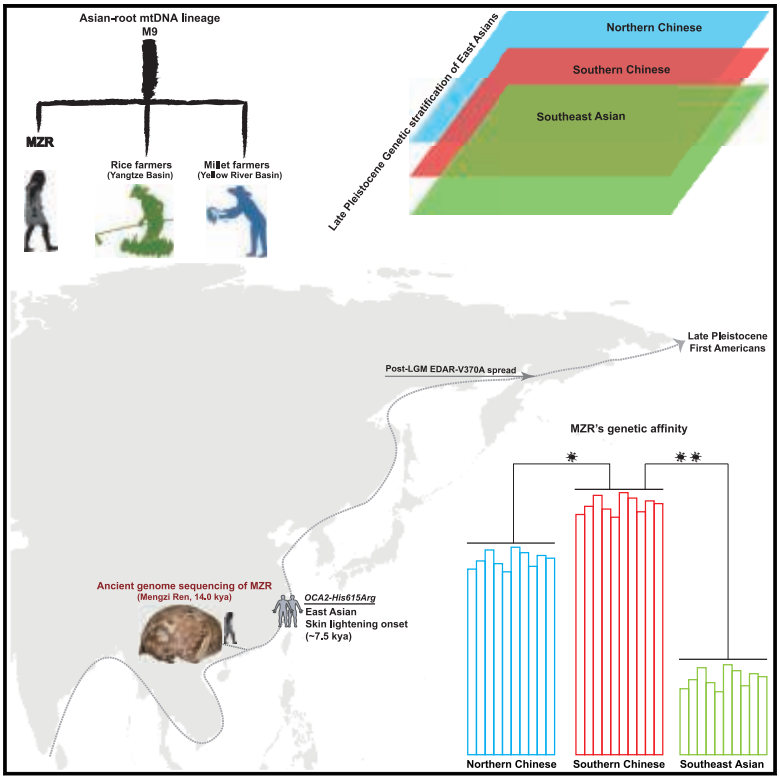
Article

Current Biology

A Late Pleistocene human genome from Southwest China

Graphical abstract



Highlights

- Genome sequencing of a Late Pleistocene human (MZR, ~14.0 kya) in Southwest China
- MZR represents an early diversified modern human lineage in East Asia
- Genetic stratification in ancient southern populations of East/Southeast Asia
- MZR deeply links to the East Asian ancestry that contributed to First Americans

Authors

Xiaoming Zhang, Xueping Ji, Chunmei Li, ..., Yanyi Huang, Yaoxi He, Bing Su

Correspondence

jxpchina@foxmail.com (X.J.), heyaoxi@mail.kiz.ac.cn (Y.H.), sub@mail.kiz.ac.cn (B.S.)

In brief

Zhang et al. conduct genome sequencing of 14,000-years-ago human remains (Mengzi Ren, MZR) unearthed in Southwest China. MZR represents an early diversified human lineage in eastern Asia where they detect a clear genetic stratification of ancient populations. MZR deeply links to the East Asian ancestry that contributed to First Americans.

Zhang et al., 2022, Current Biology *32*, 1–15
July 25, 2022
https://doi.org/10.1016/j.cub.2022.06.016

CellPress

Current Biology

CellPress

Article

A Late Pleistocene human genome from Southwest China

Xiaoming Zhang,[1,6,10] Xueping Ji,[2,3,*] Chunmei Li,[1,6,10] Tingyu Yang,[7] Jiahui Huang,[1,4] Yinhui Zhao,[1,4] Yun Wu,[3,8,9] Shiwu Ma,[5] Yuhong Pang,[7] Yanyi Huang,[7] Yaoxi He,[1,6,*] and Bing Su[1,6,11,*]

[1]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650201, China
[2]Kunming Natural History Museum of Zoology, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China
[3]Department of Paleoanthropology, Yunnan Institute of Cultural Relics and Archaeology, Kunming 650118, China
[4]Kunming College of Life Science, University of Chinese Academy of Sciences, Beijing 100049, China
[5]Mengzi Institute of Cultural Relics, Mengzi, Yunnan Province 661100, China
[6]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming 650201, China
[7]Biomedical Pioneering Innovation Center (BIOPIC) and Beijing Advanced Innovation Center for Genomics (ICG), Peking University, Beijing 100871, China
[8]School of History, Wuhan University, Wuhan 430072, China
[9]Archaeological Institute for Yangtze Civilization, Wuhan University, Wuhan 430072, China
[10]These authors contributed equally
[11]Lead contact
*Correspondence: jxpchina@foxmail.com (X.J.), heyaoxi@mail.kiz.ac.cn (Y.H.), sub@mail.kiz.ac.cn (B.S.)
https://doi.org/10.1016/j.cub.2022.06.016

SUMMARY

Southern East Asia is the dispersal center regarding the prehistoric settlement and migrations of modern humans in Asia-Pacific regions. However, the settlement pattern and population structure of paleolithic humans in this region remain elusive, and ancient DNA can provide direct information. Here, we sequenced the genome of a Late Pleistocene hominin (MZR), dated ~14.0 thousand years ago from Red Deer Cave located in Southwest China, which was previously reported possessing mosaic features of modern and archaic hominins. MZR is the first Late Pleistocene genome from southern East Asia. Our results indicate that MZR is a modern human who represents an early diversified lineage in East Asia. The mtDNA of MZR belongs to an extinct basal lineage of the M9 haplogroup, reflecting a rich matrilineal diversity in southern East Asia during the Late Pleistocene. Combined with the published data, we detected clear genetic stratification in ancient southern populations of East/Southeast Asia and some degree of south-versus-north divergency during the Late Pleistocene, and MZR was identified as a southern East Asian who exhibits genetic continuity to present day populations. Markedly, MZR is linked deeply to the East Asian ancestry that contributed to First Americans.

INTRODUCTION

Both genetic and archaeological data support the early entry of anatomically modern humans (AMHs) into southern East Asia 65 to 50 thousand years ago (kya),[1–5] and prehistoric south-to-north migration led to the current clinal structure of genetic diversity in East Asian populations.[6–8] According to the unearthed archaeological sites, these earliest AMH settlers in southern East Asia (mainland Southeast Asia [MSEA] and southern China) were ancestral to the later Hòabìnhian hunter-gatherers, who flourished in the region until ~4,000 years ago.[9,10]

Thus far, archaeological exploration in southern China has excavated numerous Late Paleolithic sites (~50.0 to 11.0 kya),[11–17] and the oldest ^{14}C date (43.5 kya) associated with Hòabìnhian pebble and flake tools currently comes from Yunnan Province of Southwest China.[18] Fortunately, some of the Late Paleolithic sites contain human remains dated to ~30.0 to 11.5 kya.[11,12,14,15,17] Among them, MZR (Mengzi Ren) from Malu Dong (Red Deer Cave) (~14.0 kya, from Yunnan Province of Southwest China) and LLR (Longlin Ren) from Laomaocao Dong (~10.5 kya, from Guangxi Province of southern China)[19] were reported to possess mosaic features of AMHs and archaic hominins based on morphological characterization.[12,20–22]

Neighboring to rainforest MSEA, Yunnan is characterized by its high biogeographic and species diversity with palaeoendemism.[23] Yunnan hosts more than 200 fossiliferous sedimentary basins documenting the evolutionary history of biodiversity, monsoon development, and regional elevation changes[24] and comprises subtropical evergreen broad-leaved coniferous forest, ranking as one of the most floristic endemic centers,[25–27] as well as the richest ethnically and linguistically diverse region in China (the 7th national census of China, 2020). Malu Dong (103°24′ E, 23°20′ N) is a partially mined cave fill located in southeastern Yunnan. It was originally excavated in 1989,[17] and a major sampling was carried out in 2008 by an international team.[20] Approximately 30 pieces of hominin remains were unearthed in the cave, including a nearly

Please cite this article in press as: Zhang et al., A Late Pleistocene human genome from Southwest China, Current Biology (2022), https://doi.org/10.1016/j.cub.2022.06.016

complete cranium calotte (MLDG-1704, the specimen studied here) and a proximal femur (MLDG-1678).[20,21] It is uncertain whether these hominin remains belong to the same individual. The calibrated radiocarbon dating sequence of the cave spans the intervals of 18,070–17,590 cal. yBP (calibrated years before present, 95% interval) to 13,415–13,165 cal. yBP (95% interval), and the hominin remains found in a series of deposits dated from 14,650–13,970 cal. yBP (95% interval) to 13,750–13,430 cal. yBP (95% interval).[20,21] We attempted to directly date the exact MZR specimen used in this study, but unfortunately, due to the poor preservation and small quantities of samples, not enough collagen fractions were recovered for radiocarbon dating. Hence, the date of 14.0 kya for the MZR is not absolutely certain, although the dating sequence of the deposits containing hominin remains is rather narrow, supporting a date in the Late Pleistocene (>11.7 kya).

Physical anthropological investigations suggest that the MZR hominin remains exhibit a combination of AMH and archaic-like traits[12,20–22] (refer to the detailed morphological descriptions in the STAR Methods section). Overall, three plausible scenarios were proposed to explain the unique morphologic characteristics of MZR. First, MZR represents a late-surviving archaic hominin population even younger than the latest *H. floresiensis* (~190–50 kya)[28] in Asia. Second, the mosaic morphologies probably result from hybridization between AMHs and unknown archaic hominin species. Third, the unusual morphologies of MZR represent the retention of ancestral polymorphisms in Paleolithic AMHs.[20–22,29] To investigate these alternative scenarios, ancient genome sequences recovered from hominin remains can serve as critical evidence in revealing the identity of MZR and the genetic diversity of the Late Pleistocene hominins in southern East Asia.

RESULTS

Both mitochondrial and nuclear genome sequences confirm that MZR is a Late Pleistocene AMH

We performed aDNA extraction and genome sequencing using the MZR cranium calotte (MLDG-1704; Figure 1). MZR is the first Late Pleistocene genome from southern East Asia (Figure 1A). Obtaining aDNA from low latitude areas in southern China is challenging due to warm and humid weather and acidic soil, which are not ideal for aDNA preservation. Additionally, among the unearthed MZR hominin remains, no ideal bone materials (such as petrous bone and teeth) were available for aDNA work, and we chose a fragment of the cranium calotte (MLDG-1704) for aDNA extraction (Figure 1B).

We employed a modified version of MYbaits (human whole-genome probes) to enrich human DNA molecules, which was previously used in studying ancient samples from Southeast Asia,[30] together with single-stranded library-based U-selection enrichment and standard shotgun sequencing. The uracil selection enables physical separation of the uracil-containing DNA strands from the non-deaminated strands during DNA library preparation.[31] In total, we performed 28 aDNA extractions using drilled bone powders from the MZR cranium and constructed 45 DNA libraries (double-stranded and single-stranded libraries). We first constructed 28 libraries without uracil-DNA glycosylase (UDG) treatment and performed small-scale sequencing. Then, we generated 17 UDG-treated libraries (so that the high cytosine deamination damage of aDNA can be repaired) for large-scale sequencing.

We checked several features of the sequencing data (non-UDG libraries) to evaluate aDNA authentication. The observed average fragmental length is 64.07 and 93.56 bp for the pre-capture and post-capture reads, respectively, which was calculated by only including the $\geq$35 bp pair-end merged and mapped sequences with a mapping quality of $\geq$25 (Figure S1A). The estimated low endogenous DNA level (0.06% on average, 0.01%–0.40%) complies with the known features of aDNA (Data S1A). We then checked the terminal damage pattern using the sequence data from the non-UDG-treated libraries; the sequencing reads showed typical high G>A and C>T substitutions at the 3' ends for the double-stranded and the single-stranded libraries, respectively. However, the expected damage pattern (a high C>T substitution) at the 5′ end is not obvious (Figures S1B and S1C), likely due to the PCR protocol used (STAR Methods). When we applied a two-round PCR protocol that employed the high-fidelity polymerase in the second-round PCR, we saw the expected damage patterns at both ends (Figure S1D). Together, these results support aDNA authentication of MZR. Among the 28 non-UDG libraries, the estimated rates of terminal damage ranged from 5.33% to 49.30%, with only three libraries having <10% rates. The estimated modern DNA contamination rates are 0.72% for nuclear DNA and 5.88% for mitochondrial DNA (mtDNA) (Data S1A and S1B), together indicating low-level modern DNA contamination.

Given the validated aDNA authentication, we used the 17 UDG-treated libraries to conduct large-scale genome sequencing. We adopted stringent filtering of the clean reads so that only the high-quality sequences were retained. We first merged the forward and reverse read pairs to recover full-length sequences, and in total we obtained approximately 1.9 billion clean merged reads from the paired-end sequence data. We mapped them to the human reference genome *hs37d5*. We then trimmed the mapped full-length reads based on their terminal damage patterns (Figure S1E) (see STAR Methods section) and remapped the trimmed reads to the reference genome. The sequences with low mapping qualities (<25) were discarded. We used the filtered and merged bam file to perform genotype calling.

Initially, we recovered 100.97 million base pairs (~0.113× coverage of the nuclear genome) using *snpAD*.[32] We also generated the "pseudohaploid" genotype using the *pileupCaller* program in *sequenceTools* (https://anaconda.org/bioconda/sequenceTools). Within the 1,240K SNP sets, we found that the genotyping results by *snpAD* and *pileupCaller* were highly consistent with each other (99.2% overlap) (Figures S1F and S1G). To further check the possible impact of artificial C to T substitutions caused by deamination damage, we performed principal component analyses (PCAs) among MZR and five modern Asian populations from the 1000 Genome Project (1KG) (http://www.1000genomes.org) using either all SNPs (2,727,839) or transversion-only SNPs (876,456). We observed the same clustering pattern in the PCA maps derived from the two SNP sets (Figure S2). Additionally, similar levels of genetic affinity of MZR with modern East Asians were detected by *f3* statistics[33] (Data S1C). Thus, the impact of deamination-induced damage is negligible in the MZR genome data.

Please cite this article in press as: Zhang et al., A Late Pleistocene human genome from Southwest China, Current Biology (2022), https://doi.org/10.1016/j.cub.2022.06.016

CellPress

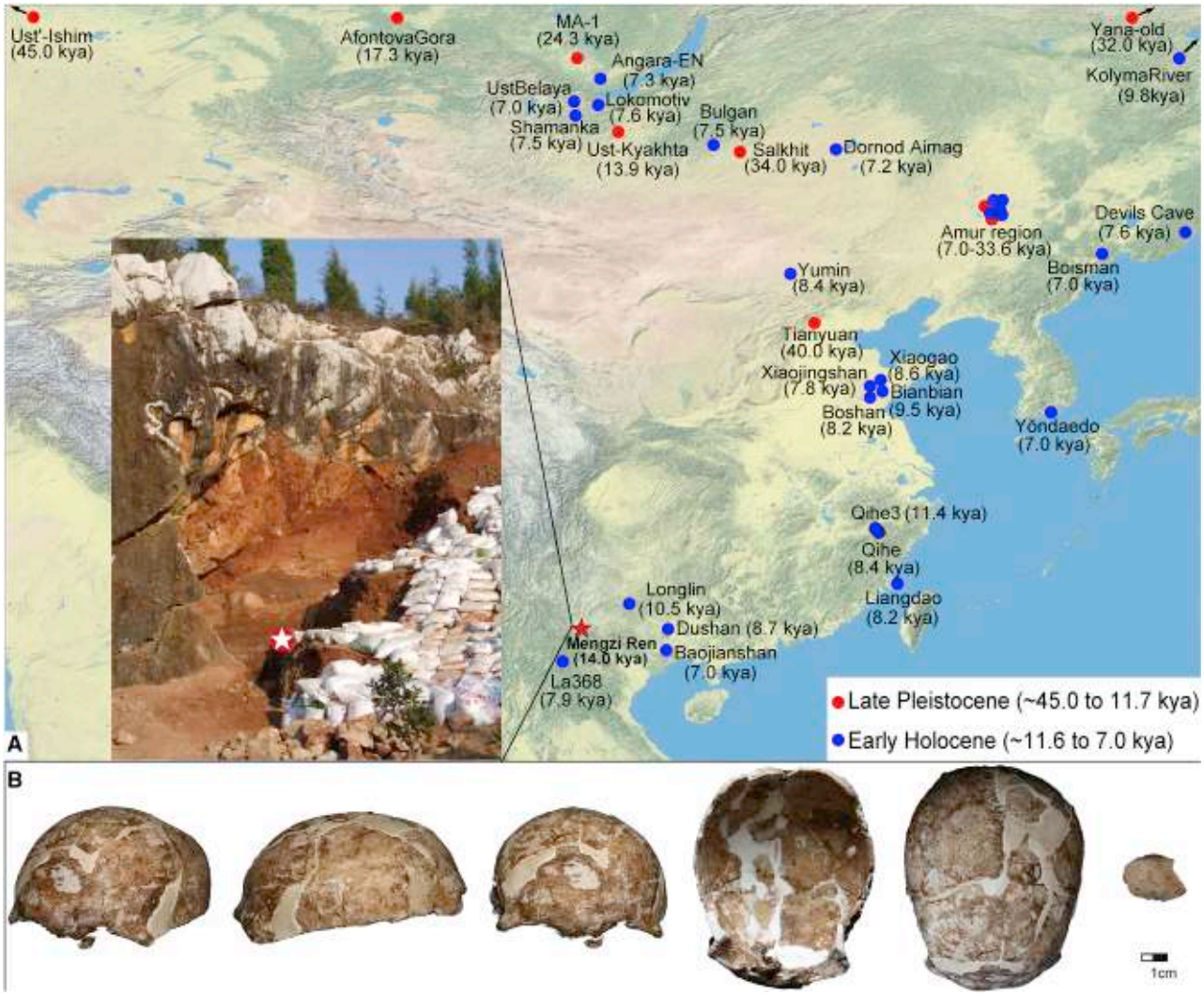


Figure 1. The Mengzi Ren (MZR) cranium in the context of the Late Pleistocene and Early Holocene sites with ancient genome data in East Asia

(A) Pictures of the excavation site called Maludong (Red Deer Cave) and the unearthed MZR cranium calotte (MLDG-1704) in this study. The hominin remains, including the cranium calotte, were discovered at the indicated spot (white star) on the site. Maludong is located in Yunnan Province of Southwest China (red star). The geographic locations of the published aDNA samples older than 7.0 kya in East Asia are indicated, and they are grouped into two time periods, the Late Pleistocene (~45.0 to 11.7 kya) and the Early Holocene (~11.6 to 7.0 kya).

(B) The multi-angle views of the MZR cranium calotte (MLDG-1704) (from left to right: diagonal forward view, left lateral view, anterior view, cranial cavity view, and superior view). The sampled specimen for aDNA extraction (a 2 × 3 cm piece) is presented at the right.

See also Data S1.

MZR was identified as a female based on the mapping ratio between the Y chromosome and autosomes (N_Y/N_{auto} = 0.0026). Fortunately, due to the high copy numbers of mitochondria in the cell, we were able to obtain on average a 125.05× sequencing depth of the mitochondrial genome and recovered 95.84% (15,880/16,569) of the mtDNA genomic sites (Data S1A and S1D). Clearly, the MZR mtDNA belongs to the AMH lineage and is assigned to a basal macro-branch of the M9 haplogroup in the global human mtDNA tree ($PhyloTree_{mt}$, Build 17) (http://www.phylotree.org/) (Figure 2A). The M9 haplogroup comprises two macro-branches (M9a'b and E) in current human populations, and the MZR mtDNA represents the third macro-branch containing one private mutation (T16304C), which is extinct in current human populations (Figure 2A; Data S1D). Macro-branch M9a'b is currently distributed in mainland East Asia with a southern origin and northward expansion at approximately 18 to 28 kya.[34] In contrast, macro-branch E is mainly distributed in island Southeast Asia (ISEA) and the Solomon Islands of Melanesia, reflecting the Neolithic expansion of Austronesian speakers[35] (Figure 2A; Data S1E). Hence, the discovery of an extinct basal M9 lineage for the MZR suggests a rich matrilineal diversity of human populations in southern East Asia during the late Pleistocene.

Please cite this article in press as: Zhang et al., A Late Pleistocene human genome from Southwest China, Current Biology (2022), https://doi.org/10.1016/j.cub.2022.06.016

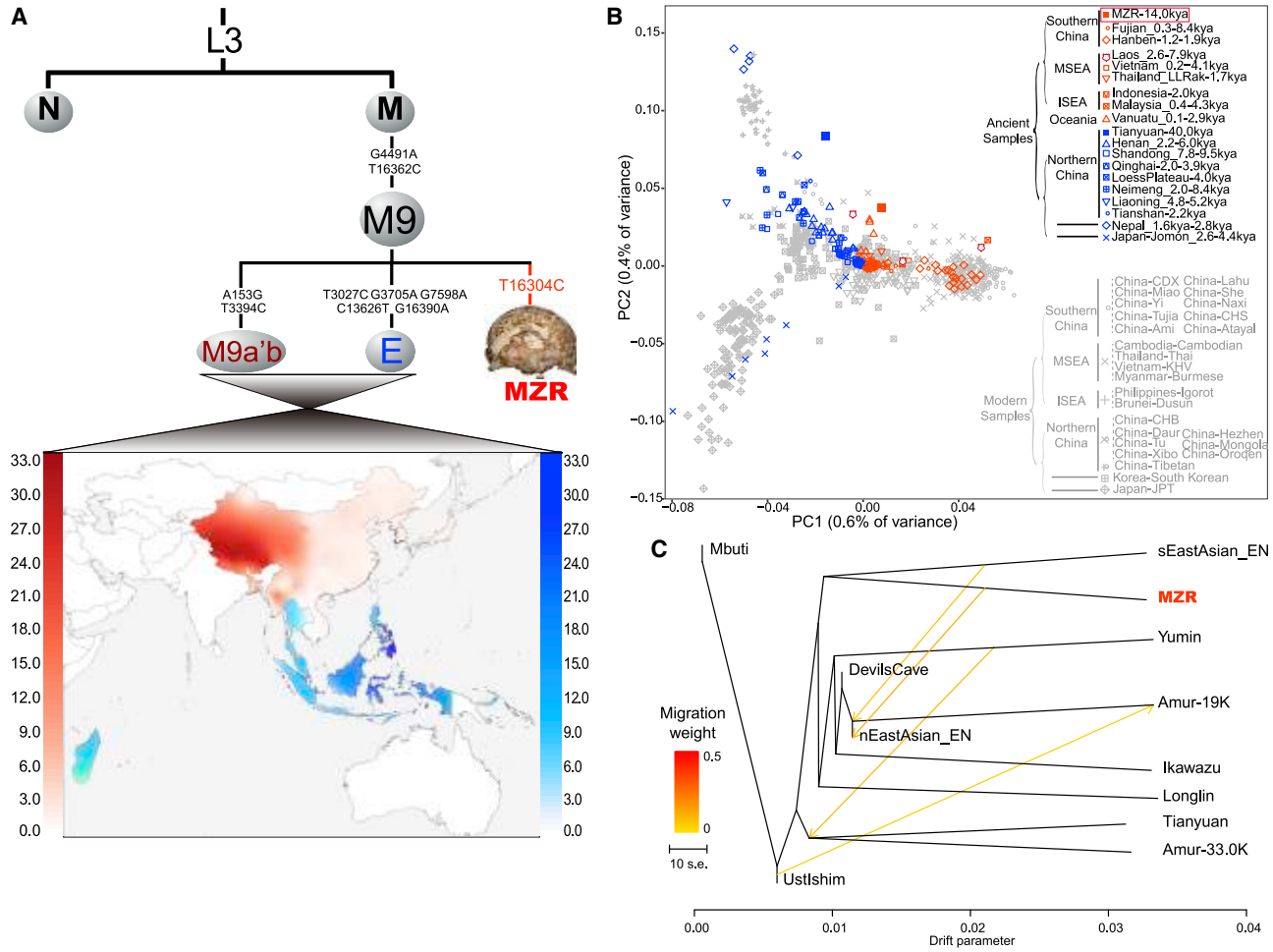


Figure 2. Genetic identity of MZR and her affinity with other ancient East Asians

(A) MZR is identified as a novel and basal mtDNA lineage belonging to the Asian-prevalent M9 haplogroup. The map shows the geographic distribution (in frequency) of the two present-day M9 macro-branches (M9a'b in red and E in blue) in East Asia.

(B) The principal component analysis (PCA) projection based on the nuclear genome sequences confirms MZR's identification of modern humans and her affinity with southern East Asians. Both ancient and modern samples are included in the PCA. We project the ancient samples onto the PC space established using the modern samples with *smartpca*.[36] MSEA, mainland Southeast Asia; ISEA, islands Southeast Asia; CHB, Han Chinese in Beijing, China; CHS, Southern Han Chinese in Guangzhou, China; CDX, Chinese Dai in Xishuangbanna, China; KHV, Kinh in Ho Chi Minh City, Vietnam; JPT, Japanese in Tokyo, Japan.

(C) The maximum likelihood tree covering the main ancient East Asian samples (≥7.0 kya) (with 4 migration edges). sEastAsian_EN: Early Neolithic southern East Asians, including Qihe3, Baojianshan5, Dushan, and Liangdao2. nEast Asian_EN: Early Neolithic northern East Asians, including Xiaogao, Bianbian, and Boshan. MZR clusters with the southern East Asians.

See also Figures S2A–S2C and S3A–S3D and Data S1.

Consistently, the nuclear genome sequences also support MZR as an AMH. In the PCA map covering both modern and ancient samples from East Asia, the MZR falls in the variation range of modern humans and is close to southern East Asians (Figures 2B and S2). The southern East Asian affiliation of MZR contrasts with the known affinity of Tianyuan (40.0 kya)[37] with northern East Asians (Figure 2B; Data S1F). Consistently, the *TreeMix*[38] analysis indicates that among the major East Asia aDNA samples (40.0–7.0 kya), MZR clusters with Early Neolithic coastal southern East Asians (sEastAsia_EN, including Qihe3 [11.5 kya], Liangdao2 [7.5 kya], Baojianshan5 [7.4 kya], and Dushan4 [8.7 kya]),[19,39] and they form the southern clade, clearly separated from the northern clade covering Early Neolithic coastal northern East Asians (nEastAsia_EN, including Bianbian [9.5 kya], Boshan [8.2 kya], and Xiaogao [8.6 kya]), Yumin (8.4 kya), DevilsCave (7.6 kya), and Amur (19.0 kya) (Figures 2C and S3). Therefore, the genome of MZR shows an affinity with southern East Asians of similar dates and implies that some degree of genetic divergence between southern and northern East Asians was likely present during the Late Pleistocene, though a low gene flow from MZR to nEastAsian_EN was detected (Figure 2C). This observation is also supported by the result of the *D* (MZR, UKY; ancient East Eurasia [≥7.0 kya], Mbuti) test, where a clinal south-to-north divergence in East Asia is detected. MZR shares more alleles with ancient southern East Asians (such as Liangdao2), while UKY is relatively close to northern East Asians (Figure 3A; Data S1G). In addition, all the ancient samples (14.0–7.5 kya) in southern East Asia exhibit negative correlations with latitude when compared

Please cite this article in press as: Zhang et al., A Late Pleistocene human genome from Southwest China, Current Biology (2022), https://doi.org/10.1016/j.cub.2022.06.016

CellPress

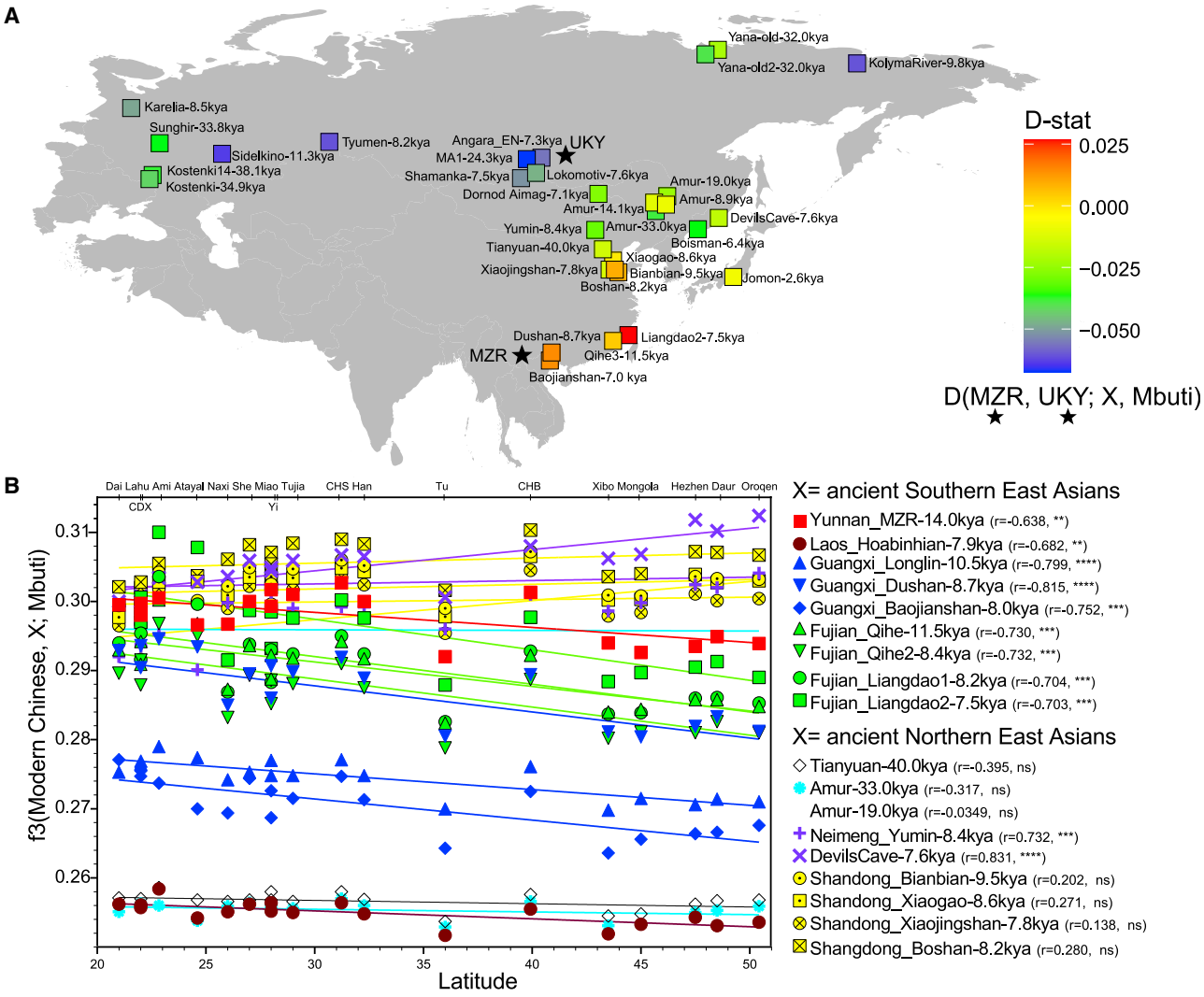


Figure 3. The Late Pleistocene south-versus-north population divergence in East Asia

(A) The plot of the *D* (MZR, UKY; ancient East Asian/Siberian [above 7.0 kya], Mbuti) test to visualize the south-versus-north population divergence in East Asia. UKY: the Ust-Kyakhta-3 sample (13.9 kya) is located south of Lake Baikal,[40] which possesses major East Asian genetic components.

(B) Correlation of the latitudes of the modern Chinese populations with the *f3* values of ancient East Asians (≥7.0 kya). In brief, the affinity degree of ancient southern East Asians shows a negative correlation with the latitude of modern Chinese, but no correlation with most of the ancient northern East Asians was observed, indicating the Late Pleistocene south-north population divergence in East Asia. The p values of the correlation coefficients are indicated by asterisks (**p < 0.01, ***p < 0.001, ****p < 0.0001). ns, not significant.

See also Figure S5 and Data S1.

to modern Chinese, while most of the ancient samples (40.0–7.6 kya) in northern East Asia show no correlation, and two samples (Yumin and DevilsCave) show positive correlations with latitude (Figure 3B; Data S1H). This pattern is consistent with the above view that some degree of genetic divergence between southern and northern East Asians was present during the Late Pleistocene. Of note, the clustering pattern of LLR and Ikawazu (the 2.6 kya Jomon individuals from Japan) is slightly different from the previous report,[19] possibly due to the high missing rates (69.14% and 18.09%, respectively) in the merged genome dataset of these two samples.

Given the AMH identity of MZR, to evaluate the level of introgression from archaic humans, using African (Mbuti) as the control group (assuming no introgression), we performed an *f4-ratio* test with ADMIXTOOLS.[33] The estimated introgression levels from archaic humans in the MZR genome are ~1.29% for Denisovan ancestries and ~1.27% for Neanderthal ancestries, similar to the reported introgression levels in current East Asians (1.20%–2.90% on average)[37] (Data S1I). In addition, taking chimpanzees as the outgroup, we conducted the *D* (East Asian, MZR; Denisovan/Neanderthal, chimpanzee) test to compare the introgression-ratio difference between the MZR and modern East Asians, and no difference was detected (Data S1J). Consistently, LLR was also reported as an AMH with similar low introgression levels from archaic humans.[19] Thus, these results tend to refute the proposed scenarios of

Please cite this article in press as: Zhang et al., A Late Pleistocene human genome from Southwest China, Current Biology (2022), https://doi.org/10.1016/j.cub.2022.06.016

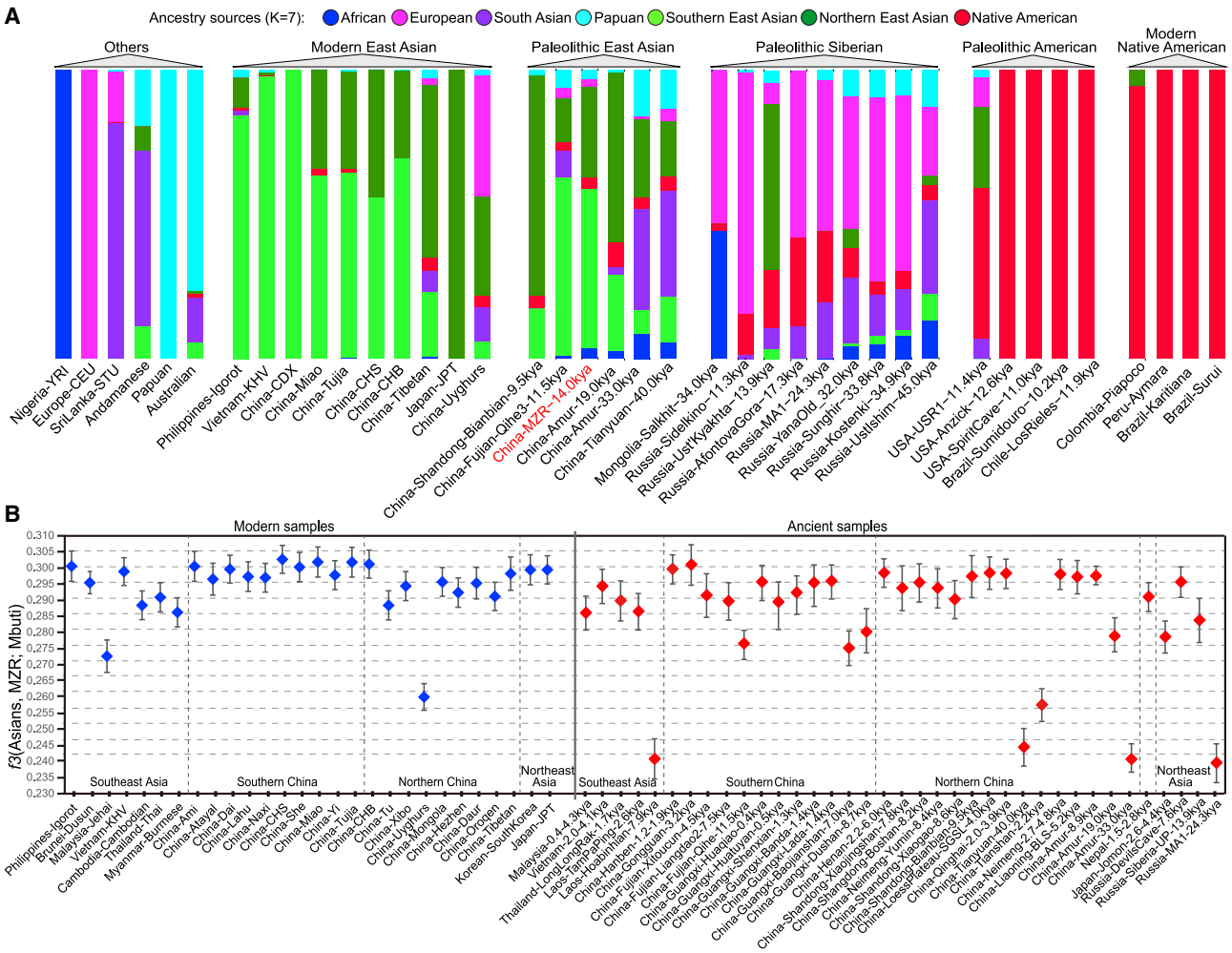


Figure 4. The genetic structure and population affinity of MZR in comparison with worldwide ancient and current human populations

(A) The genetic structure of MZR and other world populations in view of seven regional ancestry sources (K = 7). The MZR contains a large portion of southern East Asian components (light green) and a relatively small portion of northern East Asian components (dark green).

(B) The *f3* (East Asian, MZR; Mbuti) plot reveals the genetic affinity of MZR to ancient and modern East Asians (blue, modern samples; red, ancient samples). At least 20,000 SNPs were included when conducting pairwise comparisons. In general, MZR shows the highest affinity with southern Chinese and exhibits low affinity with both ancient (Hòabìnhian) and modern Southeast Asians.

See also Figures S4A–S4D and Data S1.

either surviving archaic hominins or hybridization between AMHs and unknown archaic hominins for MZR and LLR. Rather, the genome features of these two samples with "unusual" morphologies may reflect a rich diversity of Paleolithic AMHs living in southern East Asia. It should be noted that due to the low coverage of the sequenced MZR genome, we cannot completely rule out the possible existence of archaic alleles in the MZR genome introgressed from Neanderthal/Denisovan or unknown archaic hominins that may contribute to the morphological features of MZR.

Collectively, both mtDNA and nuclear genome sequences demonstrate that MZR is an AMH. Her mtDNA represents an extinct basal lineage, and her nuclear genome harbors deeply diverged Asian AMH ancestries, reflecting a rich diversity of ancient populations during the Late Pleistocene in southern East Asia.

Inferring population history of East Asians during Late Pleistocene based on the genomic data of MZR and other ancient samples

Decoding aDNA of geographically diverse human remains is highly informative in understanding population history. Compared to the systematic aDNA dissections in West Eurasia,[41] aDNA studies in East Asia are still limited.[19,30,39,42–53] The reported Late Pleistocene studies of aDNA human genomes covered only northern China,[37,49] and MZR is the first Late Pleistocene genome from southern East Asia. By integrating the genomes of current global populations and the published ancient genomes, we conducted a detailed genome structure analysis using the ADMIXTURE tools[54] (Figures 4A and S4). Consonant with the PCA and *TreeMix* results, the major genetic component in MZR belongs to southern East Asians (light green in Figure 4A), and so does Qihe3 (11.5 kya) from Fujian.[19] In contrast, Amur (19.0 kya) and Bianbian (9.5 kya)

Please cite this article in press as: Zhang et al., A Late Pleistocene human genome from Southwest China, Current Biology (2022), https://doi.org/10.1016/j.cub.2022.06.016

majorly possess a northern East Asian component, supporting the proposed south-north divergence of East Asians during the Late Pleistocene and also in line with a recent report.[49] Of note, all early post-LGM (Last Glacial Maximum) Paleo-Siberians contain appreciable proportions of the Native American component (red), confirming Siberia as the outpost of the earliest migration to America (Figures 4A and S4A).

We next performed an *f3* analysis[33] to reveal the Late Pleistocene population relationship and MZR's connection to Early Holocene and current East Asian populations. At least 20,000 SNPs were included when conducting pairwise comparisons, and the African population (Mbuti) was used as an outgroup. In line with the above results (Figures 2B, 2C, and 4A), the *f3* (modern East Asians, MZR; Mbuti) test suggests that among the modern samples, MZR is closer to southern Chinese than to northern Chinese, while among the ancient samples, this contrast is less obvious though the highest *f3* value still occurs in southern populations (Gongguan, 3.2 kya)[48] (Figure 4B; Data S1K). Notably, although the geographic location of MZR is close to Southeast Asia, MZR shows significantly less affinity to both modern and ancient Southeast Asians ($|Z| > 7.0$; Data S1K and S1L), an indication of already structured and diversified ancient populations in southern East Asia, consistent with the mtDNA data (Figure 2A; Data S1D).

In addition to the earliest southern settlement of AMHs in East Asia, ancient migration (40–18 kya) into East Asia via the "Northern Route" from West Eurasia was previously proposed. The "Northern Route" hypothesis would also explain where the subtle shared ancient north Eurasian (ANE) ancestry came from that is then also shared with Native Americans. In addition, the "Northern Route" may also contribute to the south-versus-north divergence of East Asians. This is supported by both archaeological and genetic evidence,[55–58] although the contribution of the "Northern Route" to current East Asians is relatively minor (6.78%) based on the Y chromosome data.[56,59] To test the source of the "Northern Route," we calculated population divergence levels using pairwise *Fst*, and we found that Central Asians and Siberians are the best candidates who show the lowest *Fst* values compared to East Asians, especially to Altaic speakers in northern China (Figure S5; Data S1M). The result is expected considering the geographic proximity of Central Asia and Siberia to northern China, and it is also in line with the inferred migratory route by the reported archaeological and genetic evidence.[55,56]

Lastly, the time series aDNA data can be used to track the emergence and spreading pattern of adaptive sequence variants. By utilizing the published aDNA data, we reconstructed the spatial-temporal distribution of an East Asian-specific variant (*OCA2-His615Arg*) that contributes to skin lightening due to local Darwinian positive selection (Figure 5, left panel).[60–63] It turned out that all the Late Pleistocene individuals (e.g., MZR, Tianyuan, Amur-33K, Amur-19K, and UKY) lack the derived allele (*OCA2-615Arg*). The first presence of the adaptive allele (*OCA2-615Arg*) was in Liangdao 2–7.5 kya from coastal southern China,[39] and it quickly elevated to medium frequency (25.67%, 29/113), mainly in coastal East Asia, and then spread to northern East Asia ~3,500 years ago, and finally became dominant (~60.00%) in current East Asians (Figure 5, left panel; Data S1N). This pattern suggests that the selective event in East Asians likely occurred in the Late Holocene epoch, coinciding the proposed quasi-exponential population growth during that time.[64–68] However, another East Asian-specific variant (*EDAR-V370A*)[69] exhibits a distinct pattern (Figure 5, right panel; Data S1O).

Tracing the complex migratory histories of AMHs to the Americas

We applied the outgroup *f3* (global Late Pleistocene/Early Holocene populations, MZR; Mbuti) test to determine the genetic affinity of MZR to global populations (Figure 6A; Data S1P–S1R). Among the Late Pleistocene samples (45.0–11.7 kya), MZR exhibits the closest affinity with the Paleo-Siberian UKY (13.9 kya) (*f3* = 0.2839) (closely related to First Americans[40]) and First Americans (maximum *f3* = 0.2792), even closer than to Tianyuan (0.2445) and the Amur samples (Amur-33.0kya [0.2409] and Amur-19.0K [0.2792]). The D tests also indicate that MZR/UKY and MZR/Amur-19.0K are cladal with respect to First Americans (Data S1Q), suggesting the East Asian contribution to Native Americans likely originated prior to the south-versus-north East Asian divergence. Moreover, we performed the *D* (MZR, X; First Americans, Mbuti) tests to compare the affinity of First Americans with MZR, the ancient coastal East Asians, and Paleo-Siberians (Figures 6B and S6A–S6D; Data S1Q), and we observed that First Americans (USR1, 11.4 kya; Spirit Cave, 11.0 kya; Los Rieles, 11.9 kya; Sumidouro, 10.4 kya) all exhibited higher affinity with MZR than with the late Hòabìnhian populations from Southeast Asia, the Jomon population (hunter-fishers) from Japan, and the pre-LGM Late Pleistocene UstIshim (45.0 kya), Tianyuan (40.0 kya), Salkhit (34.0 kya), Sunghir (33 kya), YanaOld (32.0 kya), and MA-1 (24.3 kya). Hence, by connecting with the Early Holocene coastal East Asians (Liangdao2) (Figure 3A) and nEastAsians_EN (Figures 2C and S3), MZR is linked deeply and indirectly to the East Asian ancestry that contributed to First Americans. Consistently, the post-LGM samples (Amur-14.1K, Amur-6.3K, KolymaRiver, DevilsCave, Shamanka, and UstBelaya) from the Amur region and the Far East region of Siberia all show a close affinity to First Americans (the $|Z|$ scores of *D* [MZR, the listed post-LGM samples; First Americans, Mbuti] test ≈3 or >3) (Data S1Q), supporting the proposed migratory route by way of the Far East region of Siberia, and from there First Americans crossed the Bering Strait.

To further test the contribution of the Late Pleistocene East Asians to the earliest Native Americans, we checked the spatial-temporal distribution of an East Asian-specific variant (*EDAR-V370A*) with an estimated occurrence of 30 kya based on genetic data of current populations.[69] Among the ancient samples, the earliest presence of this variant was in Amur-19.0K from northern East Asia,[49] followed by UKY-13.9kya, Amur-14.5K, and Amur-14.1K during the Late Pleistocene (Figure 5E; Data S1O). In America, the earliest presence was in LosRieles (12.0 kya) from coastal Chile of South America. Subsequently, *EDAR-V370A* was elevated to extremely high frequencies in both East Asia (89.41%, 76/85) and America (93.33%, 28/30) during the Early Holocene (11.6–5.0 kya) (Figure 5F) and continued to maintain high frequency in the Late Holocene (Figure 5G) and modern time (Figure 5H). Hence, the spatial-temporal distribution of *EDAR-V370A* supports a clear contribution of the Late Pleistocene East Asians to the earliest

Please cite this article in press as: Zhang et al., A Late Pleistocene human genome from Southwest China, Current Biology (2022), https://doi.org/10.1016/j.cub.2022.06.016

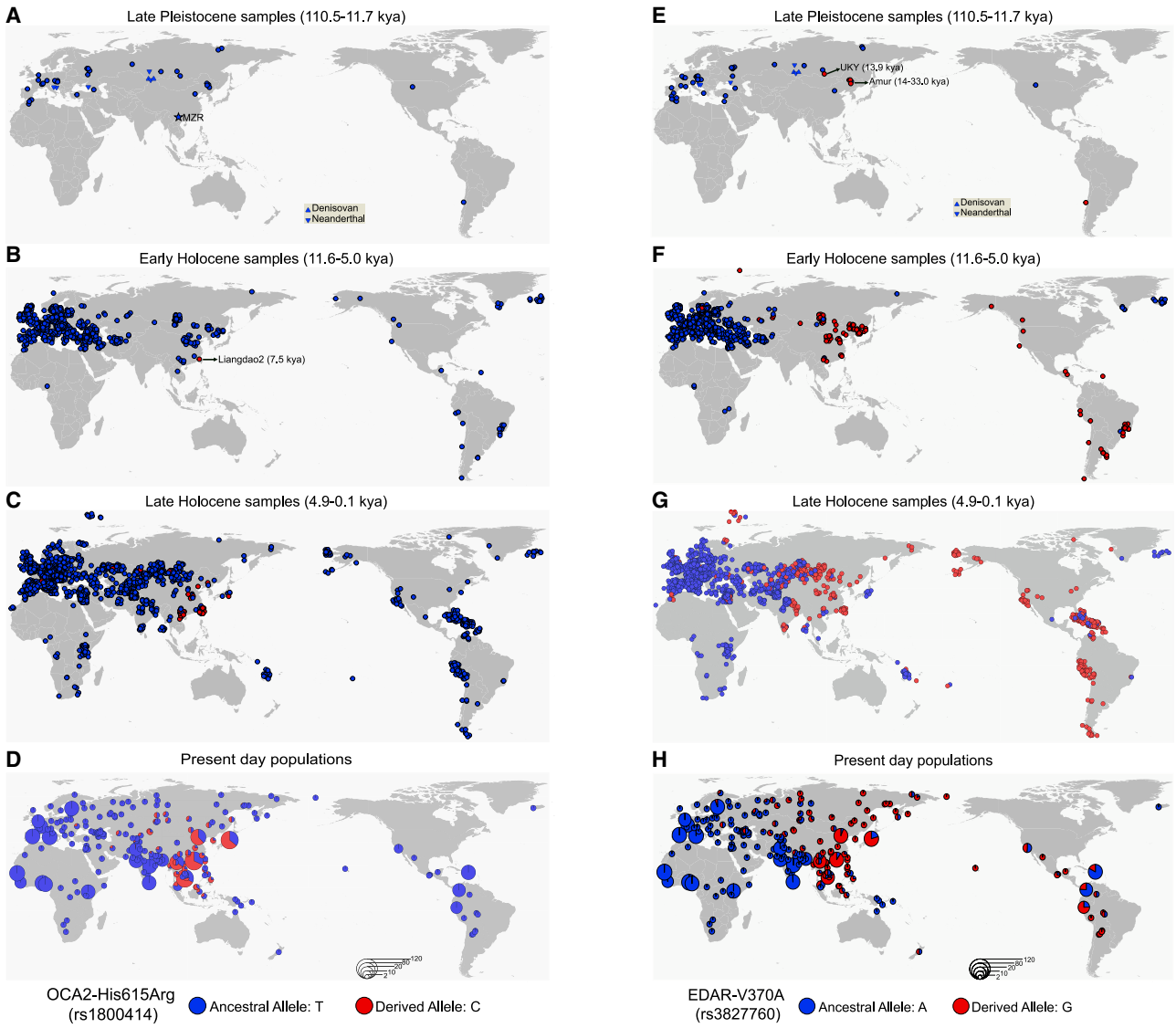


Figure 5. The spatiotemporal distribution of two East Asian-specific adaptive variants: *OCA2-His615Arg* (left) and *EDAR-V370A* (right)

We grouped the samples into four time windows according to the ages of the samples.

(A and E) The Late Pleistocene samples (110.5 to 11.7 kya). The Denisovan and Neanderthal samples are included (the upward-pointing and the downward-pointing triangles, respectively). MZR is indicated with the star.

(B and F) The Early Holocene samples (11.6 to 5.0 kya).

(C and G) The Late Holocene samples (4.9 to 0.1 kya).

(D and H) The worldwide present-day populations.

The red color refers to the adaptive allele (*OCA2-615Arg* and *EDAR-V370A*), and the blue color refers to the ancestral allele (*OCA2-615His* and *EDAR-370 V*). The samples with missing genotype data for the two variants from the Allen Ancient DNA Resource (AADR) database were excluded (Data S1N and S1O). The *OCA2-HiS615Arg* was not detected in the Late Pleistocene samples (A), and it emerged the earliest in the 7.5 kya Liangdao2 individual from Fujian, coastal eastern China (B). *OCA2-HiS615Arg* was elevated to medium frequency (25.67%) during the Late Holocene, became popular in coastal East Asia (C), and expanded to high frequencies (>50%) in modern East Asians (D). *EDAR-V370A* emerged the earliest in Amur-19K, Amur-14.5K, and UKY (13.9 kya) in northern East Asia and in the LosRieles (12.0 kya) samples from coastal Chile of South America (E). It was quickly elevated to extremely high frequency in broad East Asia (89.41%) and America (93.33%) during the Early Holocene (F). *EDAR-V370A* slightly expanded to West Eurasia, and it also appeared at a relatively low frequency in some Central American populations, likely due to the known admixture with Africans and Europeans ~500 years ago during the Late Holocene (G). Likewise, the appearance of *EDAR-V370A* in Oceanians reflects the historic Austronesian dispersal from East Asia to the Pacific Islands about 2,000 years ago (G). *EDAR-V370A* is currently present in Central Asia and the Caucasus, where the west-east admixture occurred (H).

Please cite this article in press as: Zhang et al., A Late Pleistocene human genome from Southwest China, Current Biology (2022), https://doi.org/10.1016/j.cub.2022.06.016

CellPress

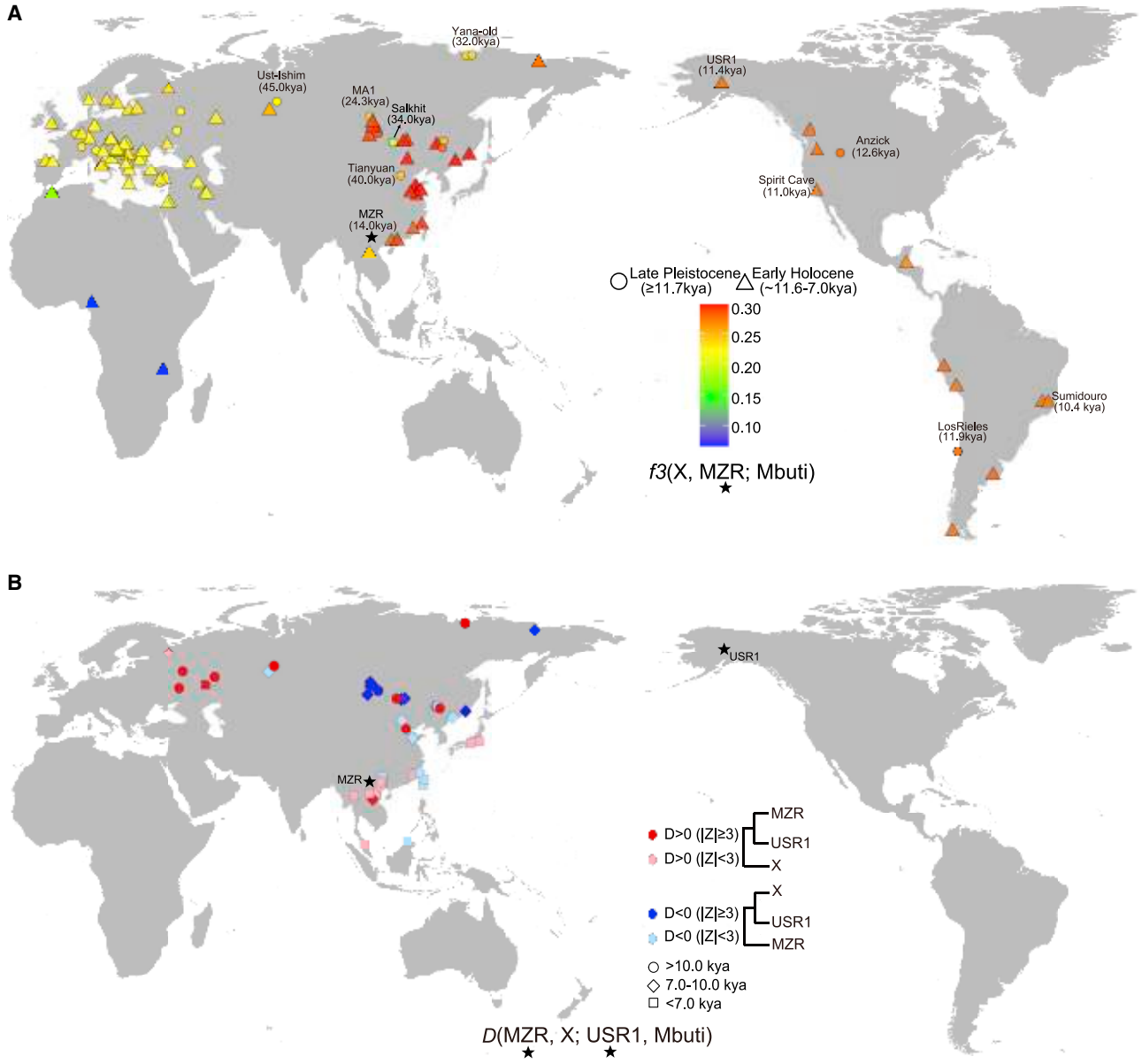


Figure 6. The genetic affinity of MZR with the first Native Americans

(A) The *f3* (global Late Pleistocene and Early Holocene samples, MZR; Mbuti) tests indicate a relatively close relationship of MZR with the first Native Americans. The *f3* values are presented in a color gradient.

(B) The *D* (MZR, ancient coastal East Asians and Paleo-Siberians; USR1, Mbuti) tests indicate the affinity of the first Native American (USR1) with MZR. In brief, USR1 is closer to MZR than to some of the ancient coastal East Asian, the late Hòabìnhian (from Southeast Asia), and Jomon (from Japan) populations, as well as to the Paleo-Siberians (above 20 kya) and Tianyuan.

See also Figures S6A–S6E and Data S1.

peopling of America. Consistently, for the *OCA2-His615Arg* variant, which occurred in the Late Holocene, we did not see its presence in either ancient or modern Native Americans (Figures 5A–5D).

When all pre-Colombian Native Americans (≥500 years ago) were included in the *f3* analysis, we did not observe obvious geographic bias of affinity with MZR, although northern Native Americans are slightly closer to MZR than southern Native Americans but statistically not significant (|*Z*| = 0.724) (Figure S6E; Data S1S). This is in line with the proposed bottleneck effect leading to genetic homogeneity of Native Americans,[70] as well as the more infrequent population movement and replacement in the Americas than in Eurasia and Africa.[71]

DISCUSSION

Southern East Asia harbors rich archaeo-anthropological sites with rich morphological diversities, including the ~100 kya

Please cite this article in press as: Zhang et al., A Late Pleistocene human genome from Southwest China, Current Biology (2022), https://doi.org/10.1016/j.cub.2022.06.016

Zhiren Cave[72] and ~120–80 kya Fuyan Cave[73] in southern China (although a recent study suggested much younger dates for these two sites[74]), the ~190–50 kya *H. floresiensis*,[28] and the 67 kya *H. luzonensis* in Southeast Asia.[75] Indeed, based on the published mitochondrial genome data, the matrilineal lineage diversity of the Late Pleistocene hominins in eastern Eurasia is quite high, including the reported Ust'-Ishim (45.0 kya, novel N*),[76] Salkhit (34.0 kya, independent novel N*),[77,78] Tianyuan (40.0 kya, basal B*),[79] Yana-old (32.0 kya, U),[80] MA-1 (24.3 kya, novel U*),[81] LLR (11.0 kya, M27d), and MZR (14.0 kya, basal M9*) in this study, many of which were lost during the post Pleistocene (<11.7 kya) human evolutionary histories. In addition, the hominin fossils from these archaeological sites, such as MZR and LLR, all exhibit rich physical anthropological diversities, some of which were thought to overlap with the morphological characteristics of archaic hominins and triggered proposals of different scenarios of human evolution in this area.[12,20–22,29,82]

In this study, we provide compelling evidence that the Late Pleistocene MZR from Malu Dong in Southwest China is an AMH. The nuclear genome data indicate that MZR represents an early diversified AMH lineage in East Asia. The mtDNA of MZR belongs to one of the root matrilineal lineages of AMHs in southern East Asia. Identified as a novel basal M9 lineage, MZR may represent one of the extinct pioneer hunter-gatherers ancestral to millet and rice farmers in China who emerged in the Yellow River and Yangtze River valley during the Early Neolithic period.[83,84] Additionally, we observe obvious stratification and substructure of ancient human populations between southern China and mainland Southeast Asia, an indication of already diversified genetic backgrounds of the Late Pleistocene populations in southern East Asia.

Importantly, we observed similar levels of introgression of Denisovan and Neanderthal ancestries in the MZR compared with current East Asians, consistent with the reported introgression levels in the LLR.[19] Thus, these observations are against the proposal of hybridization between AMHs and archaic hominins in explaining the unusual morphologies of MZR. However, due to the low coverage of the reported MZR nuclear genome, we cannot completely rule out the possibility that MZR may carry individual archaic alleles in key genomic regions that contribute to her unusual morphological features, and this possibility can be tested in the future when high-coverage genome data are available.

It should be noted that the MZR morphological data are indeed informative in reconstructing human morphological diversity during the Late Pleistocene in southern East Asia.[20,21] However, due to the limited human remains as well as the limited number of morphological traits, it would be hard to confidently reveal the identity of the studied subject. To this end, genome sequence data are critical for unequivocal species identification, quantification of genetic introgression, and reconstruction of population history.

Spatiotemporal tracing of mutations related to phenotypic changes in human populations can help reconstruct the prehistoric patterns of how natural selection has shaped these adaptive events. We observed that the *OCA2-HiS615Arg* (rs1800414) variant, a key adaptive mutation causing skin lightening in East Asians, initially emerged in the southern coastal region of China during the early Neolithic (Liangdao2-7.5kya). The rapid dispersal of this variant during the past 4,000 years in East Asia is consistent with the proposed Darwinian positive selection on the adaptive allele (*OCA2-615Arg*), leading to skin lightening in East Asians to cope with the relatively low UV radiation in high-latitude areas. Interestingly, the rapid explosion of *OCA2-615Arg* coincides with the known major population expansion in China during the Late Holocene epoch.[64–67] It should be noted that due to the limited aDNA data in East Asia, the inferred time of selection onset for *OCA2-HiS615Arg* is a rough estimation. With more aDNA data available in the future, we expect more accurate time estimation and high-resolution spatial-temporal tracking of adaptive genetic variants in East Asia.

Consistent with the dating of MZR (14.0 kya), following the end of the LGM (26.5–19.0 kya)[85] and the earliest securely dated sites in Beringia (15–14 kya),[86] we demonstrate that MZR has a higher affinity to First Americans than to Tianyuan (40 kya) and all the pre-LGM Late Pleistocene Siberians. MZR, Amur-19K, and UKY are cladal with respect to First Americans, while Amur-14kya and UKY exhibit a higher affinity to the Americans compared to MZR. Thus, MZR is linked deeply and indirectly to the ancestry that contributed to First Americans. We speculate that during the Late Pleistocene, there was an express northward expansion of AMHs starting in southern East Asia through the coastal line of China, possibly by way of the Japanese Islands, and eventually crossing the Bering Strait and reaching the Americas. However, the scenario that MZR shows a higher affinity to Americans compared to Jomon likely reflects that the 2.6 kya Jomon population does not represent the early post-LGM humans who settled in the Japanese archipelago. The proposed migratory route along the east coast of East Asia by way of the Japanese Islands is supported by a recent finding of a Paleolithic site (~16 kya) at Cooper's Ferry of western Idaho, USA, where they found the use of unfluted stemmed projectile point technologies before the appearance of the Clovis Paleoindian tradition.[87] Notably, it exhibits early cultural connections with Paleolithic nonfluted projectile point traditions in Japan.[87] The bifacial point and backed blade technologies (~22–16 kya) in Honshu, Japan, lend technological correlates to the shared ecological and geographical factors with the Americas.[88]

This scenario is also supported by the current distribution pattern of the ancient Y-chromosome lineage Hg C in coastal East Asia, Siberia, and North America.[59] Approximately 40 kya, stemming from southern East Asia, the Hg C carriers started a northward expansion along the coastal regions of mainland China, the Korean Peninsula, and the Japanese archipelago, reaching Siberia ~15 kya, and finally made their way to the Americas.[59] In addition, unlike all other East Asian populations, the indigenous Ainu people in northern Japan and Sakhalin Island, Russia, show a closer genetic affinity with northeastern Siberians than with central Siberians.[89] Hence, the Japanese Islands may serve as the midway station along the proposed migratory route, and aDNA data of Late Pleistocene human remains from Japan will be highly informative in testing the proposed coastal route. Finally, the spatial-temporal distribution of the East Asian-specific *EDAR-V370A* variant, as well as its early presence in the LosRieles-12.0kya sample from coastal Chile of South America, supported a clear contribution of the Late Pleistocene East Asians to the first Americans.

In summary, we generated ancient genome sequences of MZR, a Late Pleistocene female who lived in Southwest China, one of the global biodiversity hotspots and the ice age refuge region. The aDNA data confirm that she possesses diverse genetic components and represents an early diversified population, suggesting the scenario of more diverse AMH lineages than previously thought during the Late Pleistocene in southern East Asia.[58,72,73] Our study paves the way to explore genetic explanations of morphological complexities of early hominins. MZR also shows a deep and indirect link to the ancestry that contributed to First Americans, which may help reconstruct the earliest migratory route from East Asia to the Americas.

It should be noted that the presented sequencing data of MZR have a relatively low coverage of the genome, leading to limited numbers of shared informative SNPs for cross-population analyses, which may cause bias, e.g., in the *f3* and *D-stat* statistics. Thus, more aDNA studies of Late Pleistocene hominins from East Asia are needed to delineate a clear picture of the earliest peopling of AMHs in this region.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Description of hominin remains (MZRs) from Red Deer Cave in Yunnan Province of Southwest China
- METHOD DETAILS
  - Ancient DNA extraction, library construction, small-scale sequencing and aDNA authentication evaluation
  - Nuclear DNA and mitochondrial DNA in solution capture and large-scale sequencing
  - Data processing (merging, trimming, mapping and filtering)
  - Sex determination and modern DNA contamination estimation
  - Genotyping
  - Cluster analysis
  - Genetic affinity and archaic admixture test of MZR
  - Spatial-temporal tracing of East Asian-specific variants attributed to local adaptation

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2022.06.016.

ACKNOWLEDGMENTS

We are grateful to Prof. Shuhua Xu from CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences; Prof. Hua Chen and Dr. Qi Liu from CAS Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences; and Prof. Chuanchao Wang from Department of Anthropology and Ethnology, Institute of Anthropology, School of Sociology and Anthropology, Xiamen University for data sharing and technical support. This study was supported by the National Natural Science Foundation of China 31871258 to X.Z., Strategic Priority Research Program (XDB26000000) of Chinese Academy of Sciences, and the Open Project (GREKF21-12) of the State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, CAS to X.J. It was also supported by the Yunnan provincial "Ten Thousand Talents Plan-Youth Top Talent" project (YNWR-QNBJ-2019-160) to X.Z., Yunnan provincial "Thousand Talents Plan-Young Talents" project (YNQR-QNRC-2018-091) to C.L., and the Youth Innovation Promotion Association of CAS (2020384) to Y. He.

AUTHOR CONTRIBUTIONS

B.S., X.J., and X.Z. designed the study; X.J., Y.W., and S.M. collected the sample; X.Z., C.L., J.H., T.Y., and Y.Z. conducted the experiment; X.Z., Y. He, C.L., and Y.P. analyzed the data; and X.Z., Y. He, X.J., Y. Huang, and B.S. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 24, 2021
Revised: May 27, 2022
Accepted: June 8, 2022
Published: July 14, 2022

REFERENCES

1. Jin, L., and Su, B. (2000). Natives or immigrants: modern human origin in east Asia. Nat. Rev. Genet. *1*, 126–133. https://doi.org/10.1038/35038565.
2. Macaulay, V., Hill, C., Achilli, A., Rengo, C., Clarke, D., Meehan, W., Blackburn, J., Semino, O., Scozzari, R., Cruciani, F., et al. (2005). Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. Science *308*, 1034–1036. https://doi.org/10.1126/science.1109792.
3. Sophady, H., Forestier, H., Zeitoun, V., Puaud, S., Frère, S., Celiberti, V., Westaway, K., Mourer, R., Mourer-Chauviré, C., Than, H., et al. (2016). Laang Spean cave (Battambang province): A tale of occupation in Cambodia from the Late Upper Pleistocene to Holocene. Quatern Int. *416*, 162–176. https://doi.org/10.1016/j.quaint.2015.07.049.
4. Westaway, K.E., Louys, J., Awe, R.D., Morwood, M.J., Price, G.J., Zhao, J.X., Aubert, M., Joannes-Boyau, R., Smith, T.M., Skinner, M.M., et al. (2017). An early modern human presence in Sumatra 73, 000-63, 000 years ago. Nature *548*, 322–325. https://doi.org/10.1038/nature23452.
5. Zhang, X., Qi, X., Yang, Z., Serey, B., Sovannary, T., Bunnath, L., Seang Aun, H., Samnom, H., Zhang, H., Lin, Q., et al. (2013). Analysis of mitochondrial genome diversity identifies new and ancient maternal lineages in Cambodian aborigines. Nat. Commun. *4*, 2599. https://doi.org/10.1038/ncomms3599.
6. HUGO Pan-Asian SNP Consortium, Abdulla, M.A., Ahmed, I., Assawamakin, A., Bhak, J., Brahmachari, S.K., Calacal, G.C., Chaurasia, A., Chen, C.H., Chen, Y.T., Chen, J., et al. (2009). Mapping human genetic diversity in Asia. Science *326*, 1541–1545. https://doi.org/10.1126/science.1177074.
7. Su, B., Xiao, J., Underhill, P., Deka, R., Zhang, W., Akey, J., Huang, W., Shen, D., Lu, D., Luo, J., et al. (1999). Y-chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age. Am. J. Hum. Genet. *65*, 1718–1724. https://doi.org/10.1086/302680.
8. Shi, H., Dong, Y.L., Wen, B., Xiao, C.J., Underhill, P.A., Shen, P.D., Chakraborty, R., Jin, L., and Su, B. (2005). Y-chromosome evidence of southern origin of the East Asian-specific haplogroup O3-M122. Am. J. Hum. Genet. *77*, 408–419. https://doi.org/10.1086/444436.
9. Jinam, T.A., Phipps, M.E., Aghakhanian, F., Majumder, P.P., Datar, F., Stoneking, M., Sawai, H., Nishida, N., Tokunaga, K., Kawamura, S.,

et al. (2017). Discerning the origins of the Negritos, first Sundaland people: deep divergence and archaic admixture. Genome Biol. Evol. *9*, 2013–2022. https://doi.org/10.1093/gbe/evx118.

10. Noerwidi, S. (2017). Using dental metrical analysis to determine the terminal Pleistocene and Holocene population history of Java. In New Perspectives in Southeast Asian and Pacific Prehistory, P.J. Piper, H. Matsumura, and D. Bulbeck, eds. (ANU Press), p. 92.
11. Wan, Z.W., Ma, Z.K., Yang, X.Y., Zhang, C., Zhou, G.M., Pang, C.S., and Ge, Q.S. (2012). Starch residues from shell tools from sites of Xianrendong and Diaotonghuan and its implications for paleoclimate. Quaternary Res. *32*, 256–263.
12. Curnoe, D., Ji, X., Taçon, P.S.C., and Yaozheng, G. (2015). Possible signatures of hominin hybridization from the Early Holocene of southwest China. Sci. Rep. *5*, 12408, https://doi.org/10.1038/srep12408.
13. Xia, J.D., Wang, W.D., Wu, R., and Li, J.Z. (2005). TL dating of pottery shards excavated in Xianren Dong and Zengpiyan sites. In International Symposium on Ancient Ceramic Science and Technology, pp. 472–477.
14. Jiang, J.Y. (2006). Study of Bailian Dong, Miaoyan and Xianren Dong archaeological site: case analysis of transitional period from Palaeolithic to Neolithic epoch. Prehistoric Res. 58–67.
15. Cao, B., He, L.T., and Zhang, H. (2015). A study of human fossils discovered at the Maomao Cave site, Xingyi County, Guizhou Province. Acta Anthropol. Sin. *34*, 451–460.
16. Mao, Y.Q., and Cao, Z.T. (2012). A preliminary study of the polished bone tools unearthed in 1979 from the Chuandong site in Puding County, Guizhou. Acta Anthropol. Sin. *31*, 335–343.
17. Zhang, X.Y., Zheng, L., Yang, L.C., and Bao, Z.D. (1990). Mengzi Ren hominin fossil and its culture from Yunnan. Yunnan Social Sciences, 64–68.
18. Ji, X.P., Kuman, K., Clarke, R.J., Forestier, H., Li, Y.H., Ma, J., Qiu, K.W., Li, H., and Wu, Y. (2016). The oldest Hoabinhian technocomplex in Asia (43.5 ka) at Xiaodong rockshelter, Yunnan Province, southwest China. Quatern Int. *400*, 166–174. https://doi.org/10.1016/j.quaint.2015.09.080.
19. Wang, T., Wang, W., Xie, G., Li, Z., Fan, X., Yang, Q., Wu, X., Cao, P., Liu, Y., Yang, R., et al. (2021). Human population history at the crossroads of East and Southeast Asia since 11,000 years ago. Cell *184*, 3829–3841.e21. https://doi.org/10.1016/j.cell.2021.05.018.
20. Curnoe, D., Xueping, J., Herries, A.I.R., Kanning, B., Taçon, P.S.C., Zhende, B., Fink, D., Yunsheng, Z., Hellstrom, J., Yun, L., et al. (2012). Human remains from the Pleistocene-Holocene transition of southwest China suggest a complex evolutionary history for East Asians. PLoS One *7*, e31918, https://doi.org/10.1371/journal.pone.0031918.
21. Curnoe, D., Ji, X., Liu, W., Bao, Z., Taçon, P.S.C., and Ren, L. (2015). A hominin femur with archaic affinities from the Late Pleistocene of Southwest China. PLoS One *10*, e0143332, https://doi.org/10.1371/journal.pone.0143332.
22. Ji, X.P., Curnoe, D., Bao, Z.D., Herries, A.I.R., Fink, D., Zhu, Y.S., Hellstrom, J., Luo, Y., and Tacon, P.S.C. (2013). Further geological and palaeoanthropological investigations at the Maludong hominin site, Yunnan Province, Southwest China. Chinese Sci. Bull. *58*, 4472–4485. https://doi.org/10.1007/s11434-013-6026-5.
23. Li, C., Hongjin, D., and Peng, H. (2013). Diversity and distribution of higher plants in Yunnan, China. Biodiversity *21*, 359–363. https://doi.org/10.3724/sp.j.1003.2013.05162.
24. Linnemann, U., Su, T., Kunzmann, L., Spicer, R.A., Ding, W.N., Spicer, T., Zieger, J., Hofmann, M., Moraweck, K., Gärtner, A., et al. (2018). New U-Pb dates show a Paleogene origin for the modern Asian biodiversity hot spots. Geology *46*, 3–6. https://doi.org/10.1130/g39693.1.
25. Zhu, H., Cao, M., and Hu, H.B. (2006). Geological history, flora, and vegetation of Xishuangbanna, Southern Yunnan, China. Biotropica *38*, 310–317. https://doi.org/10.1111/j.1744-7429.2006.00147.x.
26. Xiwen, L., and Walker, D. (1986). The plant geography of Yunnan Province, Southwest China. J. Biogeogr. *13*, 367–397. https://doi.org/10.2307/2844964.
27. Li, R., Kraft, N.J.B., Yang, J., and Wang, Y.H. (2015). A phylogenetically informed delineation of floristic regions within a biodiversity hotspot in Yunnan, China. Sci. Rep. *5*, 9396. https://doi.org/10.1038/srep09396.
28. Sutikna, T., Tocheri, M.W., Morwood, M.J., Saptomo, E.W., Awe, R.D., Wasisto, S., Westaway, K.E., Aubert, M., Li, B., et al. (2016). Revised stratigraphy and chronology for Homo floresiensis at Liang Bua in Indonesia. Nature *532*, 366–369. https://doi.org/10.1038/nature17179.
29. Ji, X., Curnoe, D., Taçon, P.S.C., Zhende, B., Ren, L., Mendoza, R., Tong, H., Ge, J., Deng, C., Adler, L., et al. (2016). Cave use and palaeoecology at Maludong (Red Deer Cave), Yunnan, China. J. Archaeol. Sci. Rep. *8*, 277–283. https://doi.org/10.1016/j.jasrep.2016.06.025.
30. McColl, H., Racimo, F., Vinner, L., Demeter, F., Gakuhari, T., Moreno-Mayar, J.V., van Driem, G., Gram Wilken, U., Seguin-Orlando, A., de la Fuente Castro, C., et al. (2018). The prehistoric peopling of Southeast Asia. Science *361*, 88–92. https://doi.org/10.1126/science.aat3628.
31. Gansauge, M.T., and Meyer, M. (2014). Selective enrichment of damaged DNA molecules for ancient genome sequencing. Genome Res. *24*, 1543–1549. https://doi.org/10.1101/gr.174201.114.
32. Prüfer, K. (2018). snpAD: an ancient DNA genotype caller. Bioinformatics *34*, 4165–4171. https://doi.org/10.1093/bioinformatics/bty507.
33. Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. Genetics *192*, 1065–1093. https://doi.org/10.1534/genetics.112.145037.
34. Peng, M.S., Palanichamy, M.G., Yao, Y.G., Mitra, B., Cheng, Y.T., Zhao, M., Liu, J., Wang, H.W., Pan, H., Wang, W.Z., et al. (2011). Inland postglacial dispersal in East Asia revealed by mitochondrial haplogroup M9a'b. BMC Biol. *9*, 2. https://doi.org/10.1186/1741-7007-9-2.
35. Duggan, A.T., Evans, B., Friedlaender, F.R., Friedlaender, J.S., Koki, G., Merriwether, D.A., Kayser, M., and Stoneking, M. (2014). Maternal history of Oceania from complete mtDNA genomes: contrasting ancient diversity with recent homogenization due to the Austronesian expansion. Am. J. Hum. Genet. *94*, 721–733. https://doi.org/10.1016/j.ajhg.2014.03.014.
36. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. *19*, 1655–1664. https://doi.org/10.1101/gr.094052.109.
37. Yang, M.A., Gao, X., Theunert, C., Tong, H., Aximu-Petri, A., Nickel, B., Slatkin, M., Meyer, M., Pääbo, S., Kelso, J., and Fu, Q. (2017). 40,000-year-old individual from Asia provides insight into early population structure in Eurasia. Curr. Biol. *27*, 3202–3208.e9. https://doi.org/10.1016/j.cub.2017.09.030.
38. Pickrell, J.K., and Pritchard, J.K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. *8*, e1002967, https://doi.org/10.1371/journal.pgen.1002967.
39. Yang, M.A., Fan, X., Sun, B., Chen, C., Lang, J., Ko, Y.C., Tsang, C.H., Chiu, H., Wang, T., Bao, Q., et al. (2020). Ancient DNA indicates human population shifts and admixture in northern and southern China. Science *369*, 282–288. https://doi.org/10.1126/science.aba0909.
40. Yu, H., Spyrou, M.A., Karapetian, M., Shnaider, S., Radzevičiūtė, R., Nägele, K., Neumann, G.U., Penske, S., Zech, J., Lucas, M., et al. (2020). Paleolithic to Bronze Age Siberians reveal connections with First Americans and across Eurasia. Cell *181*, 1232–1245.e20. https://doi.org/10.1016/j.cell.2020.04.037.
41. Marciniak, S., and Perry, G.H. (2017). Harnessing ancient genomes to study the history of human adaptation. Nat. Rev. Genet. *18*, 659–674. https://doi.org/10.1038/nrg.2017.65.
42. Lipson, M., Cheronet, O., Mallick, S., Rohland, N., Oxenham, M., Pietrusewsky, M., Pryce, T.O., Willis, A., Matsumura, H., Buckley, H., et al. (2018). Ancient genomes document multiple waves of migration in Southeast Asian prehistory. Science *361*, 92–95. https://doi.org/10.1126/science.aat3188.
43. Ning, C., Wang, C.C., Gao, S., Yang, Y., Zhang, X., Wu, X., Zhang, F., Nie, Z., Tang, Y., Robbeets, M., et al. (2019). Ancient genomes reveal

Yamnaya-related ancestry and a potential source of Indo-European speakers in Iron Age Tianshan. Curr. Biol. *29*, 2526–2532.e4. https://doi.org/10.1016/j.cub.2019.06.044.

44. Ning, C., Li, T., Wang, K., Zhang, F., Li, T., Wu, X., Gao, S., Zhang, Q., Zhang, H., Hudson, M.J., et al. (2020). Ancient genomes from northern China suggest links between subsistence changes and human migration. Nat. Commun. *11*, 2700. https://doi.org/10.1038/s41467-020-16557-2.
45. Ding, M., Wang, T., Ko, A.M.S., Chen, H., Wang, H., Dong, G., Lu, H., He, W., Wangdue, S., Yuan, H., et al. (2020). Ancient mitogenomes show plateau populations from last 5200 years partially contributed to present-day Tibetans. Proc. Biol. Sci. *287*, 20192968, https://doi.org/10.1098/rspb.2019.2968.
46. Zhang, X., Li, C., Zhou, Y., Huang, J., Yu, T., Liu, X., Shi, H., Liu, H., Chia, S., Huang, S., et al. (2020). A matrilineal genetic perspective of hanging coffin custom in Southern China and Northern Thailand. iScience *23*, 101032, https://doi.org/10.1016/j.isci.2020.101032.
47. Bai, F., Zhang, X., Ji, X., Cao, P., Feng, X., Yang, R., Peng, M., Pei, S., and Fu, Q. (2020). Paleolithic genetic link between Southern China and Mainland Southeast Asia revealed by ancient mitochondrial genomes. J. Hum. Genet. *65*, 1125–1128. https://doi.org/10.1038/s10038-020-0796-9.
48. Wang, C.C., Yeh, H.Y., Popov, A.N., Zhang, H.Q., Matsumura, H., Sirak, K., Cheronet, O., Kovalev, A., Rohland, N., Kim, A.M., et al. (2021). Genomic insights into the formation of human populations in East Asia. Nature *591*, 413–419. https://doi.org/10.1038/s41586-021-03336-2.
49. Mao, X., Zhang, H., Qiao, S., Liu, Y., Chang, F., Xie, P., Zhang, M., Wang, T., Li, M., Cao, P., et al. (2021). The deep population history of northern East Asia from the Late Pleistocene to the Holocene. Cell *184*, 3256–3266.e13. https://doi.org/10.1016/j.cell.2021.04.040.
50. Carlhoff, S., Duli, A., Nägele, K., Nur, M., Skov, L., Sumantri, I., Oktaviana, A.A., Hakim, B., Burhan, B., Syahdar, F.A., et al. (2021). Genome of a middle Holocene hunter-gatherer from Wallacea. Nature *596*, 543–547. https://doi.org/10.1038/s41586-021-03823-6.
51. Zhang, F., Ning, C., Scott, A., Fu, Q., Bjørn, R., Li, W., Wei, D., Wang, W., Fan, L., Abuduresule, I., et al. (2021). The genomic origins of the Bronze Age Tarim Basin mummies. Nature *599*, 256–261. https://doi.org/10.1038/s41586-021-04052-7.
52. Robbeets, M., Bouckaert, R., Conte, M., Savelyev, A., Li, T., An, D.I., Shinoda, K.I., Cui, Y., Kawashima, T., Kim, G., et al. (2021). Triangulation supports agricultural spread of the Transeurasian languages. Nature *599*, 616–621. https://doi.org/10.1038/s41586-021-04108-8.
53. Cooke, N.P., Mattiangeli, V., Cassidy, L.M., Okazaki, K., Stokes, C.A., Onbe, S., Hatakeyama, S., Machida, K., Kasai, K., Tomioka, N., et al. (2021). Ancient genomics reveals tripartite origins of Japanese populations. Sci. Adv. *7*, eabh2419, https://doi.org/10.1126/sciadv.abh2419.
54. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. *19*, 1655–1664. https://doi.org/10.1101/gr.094052.109.
55. Li, F., Vanwezer, N., Boivin, N., Gao, X., Ott, F., Petraglia, M., and Roberts, P. (2019). Heading north: Late Pleistocene environments and human dispersals in central and eastern Asia. PLoS One *14*, e0216433, https://doi.org/10.1371/journal.pone.0216433.
56. Zhong, H., Shi, H., Qi, X.B., Duan, Z.Y., Tan, P.P., Jin, L., Su, B., and Ma, R.Z. (2011). Extended Y chromosome investigation suggests postglacial migrations of modern humans into East Asia via the northern route. Mol. Biol. Evol. *28*, 717–727. https://doi.org/10.1093/molbev/msq247.
57. Madsen, D.B., Jingzen, L., Brantingham, P.J., Xing, G., Elston, R.G., and Bettinger, R.L.J.A. (2001). Dating Shuidonggou and the Upper Palaeolithic blade industry in North China. Antiquity *75*, 706–716. https://doi.org/10.1017/s0003598x00089213.
58. Matsumura, H., Hung, H.C., Higham, C., Zhang, C., Yamagata, M., Nguyen, L.C., Li, Z., Fan, X.C., Simanjuntak, T., Oktaviana, A.A., et al. (2019). Craniometrics reveal "two layers" of prehistoric human dispersal in Eastern Eurasia. Sci. Rep. *9*, 1451. https://doi.org/10.1038/s41598-018-35426-z.
59. Zhong, H., Shi, H., Qi, X.B., Xiao, C.J., Jin, L., Ma, R.Z., and Su, B. (2010). Global distribution of Y-chromosome haplogroup C reveals the prehistoric migration routes of African exodus and early settlement in East Asia. J. Hum. Genet. *55*, 428–435. https://doi.org/10.1038/jhg.2010.40.
60. Yang, Z., Zhong, H., Chen, J., Zhang, X., Zhang, H., Luo, X., Xu, S., Chen, H., Lu, D., Han, Y., et al. (2016). A genetic mechanism for convergent skin lightening during recent human evolution. Mol. Biol. Evol. *33*, 1177–1187. https://doi.org/10.1093/molbev/msw003.
61. Chaplin, G. (2004). Geographic distribution of environmental factors influencing human skin coloration. Am. J. Phys. Anthropol. *125*, 292–302. https://doi.org/10.1002/ajpa.10263.
62. Edwards, M., Bigham, A., Tan, J., Li, S., Gozdzik, A., Ross, K., Jin, L., and Parra, E.J. (2010). Association of the OCA2 polymorphism His615Arg with melanin content in east Asian populations: further evidence of convergent evolution of skin pigmentation. PLoS Genet. *6*, e1000867, https://doi.org/10.1371/journal.pgen.1000867.
63. Norton, H.L., Kittles, R.A., Parra, E., McKeigue, P., Mao, X., Cheng, K., Canfield, V.A., Bradley, D.G., McEvoy, B., and Shriver, M.D. (2007). Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. Mol. Biol. Evol. *24*, 710–722. https://doi.org/10.1093/molbev/msl203.
64. Ren, X., Xu, J., Wang, H., Storozum, M., Lu, P., Mo, D., Li, T., Xiong, J., and Kidder, T.R. (2021). Holocene fluctuations in vegetation and human population demonstrate social resilience in the prehistory of the Central Plains of China. Environ. Res. Lett. *16*, 055030. https://doi.org/10.1088/1748-9326/abdf0a.
65. Wang, C., Lu, H., Zhang, J., Gu, Z., and He, K. (2014). Prehistoric demographic fluctuations in China inferred from radiocarbon data and their linkage with climate change over the past 50,000 years. Quaternary Sci. Rev. *98*, 45–59. https://doi.org/10.1016/j.quascirev.2014.05.015.
66. Hosner, D., Wagner, M., Tarasov, P.E., Chen, X., and Leipe, C. (2016). Spatiotemporal distribution patterns of archaeological sites in China during the Neolithic and Bronze Age: an overview. Holocene *26*, 1576–1593. https://doi.org/10.1177/0959683616641743.
67. Li, J., Yan, H., Dodson, J., Xu, Q., Sun, A., Cheng, B., Li, C., Ni, J., Zhang, X., and Lu, F. (2018). Regional-scale precipitation anomalies in Northern China during the Holocene and possible impact on prehistoric demographic changes. Geophys. Res. Lett. *45*, 10. https://doi.org/10.1029/2018gl078873.
68. Leipe, C., Long, T., Sergusheva, E.A., Wagner, M., and Tarasov, P.E. (2019). Discontinuous spread of millet agriculture in eastern Asia and prehistoric population dynamics. Sci. Adv. *5*, eaax6225, https://doi.org/10.1126/sciadv.aax6225.
69. Kamberov, Y.G., Wang, S., Tan, J., Gerbault, P., Wark, A., Tan, L., Yang, Y., Li, S., Tang, K., Chen, H., et al. (2013). Modeling recent human evolution in mice by expression of a selected EDAR variant. Cell *152*, 691–702. https://doi.org/10.1016/j.cell.2013.01.016.
70. Gravel, S., Zakharia, F., Moreno-Estrada, A., Byrnes, J.K., Muzzio, M., Rodriguez-Flores, J.L., Kenny, E.E., Gignoux, C.R., Maples, B.K., Guiblet, W., et al. (2013). Reconstructing Native American migrations from whole-genome and whole-exome data. PLoS Genet. *9*, e1004023, https://doi.org/10.1371/journal.pgen.1004023.
71. David, R. (2018). Who We Are and How We Got Here: Ancient DNA and the New Science of the Human Past (Vintage Books), pp. 174–176.
72. Liu, W., Jin, C.Z., Zhang, Y.Q., Cai, Y.J., Xing, S., Wu, X.J., Cheng, H., Edwards, R.L., Pan, W.S., Qin, D.G., et al. (2010). Human remains from Zhirendong, South China, and modern human emergence in East Asia. Proc. Natl. Acad. Sci. USA *107*, 19201–19206. https://doi.org/10.1073/pnas.1014386107.
73. Liu, W., Martinón-Torres, M., Cai, Y.J., Xing, S., Tong, H.W., Pei, S.W., Sier, M.J., Wu, X.H., Edwards, R.L., Cheng, H., et al. (2015). The earliest unequivocally modern humans in southern China. Nature *526*, 696–699. https://doi.org/10.1038/nature15696.

74. Sun, X.F., Wen, S.Q., Lu, C.Q., Zhou, B.Y., Curnoe, D., Lu, H.Y., Li, H.C., Wang, W., Cheng, H., Yi, S.W., et al. (2021). Ancient DNA and multimethod dating confirm the late arrival of anatomically modern humans in southern China. Proc. Natl. Acad. Sci. USA *118*. e2019158118. https://doi.org/10.1073/pnas.2019158118.

75. Détroit, F., Mijares, A.S., Corny, J., Daver, G., Zanolli, C., Dizon, E., Robles, E., Grün, R., and Piper, P.J. (2019). A new species of Homo from the Late Pleistocene of the Philippines. Nature *568*, 181–186. https://doi.org/10.1038/s41586-019-1067-9.

76. Fu, Q., Li, H., Moorjani, P., Jay, F., Slepchenko, S.M., Bondarev, A.A., Johnson, P.L.F., Aximu-Petri, A., Prüfer, K., de Filippo, C., et al. (2014). Genome sequence of a 45,000-year-old modern human from western Siberia. Nature *514*, 445–449. https://doi.org/10.1038/nature13810.

77. Devièse, T., Massilani, D., Yi, S., Comeskey, D., Nagel, S., Nickel, B., Ribechini, E., Lee, J., Tseveendorj, D., Gunchinsuren, B., et al. (2019). Compound-specific radiocarbon dating and mitochondrial DNA analysis of the Pleistocene hominin from Salkhit Mongolia. Nat. Commun. *10*, 274. https://doi.org/10.1038/s41467-018-08018-8.

78. Massilani, D., Skov, L., Hajdinjak, M., Gunchinsuren, B., Tseveendorj, D., Yi, S., Lee, J., Nagel, S., Nickel, B., Devièse, T., et al. (2020). Denisovan ancestry and population history of early East Asians. Science *370*, 579–583. https://doi.org/10.1126/science.abc1166.

79. Fu, Q., Meyer, M., Gao, X., Stenzel, U., Burbano, H.A., Kelso, J., and Pääbo, S. (2013). DNA analysis of an early modern human from Tianyuan Cave, China. Proc. Natl. Acad. Sci. USA *110*, 2223–2227. https://doi.org/10.1073/pnas.1221359110.

80. Sikora, M., Pitulko, V.V., Sousa, V.C., Allentoft, M.E., Vinner, L., Rasmussen, S., Margaryan, A., de Barros Damgaard, P., de la Fuente, C., Renaud, G., et al. (2019). The population history of northeastern Siberia since the Pleistocene. Nature *570*, 182–188. https://doi.org/10.1038/s41586-019-1279-z.

81. Raghavan, M., Skoglund, P., Graf, K.E., Metspalu, M., Albrechtsen, A., Moltke, I., Rasmussen, S., Stafford, T.W., Orlando, L., Metspalu, E., et al. (2014). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature *505*, 87–91. https://doi.org/10.1038/nature12736.

82. Curnoe, D., Zhao, J.X., Aubert, M., Fan, M., Wu, Y., Baker, A., Mei, G.H., Sun, X.F., Mendoza, R., Adler, L., et al. (2019). Implications of multimodal age distributions in Pleistocene cave deposits: a case study of Maludong palaeoathropological locality, southern China. J. Archaeol. Sci. Rep. *25*, 388–399. https://doi.org/10.1016/j.jasrep.2019.04.020.

83. Li, Y.C., Tian, J.Y., Liu, F.W., Yang, B.Y., Gu, K.S.Y., Rahman, Z.U., Yang, L.Q., Chen, F.H., Dong, G.H., and Kong, Q.P. (2019). Neolithic millet farmers contributed to the permanent settlement of the Tibetan Plateau by adopting barley agriculture. Natl. Sci. Rev. *6*, 1005–1013. https://doi.org/10.1093/nsr/nwz080.

84. Ko, A.S., Chen, C.Y., Fu, Q., Delfin, F., Li, M., Chiu, H.L., Stoneking, M., and Ko, Y.C. (2014). Early Austronesians: into and out of Taiwan. Am. J. Hum. Genet. *94*, 426–436. https://doi.org/10.1016/j.ajhg.2014.02.003.

85. Clark, P.U., Dyke, A.S., Shakun, J.D., Carlson, A.E., Clark, J., Wohlfarth, B., Mitrovica, J.X., Hostetler, S.W., and McCabe, A.M. (2009). The Last Glacial Maximum. Science *325*, 710–714. https://doi.org/10.1126/science.1172873.

86. Goebel, T. (2002). The "Microblade Adaptation" and Recolonization of Siberia during the Late Upper Pleistocene (Archeological Papers of the American Anthropological Association).

87. Davis, L.G., Madsen, D.B., Becerra-Valdivia, L., Higham, T., Sisson, D.A., Skinner, S.M., Stueber, D., Nyers, A.J., Keen-Zebert, A., Neudorf, C., et al. (2019). Late Upper Paleolithic occupation at Cooper's Ferry, Idaho, USA, ~16,000 years ago. Science *365*, 891–897. https://doi.org/10.1126/science.aax9830.

88. Uchiyama, J., Gillam, J.C., Savelyev, A., and Ning, C. (2020). Populations dynamics in Northern Eurasian forests: a long-term perspective from Northeast Asia. Evol. Hum. Sci. *2*, e16. https://doi.org/10.1017/ehs.2020.11.

89. Jeong, C., Nakagome, S., and Di Rienzo, A. (2016). Deep history of East Asian populations revealed through genetic analysis of the Ainu. Genetics *202*, 261–272. https://doi.org/10.1534/genetics.115.178673.

90. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

91. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

92. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L.F., and Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics *29*, 1682–1684. https://doi.org/10.1093/bioinformatics/btt193.

93. Renaud, G., Stenzel, U., and Kelso, J. (2014). leeHom: adaptor trimming and merging for Illumina sequencing reads. Nucleic Acids Res. *42*, e141. https://doi.org/10.1093/nar/gku699.

94. Peyrégne, S., and Peter, B.M. (2020). AuthentiCT: a model of ancient DNA damage to estimate the proportion of present-day DNA contamination. Genome Biol. *21*, 246. https://doi.org/10.1186/s13059-020-02123-y.

95. Renaud, G., Slon, V., Duggan, A.T., and Kelso, J. (2015). Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. Genome Biol. *16*, 224. https://doi.org/10.1186/s13059-015-0776-0.

96. Kaifu, Y., Aziz, F., and Baba, H. (2005). Hominid mandibular remains from Sangiran: 1952-1986 collection. Am. J. Phys. Anthropol. *128*, 497–519. https://doi.org/10.1002/ajpa.10427.

97. Shen, G., Gao, X., Gao, B., and Granger, D.E. (2009). Age of Zhoukoudian Homo erectus determined with (26)Al/(10)Be burial dating. Nature *458*, 198–200. https://doi.org/10.1038/nature07741.

98. Bürklein, S., Breuer, D., and Schäfer, E. (2011). Prevalence of taurodont and pyramidal molars in a German population. J. Endod. *37*, 158–162. https://doi.org/10.1016/j.joen.2010.10.010.

99. Kupczik, K., and Hublin, J.J. (2010). Mandibular molar root morphology in Neanderthals and Late Pleistocene and recent Homo sapiens. J. Hum. Evol. *59*, 525–541. https://doi.org/10.1016/j.jhevol.2010.05.009.

100. Allentoft, M.E., Sikora, M., Sjögren, K.G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P.B., Schroeder, H., Ahlström, T., Vinner, L., et al. (2015). Population genomics of Bronze Age Eurasia. Nature *522*, 167–172. https://doi.org/10.1038/nature14507.

101. Korlević, P., Gerber, T., Gansauge, M.T., Hajdinjak, M., Nagel, S., Aximu-Petri, A., and Meyer, M. (2015). Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. Biotechniques *59*, 87–93. https://doi.org/10.2144/000114320.

102. Lindo, J., Huerta-Sánchez, E., Nakagome, S., Rasmussen, M., Petzelt, B., Mitchell, J., Cybulski, J.S., Willerslev, E., DeGiorgio, M., and Malhi, R.S. (2016). A time transect of exomes from a Native American population before and after European contact. Nat. Commun. *7*, 13175, https://doi.org/10.1038/ncomms13175.

103. Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., Furtwängler, A., Haak, W., Meyer, M., Mittnik, A., et al. (2016). The genetic history of Ice Age Europe. Nature *534*, 200–205. https://doi.org/10.1038/nature17993.

104. Meyer, M., Arsuaga, J.L., de Filippo, C., Nagel, S., Aximu-Petri, A., Nickel, B., Martínez, I., Gracia, A., de Castro, J.M.B., Carbonell, E., et al. (2016). Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. Nature *531*, 504–507. https://doi.org/10.1038/nature17405.

105. Genomes Project, C., Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., et al. (2015). A global reference for human genetic variation. Nature *526*, 68–74. https://doi.org/10.1038/nature15393.

106. Pagani, L., Lawson, D.J., Jagoda, E., Mörseburg, A., Eriksson, A., Mitt, M., Clemente, F., Hudjashov, G., DeGiorgio, M., Saag, L., et al. (2016). Genomic analyses inform on migration events during the peopling of Eurasia. Nature *538*, 238–242. https://doi.org/10.1038/nature19792.

107. Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., Tandon, A., et al. (2016). The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature *538*, 201–206. https://doi.org/10.1038/nature18964.

108. Wong, L.P., Lai, J.K.H., Saw, W.Y., Ong, R.T.H., Cheng, A.Y., Pillai, N.E., Liu, X.Y., Xu, W.T., Chen, P., Foo, J.N., et al. (2014). Insights into the genetic structure and diversity of 38 South Asian Indians from deep whole-genome sequencing. PLoS Genet. *10*, e1004377. https://doi.org/10.1371/journal.pgen.1004377.

109. Wong, L.P., Ong, R.H., Poh, W.T., Liu, X., Chen, P., Li, R., Lam, K.Y., Pillai, N.E., Sim, K.S., Xu, H., et al. (2013). Deep whole-genome sequencing of 100 Southeast Asian Malays. Am. J. Hum. Genet. *92*, 52–66. https://doi.org/10.1016/j.ajhg.2012.12.005.

110. Lu, D.S., Lou, H.Y., Yuan, K., Wang, X.J., Wang, Y.C., Zhang, C., Lu, Y., Yang, X., Deng, L.A., Zhou, Y., et al. (2016). Ancestral origins and genetic history of Tibetan highlanders. Am. J. Hum. Genet. *99*, 580–594. https://doi.org/10.1016/j.ajhg.2016.07.002.

111. Mondal, M., Casals, F., Xu, T., Dall'Olio, G.M., Pybus, M., Netea, M.G., Comas, D., Laayouni, H., Li, Q.B., Majumder, P.P., et al. (2016). Genomic analysis of Andamanese provides insights into ancient human migration into Asia and adaptation. Nat. Genet. *48*, 1066–1070. https://doi.org/10.1038/ng.3621.

112. Pickrell, J.K., and Pritchard, J.K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. *8*, e1002967. https://doi.org/10.1371/journal.pgen.1002967.

113. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575. https://doi.org/10.1086/519795.

114. Adhikari, K., Fuentes-Guajardo, M., Quinto-Sánchez, M., Mendoza-Revilla, J., Camilo Chacón-Duque, J., Acuña-Alonzo, V., Jaramillo, C., Arias, W., Lozano, R.B., Pérez, G.M., et al. (2016). A genome-wide association scan implicates DCHS2, RUNX2, GLI3, PAX1 and EDAR in human facial variation. Nat. Commun. *7*, 11616, https://doi.org/10.1038/ncomms11616.

115. Tan, J., Yang, Y., Tang, K., Sabeti, P.C., Jin, L., and Wang, S. (2013). The adaptive variant EDARV370A is associated with straight hair in East Asians. Hum. Genet. *132*, 1187–1191. https://doi.org/10.1007/s00439-013-1324-1.

116. Donnelly, M.P., Paschou, P., Grigorenko, E., Gurwitz, D., Barta, C., Lu, R.B., Zhukova, O.V., Kim, J.J., Siniscalco, M., New, M., et al. (2012). A global view of the OCA2-HERC2 region and pigmentation. Hum. Genet. *131*, 683–696. https://doi.org/10.1007/s00439-011-1110-x.

Please cite this article in press as: Zhang et al., A Late Pleistocene human genome from Southwest China, Current Biology (2022), https://doi.org/10.1016/j.cub.2022.06.016

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Mengzi Ren cranial fossil from Malu Dong, Yunnan, China	This study	MZR
Chemicals, peptides, and recombinant proteins		
Proteinase K	Sigma Aldrich	Cat#P6556
HPLC Water	Sigma Aldrich	Cat#E5134
Tris	Sigma Aldrich	Cat#T1503
Silicon Dioxide	Sigma Aldrich	Cat#S5631
Magnetic Particle Stand	Life Technologies	Cat#123-21D
MinElute PCR Purification Kit	Qiagen	Cat#28006
Dynabeads MyOne streptavidin C1	Life Technologies	Cat#65001
Agencourt AMPure XP-PCR purification kit	Agilent Technologies	Cat#A63880
USER enzyme	New England Biolabs	Cat#M5505L
NEBNext Ultra II DNA Library Prep Kit for Illumina	New England Biolabs	Cat#E7645
KAPA HiFi Hot start Polymerase	KAPA BIOSYSTEMS	Cat#KK2602
High Sensitivity DNA Chips (Bioanalyzer 2100)	Agilent Technologies	Cat#5067-4626
MYbaits WGE	Arbor Biosciences	Cat#302496
MYbaits Mito	Arbor Biosciences	Cat#303096
Deposited data		
Analyzed data for MZR (Genome Sequence Archive)	This paper	GSA: PRJCA004331
Software and algorithms		
BWA	Li and Durbin[90]	http://bio-bwa.sourceforge.net/
Samtools	Li et al.,[91]	http://www.htslib.org
mapDamage2.07	Jónsson et al.,[92]	https://geogenetics.ku.dk/publications/mapdamage/
Haplogrep2.0	N/A	https://haplogrep.uibk.ac.at
leeHom	Renaud et al.,[93]	https://github.com/grenaud/leeHom
AuthentiCT	Peyrégne and Peter[94]	https://github.com/StephanePeyregne/AuthentiCT
ContamMix	Fu et al.,[79]	https://github.com/DReichLab/adna-workflow
Schmutzi	Renaud et al.,[95]	https://bioinf.eva.mpg.de/schmutzi/
snpAD	Prüfer et al.,[32]	http://bioinf.eva.mpg.de/snpAD/
EIGENSOFT	Alexander et al.,[36]	https://github.com/argriffing/eigensoft
TreeMix	Pickrell and Pritchard[38]	https://github.com/Zolmeister/treemix
ADMIXTURE	Alexander et al.,[36]	http://dalexander.github.io/admixture/download.html
ADMIXTOOLS	Patterson et al.,[33]	https://github.com/DReichLab/AdmixTools

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact: Bing Su (sub@mail.kiz.ac.cn)

Materials availability

This study did not generate new reagents.

Please cite this article in press as: Zhang et al., A Late Pleistocene human genome from Southwest China, Current Biology (2022), https://doi.org/10.1016/j.cub.2022.06.016

Data and code availability

The raw files (both genomic and mitochondrial DNA) and genotype data reported in this study have been deposited in the Genome Sequence Archive in National Genomics Data Center, Beijing Institute of Genomics (China National Center for Bioinformation), Chinese Academy of Sciences and are publicly accessible at https://bigd.big.ac.cn/gsa-human/browse/HRA002402. All data is available in the main text or the supplementary materials to reproduce or extend the analysis. The MZR cranium calotte is housed at the Mengzi Institute of Cultural Relics, Mengzi city 661100, Yunnan province, China, and can be accessed under request to X.J. and S.M.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Description of hominin remains (MZRs) from Red Deer Cave in Yunnan Province of Southwest China

Malu Dong (Red Deer Cave, 103°24′ E, 23°20′ N) is located at the hillside ~7 km southwest of Mengzi city, Yunnan Province, China, bordering the northern edge of tropical Southeast Asia (Figure 1A). The original excavation of Malu Dong was carried out in 1989 and ~30 pieces of human remains were found at that time, including a nearly complete cranium calotte (MLDG-1704), a proximal femur (MLDG-1678), a burnt cranial vault bone (MLDG-1705), hemi-mandible lacking dentition (MLDG-1679), a proximal ulna (MLDG-1710), a maxillary fragment (MLDG-1713), a possible femoral head fragment (MLDG-1717) and several small cranial vault fragments.[17] Because these human remains include two cranial vaults (MLDG-1704 and MLDG-1705),[22] suggesting at least two different individuals. It is uncertain if the samples (MLDG-1704 and MLDG-1678) showing mosaic features of AMH and archaic humans are from the same individual. In 2008, an international archaeological team (including Prof. X.P. J in this study) operated a new round of excavation (scale of 50×50×370 cm) at Malu Dong, and additional human remains (including an isolated partial I2, MLDG-1751) as well as a large number of samples for dating and archaeomagnetic analysis were collected from the identified 11 distinct stratigraphic aggregates with maximum depth of ~370 cm.[20] Calibrated radiocarbon ages from charcoal samples showed that the entire sequence spans the interval of 18,070-17,590 cal. yBP (95% interval) to 13,415-13,165 cal. yBP (95% interval), and all the 30 hominin remains were recovered from a series of deposits dated from 14,650-13,970 cal. yBP (95% interval) to 13,750-13,430 cal. yBP (95% interval). An almost complete hominin calotte remain (No. MLDG-1704) (Figure 1) was found in the deposits of depth between 235 cm to 200 cm, dated ~ 14,000 cal. yBP.[20] The MLDG-1704 calotte retained nearly complete frontal and paired parietal bones, but its base and facial skeletons were missing, as cut-marks seen along the walls of the vault and on the zygomatic process. It was proposed that the cut-marks were caused by anthropogenic alteration. The CT-scans reconstructed endocranial volume (ECV) of the MLDG-1704 calotte is ~1,327 cm^3, a reduced ECV in comparison with the mean of East Asian early *H. sapiens* (EAEHS) (1,407 ± 146 cm^3) and Neanderthal (1,407 ± 172 cm^3). The parietal bones of MLDG-1704 (chord 107 mm, arc 123 mm) are significantly short than African early *H. sapiens* (chord 125 mm, arc 135 mm), European early *H. sapiens* (EUEHS, chord 120±7 mm) and EAEHS (chord 117±4 mm), while resembles Neanderthal (chord 108±4 mm, arc 115±5 mm). The frontal constriction index (ratio of minimum/maximum frontal breadth) of MLDG-1704 (76%) is unusually low, falling into the low boundary of EAEHS (76-89%), and smaller than the mean value of *H. erectus* (82±5%), Neanderthal (86±5%) and EUEHS (82±2%). Collectively, the multivariate cluster analysis assign MLDG-1704 within the *H. sapiens* convex hull, but near the edge of the *H. erectus* range (close to the late LPHO (Lower Paleolithic *Homo*) Sangiran-17[96] and Zhoukoudian-3 (700,000-year-old from northern China)[97] *H. erectus* samples).[20] Furthermore, the CT-scans of the maxillary third molar of MLDG-1747 specimen and *in situ* M_3 molar of MLDG-1679 specimen indicate both of them possess Taurodontism, which is rare among recent Europeans and EUEHS,[98] but is known as a distinguishing feature of Neanderthal.[99] Particularly, a partial proximal femur (MLDG-1678) exhibits highly unusual morphologies, the combination of its posterior and massive sizes are rare in AMH; the anteroposterior diameter, mediolateral diameter, total area, cortical area, neck-shaft angle, and the reconstructed body mass (50 kg), all showed exceptionally small, and mostly resemble archaic hominins especially the LPHO. The femoral multivariate Neighbor-Joining phylogenetic analyses clustered MLDG-1678 with archaic hominins.[21]

METHOD DETAILS

Ancient DNA extraction, library construction, small-scale sequencing and aDNA authentication evaluation

Using cranial fragmental material from MLDG-1704 calottes (Figure 1B), we performed aDNA extraction following a recently modified protocol that was customized to enrich short fragment DNA,[100] together with sodium hypochlorite treatment methods[101] in the aDNA dedicated laboratory of Comparative Genomic Group (CGG) at Kunming Institute of Zoology, Chinese Academy of Sciences. We first conducted 2 hours UV irradiation on the sample and removed a layer of surface using a sterile dentistry trill. Then, we irradiated the sample again with 1 hour UV radiation. We drilled out ~30-50 mg bone powder for every extraction. We obtained a total of 28 extractions. We first constructed 28 libraries with no uracil-DNA-glycosylase (UDG) treatment so that the terminal damage (mostly cytosine deamination) of aDNA could be maintained. These libraries included 9 single-stranded and 19 double-stranded libraries (SSL and DSL). We performed small-scale sequencing for these libraries to evaluate the terminal substitution patterns of the sequencing reads. We used the KAPA HiFi HotStart Uracil+ ReadyMix polymerase (KAPABIOSYSTEMS: Cat#KK2802) (DSL), and the AccuPrime Pfx DNA polymerase (Invitrogen: Cat# 12344-024) (SSL), for 12-15 cycles of amplifications of the shotgun sequencing libraries, and after enrichment, we performed 8-10 cycles using the KAPA HiFi HotStart ReadyMix (KAPABIOSYSTEMS: Cat#KK2612) to amplify the libraries to reach enough amount for shotgun sequencing using the Illumina HiSeq X Ten (PE-150) platform (https://www.illumina.com.cn/systems/sequencing-platforms/hiseq-x.html). In total, 346,400,590 sequencing reads were obtained.

Please cite this article in press as: Zhang et al., A Late Pleistocene human genome from Southwest China, Current Biology (2022), https://doi.org/10.1016/j.cub.2022.06.016

For the raw sequence data, we merged the forward and reverse read pairs that overlapped with each other by at least 11 bp to recover full-length sequences using the *leehom* tool.[93] The non-merged sequences and sequences < 35 bp in length were discarded. These recovered full-length sequences were then mapped to the human reference genome *hs37d5* using *BWA*[90] with the following parameters: aln -l 16500 -n 0.01 -o 2 -t 10. The SAM/BAM alignment files were processed, and PCR duplicates were removed using *SAMtools*[91] and *Picard* (http://picard.sourceforge.net/). The unmapped reads and reads with mapping quality scores < 25 were removed.

Using the final filtered bam files, for aDNA authentication evaluation, we assessed several features of the shotgun sequencing data, including the fragmental length and the endogenous DNA level (Figure S1A; Data S1A). We also measured and visualized the terminal damage pattern of the shotgun sequencing data using mapDamage2.07.[92] Since we used the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England BioLabs), and one-round PCR strategy to construct double-stranded libraries, as shown in recent reports,[51,102] these libraries showed one end (3') elevated terminal damage pattern (Figure S1B). Similarly, the single-stranded libraries showed one end (3') elevated terminal damage pattern, as we used only one-round PCR for library amplification (Figure S1C). However, in order to test the commonly applied two-round PCR protocol, we generated a new MZR extraction, and we rebuilt a double-stranded library using the aforementioned NEBNext Ultra II DNA Library Prep Kit (Cat#E7645). For the first-round PCR, we applied 8 cycles of PCR using KAPA HiFi HotStart Uracil+ ReadyMix (Cat#KK2802), then the library was purified using PB buffer with QIAGEN MinElute spin columns (Cat# 28006), followed by a second-round PCR of 10 cycles using the NEBNext Ultra II Q5 Master Mix (from the Cat#E7645), and again purified using PB buffer with QIAGEN MinElute spin columns. We observed a clear terminal damage pattern for both ends (5' C>T and 3' G>A) (Figure S1D), suggesting that this modification can help maintain the expected damage patterns.

Nuclear DNA and mitochondrial DNA in solution capture and large-scale sequencing

After proving the aDNA authentication of MZR, we then constructed 17 additional UDG-treated libraries (13 DSL and 4 SSL),[31,51] which were subjected to whole genome and mtDNA in solution capture. The modified version of the MYbaits enrichment protocol was used, which was customized for samples from low-latitude Southeast Asia.[30] These enriched libraries were used in large-scale sequencing using Illumina HiSeq X Ten (PE-150). In total, we obtained 1,558,057,324 sequencing reads. With this enrichment method, the mapping rates increased ~23.5-fold on average (Data S1A).

Data processing (merging, trimming, mapping and filtering)

Using the aforementioned method, we combined the data of all libraries and merged the double-stranded reads to recover full-length sequences, and then we evaluated the sequences via *FastQC* (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). To reduce the impact of deamination-induced damage in mapping and SNP calling, based on the mismatch pattern at the terminal positions of the 5' and 3' ends of the recovered full-length sequences (including both the large-scale and shotgun sequencing data), we trimmed 1 to 2 bp at the 5' end and 2 to 17 bp at the 3' end to obtain high-quality sequences. For example, the MZR_cls_02_Cap2 library was trimmed to 9 bp and 2 bp at the 3' and 5' ends, respectively (Figure S1E; Data S1A). We remapped the trimmed sequences to the reference genome *hs37d5*. The unmapped reads, the reads with length < 35 bp, and the reads with mapping quality score < 25 were removed. The final filtered bam files were merged and used for genotype calling and downstream analyses.

Sex determination and modern DNA contamination estimation

The ratio of mapped reads to autosomes and Y chromosomes can be used to infer the sex of the analyzed sample (female: $N_Y/N_{auto}<0.05$; male: $N_Y/N_{auto}>0.2$),[103] and MZR (MLDG-1704) turned out to be female based on the mapping ratio ($N_Y/N_{auto}=0.0026$). We estimated modern DNA contamination rates of both nuclear and mitochondrial DNA. For the nuclear genome, because MLDG-1704 is a female, the commonly used contamination estimating method customized to the X-chromosome[81] is not applicable. We therefore applied two other methods, including a recently published method through assessing the conditional C to T mismatch frequencies[78,104] and the *AuthentiCT* tool[94] customized for the single-stranded libraries. The estimated total contamination rate for the nuclear genome was ~0.72%, and it was ~0.1% for the SSL dataset (Data S1A and S1B). For mtDNA, we used two methods to estimate contamination. The *ContamMix* tool determines the contamination rate by comparing the sample's mtDNA fragments to its consensus genome and representative 311 modern worldwide sequences,[79] and the *Schmutzi* tool jointly estimates modern contamination based on sequence deamination patterns and fragment length distributions.[95] With the two methods, the estimated contamination levels for mtDNA are ~5.88% or 8.50%. However, if the very rare mutations C16400T (88/73,524) and C16404T (0/73,524) (the Phylotree database at https://www.phylotree.org/ and the human mtDNA genome sequences from NCBI (https://www.ncbi.nlm.nih.gov/)) were considered as deamination, the contamination rate would become zero based on the *ContamMix* tool (Data S1A).

Genotyping

The *snpAD* tool[32] jointly estimates genotype frequencies and sequencing errors along sequences to allow for accurate genotyping for aDNA sequences. Using the final filtered and merged bam files, we applied *snpAD* in calling the diploid genotypes of MZR, as well as Tianyuan and LLR, under default parameters. As the tool instructions suggested (https://bioinf.eva.mpg.de/snpAD/), the nonuracil-treated libraries and the uracil-treated libraries were analyzed separately using different parameters. To mask the complex genome

Please cite this article in press as: Zhang et al., A Late Pleistocene human genome from Southwest China, Current Biology (2022), https://doi.org/10.1016/j.cub.2022.06.016

regions, all calls were filtered for mapability before genotyping based on Heng Li's mapability track (https://bioinf.eva.mpg.de/map35/100/). With this method, we retrieved ~100,969,036 base pairs for MZR.

We also employed the '*PileupCaller*' program in *sequenceTools* (https://anaconda.org/bioconda/sequenceTools) with the strategy of randomly sampling one sequence for each SNP covered and generated the 'pseudohaploid' genotypes of MZR. To be comparable, we generated a "pileup" file at all positions of the 1,240k chip set (63,702 loci) as target genotyping positions, and only base quality higher than 30 positions was included for SNP calling. The called SNPs with the haploid genotype of MZR in the 1,240k chip set are 63,702. We validated that the two genotype sets (diploid and haploid) are highly consistent (99.2% overlap) (Figures S1F and S1G), indicating that our stringent filtering strategy is robust for calling high-confidence SNPs. For downstream population analyses, we used haploid genotype data.

Cluster analysis

Since MZR was identified as a novel basal mitochondrial M9 lineage (Data S1D), we compiled global M9 lineage data (Data S1E) and mapped the geographic distribution of current global populations using *Surfer16.0* (Golden Software, LLC) (Figure 2A). To explore the population relationship between MZR and ancient and modern global populations, we first merged the data of MZR with global modern humans, including the 1,000 Genomes Project (1KG) Phase 3,[105] EGDP,[106] SGDP,[107] SSIP,[108] SSMP,[109] Tibetan[110] and Andamanese,[111] comprising a total of 56 populations and 2,616 samples (Data S1F). In addition, to compare ancient global populations and include more modern populations, we further merged them with the compiled global genotype data available at Dr. David Reich's lab (https://reich.hms.harvard.edu/datasets). Overall, the number of overlapping SNPs of MZR with the 1KG dataset was 2,727,839, and there were 125,765 overlapping SNPs with all the collected modern global populations.

To relate MZR to both ancient and modern East Asians, principal component analysis (PCA) was performed using *Smartpca* in *EIGENSOFT-v6.1.4*,[36] which included MZR and 792 East Asian samples (47,087 SNPs, genotype rate: 82.78%). We projected the ancient samples onto the PC1 and PC2 variations of the modern samples using the *Smartpca* program with the default options except lsqproject: YES and numoutlieriter: 0 (Figure 2B). We validated the impact of deamination-induced damage by comparing the PCA using all SNPs or transversion SNPs with MZR and 1KG-East Asians (Figure S2). To infer the phylogenetic position of MZR among ancient East Asian samples ($\geq$ 7.0 kya), using *TreeMix*[112] (parameters: -noss -bootstrap -k 100 -root Mbuti –m (1-4)), we constructed a maximum likelihood tree (ML tree) (34,858 SNPs, genotype rate: 90.51%). As used in recent reports,[39,49] we applied the terms sEastAsian_EN (including Qihe3, Baojianshan5, Dushan and Liangdao2) and nEastAsian_EN (including Xiaogao, Bianbian and Boshan) to represent regional populations in East Asia (Figures 2C and S3). To explore the individual ancestries and population structure of MZR, we employed *ADMIXTURE*[36] with cross-validation (CV) using three different datasets (Figures 4A and S4). We first ran *ADMIXTURE* with 949 samples (44,236 SNPs, genotype rate: 97.64%), and then we verified the pattern by including two sample sets using different SNP sets with no missing data: MZR with 1KG populations (58,622 SNPs, genotype rate: 100%) and MZR, Tianyuan with 1KG populations (40,218 SNPs, genotype rate: 100%) (Figure S4). We evaluated the results of the *ADMIXTURE* analyses by setting K values (K=2 to K=15), and the results with the lowest CV values are presented (Figures 4A and S4). In addition, to search for the possible sources of external gene flow to modern Chinese, we tested the population genetic differentiation levels measured by *Fst* using *PLINK2.0*.[113] For the modern populations in China, we used populations with a sample size $\geq$ 10, and for the external comparison populations, we selected the major populations of Europe, Central Asia, Siberia, South Asia, Andaman Islands, and Papuan, and we included both modern and ancient populations with a sample size $\geq$ 10 (Figure S5; Data S1M).

Genetic affinity and archaic admixture test of MZR

Using ADMIXTOOLS,[33] we applied *f3* (modern/ancient Asians, MZR; Mbuti) to test the shared genetic drift of MZR with other Asian populations (Mbuti as outgroup) (Figure 4B; Data S1J). We validated the impact of deamination-induced damage by comparing the *f3* values using all SNPs or transversion SNPs with MZR and 1KG samples (Data S1C). In addition, we tested *f3* (modern Chinese; ancient East Asian ($\geq$ 7.0 kya); Mbuti) and detected the correlation of the *f3* value with the geographic locations (latitude) of modern populations in China (Figure 3A; Data S1L). We tested *f3* (global Late Pleistocene/Early Holocene populations, MZR; Mbuti) to detect allele sharing of MZR with worldwide populations (Figure 6A; Data S1O). Moreover, we applied *D*(MZR, ancient coastal East Asians/Paleo-Siberians; first Native Americans, Mbuti) to compare the allele sharing level of the first Native Americans with MZR, coastal East Asians and Paleo-Siberians (Figures 6B and S6A–S6D; Data S1P); We applied *D*(MZR, UKY; ancient East Asian/Siberian ($\geq$ 7.0 kya), Mbuti) to test the Late Pleistocene south-north divergence in East Asia (Figure 3A; Data S1G). The UKY sample was unearthed from the Ust-Kyakhta-3 site south to Lake Baikal and was dated 13.9 kya. The analysis of the UKY genome indicated a major component of East Asia[40] (Figures 4A and S4A). We calculated the Denisovan and AltaiNeanderthal introgression ratios in the MZR, Papuan, European and modern East Asian populations using the published method,[33] and the results are presented in Data S1H. To compare the archaic introgression ratio differences between the MZR and modern East Asians, we also applied *D*(MZR, East Asian; Denisovan/Neanderthal, Chimpanzee), and the results showed no significant difference (Data S1I).

Spatial-temporal tracing of East Asian-specific variants attributed to local adaptation

Using both the *snpAD* and *pileupCaller* genotype data, we initially checked the two representative East Asian-specific variations *EDAR-V370A* (rs3827760: A→G)[69,114,115] and *OCA2-HiS615Arg* (rs1800414: T→C)[60,62,116] in MZR, Tianyuan and LLR and found

Please cite this article in press as: Zhang et al., A Late Pleistocene human genome from Southwest China, Current Biology (2022), https://doi.org/10.1016/j.cub.2022.06.016

that all of them possess the ancestral alleles of the two variants (the genotype of MZR is missing in *EDAR-V370A*), contrasting with the prevalence with high frequency in modern East Asia. Then, we checked the two variants the global aDNA database (https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data) and in collected worldwide modern human populations (Data S1N and S1R), the genotype missing samples were excluded) to construct the spatial-temporal distribution map of the two variants (Figure 5).