Ancient DNA indicates human population shifts and admixture in northern and southern China

Melinda A. Yang1,2,3, Xuechun Fan4,5, Bo Sun6, Chungyu Chen7, Jianfeng Lang8, Ying-Chin Ko9, Cheng-hwa Tsang10, Hunglin Chiu10, Tianyi Wang1,2,11, Qingchuan Bao12, Xiaohong Wu13, Mateja Hajdinjak14, Albert Min-Shan Ko1, Manyu Ding1,2,15, Peng Cao1,2, Ruowei Yang1,2, Feng Liu1,2, Birgit Nickel13, Qingyan Dai1,2, Xiaotian Gao1,2,15, Yinqiu Cui17, David Reich19,20,21,22, Mark Stoneking14, Qiaomei Fu1,2,15, 16

1Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, Institute of Vertebrate Paleontology and Paleoanthropology, CAS, Beijing 100044, China. 2Center for Excellence in Life and Paleoenvironment, Chinese Academy of Sciences, Beijing 100044, China. 3Department of Biology, University of Richmond, Richmond, VA 23173, USA. 4International Research Center for Austronesian Archaeology, Pingtang 350000, China. 5Fujian Museum, Fuzhou 350001, China. 6Shandong Provincial Institute of Cultural Relics and Archaeology, Jinan 250012, China. 7Institute of History and Philology, Academia Sinica, Taipei 11529, Taiwan. 8School of History and Culture, Shandong University, Jinan 250010, China. 9Environment-omics-Disease Research Center, China Medical University and Hospital, Taichung 40402, Taiwan. 10Institute of Anthropology, National Tsinghua University, Hsinchu 30013, Taiwan. 11School of Cultural Heritage, Northwest University, Xi’an 710069, China. 12The Institute of Cultural Relics and Archaeology, Inner Mongolia Autonomous Region 010011, China. 13School of Archaeology and Museology, Peking University, Beijing 100871, China. 14Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany. 15University of Chinese Academy of Sciences, Beijing 100044, China. 16Shandong Museum, Jinan 250014, China. 17School of Life Sciences, Jinlin University, Changchun 130023, China. 18Institute of Cultural Heritage, Shandong University, Qingdao 266237, China. 19Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. 20Department of Human Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA. 21Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA. 22Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA.

*Corresponding author. Email: fuqiaomei@ivpp.ac.cn

Human genetic history in East Asia is poorly understood. To clarify population relationships, we obtained genome wide data from 26 ancient individuals from northern and southern East Asia spanning 9,500-300 years ago. Genetic differentiation was higher in the past than the present, reflecting a major episode of admixture involving northern East Asian ancestry spreading across southern East Asia after the Neolithic, transforming the genetic ancestry of southern China. Mainland southern East Asian and Taiwan Strait island samples from the Neolithic show clear connections with modern and ancient samples with Austronesian-related ancestry, supporting a southern China origin for proto-Austronesians. Connections among Neolithic coastal groups from Siberia and Japan to Vietnam indicate that migration and gene flow played an important role in the prehistory of coastal Asia.

East Asia, which consists of China, Mongolia, the Korean peninsula, and nearby islands, is home to almost one quarter of the world’s population and harbors a diverse array of ethnic groups and linguistic backgrounds (1). However, the genetic history of East Asia, especially in China, is poorly understood. Patterns of genetic relatedness among present-day East Asians run along a north to south cline (2–4), and high levels of genetic drift in East Asia suggest that East Asian populations underwent strong population bottlenecks prior to the Holocene and to a greater degree than Europeans (5). Ancient DNA studies have identified how East Asian ancestry impacted populations in Southeast Asia (6, 7), the Eastern Steppe (8), and northeastern Siberia (9). These studies also indicate that ancient individuals from the Southwest Pacific Islands share a close relationship with present-day Taiwanese Austronesians, who in turn share a close relationship with mainland East Asians (10).

In most studies to date, present-day East Asians (e.g., Han or Dai) have been used to represent East Asian ancestry in modeling studies. However, the archaeological record highlights that East Asians may have been more diverse in the past than today (11, 12). This genetic diversity is not well-studied, largely due to lack of sampling, making it difficult to characterize past population structure in northern and southern East Asia and limiting inferences of how past populations impacted extant East Asians.

A craniometric study on past and present humans suggested that human history in Asia is characterized by two ‘layers’ of ancestry – a ‘first layer’ composed of pre-Neolithic hunter-gatherers with a ‘second layer’ of northern East Asians spreading across Asia from the Early Neolithic to the present (13), contributing ancestry to many East Asians today. Obtaining genetic data from Neolithic East Asians, particularly those from China, would aid in resolving the role they played in forming the genetic patterns of East Asians today.
Genome-wide data from Neolithic East Asians shows their close relationship to East Asians today

To gain insights into the genetic history of East Asians during the Neolithic, we sampled genetic material from ancient individuals across East Asia dating to 9,500-300 calibrated years before present (9.5-0.3 kBP, Fig. 1, A and B) (14). In northern East Asia (defined as north of the Qinling-Huaihe line, Fig. 1A), we sampled from the northern Chinese provinces of Inner Mongolia and Shandong. In southern East Asia, we sampled from the southern Chinese province of Fujian in mainland East Asia, as well as two Taiwan Strait islands (Fig. 1A, Table 1, and table S1). We used large scale ancient nuclear DNA capture techniques (15) to enrich for endogenous DNA at 1.2 million single nucleotide polymorphisms (SNPs) (16) in 26 individuals. 24 individuals passed our analysis filters. These filters include (1) identification of characteristic ancient DNA damage signatures suggesting the presence of endogenous DNA (17), (2) assessing modern human contamination rates to determine for each sample whether to include all fragments (≥3% contamination) or only those with the characteristic damage signature (>3% contamination) (18, 19), and (3) exclusion of individuals where a contamination rate could not be estimated (14). In total, we obtained genetic information from 24 individuals from 11 sites, sequenced to between 0.01 - 7.60x fold coverage at the targeted SNPs (Table 1), with 16 individuals sampled from southern East Asia and 8 individuals sampled from northern East Asia.

To determine whether any individuals possessed ancestry deeply diverged from present-day East Asians, we first asked the extent to which they shared ancestry with ancient Asians previously sampled who separated early from the common ancestor of East Asians (‘early Asians’, e.g., 8,000-4,000-year-old Hōabinhians from Laos and Malaysia in Southeast Asia (7), the 3,000-year-old Ikawazu individual from Japan (7), and the 40,000-year-old Tianyuan individual from Beijing, China (20) (table S1). In a principal component analysis (PCA) including ancient and present-day Asians (14), all Neolithic East Asians cluster with populations of East Asian ancestry (Fig. 1C and fig. S1), who include present-day East Asians and Neolithic Asians from Siberia, Tibet, Southeast Asia, and the Southwest Pacific who possess primarily East Asian-related ancestry (8–10, 21). Notably, this includes Early Neolithic southern East Asians (Qihe, Liangdao), who possess cranial morphology that clustered with ‘early Asians’ (13). Thus, our results fail to support the version of the ‘two layer’ model in which these individuals are included in the first layer. Consistent results were obtained in an outgroup f₃ analysis, where Neolithic East Asians share more genetic similarity with Neolithic Siberians, Tibetans, and Southwest Pacific Islanders (f₃ = 0.28-0.32) than with ‘early Asians’ (f₃ = 0.25-0.26, fig. S2).

A direct comparison of Neolithic East Asians to present-day East Asians and ‘early Asians’ in a symmetry test shows that Neolithic East Asians tend to be more closely related to present-day East Asians than to any ‘early Asian’, i.e., most f₃(Mbuti, X; present-day East Asian, ‘early Asian’) values are significantly less than zero (Han: −22.1<Z<−4.4, table S2) and most f₄(Mbuti, ‘early Asian’; present-day East Asian, X) values are ≈0 (Han: −2.4<Z<2.6, table S2). Thus, the newly sampled individuals, including those from Early Neolithic southern East Asia, are genetically closest to modern populations of East Asian ancestry.

We additionally investigated patterns of archaic admixture in the Neolithic East Asians and found them to be similar to present-day East Asians (table S3). Neolithic East Asians also do not show consistent connections with a 35,000-year-old individual from Belgium, nor do they show connections to any South Americans – patterns that differ from Tianyuan (14, 20) but are similar to present-day East Asians (fig. S3). Previous ancient DNA studies found deeply divergent ancestries in pre-Neolithic Asia (7, 20) associated with a potential ‘first layer’ of populations in Asia (13). However, populations from the Early Neolithic in both northern and southern China show different patterns that suggest they do not belong to the ‘first layer’ and already carry ancestry primarily associated with present-day East Asians.

Population division between northern and southern East Asians in the Early Neolithic

In the PCA, our newly sampled individuals separate along geographic lines. Coastal Neolithic southern East Asians cluster together and close to present-day southern East Asians, while coastal Neolithic northern East Asians cluster together and close to present-day northern East Asians (Fig. 1C). These results suggest population structure between northern and southern East Asia since the Early Neolithic.

To determine genetic relationships in Neolithic East Asia, we first used a pairwise comparison of outgroup f₃-values (22) to assess their relationship to other Neolithic populations with East Asian-related ancestry (14). Neolithic southern East Asians share high genetic similarity to each other, Neolithic Southeast Asians, and Austronesian-related islanders from the Southwest Pacific (Fig. 1D), a pattern also observed in the PCA (Fig. 1C). In an f₃-analysis (fig. S4A), Neolithic southern East Asians and the 3,000-year-old Austronesian-related islanders from the Southwest Pacific (10) consistently share a closer genetic relationship with each other than with coastal and inland Neolithic northern East Asians, Siberians (8, 9), and Tibetans (21), though some Late Neolithic southern East Asians share a connection to northern East Asians (fig. S4B), which we below suggest is related to admixture. In a maximum likelihood phylogeny allowing migration events (23), Neolithic southern East Asians group together relative to
Neolithic northern East Asians, Siberians, and Tibetans (Fig. 2A and figs. S5 and S6) in 99.3-100% of the bootstrap replicates (bs). Thus, in mainland southern East Asia and Taiwan Strait islands, we find shared southern East Asian ancestry that differs from that observed in Neolithic northern East Asians – a pattern that persists even with the evidence of admixture that we describe below.

In northern East Asians dating to the Early Neolithic, we find shared ancestry not found in Early Neolithic southern East Asians, but present in Neolithic Siberians and Tibetans. For instance, in a maximum likelihood phylogeny, Neolithic northern East Asians, Siberians, and Tibetans group together relative to Neolithic southern East Asians and Austronesian-related islanders (bs = 95.3-99%, Fig. 2A and figs. S5 and S6), and they also share high genetic similarity with each other in an outgroup f3-analysis (Fig. 1D). In particular, coastal Neolithic northern East Asians clustered together, while Yumin, an inland Neolithic northern East Asian, clustered with Neolithic Siberians from the Eastern Steppe and the Primorye region of Far East Russia (8, 9). Most Neolithic northern East Asians and Siberians share a closer genetic relationship to each other than to Neolithic southern East Asians (figs. S7 and S8). In a model developed using all f2, f3, and f4-statistics (Fig. 1C), we similarly observe that northern and southern East Asian-related ancestry separate into two lineages (Fig. 2B and figs. S16 to S18).

Though the Neolithic is highlighted by distinct northern and southern East Asian-related ancestries, gene flow also impacted populations during this time. In the PCA, the inland Early Neolithic northern East Asian Yumin does not cluster with other northern East Asians as Neolithic Siberians and coastal northern East Asians do, but rather, Yumin falls in between northern and Tibetan present-day East Asians (Fig. 1C). Additionally, two patterns improve the fit of the maximum likelihood phylogeny (Fig. 2A and figs. S5 and S6). First, Upper Paleolithic northern Siberians influenced the ancestry of Neolithic Siberians (i.e., Yana, Fig. 2A), likely mediated by populations in Siberia closely related to Native Americans (Paleosiberians, table S1) as previously observed (9) and confirmed in f1-analyses showing Paleo-Asian connections primarily to Neolithic Siberians (table S4). Second, some coastal Early Neolithic northern East Asians show affinities to coastal Late Neolithic southern East Asians when compared to more inland Neolithic populations in Asia (northern East Asian, Siberian, and Tibetan, figs. S7 and S8). Coastal Early Neolithic northern East Asians show these connections to Early Neolithic southern East Asians as well (fig. S8B and fig. S9). These connections show that the population relationships among coastal northern and southern Neolithic East Asians cannot be easily clarified without the inclusion of admixture (Fig. 2, B to D, and fig. S10).

**Growing influence of northern East Asian ancestry in southern East Asia**

In a spatiotemporal analysis, we next assessed differences in East Asian ancestry between the Neolithic and today. A major feature of Neolithic East Asian ancestry described above is a northern ancestry, represented by Neolithic northern East Asians, Siberians, and Tibetans, and a southern ancestry, represented by Neolithic southern East Asians and Austronesian-related islanders. A test of genetic differentiation (Fst) shows that coastal Neolithic northern East Asians and coastal Neolithic southern East Asians are more highly differentiated from each other (Fst=0.042, sd=0.004) than are present-day northern and southern East Asians (Fst=0.023, sd=0.01, fig. S11). This difference indicates that East Asians today are more genetically homogeneous than Neolithic East Asians.

To determine what factors led to decreased genetic differentiation between present-day northern and southern East Asians relative to the Neolithic, we used an f2-analysis to set up a symmetry test (Fig. 3, A to C), where we compared ancient and present-day individuals against the oldest northern (Bianbian) and southern (Qihe) East Asians sampled. That is, we tested f2(Mbuti, X; Bianbian, Qihe) to determine northern and southern East Asian affinities across time and space. In the Early and Late Neolithic, there is clear spatial separation, with northern populations sharing a closer relationship to Bianbian and southern populations sharing a closer relationship to Qihe (Fig. 3, A and B, and table S5A). This pattern persists when we substitute other Early Neolithic southern East Asians for Qihe and other Early Neolithic northern East Asians for Bianbian (fig. S12 and table S5, B and C). We then used qpAdm (15) to develop mixture models (Fig. 3, D to F) to estimate ancestry proportions in Neolithic populations (14), where Neolithic northern and southern East Asians each belong to a distinct ancestry associated with their location in northern or southern East Asia (Fig. 3, D and E, and table S6A).

In contrast, patterns in present-day East Asians show that the main factor reducing genetic differentiation in present-day East Asians is increased northern East Asian-related ancestry in southern East Asia. In present-day East Asians, we observe a dramatic change where instead of affinities based on location, all East Asians (including mainland southern East Asians) share an affinity with Neolithic northern East Asians, as shown by f2(Mbuti, present-day East Asian; Qihe, Bianbian)>0 (3<Z<8.8, Fig. 3C and table S5A). Estimates of ancestry proportions in our mixture models show that this is due to increased northern East Asian-related ancestry in southern mainland East Asia, with estimates of northern East Asian-related ancestry ranging from 21-55% (Fig. 3F, fig. S13, and table S6B). Southern East Asian ancestry also extends north as well, as it is found in Han populations from northern China (36-41%, fig. S13 and table S6B), as well as some...
northern East Asians (e.g., Xibo, Korean, 35-36%, Fig. 3F and table S6B). Paleosiberian-related ancestry from Siberia also greatly impacts recent East Asians, with the exception of a 300-year-old individual from coastal mainland southern East Asia, island populations, and Tibetans (Fig. 3F and table S6). This lack of Paleosiberian-related ancestry in the peripheral edges of East Asia suggests that different types of gene flow from north to south occurred in East Asia.

While we are unable to estimate the timing of gene flow events, patterns in the f4-comparisons suggest that population movement may have already started impacting East Asians by the Late Neolithic. In the test f4(Mbuti, Bianbian; Kolyma, Neolithic southern East Asians), we would expect similar results for all Neolithic southern East Asians if none of them had northern East Asian ancestry. Instead, the Late Neolithic southern East Asians share a connection to the coastal northern East Asian Bianbian that Early Neolithic southern East Asians do not share (fig. S14). Other tests of admixture support this finding: Admixture Graph (Fig. 3, C and D) and ADMIXTURE (fig. S10) results both estimate northern East Asian-related ancestry in Late Neolithic southern East Asians that is reduced in Early Neolithic southern East Asians (14).

We also investigated whether the northern East Asian affinity is closer to inland (i.e., Yumin) or coastal (i.e., Bianbian) northern East Asians. In an f3-test of symmetry, all Neolithic East Asians and coastal Neolithic Siberians share a closer relationship to Bianbian rather than Yumin, whereas all inland Neolithic Siberians and Tibetans do not (fig. S15 and table SSD). This suggests that the northern East Asian ancestry found in all present-day mainland East Asians is primarily related to populations along the lower reaches of the Yellow River. These observations are consistent with archaeological and historical studies that argue for an origin of the Han ethnic group in northern China, along the Yellow River (24).

**Proto-Austronesian origins in southern China and coastal connections**

Austronesian-speaking populations extend from Taiwan to the Southwest Pacific and to Madagascar. Mainland southern China has been proposed to be the likely origin of the proto-Austronesian population that entered Taiwan, based on geographic proximity across the Taiwan Strait and the occurrence of similar artifacts at archaeological sites from both regions (II, 22, 25, 26). Furthermore, genetic modeling with mitochondrial DNA (mtDNA) genomes from present-day Southeast Asians (27) also suggest a southern Chinese origin. The mtDNA of the 8,320-8,060-year-old Liangdaol individual belongs to haplogroup E1, which is common in Austronesian-speaking populations in Taiwan, the Philippines, and Indonesia today, and is most similar to mtDNA found in the aboriginal Formosans from Taiwan (28).

Our data also support a proto-Austronesian origin in mainland southern China. Neolithic southern East Asians consistently fall on a cline leading to present-day Austronesians rather than other present-day southern East Asian populations in a PCA (Fig. 1C), and they share a connection with present-day Austronesians (Ami) relative to other present-day southern East Asians, e.g., f4(Mbuti, coastal sEastAsia_EN/LN; Dai, Ami)>0 (5.2<Z<12.7, table S7A). This connection to present-day Austronesians extending across all Neolithic samples tested from mainland southern China and Taiwan Strait islands supports a coastal southern China origin for proto-Austronesians (29).

We further investigated connections between the ancient samples and 3,000-year-old Southwest Pacific islanders from Vanuatu (10), who share a close relationship to present-day Austronesians. The Neolithic samples from mainland southern China and the Taiwan Strait islands group with the ancient Vanuatu samples in a maximum-likelihood tree (b.s. = 82.3-100%, Fig. 2A and fig. S5) and share high genetic similarity with the Vanuatu in an outgroup f3-analysis (Fig. 1D). Moreover, a direct f4-comparison demonstrates that Late Neolithic southern East Asians share a close genetic relationship with these Austronesian-related Southwest Pacific islanders (fig. S4A), with no excess connections from either group to Neolithic northern East Asians (fig. S4B). These results further support the proposed connection between Neolithic southern East Asians and proto-Austronesians (29). A lack of genetic isolation between coastal populations can be observed all along the coast of East and Southeast Asia. While most Neolithic Southeast Asians are a mixture of Hōabinhian-related and southern East Asian-related ancestry, a 4,000-year-old population from Vietnam shares an especially close relationship to coastal Late Neolithic southern East Asians (table S7B). Strikingly, these coastal connections extend further north. A 2,700-year-old Jōmon individual from the Japanese archipelago that separated early from present-day East Asians was suggested to share genetic affinity to Hōabinhians (7), but comparisons including Neolithic East Asians do not show this pattern (table S8). Instead we find that the Jōmon individual shows affinities to several coastal Neolithic populations in Siberia, as well as southern East Asia (table S9). The patterns demonstrated here show that coastal regions were areas of interconnectivity and gene flow rather than isolation. These affinities between populations along the eastern coast of Asia (and island populations off the coast) that are not shared by more inland Asian populations suggest that interactions along marine-related environments played an important role in the prehistory of coastal Asia.

**Discussion**

Our genetic survey of Neolithic northern and southern East Asians shows population differentiation during the Early
Neolithic to an extent not observed in present-day East Asians. In a craniometric analysis of ancient and present-day East Asians, it was proposed that a 'second layer' of northern East Asian-related populations spread across East Asia around the Early Neolithic, replacing at least in part a 'first layer' of pre-Neolithic hunter-gatherers (30). While we did not find evidence of a 'first layer' population in coastal southern East Asia by 8,400 years ago, we do observe increased northern influences in southern East Asia between the Early Neolithic and today. Thus, the argument for the spread of a 'second layer' associated with northern East Asian ancestry is still an important model to explore in the context of East Asian prehistory.

The spread of northern East Asian ancestry led, however, to increased admixture in both directions, such that most of today's East Asians are a mixture of northern and southern East Asian ancestries. Thus, not only was there spread of northern East Asian ancestry into southern East Asia, but southern East Asian-related ancestry can be found in some present-day northern East Asians. That we do not observe admixture to this extent in the Neolithic suggests that much of the human movement that contributed to present-day East Asian genetic patterns must have occurred after the Neolithic.

Shared ancestry amongst ancient individuals on the southeast coast of mainland East Asia, islands in the Taiwan Strait, and the Southwest Pacific island Vanuatu suggest that the Austronesian expansion derived from a population that came from southern China, a pattern supported by similarities in material culture on the southeast coast of China and in Austronesian-related material culture (11), as well as in studies of uniparental markers (27, 28). Moreover, gene flow among coastal populations in East Asia is a common trend, as different coastal connections can be observed in Neolithic populations from as far north as coastal Siberia and the Japanese archipelago to as far south as coastal Vietnam.

In individuals from northern and southern China dating to 9.5-0.3 kBP, we observe a close genetic relationship between East Asians, albeit with shifts in ancestry indicative of population movement and admixture during the Neolithic. With increased genetic sampling from the Paleolithic, as well as populations further inland in central China, we anticipate being able to further clarify the relationships among Paleolithic hunter-gatherers, Neolithic farmers, and present-day populations of East Asia.

REFERENCES AND NOTES


SUPPLEMENTARY MATERIALS

4 November 2019; accepted 1 May 2020
Published online 14 May 2020
10.1126/science.aba0909

ACKNOWLEDGMENTS

We would like to thank J. Kelso, M. Slatkin, and J. Gardner for comments, as well as S. Pääbo and his lab from the Max Planck Institute of Evolutionary Anthropology for their help in setting up the lab at IVPP, CAS. We would additionally like to thank our reviewers for giving insightful suggestions to the manuscript. Funding: This work was supported by the Chinese Academy of Sciences (CAS, XDB26000000), National Natural Science Foundation of China (91731303, 41925009, 41630102), CAS (XDA1905010, QYZDB-SSW-DQC003), Tencent Genome Sequence Archive (HRA000123). All software used are freely available online and are referenced in (4).

Author contributions: Q.F. conceived the idea and wrote the manuscript. M.A.Y and Q.F. analyzed the data and wrote the manuscript with critical input from all authors. M.A.Y., T.W., M.D. and Q.F. wrote and edited the supplement. M.A.Y and Q.F. assembled archaeological materials and dating. M.Z., X.G. and Q.F. assembled archaeological materials and dating. Q.F., P.C., R.Y., F.L., X.F., Q.D., C.N. and Y.C. performed or supervised wet laboratory work. M.A.Y. and Q.F. performed analyses and writing at the Institute of Vertebrate Paleontology and Paleoanthropology, with additional revisions done at the University of Richmond.

Competing interests: The authors declare no competing interests.

Data and materials availability: BAM files and genotype calls for the newly sequenced individuals are available at the Genome Sequence Archive (31) in BIG Data Center (32) (https://bigd.big.ac.cn/lgsc-human; accession number: HRA000123). All software used are freely available online and are referenced in (4).

REFERENCES


Supplementary Text

Figs. S1 to S18

Tables S1 to S9

References (33–79)

MDAR Reproducibility Checklist
Fig. 1. Geographic, temporal, and genetic information for newly sampled East Asians. (A) Map showing the locations from which ancient individuals were sampled for this study. The Qinling-Huaihe line between the Huaihe River and the Qinling mountains is used to divide northern and southern East Asia. Gray shading indicates differences in elevation across the region. (B) Chronology of ancient individuals sampled in this study, where individuals date to 9,500-7,500 BP (Early Neolithic), 5,000-4,000 BP (Late Neolithic), or 300 BP (Historic). Individuals from northern East Asia all date to the Early Neolithic, while individuals from southern East Asia date to all three time periods. (C) Projection of ancient Asians onto a principal component analysis for present-day East Asians. Present-day East Asians (gray dots) form three main vertices: ‘northern East Asians’, ‘southern East Asians’, and ‘Tibetans’ that correspond to the major present-day East Asian populations. The gray text refers to language groups associated with present-day populations, and a more detailed listing of ethnic groups included can be found in table S1 or fig. S1. Ancient published samples are grouped by archeological location and numbered in gray. 1: ‘Early Asians’; 2: ancient Austronesian-related islanders from Southeast Asia and the Southwest Pacific; 3: Neolithic, Bronze Age, and historical Southeast Asians; 4: Neolithic, Bronze Age, and historical Tibetans; 5: Paleosiberian (Kolyma) and Neolithic Siberians. (D) Pairwise outgroup $f_3$-statistics for ancient East Asians, where yellow indicates high genetic similarity between pairs. The outgroup $f_3$-test takes the form $f_3(Mbuti; X, Y)$ where X and Y are Neolithic Asians listed in the rows and columns and Mbuti (a present-day population from Central Africa) is the outgroup.
Fig. 2. Modeling tree relationships between ancient Asians. (A) A maximum likelihood phylogeny allowing three migration events using the Treemix software. Aside from coastal nEastAsia_EN, sEastAsia_EN, and sEastAsia_LN which refer to the ‘Population’ column in Table 1 or table S1 and group several ancient individuals from more than one site together, all others refer to the nomenclature in the ‘Name’ column designated in Table 1 or table S1. In (B), we show an Admixture Graph that fits the data (|maxZ|=2.7), where we include two Early Neolithic southern East Asians (sEastAsia: Qihe, Liangdao2), three Early Neolithic northern East Asians (inland and coastal nEastAsia: Yumin, Boshan, Xiaogao), a Neolithic Siberian (Shamanka_EN), a Paleosiberian (Kolyma), early Asian (Tianyuan), two individuals of non-Asian ancestry (Kostenki14, Yana), and the Central African Mbuti whom we used to root the tree. Branches are denoted by solid lines with branch lengths given in units of 1000 times the $f_2$ drift distance, and admixture events are denoted by dotted lines with mixture proportions as shown (all numbers rounded to the nearest integer). Other tree models that fit the data are shown in fig. S18. In (C and D), we consider the estimated mixture proportions adding these populations onto the graph in (B). We add present-day (C) and ancient (D) East Asians to the graph in (B), allowing for each addition to be added as a single node or a mixture of two branches. We considered the tree with the best fit considering all $f_2$, $f_3$, and $f_4$-statistics comparing that individual or population to all tips in the tree. The color coding and associated key refers to the area of the tree where they attach and the estimated mixture proportions. In both (A) and (B), we find a separation between Early Neolithic northern and southern East Asians with connections between coastal populations in the north and south.
Fig. 3. Ancestry localized to the north and south in the Early Neolithic is found in admixed form across northern and southern East Asia today. (A to C) A heatmap showing $f_2$(Mbuti, X; Qihe, Bianbian), where Bianbian is the oldest northern East Asian sampled (~9,500 BP) and Qihe is the oldest southern East Asian sampled (~8,400 BP). X are East and Southeast Asians who date to the Early Neolithic (A), Late Neolithic (B), or present-day (C). Green indicates more affinity to Early Neolithic northern East Asians, whereas blue indicates greater affinity to Early Neolithic southern East Asians. Numerical values can be found in table S5A. (D to F) Ancestry proportions estimated for Neolithic and present-day individuals. Possible ancestries are northern East Asian (green), southern East Asian (blue), Paleosiberian (light green), Hoabinhian (orange), and Jōmon (light blue). Proportions were determined using qpAdm, with representative sources and outgroups described in (14). During the Neolithic, there was a division between East Asians north and south of the Qinling-Huaihe line (dark gray line on map) in ancestry, but in the present-day, ancestries previously localized to only northern or southern East Asia can be found in appreciable frequencies in both regions. Results for present-day Han are in fig. S13 and point values are in table S6.
Table 1. Ancient individuals sampled in this study.

<table>
<thead>
<tr>
<th>ID</th>
<th>Sample ID</th>
<th>Population</th>
<th>Name&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Skeletal Element</th>
<th>Date (cal BP)</th>
<th>Region</th>
<th>Lat</th>
<th>Lon</th>
<th>Library style</th>
<th>Sex</th>
<th>C−T%</th>
<th>Contamination</th>
<th>MtDNA</th>
<th>Cor&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M1</td>
<td>inlandEastAsia_EN</td>
<td>Yunnin</td>
<td>Teeth</td>
<td>8,415-8,335</td>
<td>Inner Mongolia</td>
<td>42</td>
<td>114.2</td>
<td>DS</td>
<td>F</td>
<td>17</td>
<td>-</td>
<td>0.028</td>
<td>7.25</td>
</tr>
<tr>
<td>2</td>
<td>Blue Shan</td>
<td>coastal EastAsia_EN</td>
<td>Blue Shan</td>
<td>Petros</td>
<td>9,544-9,480</td>
<td>Shandong</td>
<td>36.1</td>
<td>118.5</td>
<td>SS-ES</td>
<td>M</td>
<td>11</td>
<td>0.01</td>
<td>0.014</td>
<td>2.21</td>
</tr>
<tr>
<td>3</td>
<td>IS11</td>
<td>coastal EastAsia_EN</td>
<td>Boshan</td>
<td>Teeth</td>
<td>8,320-8,408</td>
<td>Shandong</td>
<td>36.5</td>
<td>117.9</td>
<td>SS-UDG</td>
<td>M</td>
<td>10</td>
<td>0.03</td>
<td>0.07</td>
<td>7.36</td>
</tr>
<tr>
<td>4</td>
<td>M7</td>
<td>coastal EastAsia_EN</td>
<td>Xiaoxiang</td>
<td>Petros</td>
<td>7,872-7,721</td>
<td>Shandong</td>
<td>36.5</td>
<td>117.9</td>
<td>DS</td>
<td>F</td>
<td>31</td>
<td>-</td>
<td>0.018</td>
<td>0.21</td>
</tr>
<tr>
<td>5</td>
<td>M6</td>
<td>coastal EastAsia_EN</td>
<td>Xiaoxiang</td>
<td>Petros</td>
<td>7,995-7,766</td>
<td>Shandong</td>
<td>36.5</td>
<td>117.4</td>
<td>DS</td>
<td>F</td>
<td>24</td>
<td>-</td>
<td>0.029</td>
<td>0.04</td>
</tr>
<tr>
<td>6</td>
<td>M4</td>
<td>coastal EastAsia_EN</td>
<td>Xiaoxiang</td>
<td>Petros</td>
<td>7,877-7,735</td>
<td>Shandong</td>
<td>36.5</td>
<td>117</td>
<td>DS</td>
<td>M</td>
<td>21</td>
<td>-</td>
<td>0.019</td>
<td>0.05</td>
</tr>
<tr>
<td>7</td>
<td>M3</td>
<td>coastal EastAsia_EN</td>
<td>Xiaoxiang</td>
<td>Petros</td>
<td>8,777-8,391</td>
<td>Shandong</td>
<td>37.9</td>
<td>117.6</td>
<td>DS</td>
<td>F</td>
<td>30</td>
<td>-</td>
<td>0.018</td>
<td>7.6</td>
</tr>
<tr>
<td>8</td>
<td>Q102</td>
<td>coastal EastAsia_EN</td>
<td>Qu102</td>
<td>Petros</td>
<td>8,429-8,359</td>
<td>Fujian</td>
<td>25.4</td>
<td>117.6</td>
<td>SS</td>
<td>F</td>
<td>34</td>
<td>-</td>
<td>0.032</td>
<td>0.045</td>
</tr>
<tr>
<td>9</td>
<td>Q101</td>
<td>coastal EastAsia_EN</td>
<td>Qu101</td>
<td>Petros</td>
<td>8,429-8,359</td>
<td>Fujian</td>
<td>25.4</td>
<td>117.6</td>
<td>SS</td>
<td>F</td>
<td>34</td>
<td>-</td>
<td>0.032</td>
<td>0.045</td>
</tr>
<tr>
<td>10</td>
<td>M1</td>
<td>island EastAsia_EN</td>
<td>Lingshui01</td>
<td>Petros</td>
<td>8,320-8,050</td>
<td>Lingshui</td>
<td>26.3</td>
<td>120.2</td>
<td>SS-UDG</td>
<td>M</td>
<td>12</td>
<td>0.03</td>
<td>0.015</td>
<td>2.72</td>
</tr>
<tr>
<td>11</td>
<td>M2</td>
<td>island EastAsia_EN</td>
<td>Lingshui02</td>
<td>Petros</td>
<td>7,590-7,560</td>
<td>Lingshui</td>
<td>26.3</td>
<td>120.2</td>
<td>SS-UDG</td>
<td>F</td>
<td>19</td>
<td>-</td>
<td>0.005</td>
<td>1.08</td>
</tr>
<tr>
<td>12</td>
<td>Suogang11</td>
<td>inlandEastAsia_EN</td>
<td>Suogang</td>
<td>Petros</td>
<td>4,800-4,390*8</td>
<td>Penglai</td>
<td>23.5</td>
<td>119.6</td>
<td>SS</td>
<td>M</td>
<td>8</td>
<td>-0.02</td>
<td>0.022</td>
<td>0.04</td>
</tr>
<tr>
<td>13</td>
<td>Suogang13</td>
<td>inlandEastAsia_EN</td>
<td>Suogang</td>
<td>Petros</td>
<td>4,800-4,390*4</td>
<td>Penglai</td>
<td>23.5</td>
<td>119.6</td>
<td>SS</td>
<td>M</td>
<td>10</td>
<td>0.11</td>
<td>0.28</td>
<td>0.03</td>
</tr>
<tr>
<td>14</td>
<td>L5705</td>
<td>coastal EastAsia_EN</td>
<td>Xinmou</td>
<td>Petros</td>
<td>4,419-4,246</td>
<td>Fujian</td>
<td>26.2</td>
<td>119.1</td>
<td>SS</td>
<td>F</td>
<td>33</td>
<td>-</td>
<td>0.022</td>
<td>0.06</td>
</tr>
<tr>
<td>15</td>
<td>L5700</td>
<td>coastal EastAsia_EN</td>
<td>Xinmou</td>
<td>Petros</td>
<td>4,550-4,417</td>
<td>Fujian</td>
<td>26.2</td>
<td>119.1</td>
<td>SS</td>
<td>F</td>
<td>28</td>
<td>-</td>
<td>0.027</td>
<td>0.43</td>
</tr>
<tr>
<td>16</td>
<td>L5692</td>
<td>coastal EastAsia_EN</td>
<td>Xinmou</td>
<td>Petros</td>
<td>4,350-4,174</td>
<td>Fujian</td>
<td>26.2</td>
<td>119.1</td>
<td>SS</td>
<td>M</td>
<td>24</td>
<td>-0.02</td>
<td>0.058</td>
<td>0.01</td>
</tr>
<tr>
<td>17</td>
<td>L5706</td>
<td>coastal EastAsia_EN</td>
<td>Xinmou</td>
<td>Petros</td>
<td>4,557-4,406</td>
<td>Fujian</td>
<td>26.2</td>
<td>119.1</td>
<td>SS</td>
<td>M</td>
<td>25</td>
<td>-0.02</td>
<td>0.043</td>
<td>0.04</td>
</tr>
<tr>
<td>18</td>
<td>L5704</td>
<td>coastal EastAsia_EN</td>
<td>Xinmou</td>
<td>Petros</td>
<td>4,500-4,232</td>
<td>Fujian</td>
<td>26.2</td>
<td>119.1</td>
<td>SS</td>
<td>M</td>
<td>23</td>
<td>0.11</td>
<td>0.11</td>
<td>0.04</td>
</tr>
<tr>
<td>19</td>
<td>L5703</td>
<td>coastal EastAsia_EN</td>
<td>Xinmou</td>
<td>Petros</td>
<td>4,644-4,390</td>
<td>Fujian</td>
<td>26.2</td>
<td>119.1</td>
<td>SS</td>
<td>M</td>
<td>22</td>
<td>0.12</td>
<td>0.074</td>
<td>0.02</td>
</tr>
<tr>
<td>20</td>
<td>L5702</td>
<td>coastal EastAsia_EN</td>
<td>Xinmou</td>
<td>Petros</td>
<td>-</td>
<td>Fujian</td>
<td>26.2</td>
<td>119.1</td>
<td>SS</td>
<td>M</td>
<td>13</td>
<td>-</td>
<td>0.01</td>
<td>0.04</td>
</tr>
<tr>
<td>21</td>
<td>L5701</td>
<td>coastal EastAsia_EN</td>
<td>Xinmou</td>
<td>Petros</td>
<td>4,413-4,240</td>
<td>Fujian</td>
<td>26.2</td>
<td>119.1</td>
<td>SS</td>
<td>M</td>
<td>28</td>
<td>0.07</td>
<td>0.039</td>
<td>0.04</td>
</tr>
<tr>
<td>22</td>
<td>L7415</td>
<td>coastal EastAsia_EN</td>
<td>Tanchishan</td>
<td>Petros</td>
<td>4,419-4,246</td>
<td>Fujian</td>
<td>26.1</td>
<td>119.2</td>
<td>SS</td>
<td>M</td>
<td>23</td>
<td>0.05</td>
<td>0.04</td>
<td>351812</td>
</tr>
<tr>
<td>23</td>
<td>L7417</td>
<td>coastal EastAsia_EN</td>
<td>Tanchishan</td>
<td>Petros</td>
<td>4,350-4,174</td>
<td>Fujian</td>
<td>26.1</td>
<td>119.2</td>
<td>SS</td>
<td>F</td>
<td>24</td>
<td>-0.05</td>
<td>0.14</td>
<td>144849</td>
</tr>
<tr>
<td>24</td>
<td>L7416</td>
<td>coastal EastAsia_EN</td>
<td>Tanchishan</td>
<td>Petros</td>
<td>-</td>
<td>Fujian</td>
<td>26.2</td>
<td>119.1</td>
<td>SS</td>
<td>M</td>
<td>11</td>
<td>-</td>
<td>0.01</td>
<td>8422</td>
</tr>
<tr>
<td>25</td>
<td>L5698</td>
<td>coastal EastAsia_EN</td>
<td>Tanchishan</td>
<td>Petros</td>
<td>4,410-4,225</td>
<td>Fujian</td>
<td>26.2</td>
<td>119.1</td>
<td>SS</td>
<td>F</td>
<td>25</td>
<td>0.0</td>
<td>0.02</td>
<td>2092</td>
</tr>
<tr>
<td>26</td>
<td>L5696</td>
<td>coastal EastAsia_EN</td>
<td>Tanchishan</td>
<td>Petros</td>
<td>-</td>
<td>Fujian</td>
<td>26.2</td>
<td>119.1</td>
<td>SS</td>
<td>M</td>
<td>18</td>
<td>0.16</td>
<td>0.09</td>
<td>0.02</td>
</tr>
<tr>
<td>27</td>
<td>L5694</td>
<td>Coastal EastAsia_H</td>
<td>Cheyun</td>
<td>Petros</td>
<td>334-281</td>
<td>Fujian</td>
<td>25.6</td>
<td>117.3</td>
<td>SS</td>
<td>M</td>
<td>11</td>
<td>0.01</td>
<td>0.005</td>
<td>0.38</td>
</tr>
</tbody>
</table>

*The bone was not directly radiocarbon dated. For the Suogang site, shells were radiocarbon dated to 4,633-4,287 cal BP and 4,793-4,407 cal BP. & radiocarbon date published elsewhere

<sup>a</sup>Based on the 1.2M SNP Panel.

<sup>A</sup>Alleles were determined using the damage-restricted library.

Red highlighting - high contamination; yellow highlighting - Sample was excluded from demographic analysis due to low number of SNPs.

<sup>a</sup>Labels used in analyses, which is the same as the site name. If the same name, they were grouped together.
Ancient DNA indicates human population shifts and admixture in northern and southern China

Melinda A. Yang, Xuechun Fan, Bo Sun, Chungyu Chen, Jianfeng Lang, Ying-Chin Ko, Cheng-hwa Tsang, Hunglin Chiu, Tianyi Wang, Qingchuan Bao, Xiaohong Wu, Mateja Hajdinjak, Albert Min-Shan Ko, Manyu Ding, Peng Cao, Ruowei Yang, Feng Liu, Birgit Nickel, Qingyan Dai, Xiaotian Feng, Lizhao Zhang, Chengkai Sun, Chao Ning, Wen Zeng, Yongsheng Zhao, Ming Zhang, Xing Gao, Yinqiu Cui, David Reich, Mark Stoneking and Qiaomei Fu

published online May 14, 2020