

Ancient DNA reveals the population interactions and a Neolithic patrilineal community in Northern Yangtze Region

Received: 21 October 2024

Accepted: 27 August 2025

Published online: 30 September 2025

 Check for updates

Tingyu Yang^{1,2,9}, Jianing He^{3,9}, Chunmei Li^{1,4,9}, Chao Ning^{1,3,9}, Tianming Zhang¹, Jincheng Wang¹, Hai Zhang^{1,3}, Guanbo Wang^{1,5}, Xiaowen Jia³, Chi Zhang³✉, Shi Yan⁶, Xiannian Zhang¹, Fan Wu¹, Chaodong Wu^{1,7}, Xiaohong Wu^{1,3}✉, Yuhong Pang¹✉ & Yanyi Huang^{1,2,5,8}✉

The scarcity of ancient genomes of early rice cultivators has hindered our understanding of their genetic profiles and dynamic interactions with millet cultivators. In this study, we analyzed 58 ancient genomes from Baligang, a long-term settlement situated on the northern rim of the Middle Yangtze River Basin, spanning from the Middle Neolithic to the Late Bronze Age. This chronostratified genetic dataset offers comprehensive genomic insights into populations engaged in early rice cultivation, enhancing our understanding of East Asian demographic history. Our study revealed that Baligang experienced successive waves of admixture over time, with ~4,200 BP as a critical transition point, marked by a significant genomic influx of southern East Asian populations. We also identify the detailed kinship relations within multi-generational secondary burial and the existence of patrilineal communities dating back five millennia, offering fresh perspectives on early social structure in pre-historic China.

In East Asia, the Yangtze River Basin in the south and Yellow River Basin in the north served as two primary loci of population expansion in the Holocene^{1–3}. Their distinct geographic, climatic, and land conditions led to divergent agricultural practices, cultural development, and demographic transition throughout prehistoric and historical periods^{4,5}.

In the north, the Neolithic Yangshao (YS, ~7000–5000 years before present (BP)) and Longshan (LS, ~4500–3800 BP) cultures, based on millet agriculture, maintained a long-standing and influential presence in the middle Yellow River Basin^{6–10} (Fig. 1a). In contrast, in the south, societal complexity developed alongside rice cultivation,

following the crop's domestication in the middle and lower Yangtze River Basin and its subsequent establishment as a staple^{3,11}. Two main cultures emerged in this region –Qujialing (QJL, ~4900–4500 BP), and Shijiahe (SJH, ~4500–4200 BP), which also left a profound influence on the genesis of social structure, evidenced by large-scale religious ceremonies and meticulously organized cities, from the Qujialing cultural period onward^{12,13} (Fig. 1a). Previous research has indicated a pattern of ongoing interaction and overlapping territories in the region bridging the Yellow River and Yangtze River, where cultural transformations were closely entwined with agricultural development^{10,14}. However, the details of population interplay remain largely uncharted since no

¹Biomedical Pioneering Innovation Center, Peking University, Beijing, China. ²Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China.

³School of Archaeology and Museology, Peking University, Beijing, China. ⁴Bio-X Center for Interdisciplinary Innovation, Yunnan University, Kunming, Yunnan, China. ⁵Institute for Cell Analysis, Shenzhen Bay Laboratory, Guangdong, China. ⁶School of Ethnology and Sociology, Minzu University of China, Beijing, China. ⁷Institute of Oil and Gas, Peking University, Beijing, China. ⁸College of Chemistry and Molecular Engineering, Beijing National Laboratory for Molecular Sciences, Peking University, Beijing, China. ⁹These authors contributed equally: Tingyu Yang, Jianing He, Chunmei Li, Chao Ning.

✉ e-mail: cscacz@pku.edu.cn; wuxh@pku.edu.cn; pangyuhong@pku.edu.cn; yanyi@pku.edu.cn

ancient human genome has been obtained from the Yangtze River region so far.

In this paper, we focus on Baligang (Fig. 1a–c), a Neolithic–Bronze Age archaeological site spanning approximately 8500 to 2500 BP, located on the northern edge of the middle Yangtze River Basin. This region marks an intermediate zone between the Yellow River and Yangtze River Basins^{5,15–17}. Stratified archaeological evidence suggests that Baligang underwent alternating transitions influenced by both northern and southern cultures^{5,17}. Baligang provides an ideal context for broadening our understanding of Neolithic East Asia for several reasons. Firstly, Baligang provides a unique opportunity for extracting genomic components associated with ancient rice cultivators of the Yangtze River Basin. Secondly, as a continuous human settlement since the Middle Neolithic, Baligang allows us to explore the dynamic interplay between population movements, cultural transitions, and agricultural dispersal over time. Furthermore, the site features several large-scale secondary burials dated to the late Yangshao cultural period (~5000 BP), where the remains of all individuals were exhumed from their primary graves and reburied collectively. The motives and social organization underlying this widespread burial practice of the period remain to be explored through archaeogenomic approaches. Moreover, given the relative sparsity of studies on ancient kinship structure in East Asia¹⁸, genetic kinship analysis of the secondary burials containing multiple generations of individuals at the Baligang site^{15,19–21}, may provide valuable insights into the Neolithic social organization in the region.

In our study, we elucidate population dynamics through the genomic analysis of 58 ancient individuals from the Baligang population, spanning from the Middle Neolithic to Late Bronze Age. The analyzed individuals display multiple cultural affiliations associated with both Yellow River societies in northern East Asia and Yangtze River societies in southern East Asia. Our findings demonstrate how the Early Neolithic East Asian populations contributed to the Baligang gene pool, revealing sustained bidirectional gene flow between northern and southern ancestral components throughout this extended temporal framework. These genetic interactions likely facilitated the concurrent transmission of agricultural practices and material cultures across regions. Furthermore, we provide kinship information for large-scale secondary burial practices and the existence of a patrilineal social structure in East Asia dating back to approximately five thousand years ago.

Results

The Baligang site and its ancient genomes

The Baligang (BLG) site, situated on the northern periphery of the middle Yangtze River region, is a village-sized ruin nestled beside the Tuan River, which is part of the Han River water system, the largest tributary of the Yangtze River (Fig. 1a–c). Ongoing excavations since 1991 have revealed that Baligang was a longstanding human settlement dating back to the early Neolithic era (~8500 BP). Baligang comprises six major cultural deposits (pre-YS, YS, QJL, SJH, LS, and Zhou), spanning the Early Neolithic (EN), Middle Neolithic (MN), Late Neolithic (LN) and Late Bronze Age (LBA) periods in the order of time. Human skeletons (>700 individuals) were unearthed from the top five cultural deposits. Rice was identified as the main crop in the Neolithic period based on Baligang archaeobotanical evidence, and the carbon isotopic analysis also showed a preference for a high proportion of the C3 diet in the pre-LS cultural period (Fig. 1d; Supplementary Fig. 1).

An intriguing development during the Late Neolithic period (around 5000 BP) is the appearance of well-organized, suite-like housing complexes and large secondary burials, may indicate specific patterns of social customs such as descent organization, community cohesion and ritualistic burial practices. Notably, among the secondary burials at Baligang, grave M13 stands out as the largest, containing the remains of over 90 individuals¹⁵. The purposeful accumulation of

pig lower jaws in grave M13 suggests that this might be tied to a Neolithic spiritual event; thus, obtaining the demographic structure of M13 community through ancient genomic analyses could provide insights into social organizations.

We screened bone samples of 103 individuals from Baligang, dated between 6100 and 2500 calibrated years before present (cal BP) and retrieved genome-wide aDNA data for 80 individuals. Among them, 58 individuals with preserved endogenous human DNA exceeding 1% were selected for deep sequencing, with depths ranging from 0.01 to 1.39× and the authenticity confirmed (Supplementary Data 1). We generated pseudo-haploid genotypes for these individuals by randomly sampling a single high-quality base of 1.24 million variants for downstream population genetic analyses^{22,23} (Supplementary Data 2) and estimation of genetic relationships. Using qpWave²⁴ (Supplementary Fig. 2a), we observed that individuals from all periods were genetically consistent with individuals from the same time period, except four individuals that did not cluster with others of the same period: BLG-M13-17 (LN-YS_o1) and BLG-M13-39A (LN-YS_o2) from LN-YS, BLG-M208 (LN-SJH_o) from LN-SJH, and BLG-M198 (LBA-Zhou_o) from LBA-Zhou (Supplementary Fig. 2b). Based on this result, we divided the individuals into six groups according to their respective chronological and archaeological contexts: MN-YS ($n=9$), LN-YS ($n=30$), LN-QJL ($n=2$), LN-SJH ($n=3$), LN-LS ($n=6$), and LBA-Zhou ($n=4$), with the four abovementioned individuals treated as separate outliers in the subsequent analyses. Besides, three individuals of LN-YS group (M13-76, M13-88, and M13-89) were excluded from group-based population genetic analyses due to close kinship (first-degree relatedness) and relatively lower autosomal coverage.

Genetic structure of Baligang individuals

Principal component analysis (PCA)^{25,26} was conducted on current human populations, specifically focusing on 39 contemporary East Asian populations^{22,27,28}, and the ancient genomes were projected onto the dominant variations (Fig. 2a; Supplementary Fig. 3a). Baligang individuals primarily clustered within the modern Han variation and were positioned nearest to the Neolithic populations in the middle and lower reaches of the Yellow River region²⁹, but showed a distinct southward shift. Considering Baligang's location in the pan-Yangtze River basin, we also applied a PCA oriented towards the spectrum of contemporary southern East Asian diversity. The results indicate that Baligang individuals cluster alongside present-day Hmong-Mien and Sino-Tibetan speakers (Supplementary Fig. 3b).

Pairwise comparison of outgroup f_3 values^{22,30} affirms that, despite the genetic affinity between the ancient Yellow River populations and the Baligang individuals, the latter show higher allele sharing with southern East Asians, which was also confirmed by F_4 statistics (Fig. 2b; Supplementary Fig. 3c; Supplementary Fig. 4). Unsupervised ADMIXTURE³¹ analysis also suggests that Baligang individuals generally carry a higher proportion of southern East Asian components than ancient Yellow River populations, especially after the LN-SJH period (Fig. 2c; Supplementary Fig. 5). Additionally, these analyses collectively identify the presence of the four outliers, thereby corroborating our initial qpWave results.

Chronological interplay between Northern and Southern Neolithic East Asians

Baligang documents the continuous change in human population genetics at a specific location over approximately four millennia, forming 3–4 meters of sediment with multiple physically stacked layers of northern (YS and LS) and southern (QJL and SJH) cultural artifacts. Utilizing F_4 statistics³², we explored various sources (Fig. 2d; Supplementary Fig. 6) to analyze gene flow directions between successive periods, and further employed qpAdm²⁴ modelling to estimate ancestry proportions in Baligang individuals (Fig. 3a; Supplementary Fig. 7; Supplementary Data 3).

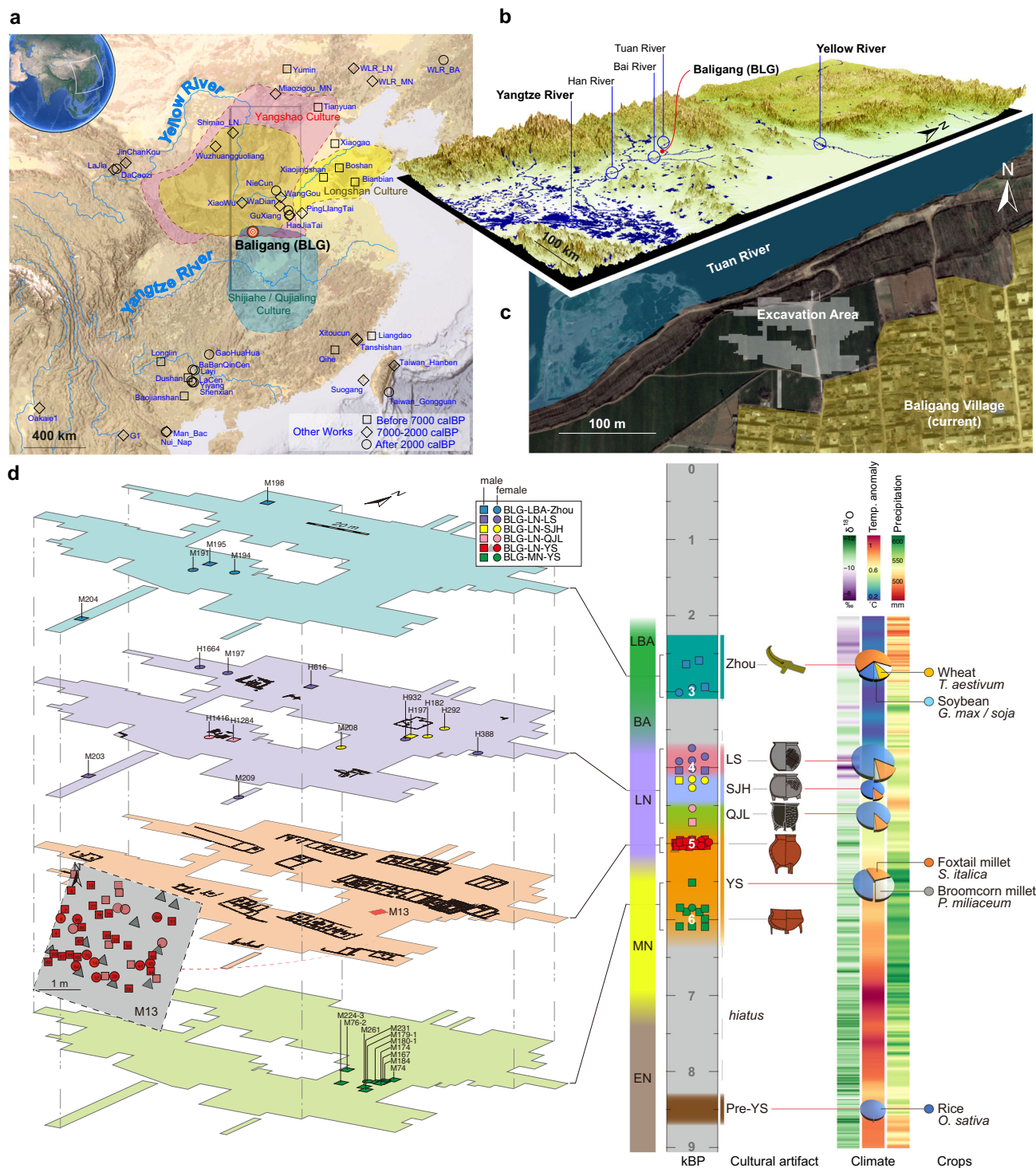


Fig. 1 | Geographic placement and chronology of Baligang subjects. **a** The locations of the Baligang (BLG) site and previously published individuals. The map also highlights the primary regions where Yangshao (YS), Longshan (LS), and Shijiahe (SJH)/Qujialing (QJL) cultures predominated. The base map was obtained from the World Terrain Base domain map dataset (<https://www.arcgis.com/home/item.html?id=c61ad8ab017d49e1a82f580ee1298931>) and created with ArcGIS pro v3.0. Sources: Esri, USGS, NOAA. **b** Geographical setting of Baligang and its surrounding areas, including the regions flanking the Yellow River and the Yangtze River, as well as the significant rivers that connected Baligang to the Yangtze River. The basemap was plotted using the open-source data “SRTMDEM 90-meter resolution original elevation data”, The data set is provided by Geospatial Data Cloud site, Computer Network Information Center, Chinese Academy of Sciences. (<http://www.gscloud.cn>). **c** The current satellite image of the Baligang area,

situated on the southern bank of the Tuan River, a tributary of the Han River. The map is provided by National platform for common geospatial information services (<https://www.tianditu.gov.cn>). **d** The principal architectural features of excavated Baligang ancient dwellings, portrayed on a map that includes four sedimentary layers: one from the Middle Neolithic (MN), two from the Late Neolithic (LN) period, and one from the Late Bronze Age (LBA). The map also marks the sample names and locations of each thoroughly sequenced individual, with explicitly identified biological sex information. The radiocarbon dating of Baligang individuals sampled in this study is displayed in chronological order, supplemented by representative cultural artifacts, significant climate indicators (oxygen levels, temperature, precipitation)³⁹, and the distribution and size of agricultural types (the pie chart size corresponds to the number of unearthed crop remains)¹⁷.

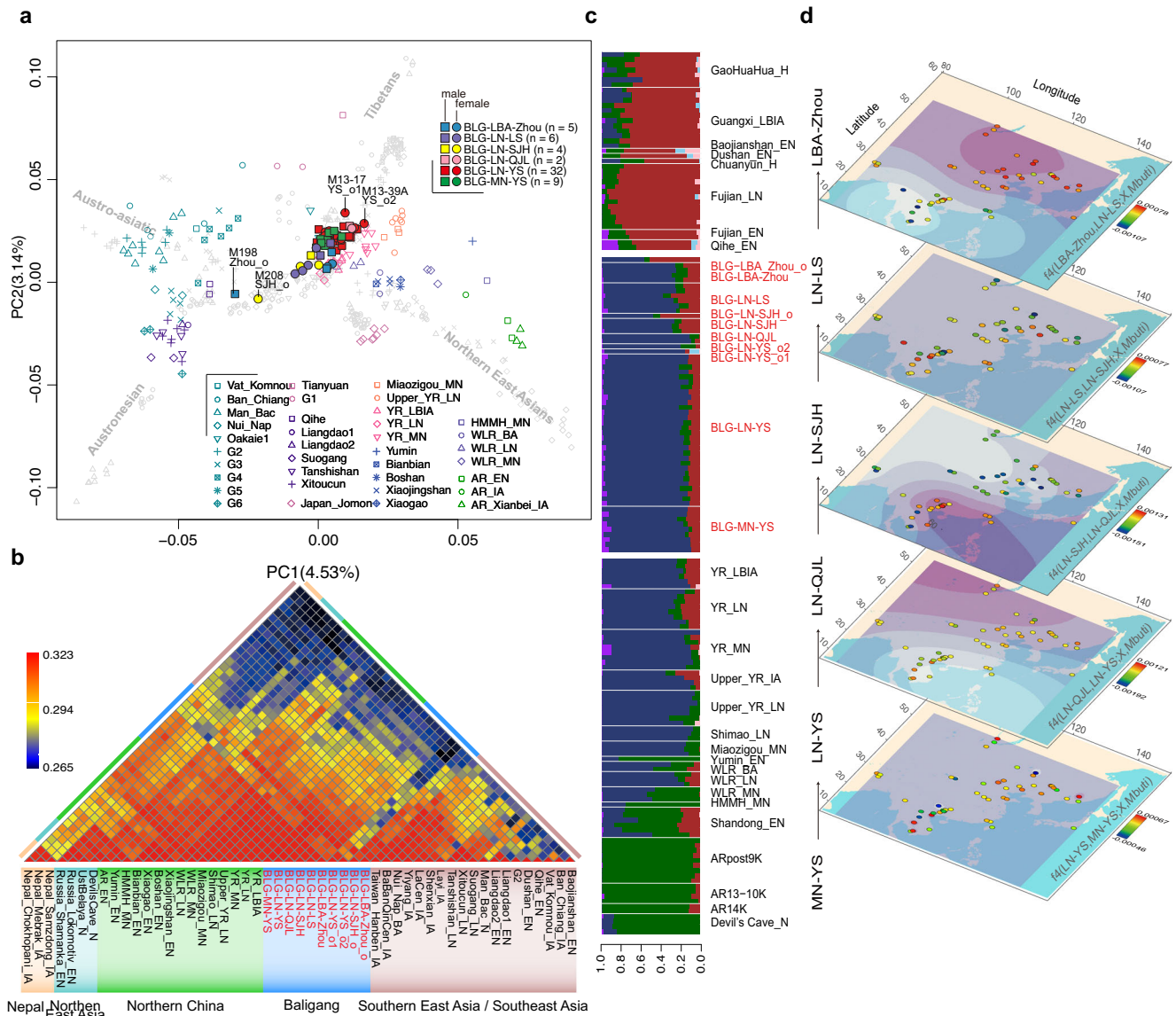


Fig. 2 | Dynamic genetic structure of ancient Baligang populations. a Principal component analysis (PCA) with ancient Asians (colored symbols), including samples from previous studies) projected onto present-day East Asians' (grey symbols) diversity. Present-day East Asians in the background form four major trajectories to populations labeled as Tibetans, Northern East Asians, Austronesians, and Austroasiatic, respectively. **b** The pairwise genetic similarity amongst ancient East Asians, incorporating the newly examined Baligang samples, was assessed via outgroup f_3 -statistics ($f_3(\text{Mbuti}; X, Y)$). Source data are provided as a Source Data file. **c** An ADMIXTURE analysis ($K = 6$) of various ancient Asian groups. Colors green, blue, and red denote northern East Asian, central China, and southern East Asian

related genetic components, respectively. **d** Genetic affinities to ancient East and Southeast Asia populations between successive Baligang groups using F_4 statistics with the form $f_4(\text{later period, earlier period}; X, \text{Mbuti})$. Red suggests a stronger affinity to the later period, blue to the earlier period. Geospatial patterns of gene flow were visualized through kriging distribution, which was derived from the f_4 statistical values above. The kriged interpolated gradient, ranging from deeper purple (designating higher interpolation values) to lighter purple (marking lower interpolation values), illustrates the direction of gene flow. Source data are provided as a Source Data file.

Our distal qpAdm modelling results reveal the presence of north-south admixture as early as the MN-YS period (6000 cal BP). From the LN-YS (5000 cal BP) to the LN-QJL period (4700 cal BP), the Baligang population experienced genetic influence from the north. Then, during the LN-SJH period (4300 cal BP), the gene flow predominantly reversed, with the southern East Asian (sEA) component, represented by Fujian_EN³³, rising to 35% of the ancestry proportion. However, during the LBA-Zhou period (2700 cal BP), the gene influx subtly shifted to the north again, diminishing the sEA component to 24% (Fig. 3a; Supplementary Fig. 7a). The proximal qpAdm modeling contrasting adjacent periods also corroborates this fluctuating pattern of gene flows (Fig. 3a; Supplementary Fig. 7b).

Interestingly, these dynamic genetic changes did not entirely synchronize with the cultural or agricultural changes. For instance, during the LN-QJL period, the unearthed rice remains were reported to significantly increase to 88%, from 42% in the preceding LN-YS period. During this transition, although the elevation of rice cultivation happened simultaneously with the northern-southern cultural change, the influx of sEA ancestry was not identified in LN-QJL individuals but delayed to the subsequent LN-SJH period.

Genetic components and transmission of ancient rice farmers

The LN-SJH period is found to be a critical division point of genetic continuity based on our F_4 statistics analysis. We conducted

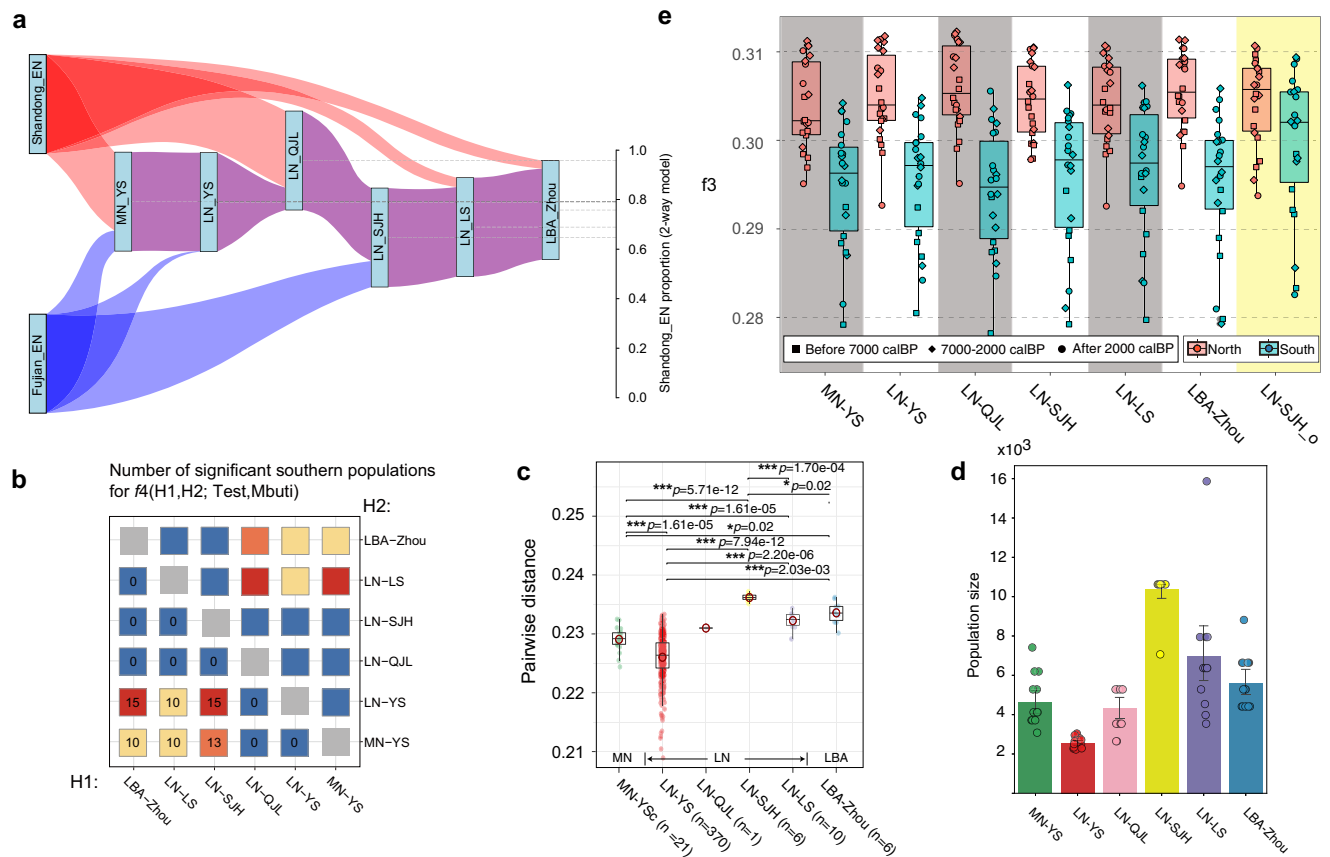


Fig. 3 | The migration and dispersal of rice farmers in Baligang and the surrounding region. a The Sankey diagram illustrates the temporal admixture model of Baligang populations, with values derived from the corresponding qpAdm proximal model. Shandong_EN (red) and Fujian_EN (blue) represent the northern and southern East Asian genetic components, respectively, while purple indicates the earlier period component in the proximal model. The height of each period (corresponding to the right axis scale) reflects the proportion of the northern East Asian component (Shandong_EN) in the distal model, which models the admixture between northern (Shandong_EN) and southern (Fujian_EN) East Asian genetic component. **b** The details of genetic shifts in Baligang populations. The heatmap displays the genetic affinity to southern East Asian populations (Test) comparing later (H1) and earlier (H2) time periods of BLG populations through f_4 (H1, H2; Test, Mbuti). Each cell in the heatmap indicates the count of southern East Asian populations showing significant value ($|Z| > 3$) for a given H1-H2 pair. The detailed f_4 statistic results for each specific H1-H2 combination are presented in Supplementary Fig. 9b. **c** The genetic diversity within each group of Baligang inhabitants, depicted as the distribution of autosome pairwise distance between sub-groups

within each period. Box plots show the median (centre line), first and third quartiles (lower and upper hinges), and whiskers extending to 1.5X interquartile range from the hinges, the mean values are marked with hollow circles. Pairwise comparisons between groups were performed using the two-sided Games-Howell test, which adjusts for unequal variances and multiple comparisons through Welch-Satterthwaite degrees of freedom, with significance levels denoted as *** $p < 0.001$, ** $p < 0.01$, and * $p < 0.05$. **d** Effective population size estimates for each group were determined using hapROH. Technical replicate values from 15 genotyping runs displayed as dots, and bar plot data represent mean values $\pm$ SEM. Source data are provided as a Source Data file. **e** The outgroup f_3 statistics are formulated as $f_3(\text{BLG, Test; Mbuti})$ for each BLG group and SJH outlier. 'Test' denotes various ancient East Asians from diverse regions. Box plots show the median, first and third quartiles (lower and upper hinges), and whiskers extending to 1.5X interquartile range from the hinges. The affinities between SJH outlier and southern East Asians were notably higher than those of their contemporaneous individuals. Information of northern ($n = 22$) and southern ($n = 22$) populations were listed in Supplementary Data 4.

pairwise comparisons across various periods and discovered a significant difference in the composition of southern East Asian ancestry before and after the LN-SJH period (Fig. 3b; Supplementary Fig. 8). PCA kernel fitting (Supplementary Fig. 3d) also revealed a clear change between the LN-QJL and the LN-SJH periods, around 4200 BP, further supporting a significant influx of southern East Asian genetic components. Interestingly, parallel genetic changes were also detected in contemporaneous Late Neolithic Yellow River populations²⁹. Additionally, we calculated the pairwise genetic differences between Baligang sub-groups and found that individuals from the LN-SJH period exhibit the highest genetic diversity (Fig. 3c). Consistently, the effective population size (N_e), estimated using the length distribution of Runs of Homozygosity (ROH) of 4–20 cM, is the largest in the LN-SJH period (Fig. 3d), indicating that a substantial degree of population mixing might have occurred during this time. Additionally, the outgroup f_3 -statistics and f_4 -statistics results showed

one outlier individual (LN-SJH_o) with particularly higher content of southern East Asian ancestry (Fig. 3e; Supplementary Data 4; Supplementary Fig. 9). We speculate the prevalence of sustained high proportion of rice planting during the LN-SJH period might be associated with the migration of rice farming-related populations into the Baligang site and its surrounding areas around that time. The ancestral information embedded in Baligang genomes may provide a unique dataset for exploring the expansion of rice cultivation.

Insights into social structure in neolithic East Asia

Secondary burials within the Yangshao culture, exemplified by grave M13, possibly originated from a ritual conducted by the living community. Nevertheless, the interrelations among these buried individuals, spanned around two hundred years (5100–4900 cal BP, Supplementary Data 1), has long been debated.

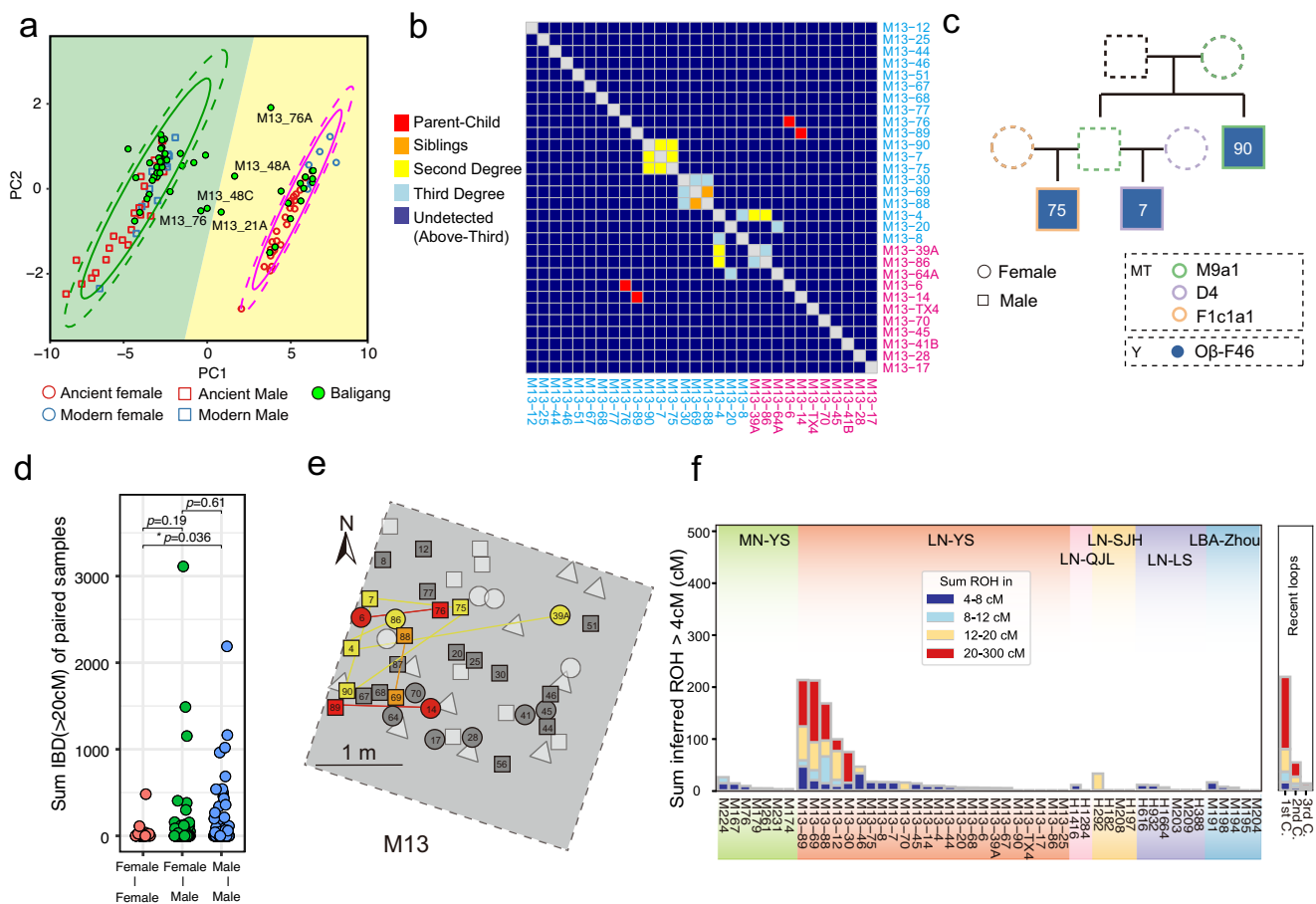


Fig. 4 | Social structure of Neolithic East Asia reflected by BLG genomes. **a** Sex determination of 52 Baligang individuals from grave M13 was conducted through PCA-based classification using an SVM classifier trained on 67 individuals of known sex. The PCA plot displays sex-specific clustering, with male individuals enclosed within green confidence ellipses and female individuals within pink confidence ellipses. The solid line represents 99%, and the dashed line represents 99.9% confidence intervals. The junction of yellow-green blocks is the boundary for classification of two sexes. Source data are provided as a Source Data file. **b** The matrix of pairwise relationship between the individuals in grave M13 ($> 0.1\times$ coverage, $n = 30$) is inferred by KIN to identify the different degrees of relatedness. **c** An example of a pedigree structure inferred from three related individuals in

grave M13, with their MT and Y haplotypes identified. **d** The distribution of IBD sharing among paired individuals, classified by genetic sex within grave M13 (26 Female-Female pairs, 115 Female-Male pairs and 114 Male-Male pairs), revealing closer genetic relatedness between males, $*p < 0.05$, from two-sided t -test. **e** The individual pairs within the second degree of relatedness are indicated in their burial positions on the grave M13 map, the color label of kinship corresponding to figure B. **f** The distribution of different lengths' ROH of BLG individuals in different periods. The distribution under different simulated conditions is shown on the right for reference. The genetic patterns observed in individuals M13-89, M13-69, and M13-88 imply a consanguineous relationship between their parents. Source data are provided as a Source Data file.

Since many individuals randomly sampled from grave M13 have low genomic coverage, we used a machine learning-based model to accurately identify their biological sexes. More than two-thirds of these individuals (52 out of the 75) could be identified as 35 males and 17 females (Fig. 4a; Supplementary Fig. 10; Supplementary Data 1), indicating a male majority in this grave. Analysis of uniparental genetic characteristics of grave M13 individuals revealed that majority mtDNA macro haplotypes (82%) belonged to haplogroups D, M, A, and R (34%, 23%, 14%, and 11%, respectively), suggesting a diverse maternal relationship had already been established in Neolithic Baligang community. However, the Y haplotypes of grave M13 males were concentrated, with all lineages falling under the branch of O Φ -F46 (Supplementary Fig. 11a; Supplementary Data 1), which also persistently existed in the Baligang site from 6000 to 4000 BP and accounted for 82% (23/28) of Baligang males.

Kinship analysis was conducted on 30 individuals of grave M13 with coverage of $0.1\times$ and above, revealing three first-degree, five second-degree, and five third-degree relationships (Fig. 4b; Supplementary Data 5). According to the mtDNA haplotypes results, two first-degree relationships were identified as father-daughter pairs with one

daughter approximately 15 years old and another daughter aged 18–25 years as identified through physical anthropology; another first-degree relationship was between a set of brothers. We further utilized a modified version of CREST³⁴ to identify two pairs of second-degree relationships as paternal half-siblings (M13-75 & M13-7, M13-4 & M13-86), and two pairs as patrilineal avuncular relationships (M13-75 & M13-90, M13-7 & M13-90, Supplementary Fig. 12; Supplementary Data 6). The genetic relationships among M13-75, M13-7, and M13-90, corroborated by their uniparentally inherited haplogroup data, provide compelling evidence for a pedigree which was organized through patrilineal lineage (Fig. 4c), while new mtDNA lineages were contributed by the incoming females in each generation, suggesting a pattern of female exogamy.

Furthermore, using ancIBD³⁵, we conducted a thorough analysis of the distant genetic relatedness between all paired individuals, identifying numerous third to sixth-degree pairs sharing a cumulative IBD exceeding 100 cM (Supplementary Fig. 11b–d; Supplementary Data 7). The results showed the prevalence of closer kinship between adult males compared to adult females, with the latter showing a nearly entire absence of close relatedness (Fig. 4d), further suggesting

a prevailing patrilineality. This is also supported by the uniformity of Y haplotype among grave M13 males contrasted with the high diversity of mtDNA haplotypes³⁶.

However, due to constraints in sampling rate and aDNA preservation, as well as the uncertainty on how the human remains in grave M13 were selected for burial, direct inference of detailed kinship structures appears exceptionally difficult. Therefore, we conducted simulations of populations under various kinship structure scenarios and assessed the similarity of IBD-sharing profiles between the sampled individuals from simulated populations and the actual data from M13 (Supplementary Fig. 13a). Among various kinship structure settings, simulated populations demonstrating the highest similarities with the M13 dataset consistently display higher centralities indicative of near-single lineage structures (Supplementary Fig. 13b). This implies that M13 individuals may come from a population composed of a dominant single lineage. Furthermore, we quantitatively estimated the M13 population to be with a centrality >0.8 at a posterior probability of >78% (Supplementary Fig. 13c–d), favoring the assumption that the population consists of a dominant single lineage. In addition, our simulation results generally favor the assumption that M13 individuals likely come from a population of over 200 persons, with an estimated posterior probability exceeding 87% (Supplementary Fig. 13e–g).

Besides, combined with individual buried locations, no discernible spatial organization pattern was observed in the placement of genetically related individuals within the pit, (Fig. 4e), implying that the exact lines of genealogy may be either unknown or unimportant in this secondary funerary practice. Additionally, from the ROH distribution, we also found evidence of consanguineous marriage, although not common, indicating that there might not have been a mechanism in place to strictly prevent reproduction between closely related individuals during that time (Fig. 4f).

Discussion

Previously published ancient genome data have suggested that the north-south separation of ancient East Asians occurred no later than the Early Neolithic^{33,37}. However, this genetic differentiation has significantly diminished in present-day East Asians due to extensive mixture between northern and southern East Asian ancestries. Our chronological data indicated that the wide-ranging occurrence of genetic mixing between northern and southern East Asian ancestries had at least reached the northern periphery of the middle Yangtze River basin as early as the Middle Neolithic age. The genetic profile observed in Baligang individuals suggests that there has been no significant population replacement events seen in contemporaneous Europe³⁸. However, coinciding with the highest population diversity, there was an obvious genetic shift towards southern East Asians during the LN-SJH period around 4200 years ago. One possible speculation may be related to an unusual period of climate anomalies, although there still exists paleoenvironmental debate^{39–44}, that led to various extreme climate events and long-term significant landscape transformations^{45–48}, and greatly influenced crop cultivation and human activities. The $\delta^{18}\text{O}$ record from stalagmites and precipitation data surrounding the Baligang site (31°–36.5°N in China) indicate a wet phase climate around 4200 BP⁴⁶ (Fig. 1d), making the area potentially a suitable destination for massive migration of rice farmers from the Yangtze River basin. The parallel changes supported by the genetic evidence of an influx of southern East Asian ancestry into ancient Yellow River populations around the Late Neolithic²⁹, may suggest the widespread climatic impacts on environmental and migration patterns at that time. Furthermore, the significantly higher southern East Asian ancestry in LN-SJH_o may suggest that this outlier individual represents an early southern immigrant. More explorations of ancient environmental research and associated archaeological studies are needed to clarify the interaction between environmental effects and human behavior. In addition, we acknowledge a limitation in our study that the

LN-QJL period is represented by only a small number of analyzed individuals. More archaeogenomic evidence from this region in the future will help to draw more definitive conclusions about the transitional period.

Archaeological evidence suggests that northern and southern cultures alternated at different time periods of Baligang site, with corresponding fluctuations in the proportion of rice cultivation. Our genetic data complement this pattern, revealing a coherent trajectory of population dynamics. It is worth noting that evidence from different fields may differ in timing and intensity. For instance, genetic changes in LN-QJL and LN-LS periods appears to have a potential time lag compared to cultural and agricultural changes, which warrants further investigation, including limitations on sample size in certain periods. Importantly, despite these subtle differences in timing, the overall pattern of cultural, agricultural, and genetic transitions displays remarkable consistency, strengthening our confidence in the overall interpretation of the data.

Additionally, the archaeobotanical results indicate a significant increase in the proportion of rice in the botanical assemblage at Baligang site as early as the LN-QJL period, with a sustained high proportion during the LN-SJH period, followed by a slight decline in the LN-LS period (Fig. 1d). This supports the genetic change found in Baligang population that the affinity toward rice farmers from south China increased before 4200 BP. However, our isotopic analysis shows the $\delta^{13}\text{C}$ values remain relatively consistent between the YS and QJL-SJH periods (Supplementary Fig. 1), which does not completely align with the results from the archaeobotanical evidence. It is important to note that the $\delta^{13}\text{C}$ values for individuals in each pre-LS period, except for LN-QJL period, exhibit a wide variation (ranging from −16‰ and −20.5‰). This suggests that dietary compositions were diverse within each period at Baligang site and may reflect individual preferences for different food sources. In contrast, the archaeobotanical results provide an average representation of the botanical assemblage compositions across the entire period and site, which may account for the observed discrepancy. Further isotopic analyses of the animal and plant remains at Baligang site would enhance the interpretation of the human isotopic data.

The burial practices likely reflect the social patterns of living individuals. Archaeological evidence indicates that secondary burial was a common funerary practice during the Neolithic Yangshao cultural period of Neolithic Age in China, with a long-time span, widespread range, and phased and geographical characteristics^{19,20}. However, based solely on cultural artifacts or physical characteristics of human remains, the motivation for such burial customs is still unclear. Ancient genomic data from grave M13, representative of large secondary burial in Baligang site, provide an essential clue for deciphering the relationships among the co-buried individuals and help reconstruct the social organization in the Chinese Neolithic, however, we emphasize that these interpretations may not be generalizable to other time periods. There have been controversies in archaeological and anthropological research regarding whether communities that practice secondary burials are organized by patrilineal or matrilineal lineage, and it also remains unclear whether all the human remains collected for secondary burials are related by blood ties among the residents of this settlement⁴⁹. Our simulation results generally favor the assumption that BLG-M13 individuals exhibit a patrilineal kinship system, where family lineage and descent are traced through the male line. These deliberately selected individuals were likely members of specific patrilineal family based on blood relations, rather than a collection of unrelated community residents. Several hypotheses have been proposed to explain the practice of secondary burials in ancient Chinese society, including but not restricted to residential migration, religious beliefs, or the desire to reunite clan or family members after death^{15,49}. Combined with certain archaeological discoveries, such as graves intentionally situated between two well-ordered housing

complexes, meticulously arranged piles of human remains within the graves, and the presence of ceremonial animal jaws and heads, our genetic data of BLG-M13 individuals are more supportive of the archaeological speculation that this secondary burial practice aimed at reuniting deceased members of the family, thereby reinforcing the cohesion of the community. Nonetheless, one limitation of our simulation method is that data sparsity and unknown biases in the sampling process might significantly affect the accuracy of kinship structure estimation. For instance, as observed in our analysis using simulated data, our method often underestimates the total individual number at lower sampling proportions, since randomly drawn samples tend to be derived from small subsets of the actual population. For this reason, we can only conclude that the M13 pedigree comprises over 200 individuals with a posterior probability of 87%. Furthermore, our simulation captures only some general features of the population's kinship structure, whereas the real-world scenario is likely more complex. This is illustrated by our discovery of two father-daughter pairs in grave M13, where the daughters were near-adult or adult at the time of death. This pattern may suggest that they remained unmarried for specific reasons. Alternatively, it could indicate that female exogamy was not strictly practiced in the Baligang community during the LN-YS period, or that they had married into other patrilineal families but were mistakenly selected for secondary burial. The actual circumstances remain unclear and warrant further investigation. Overall, we provide innovative methodologies for the ascertainment of kinship and sex of individuals from complex burial sites, shedding light on the study of societal structures and burial customs of ancient communities.

Methods

Ethics and inclusion statement

The samples were obtained from well-documented archaeological excavations conducted by the school of archaeology and museology at Peking University, all necessary permits for the excavation and analysis of the ancient human remains were obtained from relevant local authorities. We ensured that all samples were ethically sourced and that the excavation and analysis followed proper archaeological and bioethical guidelines to minimize contamination and ensure respectful handling of human remains. The study protocol was approved by the ethics committee of the school of archaeology and museology at Peking University.

Radiocarbon dating and isotope analysis

In this study, we examine 103 individuals from the Baligang site. Of these, bone samples from 35 individuals underwent collagen extraction⁵⁰ and graphitization at the Radiocarbon Dating Laboratory, School of Archaeology and Museology, Peking University. Radiocarbon measurements were preformed using accelerator mass spectrometry (AMS) at the AMS Center, School of Physics, Peking University (Supplementary Data 1). The half-life of radiocarbon ($C-14$) is 5568 years. We present radiocarbon dates as years before 1950 (BP). These dates were calibrated using OxCal v4.4.4⁵¹ with the IntCal20 calibration curve⁵². All radiocarbon dates are reported with a credible interval of 95.4%, corresponding to two standard deviations.

Stable isotope analysis was conducted on bone samples from 30 Baligang individuals to obtain $\delta^{13}C$ and $\delta^{15}N$ values. This analysis was performed at the Stable Isotope Analytic Laboratory, School of Archaeology and Museology, Peking University. Stable carbon and nitrogen isotope values were measured using an IsoPrime100 mass spectrometer coupled to a vario PYRO cube element analyzer. Calibrations were made relative to the Vienna Pee Dee Belemnite (VPDB) and AIR scales, employing IAEA600 and USGS41 as standards. The precision (1σ) of repeated measurements of in-house laboratory standards under long-term monitoring in the lab was $\pm 0.15\text{‰}$ for $\delta^{13}C$ and $\pm 0.20\text{‰}$ for $\delta^{15}N$.

Ancient DNA processing

All of Baligang human samples were obtained from school of archaeology and museology at Peking University and processed in dedicated ancient DNA laboratories at Biomedical Pioneering Innovation Center, Peking University in Beijing. For petrous bone samples, dirt was first removed from the surface with a single-use grinding blade at low speed. Then the bone powder was obtained by drilling the densest osseous materials around the cochlea with disposable ball cutters⁵³. For auditory ossicle samples, we directly decontaminated the surface of them through a wash with 1% HClO and 95% ethanol then exposure to UV irradiation for 10 min on each side⁵⁴. For other skeletal bone samples, such as femur and tibia, we also first removed the surface dirt. Then, we applied a quick wash with 4% NaClO, ultra-pure water and ethanol followed by irradiation under UV for 15 min. Fine bone powder was produced by the same drilling process as petrous bones.

DNA extraction was carried out from 50–100 mg bone powder or complete auditory ossicles in the PCR hood inside dedicated ancient DNA room following the silica-based protocol⁵⁵. Qiagen PB buffer supplemented with sodium acetate, sodium chloride was used as binding buffer, and large volume spin columns commercially available from Roche (High Pure Viral Nucleic Acid Large Volume) was used for DNA binding and purification, which resulted in 65 μ l of DNA extract per sample. A 30 min pre-digestion step was implemented before DNA extraction to remove surface contaminations⁵⁶. At least one negative control was also included through each DNA extraction process to ensure no cross contamination.

Double-stranded library preparation method was used to DNA extracts based on a single-tube strategy⁵⁷ by NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB) with some optimizations as below: (1) Y-shaped adapter at a final concentration of 0.1–0.4 μ M in 1/4 reaction system (10 times the original concentration) was used to greatly improve ligation efficiency and save cost; (2) Two-round PCR amplifications after adapter ligation was adopted without inter-reaction purification until first round of amplification step in order to reduce considerable DNA loss and adapter-dimers as well; (3) Short PCR primers were designed for first round of amplification using AmpliTaq Gold 360 master mix (Life Technology) followed by purification with VAHTS DNA Clean beads (Vazyme) to remove adapter-dimers. The resulted short-length libraries were then converted to full-length libraries for sequencing through second round of amplification with double indexed P5-P7 primers using Q5 master mix (New England Biolabs). A negative control was included for each batch of library preparation. We first used 5 μ l of DNA extract without Uracil DNA Glycosylase (UDG) pre-treatment for library screening on Illumina Nextseq 500 platform with paired-end 2×76 cycles. Samples with terminal damage pattern of typical aDNA were then picked up for further preparation of half-UDG-treated libraries from 10 μ l of DNA extract⁵⁸ except for some M13 individuals (Supplementary Data 1). Lastly, libraries quantified by Fragment Analysis were sequenced on Illumina Novaseq platform with paired-end 2×150 cycles or HiSeq2500 platform with paired-end 2×100 cycles. Base calling and filtering of exactly correct index sequences were performed with CASAVA 1.8.2 and low-quality reads (average or 50% Phred quality value ≤ 20) were discarded.

Mitochondrial DNA enrichment

RNA Probes targeted to human mitochondrial genome and reagents for capture were all from iGeneTech Co. Ltd in China. DNA library of 500 ng was used for mitochondrial enrichment by one round of in-solution hybridization following the manufacturer instructions that the hybridization condition was set to 50 °C for 24 h. Subsequently, the captured libraries were amplified with P1-P2 primers for 12 cycles using Q5 High-Fidelity Polymerase and purified by VAHTS DNA Clean beads. Lastly, libraries quantified by Fragment Analysis were sequenced on an Illumina Nextseq 500 platform with paired-end 2×76 cycles (2×8 for

the indices) and CYGNUS S100 platform with 1 × 100 cycles (1 × 8 for the indices).

Data analysis

Sequencing data analysis. The paired-end sequencing reads were merged into a single sequence by at least 11 bp overlap between forward and reverse read with no more than 2 mismatches. Low quality reads in either end of the paired-end reads were discarded. The merged reads between 35 and 189 bp length were used to align to the human reference genome (GRCh37d5) by BWA-samse (v0.7.15-r1140)⁵⁹ with seed disabled parameters (-l 16500 -n 0.01 -o 2). Mapping quality less than 30 were filtered by Samtools⁶⁰ (v1.9). PCR duplications were removed by Picard markduplicate(v2.21.8) (<http://broadinstitute.github.io/picard>). We realigned the Indel region with GATK3.5 Indel-Realigner and chose uniq-map reads for downstream analysis. For mitochondrial DNA, the merged reads obtained from mitochondrial capture sequencing were aligned to revised Cambridge Reference Sequence (rCRS)⁶¹ using BWA aln/samse algorithm and next CircularMapper⁶² (v1.93.4). Duplicate reads were removed using Dedup 0.12.3 (<https://github.com/apeltzer/DeDup>) and the consensus sequences were generated using Schmutzi v1.5.6⁶³. Terminal damage patterns were calculated with mapDamage 2.1.0⁶⁴, our analysis revealed that for nearly all libraries, the damage rate fell below 0.01 beyond the last 3 base pairs at the read ends (Supplementary Fig. 14), so we masked the last 3 base pairs at both ends of the reads by converting them to 'N' to remove the interference of damaged signal.

Genotyping and dataset integration. We randomly sampled a single high-quality base from reads with base quality and mapping quality over 30 as pseudodiploid genotypes using the pileupCaller (sequenceTools, v1.2.2) program (<https://github.com/stschiff/sequenceTools>). Then we combined our genotypes with published ancient individuals (Supplementary Data 2). We further intersected the dataset with two sets of world-wide genotype panels for downstream population analysis, one is based on the Affymetrix HumanOrigins Axiom Genome-wide Human Origins array ("HumanOrigins"; 593,124 autosomal SNPs)²² and the other integrates the panel of 1240k capture SNPs ("1240 K"; 1,233,553 autosomal SNPs) from the Simons Genome Diversity Panel (SGDP)²³ and the Human Genome Diversity Project (HGDP)-shotgun data⁶⁵ for higher distinguishability.

Authenticity and contamination estimation of aDNA. Firstly, ≥10% of terminal damage rate calculated by MapDamage 2.0 was expected to be potentially authentic ancient DNA signal⁶⁶. Then, we applied two criteria to estimate contamination rates. 1. We compared mtDNA fragments of each individual with 311 mitogenome sequences from worldwide populations using ContamMix⁶⁷ (using 3 bp terminal trimmed bams). If >5% of the fragments matched one of the worldwide sequences better than the consensus, we treated those libraries as contaminated. 2. We additionally evaluated the nuclear genome contamination rate for males based on X chromosome data by ANGSD⁶⁸ v0.920 which considers the sequencing errors and mismatches on polymorphic positions. We considered those libraries as contaminated if estimated contamination rate >5%.

Assignment of uniparental Haplogroup. To determine the mtDNA lineages that our ancient individuals belong to, we employed mtDNable (<http://www.mtDNAcommunity.org/downloads.aspx>), Pylo-Treem (<http://www.phylotree.org>) and Mitotool (<http://www.mitotool.org/index.html>) to define their exact haplogroups, based on phylogenetic tree build 17⁶⁹ (<http://www.phylotree.org>). We also referred to HaploGrep2 (v2.4)⁷⁰ for mtDNA haplogroup determination. Substitutions defining the assigned haplogroup were also manually checked. We determined the male Y chromosome haplogroup by examining a set of positions on the diagnostic positions on the ISOGG

database (www.isogg.org) using Y_Leaf⁷¹ software, and assigned the final haplogroups by the most downstream derived SNPs (Supplementary Data 1).

Sex determination. The sex of Baligang individuals with high quality was determined by the ratio of coverage on sex chromosomes to autosomes. For individuals of low genome coverage, we proposed a machine learning model to make accurate sex determination based on the number of aligned reads (mapping quality ≥ 30, q30) from shotgun DNA sequencing and PCA (Principal Component Analysis) classification. Unique size of each chromosome was calculated referring to protocol described by Baslan et al.⁷². In brief, we generate the simulation reads (50 bp) uniformly from whole genome and align back using BWA. The number of uniquely mapped reads on each chromosome is counted as the unique size.

We first established a training dataset for this model using the shotgun sequencing data of known sex downloaded from National Center for Biotechnology Institute, 1000 genome project and European Bioinformatics Institute, which includes 48 ancient^{73,74} and 19 present-day human samples^{74,75} (31 females and 36 males). The age of ancient samples ranges from 70,000 years old (Neandertals) to 200 years old (original Australian sample). Then the ratios of q30 reads on unique size of X chromosome to each autosome were calculated then log2 transformed for each individual to form a matrix. PCA was performed using R package to generate PCA scores which were finally projected onto a two-dimensional space where the PC1 and PC2 scores were used for classification. The confidence interval (CI) for both sex types was generated using multivariate normal distribution (solid ring for 99% and dashed ring for 99.9%). The Support Vector Machine (SVM) model, a learning model, was trained using svm function in "e1071" R package with the PC1 and PC2 scores. An oblique dashed line was built by SVM, producing a classifying boundary for male and female, respectively.

We obtained shotgun sequencing data of 45 present-day humans (22 males and 23 females) from Professor Wang in Tsinghua University. These sequencing libraries were constructed from blood DNA with conventional method in the same batch. The ratios of q30 reads on unique size of X chromosome to each autosome were computed then log2 transformed, and PCA was performed. PCA scores were predicted using the PCA transformation model generated as above. By the same strategy, the PCA scores of 30 Baligang M13 individuals were processed and PCA was performed to determine their sex.

To evaluate the power of our method as a function of reads number, we down sampled q30 reads from 10 male and 10 female individuals out of 67 reference individuals to generate a set including 10,000, 8000, 4000, 2000, 1000 and 500 reads, among which Ire-8 is a 5000 year old Neolithic Scandinavian male individual, RISE240 is an early Bronze Age Eurasia female individual, and HG00096, HG00100 is a modern male and female individual, respectively. The down sampling was carried out for 100 times.

Principal components analysis. We performed principal component analysis (PCA) by smartpca (v13050) from EIGENSOFT²⁵ (v7.2.1). The eigenvectors were constructed by present-day populations listed on the Supplementary Fig. 3a, b and the ancient individuals were projected with the lsqproject: YES option. The other options were set to default except for numoutlieriter: 0, and autoshrink: YES to display all the samples including the outlier. We included several datasets in the present-day populations of principal component analysis: 26 populations from Human Origin (HO) SNP panel, southeast Asia individuals from Lipson et al.²⁸, (Htin, Mlabri), and Han and Tibetan populations from Lu et al.²⁷.

Admixture. We use the Model-Based Estimation of Ancestry algorithm ADMIXTURE³¹ (v1.3.0) to capture the genetic structure within the data.

It performs an unsupervised clustering of large multilocus genotyping datasets, each cluster represent a hypothetical ancestral component, and each individual is a mixture of the probabilities of these assumed ancestry proportions using a maximum likelihood estimate. The number of cluster “K” which represent the number of “hypothetical ancestral populations” is predefined by the researcher.

The dataset used in ADMIXTURE is from the 1240 K platform, assembling newly sequenced samples with 219 worldwide modern and ancient populations. Before ADMIXTURE analysis, the dataset was filtered to 458582 SNPs with PLINK 1.90⁷⁶ to remove SNPs with minor allele frequency lower than 0.01 ($-maf\ 0.01$) and linked SNPs with $r^2 > 0.4$ ($-indep-pairwise\ 200\ 25\ 0.4$). We investigate values of K from K3 to K10 running the Admixture with the flag $-cv$ to evaluate a “cross-validation error” for each K. We explore the structure in each K and choose to concentrate on K=6 because there was not significant change in the ancestry components constitution of Baligang above 6, and the higher values of K were not giving more information for the studied population.

F statistics. We calculated F3 statistics using qp3Pop (v412) in ADMIXTOOLS 4.1 package²⁵, and F4 statistics by qpDstat (v712) with parameter “f4 mode: YES”. Mbuti population was specified as an outgroup for all *f*-statistics analysis. SEs in F4 statistics were calculated from the default block jackknife approach.

Kinship. We assessed the degrees of kinship for every pair of individuals newly reported in this study using KIN⁷⁷ software, KIN is a hidden Markov model (HMM)-based approach to detect patterns of shared identity-by-descent (IBD) states along a pair of genomes. The kinship degrees were classified to various types based on three IBD states which refers to individuals share zero (k0), one (k1) or two (k2) chromosomes IBD at any given position. Specifically, k0, k1, k2 for identical twins are (0, 0, 1), for parent-child are (0,1,0), for siblings are (0.25, 0.5, 0.25), for second-degree relatives are (0.5,0.5,0), for third-degree relatives are (0.75, 0.25, 0), for unrelated individuals are (1,0,0). To detect IBD and infer kinship, for individuals with coverage above 0.1X, using genome windows of 10 Mb, we performed IBD inference of relative pairs across a range of kinship categories including unrelated, first through third degree relatives, as well as grandparent-grandchild, avuncular, half sibling, parent-offspring, full sibling and identical twin pairs. The likelihoods of multiple IBD sharing models were compared, assigning each relative pair to the model with the maximum likelihood. And we take results higher than 1.0 log likelihood ration as reliable output (Supplementary Data 5).

To further validate KIN results and detect more distant connections, we also estimated the genetic relatedness between those ancient individuals by shared IBD pattern. The samples were imputed using the software GLIMPSE2⁷⁸ and the data processing steps followed the standard procedures outlined in the GLIMPSE2 tutorials (<https://odelaneau.github.io/GLIMPSE2/docs/tutorials>). To perform imputation, phased haplotypes from the 1000 Genomes dataset were used as a reference panel (<http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/>). The imputed and phased data were then used as the input into the software ancIBD (v.0.3a1; <https://pypi.org/project/ancIBD/>)³⁵. All Baligang individuals with coverage of 0.15X and above were used for IBD analysis using the default settings. We collected summary statistics for each pair in which long IBD segments (>20 cM) was detected (Supplementary Fig. 11b–d; Supplementary Data 7).

All unrelated individuals were kept for subsequent population genetics analyses. For any pair of individuals identified as identical/twins or first-degree relatives, we only picked the one with the higher SNP number, therefore three individuals in M13 grave were excluded which had first-degree relationship to another individual of the same grave.

Admixture modeling with qpWave and qpAdm. We used qpWave(v400) from ADMIXTOOLS 4.1 with the parameter ‘allsnp: YES’ to model our individuals and populations as mixtures of proposed clade groups, setting a minimum threshold of 100,000 SNPs for individual modeling. For this analysis, we used the following populations as outgroup: the modern African Mbuti ($n = 4$), Levantine hunter-gather Natufian ($n = 6$), Andamanese islanders Onge ($n = 6$), Neolithic Iranian farmer Iran_N ($n = 5$), Mesolithic Eurasian hunter-gatherer Loschbour ($n = 1$), Neolithic Anatolian farmer Anatolia_N ($n = 24$), native Central American Mixe ($n = 3$), Ancient Taiwanese group Taiwan_Hanben ($n = 44$), Russia native Kamchatka Peninsula group Itelmen ($n = 1$) and 40,000-year old Paleolithic early Asian Tianyuan ($n = 1$) to distinguish ancestries related to different trajectories across East Asia. Additionally, we conducted qpAdm(v650) analysis with the parameter ‘allsnp: YES’ to identify the most probable ancestral sources and model the ancestries of admixed populations. To investigate the admixture patterns of northern East Asian and southern East Asian, a rotation strategy was employed to determine the optimal model, with further details available in Supplementary Data 3, and all the reference combinations were tested by qpWave to make sure they are distinguished source under these outgroups.

Estimating population size by shared genomic segments. First, we identified ROH within our ancient individuals newly reported in this study using hapROH packages⁷⁹ (version 0.1a8, <https://pypi.org/project/hapROH/>) in python, which uses 5008 phased haplotypes of the 1000 Genomes Project dataset⁷³ (Phase 3, release 20130502) as the reference panel for Hidden Markov Model (HMM). Referring to previous simulation, we applied the method on ancient individuals on 1240 K panel to identify ROH longer than 4 cM. We used the default parameter suited for 1240 K ancient data. Then we sum up the distribution and number of ROHs in different length of each individual (classified into 4–8 cM, 8–12 cM, 12–20 cM, >20 cM, respectively). Then we used a maximum-likelihood inference method to estimate effective population size (N_e) from ROH refer to the description in Fernandes et al.⁸⁰. It infers the N_e that maximized the likelihood for observed ROH lengths in a given set of individuals (depth > 0.15). 95% confidence intervals were showed by the error bar which correspond to values within 10 times replicates of the maximum of the log-likelihood function.

Classification of 2nd degree relationship categories based on multi-way IBD sharing. We derived a modified method based on the foundational principle of CREST³⁴ to attempt to classify 2nd degree relationship categories—grandparent-grandchild (GP), avuncular (AV), and half siblings (HS), utilizing multi-way IBD sharing with mutual relatives of 3rd or more degree. Briefly, we categorized all potential pedigree relationships that 2nd degree pairs may have with their mutual relatives into multiple cases (Supplementary Fig. 12a); Then, for each mutual relative associated with a given 2nd degree pair, we ascertained the occurrence probability of their IBD sharing statistics in any of the defined cases (Supplementary Fig. 12b). Consequently, the probability of the 2nd degree pair belonging to each specific type was computed directly using a naïve Bayesian method, leveraging the aforementioned occurrence probabilities.

To evaluate the power of our method to distinguish these 3 categories, we subjected it to a simulated test dataset and obtained the precision rates of 77.8%, 70.9%, and 71.1% for GP, AV, and HS pairs, respectively. Lastly, we applied our method to infer the most probable pedigree structure for the three BLG-M13 individuals (M13-7, M13-75 and M13-90). See supplementary notes for a detailed description of this approach.

Kinship structure estimation based on total IBD length matrix simulation. We developed a hypothesis-driven approach based on the

overall IBD sharing pattern among randomly sampled individuals from a population to estimate its kinship structure. We first employed PEDSIM⁸¹ to generate simulated populations under various kinship structure assumptions. From these populations, we randomly sampled individuals and summarized pairwise total IBD length matrices. Subsequently, by comparing the matrix similarities between sampled individuals from simulated populations and target populations (denoting real population BLG-M13 or artificially generated test populations), key kinship structure properties of target populations were approximated.

We focused on two main properties: centrality C (defined as the proportion of individuals in the largest single lineage to the total individual number N). Estimation of centrality mainly aims to discern whether a population exhibits a dominant single lineage (resembling a “family tomb” scenario) or a combination of several largely independent lineages (akin to a “public burial” scenario). Therefore, we simulated over 12,000 populations from different individual number intervals with a near-uniform centrality distribution from 0.2 to 1. For each simulated population, we randomly sampled 16 male and 8 female individuals (matching the individuals with genomic depth $> 0.15X$ in BLG-M13 for simplicity of application) and generated their pairwise total IBD length matrices with IBD segments above 20 cM. The matrices were separated by individual sex and clustered using the UPGMA algorithm. Then we calculated the matrix similarity index between the simulated and target population with a simple measurement $S(A,B) = 1 - |A-B|/(|A| + |B|)$, where $\|M\|$ represents the Frobenius norm of a matrix M (Supplementary Fig. 13a). Subsequently, we computed the mean centrality of simulated populations with top and median 10% mean matrix similarities, which are denoted as $\bar{C}_{top10\%}$ and $\bar{C}_{med10\%}$, respectively. A high value of $\bar{C}_{top10\%}$ that significantly exceeds $\bar{C}_{med10\%}$ suggests that the target population exhibits a high centrality, namely a dominant single lineage (Supplementary Fig. 13b). To quantitatively evaluate the effectiveness of this method in classifying single or multi lineage structure, over 4000 populations were independently simulated as a test population dataset, and the true positive and false positive rates under different thresholds (t_1, t_2) (where t_1 define the minimum threshold for $\bar{C}_{top10\%}$ and t_2 define the minimum threshold for $\bar{C}_{top10\%} - \bar{C}_{med10\%}$) were summarized, yielding an ROC curve with an AUC of 0.727 (Supplementary Fig. 13c). Lastly, we applied this method to the IBD matrix of BLG-M13 imputed by ancIBD, yielding $\bar{C}_{top10\%} = 0.852$ and $\bar{C}_{med10\%} = 0.547$. Then, based on the precision rate under thresholds (t_1, t_2) = (0.85, 0.28), we estimated the posterior probability that BLG-M13 population consists of a dominant single lineage (Supplementary Fig. 13d).

Estimation of the total individual number N of a population primarily involves assessing the size of the community and the extent of community member included in the burial. Therefore, we simulated over 4000 single-lineage populations for each of 6 size intervals from [50,100) to [400,500). Sample extraction, matrix generation, and similarity calculation were performed as described above. We identified the peak position of mean similarity corresponding to individual number to offer an estimate of total individual number N in the target population (Supplementary Fig. 13e). To obtain quantitative results of BLG-M13 population, we ran the same procedure using independently simulated populations spanning all the 6 size intervals as test population dataset (Supplementary Fig. 13f), giving an estimate of the posterior probability of total individual number N ranges (Supplementary Fig. 13g). See supplementary notes for a detailed methodology description.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All studied human remains are curated in the School of Archaeology and Museology, Peking University. Requests for access should be directed to Prof. Chi Zhang, the co-author of this study. The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive⁸² in National Genomics Data Center⁸³, China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences (GSA-Human: HRA007922) that are publicly accessible at <https://ngdc.cncb.ac.cn/gsa-human>. The mitochondrial DNA sequences have been deposited in GenBase⁸⁴ in National Genomic Data Center⁸³, Beijing Institute of Genomics, Chinese Academy of Sciences/China National Center for Bioinformation, under accession number C.AA111067.1 to C.AA111103.1 that is publicly accessible at <https://ngdc.cncb.ac.cn/genbase>. All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The previously reported ancient DNA datasets used in this study are available in the Allen Ancient DNA Resource v54.1 (<https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data>). The dataset used for 14C radiocarbon date calibration is IntCal20 (<http://intcal.org>). The Genome Reference Consortium Human Build 37 (GRCh37) is available through the National Center for Biotechnology Information under the BioProject accession number PRJNA31257. The revised Cambridge reference sequence is available through the National Center for Biotechnology Information under the Reference Sequence accession number NC_012920.1. The base map for Fig. 1a was obtained from the Word Terrain Base domain map dataset (<https://www.arcgis.com/home/item.html?id=c61ad8ab017d49e1a82f580ee1298931>) and created with ArcGIS pro v3.0. In Addition, Source data are provided with this paper.

Code availability

All relevant code and software applied in this work have been acknowledged and corresponding citations are provided in the Methods section. Code for the kinship structure analysis and result visualization of the BLG-M13 grave presented in this study is available at the GitHub repository https://github.com/MarcusKreol/BLG_M13_analysis.

References

- Cohen, D. J. The beginnings of agriculture in China: a multiregional view. *Curr. Anthropol.* **52**, S273–S293 (2011).
- Long, T. et al. The early history of wheat in China from (14)C dating and Bayesian chronological modelling. *Nat. Plants* **4**, 272–279 (2018).
- Zhang, J. et al. Rice’s trajectory from wild to domesticated in East Asia. *Science* **384**, 901–906 (2024).
- He, K. Y., Lu, H. Y., Zhang, J. P., Wang, C. & Huan, X. J. Prehistoric evolution of the dualistic structure mixed rice and millet farming in China. *Holocene* **27**, 1885–1898 (2017).
- Weisskopf, A., Deng, Z. H., Qin, L. & Fuller, D. Q. The interplay of millets and rice in Neolithic central China: integrating phytoliths into the archaeobotany of Baligang. *Archaeol. Res. Asia* **4**, 36–45 (2015).
- He, K. Y., Lu, H. Y., Zhang, J. P. & Wang, C. Holocene spatiotemporal millet agricultural patterns in northern China: a dataset of archaeobotanical macroremains. *Earth Syst. Sci. Data* **14**, 4777–4791 (2022).
- Wang, L. & Cui, Y. F. The spatiotemporal pattern of cultural evolution response to agricultural development and climate change from yangshao culture to bronze age in the yellow river basin and surrounding regions, North China. *Front. Earth Sc.-Switz.* **9**, 657179 (2021).

8. Feng, S. F. et al. Red beer consumption and elite utensils: the emergence of competitive feasting in the Yangshao culture, North China. *J. Anthropol. Archaeol.* **64**, 101365 (2021).
9. Peterson, C. E. & Shelach, G. Jiangzhai: social and economic organization of a Middle Neolithic Chinese village. *J. Anthropol. Archaeol.* **31**, 265–301 (2012).
10. Li, Y. P., Zhang, J. N., Zhang, X. H. & Zhao, H. T. Agriculture, the environment, and social complexity from the early to late yangshao periods (5000–3000 BC): insights from macro-botanical remains in North-Central China. *Front Earth Sci. Switz.* **9**, 662391 (2021).
11. Song, Z., Chen, J. & Zhao, Y. Rice domestication and the Yangtze River civilization. *Biodivers. Sci. (Chin. Ser.)* **26**, 346 (2018).
12. Huan, X. J. et al. The emergence of rice and millet farming in the Zang-Yi corridor of Southwest China dates back to 5000 years ago. *Front Earth Sc.-Switz.* **10**, 874649 (2022).
13. Li, Y. Y. et al. Palaeoecological records of environmental change and cultural development from the Liangzhu and Qujialing archaeological sites in the middle and lower reaches of the Yangtze River. *Quatern. Int.* **227**, 29–37 (2010).
14. Leipe, C., Long, T., Sergusheva, E. A., Wagner, M. & Tarasov, P. E. Discontinuous spread of millet agriculture in eastern Asia and prehistoric population dynamics. *Sci. Adv.* **5**, eaax6225 (2019).
15. Zhang, C. et al. Study on the burial practice of tomb M13 of the Yangshao culture at Baligang site in Dengzhou City. *Chin. Archaeol.* **20**, 132–138 (2020).
16. Weisskopf, A. A wet and dry story: distinguishing rice and millet arable systems using phytoliths. *Veg. Hist. Archaeobot.* **26**, 99–109 (2017).
17. Deng, Z. H. et al. From early domesticated rice of the middle yangtze basin to millet, rice and wheat agriculture: archaeobotanical macro-remains from Baligang, Nanyang Basin, Central China (6700–500 BC). *Plos One* **10**, e0139885 (2015).
18. Ning, C. et al. Ancient genome analyses shed light on kinship organization and mating practice of Late Neolithic society in China. *Isis* **24**, 11 (2021).
19. Gao, Q. & Lee, Y. K. A biological perspective on Yangshao Kinship. *J. Anthropol. Archaeol.* **12**, 266–298 (1993).
20. Lee, Y. K. & Zhu, N. Social integration of religion and ritual in prehistoric China. *Antiquity* **76**, 715–723 (2002).
21. Jianing, H., Nan, L. & Chi, Z. Bioarchaeology of Neolithic Yangshao human skeletal remains from Baligang, Dengzhou: A diachronic approach. *Acta Anthropol. Sin.* **41**, 686 (2022).
22. Patterson, N. et al. Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
23. Mallick, S. et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016).
24. Haak, W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
25. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *Plos Genet* **2**, 2074–2093 (2006).
26. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet* **38**, 904–909 (2006).
27. Lu, D. S. et al. Ancestral Origins and Genetic History of Tibetan Highlanders. *Am. J. Hum. Genet* **99**, 580–594 (2016).
28. Lipson, M. et al. Ancient genomes document multiple waves of migration in Southeast Asian prehistory. *Science* **361**, 92–95 (2018).
29. Ning, C. et al. Ancient genomes from northern China suggest links between subsistence changes and human migration. *Nat. Commun.* **11**, 2700 (2020).
30. Raghavan, M. et al. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87–91 (2014).
31. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19**, 1655–1664 (2009).
32. Durand, E. Y., Patterson, N., Reich, D. & Slatkin, M. Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* **28**, 2239–2252 (2011).
33. Yang, M. A. et al. Ancient DNA indicates human population shifts and admixture in northern and southern China. *Science* **369**, 282–288 (2020).
34. Qiao, Y., Sannerud, J. G., Basu-Roy, S., Hayward, C. & Williams, A. L. Distinguishing pedigree relationships via multi-way identity by descent sharing and sex-specific genetic maps. *Am. J. Hum. Genet* **108**, 68–83 (2021).
35. Ringbauer, H. et al. Accurate detection of identity-by-descent segments in human ancient DNA. *Nat. Genet* **56**, 10–11 (2024).
36. Orlando, L. A genetic window into the human social past. *Proc. Natl. Acad. Sci. USA* **120**, e2312672120 (2023).
37. Mao, X. W. et al. The deep population history of northern East Asia from the Late Pleistocene to the Holocene. *Cell* **184**, 3256–3266 (2021).
38. Liu, Y. C., Mao, X. W., Krause, J. & Fu, Q. M. Insights into human history from the first decade of ancient human genomics. *Science* **373**, 1479–1484 (2021).
39. Jia, X. et al. Agricultural adaptations to topography and climate changes in Central China during the mid- to late-Holocene. *Holocene* **31**, 1705–1715 (2021).
40. Innes, J. B., Zong, Y. Q., Wang, Z. H. & Chen, Z. Y. Climatic and palaeoecological changes during the mid- to Late Holocene transition in eastern China: high-resolution pollen and non-pollen palynomorph analysis at Pingwang, Yangtze coastal lowlands. *Quat. Sci. Rev.* **99**, 164–175 (2014).
41. Lu, P. et al. Evolution of Holocene alluvial landscapes in the north-eastern Songshan Region, Central China: Chronology, models and socio-economic impact. *Catena* **197**, 104956 (2021).
42. He, K. Y. et al. Antipodal pattern of millet and rice demography in response to 4.2 ka climate event in China. *Quat. Sci. Rev.* **295**, 107786 (2022).
43. Liu, F. G. & Feng, Z. D. A dramatic climatic transition at ~4000 cal. yr BP and its cultural responses in Chinese cultural domains. *Holocene* **22**, 1181–1197 (2012).
44. McKay, N. P. et al. The 4.2 ka event is not remarkable in the context of Holocene climate variability. *Nat. Commun.* **15**, 6555 (2024).
45. Hu, C. Y. et al. Quantification of Holocene Asian monsoon rainfall from spatially separated cave records. *Earth Planet Sc. Lett.* **266**, 221–232 (2008).
46. Chen, F. H. et al. East Asian summer monsoon precipitation variability since the last deglaciation. *Sci. Rep.-Uk* **5**, 1–11 (2015).
47. Marcott, S. A., Shakun, J. D., Clark, P. U. & Mix, A. C. A Reconstruction of regional and global temperature for the past 11,300 years. *Science* **339**, 1198–1201 (2013).
48. Zhu, X. H. et al. Late Neolithic phytolith and charcoal records of human activities and vegetation change in Shijiahe culture, Tanjialing site, China. *Plos One* **12**, e0177287 (2017).
49. Zhang, K. S. Secondary burial in ancient Chinese clan society. *Cult. Relics Cent. China (Chin. Ser.)* **01**, 41–46 (1999).
50. Yuan, S. X. et al. Comparison of different bone pretreatment methods for AMS 14C dating. *Nucl. Instrum. Meth B* **172**, 424–427 (2000).
51. Ramsey, C. B. Bayesian analysis of radiocarbon dates. *Radiocarbon* **51**, 337–360 (2009).
52. Reimer, P. J. et al. The intcal20 northern hemisphere radiocarbon age calibration curve (0–55 Cal Kbp). *Radiocarbon* **62**, 725–757 (2020).

53. Pinhasi, R., Fernandes, D. M., Sirak, K. & Cheronet, O. Isolating the human cochlea to generate bone powder for ancient DNA analysis. *Nat. Protoc.* **14**, 1194–1205 (2019).
54. Sirak, K. et al. Human auditory ossicles as an alternative optimal source of ancient DNA (vol 30, pg 427, 2020). *Genome Res.* **30**, 938–938 (2020).
55. Dabney, J. & Meyer, M. Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *Bio-techniques* **52**, 87–94 (2012).
56. Damgaard, P. B. et al. Improving access to endogenous DNA in ancient bones and teeth. *Sci. Rep.-Uk* **5**, 1–12 (2015).
57. Caroe, C. et al. Single-tube library preparation for degraded DNA. *Methods Ecol. Evol.* **9**, 410–419 (2018).
58. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil - DNA - glycosylase treatment for screening of ancient DNA. *Philos. T R. Soc. B* **370**, 20130624 (2015).
59. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
60. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
61. Andrews, R. M. et al. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet* **23**, 147–147 (1999).
62. Peltzer, A. et al. EAGER: efficient ancient genome reconstruction. *Genome Biol.* **17**, 1–14 (2016).
63. Renaud, G., Slon, V., Duggan, A. T. & Kelso, J. Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* **16**, 1–18 (2015).
64. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).
65. Li, J. Z. et al. Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319**, 1100–1104 (2008).
66. Briggs, A. W. et al. Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl Acad. Sci. USA* **104**, 14616–14621 (2007).
67. Fu, Q. M. et al. DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl. Acad. Sci. USA* **110**, 2223–2227 (2013).
68. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of next generation sequencing data. *Bmc Bioinform.* **15**, 1–13 (2014).
69. van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* **30**, E386–E394 (2009).
70. Weissensteiner, H. et al. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* **44**, W58–W63 (2016).
71. Ralf, A., González, D. M., Zhong, K. Y. & Kayser, M. Yleaf: software for human y-chromosomal haplogroup inference from next-generation sequencing data (vol 35, pg 1291, 2018). *Mol. Biol. Evol.* **35**, 1820–1820 (2018).
72. Baslan, T. et al. Genome-wide copy number analysis of single cells. *Nat. Protoc.* **7**, 1024–1041 (2012).
73. Altshuler, D. M. et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
74. Allentoft, M. E. et al. Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172 (2015).
75. Skoglund, P., Storå, J., Götherström, A. & Jakobsson, M. Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J. Archaeol. Sci.* **40**, 4477–4482 (2013).
76. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet* **81**, 559–575 (2007).
77. Popli, D., Peyregne, S. & Peter, B. M. KIN: a method to infer relatedness from low-coverage ancient DNA. *Genome Biol.* **24**, 10 (2023).
78. Rubinacci, S., Hofmeister, R. J., da Mota, B. S. & Delaneau, O. Imputation of low-coverage sequencing data from 150,119 UK Biobank genomes. *Nat. Genet* **55**, 1088–1090 (2023).
79. Ringbauer, H., Novembre, J. & Steinrücken, M. Parental relatedness through time revealed by runs of homozygosity in ancient DNA. *Nat. Commun.* **12**, 5425 (2021).
80. Fernandes, D. M. et al. A genetic history of the pre-contact Caribbean. *Nature* **590**, 103–110 (2021).
81. Caballero, M. et al. Crossover interference and sex-specific genetic maps shape identical by descent sharing in close relatives. *Plos Genet* **15**, e1007979 (2019).
82. Chen, T. et al. The genome sequence archive family: toward explosive data growth and diverse data types. *Genom. Proteom. Bioinform.* **19**, 578–583 (2021).
83. CNGB-NGDC Members and Partners. Database Resources of the National Genomics Data Center, China National Center for Bioinformation in 2024. *Nucleic Acids Res.* **52**, D18–D32 (2024).
84. Bu, C. et al. GenBase: a nucleotide sequence database. *Genom. Proteom. Bioinform.* **22**, qzae047 (2024).

Acknowledgements

We thank Prof. Zhenhua Deng for the Archaeobotanical data of Baligang. We also thank Binxiao Feng, Prof. Fuchou Tang, Prof. He Yu and Dr. Ruiqiang Li for insightful discussions. This work is supported by National Natural Science of China (T2188102 to Y.H. and T2192953 to Y.P., H.Z.), Beijing National Laboratory for Molecular Sciences (BNLMS-CXTD-202401 to Y.H.), and Beijing Advanced Innovation Center for Genomics. Jincheng Wang was supported in part by the Postdoctoral Fellowship of Peking-Tsinghua Center for Life Sciences.

Author contributions

Conceptualization: Y.H. Experiment: T.Y., J.H., C.L., J.W., X.J., X.W., and Y.P. Data analysis: T.Y., J.H., C.L., C.N., T.Z., J.W., G.W., H.Z., C.Z., S.Y., X.Z., F.W., C.W., X.W., Y.P., and Y.H. Writing: Y.H., Y.P., T.Y., C.N., T.Z., and G.W.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-63743-1>.

Correspondence and requests for materials should be addressed to Chi Zhang, Xiaohong Wu, Yuhong Pang or Yanyi Huang.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025