeto-Burman speaking neighbors. *Nat. Commun.* **13**, 1203 (2022). doi: 10.1038/s41467-022-28827-2; pmid: 35260549
4. C. Jeong *et al.*, Long-term genetic stability and a high-altitude East Asian origin for the peoples of the high valleys of the Himalayan arc. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 7485–7490 (2016). doi: 10.1073/pnas.1520844113; pmid: 27325755
5. M. A. Yang *et al.*, Ancient DNA indicates human population shifts and admixture in northern and southern China. *Science* **369**, 282–288 (2020). doi: 10.1126/science.aba0909; pmid: 32409524
6. C. Ning *et al.*, Ancient genomes from northern China suggest links between subsistence changes and human migration. *Nat. Commun.* **11**, 2700 (2020). doi: 10.1038/s41467-020-16557-2; pmid: 32483115
7. C. C. Wang *et al.*, Genomic insights into the formation of human populations in East Asia. *Nature* **591**, 413–419 (2021). doi: 10.1038/s41586-021-03336-2; pmid: 33618348
8. T. Wang *et al.*, Human population history at the crossroads of East and Southeast Asia since 11,000 years ago. *Cell* **184**, 3829–3841.e21 (2021). doi: 10.1016/j.cell.2021.05.018; pmid: 34171307
9. M. Lipson *et al.*, Ancient genomes document multiple waves of migration in Southeast Asian prehistory. *Science* **361**, 92–95 (2018). doi: 10.1126/science.aat3188; pmid: 29773666
10. H. McColl *et al.*, The prehistoric peopling of Southeast Asia. *Science* **361**, 88–92 (2018). doi: 10.1126/science.aat3628; pmid: 29976827
11. D. Lu *et al.*, Ancestral Origins and Genetic History of Tibetan Highlanders. *Am. J. Hum. Genet.* **99**, 580–594 (2016). doi: 10.1016/j.ajhg.2016.07.002; pmid: 27569548
12. E. Huerta-Sánchez *et al.*, Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* **512**, 194–197 (2014). doi: 10.1038/nature13408; pmid: 25043035
13. F. Chen *et al.*, A late Middle Pleistocene Denisovan mandible from the Tibetan Plateau. *Nature* **569**, 409–412 (2019). doi: 10.1038/s41586-019-1139-x; pmid: 31043746
14. D. Zhang *et al.*, Denisovan DNA in Late Pleistocene sediments from Baishiya Karst Cave on the Tibetan Plateau. *Science* **370**, 584–587 (2020). doi: 10.1126/science.abb6320; pmid: 33122381
15. H. Hu *et al.*, Evolutionary history of Tibetans inferred from whole-genome sequencing. *PLOS Genet.* **13**, e1006675 (2017). doi: 10.1371/journal.pgen.1006675; pmid: 28448578
16. M. A. Yang *et al.*, 40,000-year-old individual from Asia provides insight into early population structure in Eurasia. *Curr. Biol.* **27**, 3202–3208.e9 (2017). doi: 10.1016/j.cub.2017.09.030; pmid: 29033327
17. F. Rispoli, "The expansion of rice and millet farmers into Southeast Asia," in *The Oxford Handbook of Early Southeast Asia*, C. F. W. Higham, N. C. Kim, eds. (Oxford Univ. Press, 2022), pp. 339–359.
18. C. F. W. Higham, "The Neolithic of mainland Southeast Asia," in *The Oxford Handbook of Early Southeast Asia*, C. F. W. Higham, N. C. Kim, Eds. (Oxford Univ. Press, 2022), pp. 360–375.
19. D. Q. Fuller *et al.*, Consilience of genetics and archaeobotany in the entangled history of rice. *Archaeol. Anthropol. Sci.* **2**, 115–131 (2010). doi: 10.1007/s12520-010-0035-y
20. R. Blench, *Reconstructing Austroasiatic Prehistory* (Brill, 2015).
21. L. Sagart, "The Austroasiatics: East to west or west to east? in *Dynamics of Human Diversity: The Case Of Mainland Southeast Asia*, N. Enfield, Ed. (Pacific Linguistics, 2011), chap. 15, pp. 345–361.
22. G. van Driem, *Languages of the Himalayas: An Ethnolinguistic Handbook of the Greater Himalayan Region containing an Introduction to the Symbiotic Theory of Language* (Brill, 2001).
23. R. Blust, Beyond the Austronesian homeland: The Austric hypothesis and its implications for archaeology. *Trans. Am. Philos. Soc.* **86**, 117–158 (1996). doi: 10.2307/1006623
24. G. van Driem, The ethnolinguistic identity of the domesticators of Asian rice. *C. R. Palevol* **11**, 117–132 (2012). doi: 10.1016/j.crpv.2011.07.004
25. G. van Driem, "The domestications and the domesticators of Asian rice," in *Language Dispersal Beyond Farming*, M. Robbeets, A. Savelyev, Eds. (John Benjamins Publishing, 2017), pp. 183–214.
26. L. A. Reid, "New linguistic evidence for the Austric hypothesis," paper presented at the Eighth International Conference on Austronesian Linguistics, Taipei, Taiwan, 28–30 December 1997.
27. R. Dal Martello, The origins of multi-cropping agriculture in Southwestern China: Archaeobotanical insights from third to first millennium B.C. Yunnan. *Asian Archaeol.* **6**, 65–85 (2022). doi: 10.1007/s41826-022-00052-2; pmid: 35971515
28. X. Ji *et al.*, The oldest Hoabinhian technocomplex in Asia (43.5 ka) at Xiaodong rockshelter, Yunnan Province, southwest China. *Quat. Int.* **400**, 166–174 (2016). doi: 10.1016/j.quaint.2015.09.080
29. Y. Wu *et al.*, The Hoabinhian technocomplex in southwest China: Preliminary report on new discoveries in recent decades. *Anthropologie* **128**, 103234 (2024). doi: 10.1016/j.anthro.2024.103234
30. H. C. Hung, Z. Chi, H. Matsumura, L. Zhen, "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*, H. Matsumura, H. Hung, L. Zhen, K. Shinoda, Eds. (National Museum of Nature and Science, Monograph No. 47, 2017), pp. 205–228.
31. Institute of Archaeology Chinese Academy of Social Sciences, *Guilin Zengpiyan* (Cultural Relics Publishing House, 2003).

32. Q. Fu *et al.*, An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216–219 (2015). doi: 10.1038/nature14558; pmid: 26098372
33. Q. Fu *et al.*, DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 2223–2227 (2013). doi: 10.1073/pnas.1221359110; pmid: 23341637
34. L. Speidel, M. Forest, S. Shi, S. R. Myers, A method for genome-wide genealogy estimation for thousands of samples. *Nat. Genet.* **51**, 1321–1329 (2019). doi: 10.1038/s41588-019-0484-x; pmid: 31477933
35. T. Chiou-Peng, "The Dian culture in southwest China" in *The Oxford Handbook of Early Southeast Asia*, C. F. W. Higham, N. C. Kim, Eds. (Oxford Univ. Press, 2022), pp. 597–624.
36. L. Tao *et al.*, Ancient genomes reveal millet farming-related demic diffusion from the Yellow River into southwest China. *Curr. Biol.* **33**, 4995–5002.e7 (2023). doi: 10.1016/j.cub.2023.09.055; pmid: 37852263
37. M. Ma *et al.*, Forager-farmer transition at the crossroads of East and Southeast Asia 4900 years ago. *Sci. Bull.* **69**, 103–113 (2024). doi: 10.1016/j.scib.2023.10.015; pmid: 37914610
38. J. Guo *et al.*, Genomic insights into Neolithic farming-related migrations in the junction of east and southeast Asia. *Am. J. Biol. Anthropol.* **177**, 328–342 (2022). doi: 10.1002/ajpa.24434
39. H. Shang, H. Tong, S. Zhang, F. Chen, E. Trinkaus, An early modern human from Tianyuan Cave, Zhoukoudian, China. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 6573–6578 (2007). doi: 10.1073/pnas.0702169104; pmid: 17416672
40. R. Dennell, M. Martinón-Torres, J.-M. Bermúdez de Castro, G. Xing, A demographic history of Late Pleistocene China. *Quat. Int.* **559**, 4–13 (2020). doi: 10.1016/j.quaint.2020.03.014
41. W. Haak *et al.*, Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015). doi: 10.1038/nature14317; pmid: 25731166
42. I. Lazaridis *et al.*, Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014). doi: 10.1038/nature13673; pmid: 25230663
43. I. Lazaridis *et al.*, Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**, 419–424 (2016). doi: 10.1038/nature19310; pmid: 27459054
44. P. J. Brantingham *et al.*, "A short chronology for the peopling of the Tibetan Plateau" in *Developments in Quaternary Sciences* (Elsevier, 2007), vol. 9, pp. 129–150.
45. D. B. Madsen *et al.*, The Late Upper Paleolithic occupation of the northern Tibetan Plateau margin. *J. Archaeol. Sci.* **33**, 1433–1444 (2006). doi: 10.1016/j.jas.2006.01.017
46. P. Zhang *et al.*, Denisovans and Homo sapiens on the Tibetan Plateau: Dispersals and adaptations. *Trends Ecol. Evol.* **37**, 257–267 (2022). doi: 10.1016/j.tree.2021.11.004; pmid: 34863581
47. D. B. Madsen *et al.*, Early foraging settlement of the Tibetan Plateau highlands. *Archaeological Research in Asia* **11**, 15–26 (2017). doi: 10.1016/j.ara.2017.04.003
48. E. Luo, Y. Li, *Archaeology of Southwest China: Neolithic to Western Han* (Science Press, 2020).
49. Z. Zhang, *Dian Kingdom and Dian culture* (Yunnan Fine Arts Press, 1997).
50. L. Sagart, "Language families of southeast Asia" in *The Oxford Handbook of Early Southeast Asia*, C. F. W. Higham, N. C. Kim, Eds. (Oxford Univ. Press, 2022), pp. 321–338.
51. P. Sidwell, R. Blench, "The Austroasiatic Urheimat: The southeastern riverine hypothesis" in *Dynamics of Human Diversity: The Case of Mainland Southeast Asia*, N. J. Enfield, Ed. (Australian National University, 2011), pp. 315–343.
52. Z. Zhang *et al.*, The Tibetan-Yi region is both a corridor and a barrier for human gene flow. *Cell Rep.* **39**, 110720 (2022). doi: 10.1016/j.celrep.2022.110720; pmid: 35476999
53. J. Dabney *et al.*, Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 15758–15763 (2013). doi: 10.1073/pnas.1314445110; pmid: 24019490
54. G. Renaud, U. Stenzel, J. Kelso, leeHom: Adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Res.* **42**, e141 (2014). doi: 10.1093/nar/gku699; pmid: 25100869
55. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009). doi: 10.1093/bioinformatics/btp324; pmid: 19451168
56. R. M. Andrews *et al.*, Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **23**, 147 (1999). doi: 10.1038/13779; pmid: 10508508
57. D. M. Church *et al.*, Modernizing reference genome assemblies. *PLOS Biol.* **9**, e1001091 (2011). doi: 10.1371/journal.pbio.1001091; pmid: 21750661
58. S. Sawyer, J. Krause, K. Guschanski, V. Savolainen, S. Pääbo, Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLOS ONE* **7**, e34131 (2012). doi: 10.1371/journal.pone.0034131; pmid: 22479540
59. Q. Fu *et al.*, A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* **23**, 553–559 (2013). doi: 10.1016/j.cub.2013.02.044; pmid: 23523248
60. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: Analysis of Next Generation Sequencing data. *BMC Bioinformatics* **15**, 356 (2014). doi: 10.1186/s12859-014-0356-4; pmid: 25420514
61. N. Nakatsuka *et al.*, ContamLD: Estimation of ancient nuclear DNA contamination using breakdown of linkage disequilibrium. *Genome Biol.* **21**, 199 (2020). doi: 10.1186/s13059-020-02111-2; pmid: 32778142
62. P. Skoglund *et al.*, Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 2229–2234 (2014). doi: 10.1073/pnas.1318934111; pmid: 24469802
63. P. Skoglund, J. Storå, A. Götherström, M. Jakobsson, Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J. Archaeol. Sci.* **40**, 4477–4482 (2013). doi: 10.1016/j.jas.2013.07.004
64. H. Weissensteiner *et al.*, HaploGrep 2: Mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* **44**, W58–W63 (2016). doi: 10.1093/nar/gkw233; pmid: 27084951
65. M. van Oven, M. Kayser, Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* **30**, E386–E394 (2009). doi: 10.1002/humu.20921; pmid: 18853457
66. S. Rubinacci, D. M. Ribeiro, R. J. Hofmeister, O. Delaneau, Efficient phasing and imputation of low-coverage sequencing data using large reference panels. *Nat. Genet.* **53**, 120–126 (2021). doi: 10.1038/s41588-020-00756-0; pmid: 33414550
67. B. Sousa da Mota *et al.*, Imputation of ancient human genomes. *Nat. Commun.* **14**, 3660 (2023). doi: 10.1038/s41467-023-39202-0; pmid: 37339987
68. S. Rubinacci, R. J. Hofmeister, B. Sousa da Mota, O. Delaneau, Imputation of low-coverage sequencing data from 150,119 UK Biobank genomes. *Nat. Genet.* **55**, 1088–1090 (2023). doi: 10.1038/s41588-023-01438-3; pmid: 37386250
69. H. Ringbauer *et al.*, Accurate detection of identity-by-descent segments in human ancient DNA. *Nat. Genet.* **56**, 143–151 (2024). doi: 10.1038/s41588-023-01582-w; pmid: 38123640
70. M. Rivollat *et al.*, Extensive pedigrees reveal the social organization of a Neolithic community. *Nature* **620**, 600–606 (2023). doi: 10.1038/s41586-023-06350-8; pmid: 37495691
71. J. M. Monroy Kuhn, M. Jakobsson, T. Günther, Estimating genetic kin relationships in prehistoric populations. *PLOS ONE* **13**, e0195491 (2018). doi: 10.1371/journal.pone.0195491; pmid: 29684051
72. H. Ringbauer, J. Novembre, M. Steinrücken, Parental relatedness through time revealed by runs of homozygosity in ancient DNA. *Nat. Commun.* **12**, 5425 (2021). doi: 10.1038/s41467-021-25289-w; pmid: 34521843
73. N. Patterson *et al.*, Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012). doi: 10.1534/genetics.112.145037; pmid: 22960212
74. D. Liu *et al.*, Extensive ethnolinguistic diversity in Vietnam reflects multiple sources of genetic diversity. *Mol. Biol. Evol.* **37**, 2503–2519 (2020). doi: 10.1093/molbev/msaa099; pmid: 32344428
75. W. Kutanan *et al.*, Reconstructing the Human Genetic History of Mainland Southeast Asia: Insights from Genome-Wide data from Thailand and Laos. *Mol. Biol. Evol.* **38**, 3459–3477 (2021). doi: 10.1093/molbev/msab124; pmid: 33905512
76. S. Mallick *et al.*, The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016). doi: 10.1038/nature18964; pmid: 27654912
77. J. Z. Li *et al.*, Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319**, 1100–1104 (2008). doi: 10.1126/science.1153717; pmid: 18292342
78. D. Tagore, F. Aghakhanian, R. Naidu, M. E. Phipps, A. Basu, Insights into the demographic history of Asia from common ancestry and admixture in the genomic landscape of present-day Austroasiatic speakers. *BMC Biol.* **19**, 61 (2021). doi: 10.1186/s12915-021-00981-x; pmid: 33781248
79. F. Aghakhanian *et al.*, Unravelling the genetic history of Negritos and indigenous populations of Southeast Asia. *Genome Biol. Evol.* **7**, 1206–1215 (2015). doi: 10.1093/gbe/evv065; pmid: 25877615
80. N. Patterson, A. L. Price, D. Reich, Population structure and eigenanalysis. *PLOS Genet.* **2**, e190 (2006). doi: 10.1371/journal.pgen.0020190; pmid: 17194218
81. M. Lipson *et al.*, Three phases of ancient migration shaped the ancestry of human populations in Vanuatu. *Curr. Biol.* **30**, 4846–4856.e6 (2020). doi: 10.1016/j.cub.2020.09.035; pmid: 33065004
82. E. Y. Durand, N. Patterson, D. Reich, M. Slatkin, Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* **28**, 2239–2252 (2011). doi: 10.1093/molbev/msr048; pmid: 21325092
83. M. G. Llorente *et al.*, Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent. *Science* **350**, 820–822 (2015). doi: 10.1126/science.aad2879; pmid: 26449472
84. A. Seguin-Orlando *et al.*, Genomic structure in Europeans dating back at least 36,200 years. *Science* **346**, 1113–1118 (2014). doi: 10.1126/science.aaa0114; pmid: 25378462
85. V. M. Narasimhan *et al.*, The formation of human populations in South and Central Asia. *Science* **365**, eaat7487 (2019). doi: 10.1126/science.aat7487; pmid: 31488661

86. M. Lipson *et al.*, Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature* **551**, 368–372 (2017). doi: 10.1038/nature24476; pmid: 29144465
87. I. Mathieson *et al.*, Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015). doi: 10.1038/nature16152; pmid: 26595274
88. E. R. Jones *et al.*, Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* **6**, 8912 (2015). doi: 10.1038/ncomms9912; pmid: 26567969
89. M. Sikora *et al.*, The population history of northeastern Siberia since the Pleistocene. *Nature* **570**, 182–188 (2019). doi: 10.1038/s41586-019-1279-z; pmid: 31168093
90. Q. Fu *et al.*, Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445–449 (2014). doi: 10.1038/nature13810; pmid: 25341783
91. N. P. Cooke *et al.*, Ancient genomics reveals tripartite origins of Japanese populations. *Sci. Adv.* **7**, eabh2419 (2021). doi: 10.1126/sciadv.abh2419; pmid: 34533991
92. É. Harney, N. Patterson, D. Reich, J. Wakeley, Assessing the performance of qpAdm: A statistical tool for studying population admixture. *Genetics* **217**, iyaa045 (2021). doi: 10.1093/genetics/iyaa045; pmid: 33772284
93. R. Maier *et al.*, On the limits of fitting complex models of population history to *f*-statistics. *eLife* **12**, e85492 (2023). doi: 10.7554/eLife.85492; pmid: 37057893
94. J. K. Pickrell, J. K. Pritchard, Inference of population splits and mixtures from genome-wide allele frequency data. *PLOS Genet.* **8**, e1002967 (2012). doi: 10.1371/journal.pgen.1002967; pmid: 23166502
95. B. R. Baum, PHYLIP: Phylogeny Inference Package Version 3.2. Joel Felsenstein. *Q. Rev. Biol.* **64**, 539–541 (1989). doi: 10.1086/416571
96. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009). doi: 10.1101/gr.094052.109; pmid: 19648217
97. S. Purcell *et al.*, PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007). doi: 10.1086/519795; pmid: 17701901
98. M. Chintalapati, N. Patterson, P. Moorjani, The spatiotemporal patterns of major human admixture events during the European Holocene. *eLife* **11**, e77625 (2022). doi: 10.7554/eLife.77625; pmid: 35635751
99. P. Moorjani *et al.*, A genetic method for dating ancient genomes provides a direct estimate of human generation interval in the last 45,000 years. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 5652–5657 (2016). doi: 10.1073/pnas.1514696113; pmid: 27140627
100. B. M. Peter, 100,000 years of gene flow between Neandertals and Denisovans in the Altai mountains. bioRxiv 990523 [Preprint] (2020); https://doi.org/10.1101/2020.03.13.990523.
101. T. Chen *et al.*, The genome sequence archive family: Toward explosive data growth and diverse data types. *Genomics Proteomics Bioinformatics* **19**, 578–583 (2021). doi: 10.1016/j.gpb.2021.08.001; pmid: 34400360
102. X. Bai *et al.*, database Resources of the National Genomics data Center, China National Center for Bioinformation in 2024. *Nucleic Acids Res.* **52**, D18–D32 (2024). doi: 10.1093/nar/gkad1078; pmid: 38018256
103. X. Wei *et al.*, Neolithic to Bronze Age human maternal genetic history in Yunnan, China. *J. Genet. Genomics* **52**, 483–493 (2025). doi: 10.1016/j.jgg.2024.09.013; pmid: 39343094

ACKNOWLEDGMENTS

We gratefully acknowledge archaeological teams from Yunnan for their valuable support and thank Z. Chen for improving the summary figure. **Funding:** This work was supported by the National Natural Science Foundation of China (grant 41925009) and the Chinese Academy of Sciences (grant YSBR-019). L.S. was supported by a Sir Henry Wellcome Fellowship (220457/Z/20/Z). X.W. was supported by the Key National Social Science Foundation of China (grant 16ZDA144). **Author contributions:** Q.F. designed and supervised the research project. T.W., Z.Z., M.M., R.M., H.Y., Z.J., C.H., X.Li, D.Z., X.W., X.Liu, M.Z., Y.W., F.Y., R.Z., L.K., G.D., and Q.F. assembled archaeological materials and performed dating. C.P., F.L., Q.D., R.Y., W.P., and Q.F. performed or supervised wet laboratory work. X.F. and Q.F. processed the data. T.W., M.A.Y., H. S., L.S., F.B., and Q.F. analyzed the data. T.W., M.A.Y., and Q.F. wrote the manuscript. T.W., M.A.Y., M.M., H.S., L.S, and Q.F. wrote the manuscript supplements. T.W., M.A.Y, Y.L., M.S., and Q.F. revised the manuscript. All authors discussed, critically revised, and approved the final version of the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** We also generated six shotgun data for the published ancient genomes, five from the Tibetan Plateau (*2*) (Zongri5.1k, Yushu2.8k, Nagqu2.5k, Nagqu1.1k, and Shigatse0.7k) and one from the Amur River (*1*) (AR19K). The aligned sequences and genotype calls are available through the Genome Sequence Archive (*101*) in the National Genomics data Center (*102*) (https://ngdc.cncb.ac.cn/gsa-human/; accession number: PRJCA015361). The pseudohaploid genotyped data (Eigenstrat format) are available in OMIX, from the China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences (https://ngdc.cncb.ac.cn/omix/; accession number: PRJCA015361). Mitochondrial DNA (fasta and bam format) have been published (*103*). All software used are freely available online.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adq9792
Materials and Methods; Supplementary Text Sections S1 to S9; Figs. S1 to S35; Tables S1 to S3; References (*104*–*168*)

Submitted 10 June 2024; accepted 14 March 2025

10.1126/science.adq9792