HUMAN EVOLUTION

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

Melinda A. Yang[1,2,3], Xuechun Fan[4,5], Bo Sun[6], Chungyu Chen[7], Jianfeng Lang[8], Ying-Chin Ko[9], Cheng-hwa Tsang[10], Hunglin Chiu[10], Tianyi Wang[1,2,11], Qingchuan Bao[12], Xiaohong Wu[13], Mateja Hajdinjak[14], Albert Min-Shan Ko[1], Manyu Ding[1,2,15], Peng Cao[1,2], Ruowei Yang[1,2], Feng Liu[1,2], Birgit Nickel[13], Qingyan Dai[1,2], Xiaotian Feng[1,2], Lizhao Zhang[1,2], Chengkai Sun[16], Chao Ning[17], Wen Zeng[18], Yongsheng Zhao[18], Ming Zhang[1,2,15], Xing Gao[1,2,15], Yinqiu Cui[17], David Reich[19,20,21,22], Mark Stoneking[14], Qiaomei Fu[1,2,15]*

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 9500 to 300 years ago. Genetic differentiation in this region was higher in the past than the present, which reflects a major episode of admixture involving northern East Asian ancestry spreading across southern East Asia after the Neolithic, thereby 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 individuals with Austronesian-related ancestry, which supports an origin in southern China 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 before the Holocene and to a greater degree than European populations (*5*). Ancient DNA studies have identified how East Asian ancestry affected 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 Asian populations (e.g., Han or Dai) have been used to represent East Asian ancestry in modeling studies. However, the archaeological record suggests that East Asians may have been more diverse in the past than they are today (*11*, *12*). This genetic diversity is not well studied, largely owing to a lack of sampling, which makes it difficult to characterize past population structure in northern and southern East Asia and limits inferences of how past populations affected extant East Asians.

A craniometric study on past and present humans has suggested that human history in Asia is characterized by two so-called layers of ancestry: a first layer composed of pre-Neolithic hunter-gatherers with a second layer of northern East Asians who spread 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 help resolve the role they played in forming the genetic patterns of present-day East Asians.

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 9500 to 300 calibrated years before the present (cal yr B.P.) (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. Of these, 24 individuals passed our analysis filters. These filters include the following: (i) identification of characteristic ancient DNA damage signatures that suggest the presence of endogenous DNA (*17*); (ii) assessment of 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 (iii) 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- and 7.60-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 had ancestry that deeply diverged from that of present-day East Asians, we first investigated the extent to which they shared ancestry with previously sampled ancient Asians who separated early from the common ancestor of East Asians, referred to as early Asians [e.g., 8000- to 4000-year-old Hòabìnhians from Laos and Malaysia in Southeast Asia (*7*), the 3000-year-old Ikawazu individual from Japan (*7*), and the 40,000-year-old Tianyuan individual from Beijing, China (*20*)] (table S1). In a principal components analysis (PCA) that included ancient and present-day Asians (*14*), all Neolithic East Asians cluster with populations of East Asian ancestry (Fig. 1C and fig. S1), which include present-day East Asians and Neolithic Asians from Siberia, Tibet, Southeast Asia, and the Southwest Pacific who have primarily East Asian–related ancestry (*8*–*10*, *21*). Notably, this includes Early Neolithic southern East Asians (Qihe and Liangdao), who have cranial morphology that clustered with that of early Asians (*13*). Thus, our results fail to support the version of the two-layer model in which these individuals are included in the

---

[1]Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences (CAS), Beijing 100044. [2]Center for Excellence in Life and Paleoenvironment, Chinese Academy of Sciences, Beijing 100044. [3]Department of Biology, University of Richmond, Richmond, VA 23173. [4]International Research Center for Austronesian Archaeology, Pingtan 350000. [5]Fujian Museum, Fuzhou 350001. [6]Shandong Provincial Institute of Cultural Relics and Archaeology, Jinan 250012. [7]Institute of History and Philology, Academia Sinica, Taipei 11529. [8]School of History and Culture, Shandong University, Jinan 250100. [9]Environment-Omics-Disease Research Center, China Medical University and Hospital, Taichung 40402. [10]Institute of Anthropology, National Tsinghua University, Hsinchu 30013. [11]School of Cultural Heritage, Northwest University, Xi'an 710069. [12]The Institute of Cultural Relics and Archaeology, Inner Mongolia Autonomous Region 010011. [13]School of Archaeology and Museology, Peking University, Beijing 100871. [14]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig. [15]University of Chinese Academy of Sciences, Beijing 100049. [16]Shandong Museum, Jinan 250014. [17]School of Life Sciences, Jilin University, Changchun 130023. [18]Institute of Cultural Heritage, Shandong University, Qingdao 266237. [19]Department of Genetics, Harvard Medical School, Boston, MA 02115. [20]Department of Human Evolutionary Biology, Harvard University, Cambridge, MA 02138. [21]Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA 02142. [22]Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115.
*Corresponding author. Email: fuqiaomei@ivpp.ac.cn

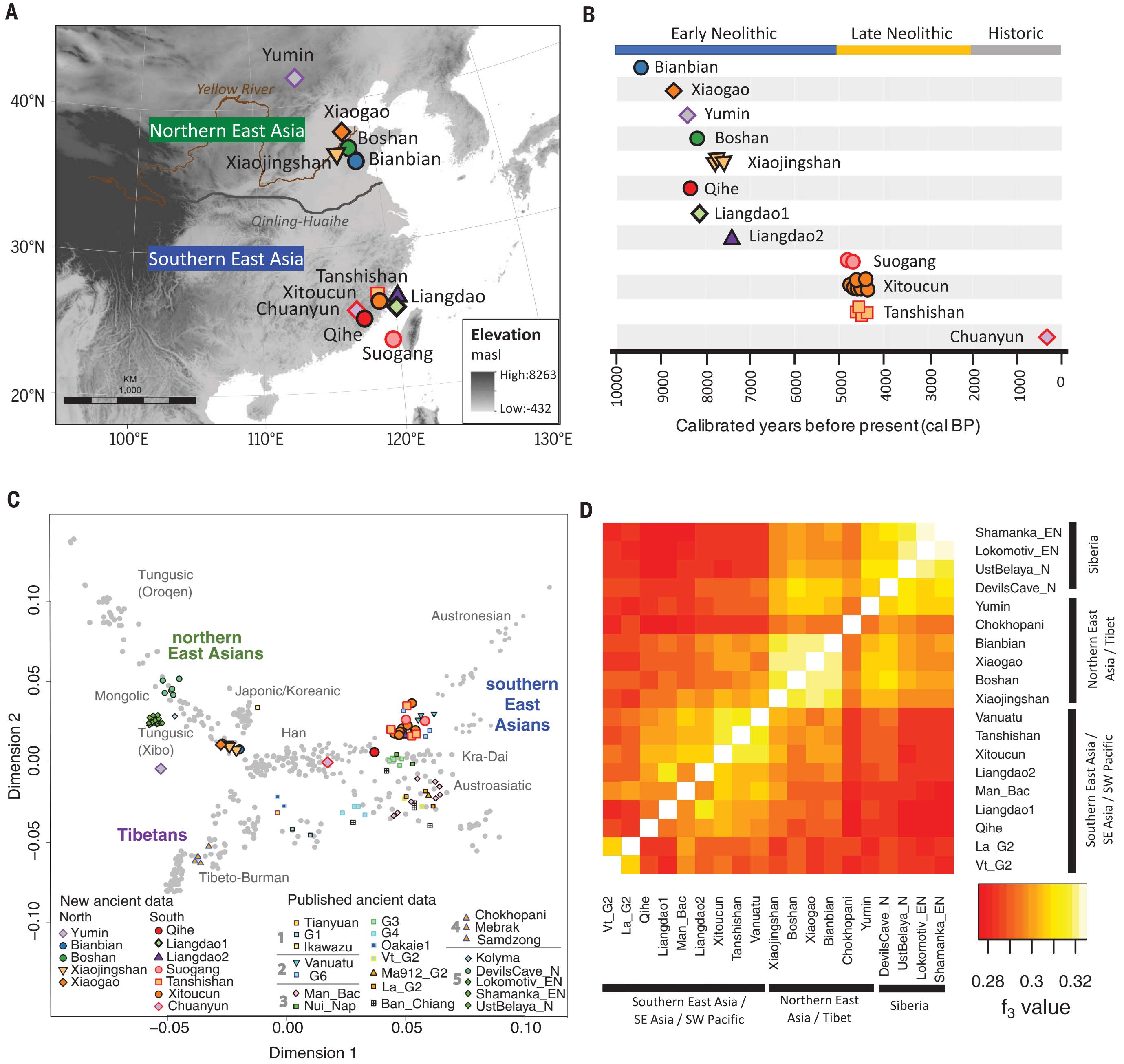

Fig. 1. Geographic, temporal, and genetic information for newly sampled East Asians. (**A**) Map showing the locations from which ancient individuals were sampled for this study. The Qinling-Huaihe line between the Huaihe River and the Qinling mountains is used to divide northern and southern East Asia. Gray shading indicates differences in elevation across the region. masl, meters above sea level. (**B**) Chronology of ancient individuals sampled in this study, where individuals date to 9500 to 7500 cal yr B.P. (Early Neolithic), 5000 to 4000 cal yr B.P. (Late Neolithic), or 300 cal yr B.P. (Historic). Individuals from northern East Asia all date to the Early Neolithic, whereas individuals from southern East Asia date to all three time periods. (**C**) Projection of ancient Asians onto a principal components 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 archaeological location and numbered in gray. 1 indicates Early Asians; 2 indicates ancient Austronesian-related islanders from Southeast Asia and the Southwest Pacific; 3 indicates Neolithic, Bronze Age, and historical Southeast Asians; 4 indicates Neolithic, Bronze Age, and historical Tibetans; and 5 indicates 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. SE Asia refers to Southeast Asia, and SW Pacific refers to Southwest Pacific.

first layer. Consistent results were obtained in an outgroup f_3 analysis, where Neolithic East Asians were found to share more genetic similarity with Neolithic Siberians, Tibetans, and Southwest Pacific Islanders ($f_3 = 0.28$ to 0.32) than with early Asians ($f_3 = 0.25$ to 0.26; fig. S2).

A direct comparison of Neolithic East Asians with 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. In other words, most f_4(*Mbuti*, *X*; *present-day East Asian*, *early Asian*) values are significantly <0 (Han: $-22.1 < Z < -4.4$; table S2) and most f_4(*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 also investigated patterns of archaic admixture in the Neolithic East Asians and found them to be similar to patterns observed in present-day East Asians (table S3). Neolithic East Asians 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 those seen in Tianyuan (*14*, *20*) but are similar to those in present-day East Asians (fig. S3). Previous ancient DNA studies have 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, which suggests that 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, whereas coastal Neolithic northern East Asians cluster together and close to present-day northern East Asians (Fig. 1C). These results suggest that population structure between northern and southern East Asia has existed since the Early Neolithic.

To determine genetic relationships in Neolithic East Asia, we first used a pairwise comparison of outgroup f_3 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 f_4 analysis (fig. S4A), Neolithic southern East Asians and the 3000-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*). Although some Late Neolithic southern East Asians share a connection to northern East Asians (fig. S4B), we suggest that this 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 to 100% of the bootstrap (bs) replicates. 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 that is not found in Early Neolithic southern East Asians but is present in Neolithic Siberians and Tibetans. For example, 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 to 99%; Fig. 2A and figs. S5 and S6), and they also share high genetic similarity with each other in an outgroup f_3 analysis (Fig. 1D). Particularly, coastal Neolithic northern East Asians clustered together, whereas 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

Table 1. Ancient individuals sampled in this study. Red highlighting indicates high contamination, and yellow highlighting indicates that the sample was excluded from demographic analysis because of a low number of SNPs. Lat, latitude; Lon, longitude; C→T%, percentage of C→T substitutions; Covg, coverage; DS, double-stranded; SS, single-stranded; SS UDG, single-stranded with uracil-DNA-glycosylase treatment.

ID	# of libraries	Sample ID	Population	Name†	Skeletal element	Date cal yr B.P.	Region	Lat	Lon	Library style	Sex	C→T%	Contamination		Covg$^{\#}$	SNPs
													Chr X	MtDNA		
Yumin	1	M1	inland nEastAsia_EN	Yumin	Teeth	8,415-8,335	Inner Mongolia	42	114.2	DS	F	17	-	0.028	7.25	876757
Bianbian	2	Bianbian	coastal nEastAsia_EN	Bianbian	Petrous	9,545-9,480	Shandong	36.1	118.5	SS,DS	M	11	0.01	0.014	2.21	701880
BS	11	BS11	coastal nEastAsia_EN	Boshan	Tibia	8,320-8,040$^{\&}$	Shandong	36.5	117.9	SS UDG	M	10	0.03	0.01	7.36	1045084
XJS1309_M7	3	M7	coastal nEastAsia_EN	Xiaojingshan	Petrous	7,872-7,721	Shandong	36.5	117.9	DS	F	31	-	0.018	0.22	191653
XJS1311_M16	3	M16	coastal nEastAsia_EN	Xiaojingshan	Petrous	7,935-7,786	Shandong	36.5	117	DS	F	24	-	0.029	0.64	387603
XJS1309_M4	4	M4	coastal nEastAsia_EN	Xiaojingshan	Petrous	7,877-7,735	Shandong	36.5	117	DS	M	21	-	0.019	0.58	386648
Xiaogao	4	M1	coastal nEastAsia_EN	Xiaogao	Petrous	8,777-8,591	Shandong	37.9	117.6	DS	F	30	-	0.018	7.6	948953
Qihe2	4	Qihe2	coastal sEastAsia_EN	Qihe	Petrous	8,428-8,359	Fujian	25.4	117.6	SS	F	34	-	0.13	0.45^	328913^
LD1	1	M1	island sEastAsia_EN	Liangdao1	Phalanx	8,320-8,060$^{\&}$	Liang Island	26.3	120.2	SS UDG	M	12	0.03	0.015	2.72	843051
LD2	1	M2	island sEastAsia_EN	Liangdao2	Phalanx	7,590-7,560$^{\&}$	Liang Island	26.3	120.2	SS UDG	F	19	-	0.005	1.68	744266
SuogangB1	7	B1	island sEastAsia_ LN	Suogang	Petrous	4,800-4,300$^{*\&}$	Penghu Island	23.5	119.6	SS	M	8	-0.02	0.12	0.04^	48386^
SuogangB3	7	B3	island sEastAsia_ LN	Suogang	Petrous	4,800-4,300$^{*\&}$	Penghu Island	23.5	119.6	SS	M	10	0.11	0.28	0.03^	30050^
L5705	1	M49	coastal sEastAsia_LN	Xitoucun	Petrous	4,419-4,246	Fujian	26.2	119.1	SS	F	33	-	0.026	0.66	434148
L5700	1	M32	coastal sEastAsia_LN	Xitoucun	Petrous	4,530-4,417	Fujian	26.2	119.1	SS	F	28	-	0.027	0.43	307394
L5692	1	M26	coastal sEastAsia_LN	Xitoucun	Petrous	4,530-4,417	Fujian	26.2	119.1	SS	M	24	-0.02	0.058	0.01^	16027^
L5706	1	M15	coastal sEastAsia_LN	Xitoucun	Petrous	4,527-4,406	Fujian	26.2	119.1	SS	M	25	0.02	0.043	0.04^	49141^
L5704	1	M13	coastal sEastAsia_LN	Xitoucun	Petrous	4,580-4,423	Fujian	26.2	119.1	SS	M	23	0.11	0.11	0.04^	47786^
L5703	1	M44	coastal sEastAsia_LN	Xitoucun	Petrous	4,644-4,500	Fujian	26.2	119.1	SS	M	22	0.12	0.074	0.02^	19114^
L5702	1	M46	coastal sEastAsia_LN	Xitoucun	Petrous	-	Fujian	26.2	119.1	SS	M	13	-	-	0.01	9407
L5701	1	M18 -2	coastal sEastAsia_LN	Xitoucun	Petrous	4,418-4,240	Fujian	26.2	119.1	SS	M	28	0.07	0.089	0.04^	47608^
L7415	1	M6	coastal sEastAsia_LN	Tanshishan	Petrous	4,419-4,246	Fujian	26.1	119.2	SS	M	22	0.03	0.03	0.41	338182
L7417	1	M20	coastal sEastAsia_LN	Tanshishan	Petrous	4,526-4,417	Fujian	26.1	119.2	SS	F	24	-	0.035	0.14^	144349^
L7416	1	M25	coastal sEastAsia_LN	Tanshishan	Petrous	-	Fujian	26.2	119.1	SS	M	11	-	-	0.01	8422
L5698	1	M26 -1	coastal sEastAsia_LN	Tanshishan	Petrous	4,410-4,225	Fujian	26.2	119.1	SS	F	25	-	0.06	0.02^	20920^
L5696	1	M12	coastal sEastAsia_LN	Tanshishan	Petrous	-	Fujian	26.2	119.1	SS	M	18	0.16	0.196	0.02^	20107^
L5694	1	Chuanyun	coastal sEastAsia_H	Chuanyun	Toe	334 281	Fujian	25.6	117.3	SS	M	11	0.01	0.005	0.38	322808

†Labels used in analyses, which is the same as the site name. If the same name, they were grouped together. #Based on a panel of 1.2 million SNPs (*14–16*). &Radiocarbon date published elsewhere. *The bone was not directly radiocarbon dated. For the Suogang site, shells were radiocarbon dated to 4633 to 4287 cal yr B.P. and 4793 to 4407 cal yr B.P. ^Alleles were determined using the damage-restricted library.

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 f_2, f_3, and f_4 statistics for populations included in the model [admixture graph; details provided in (*14*)], we similarly observe that northern and southern East Asian–related ancestry separates into two lineages (Fig. 2B and figs. S16 to S18).

Though the Neolithic is characterized by distinct northern and southern East Asian–related ancestries, gene flow also affected 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; 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 that are closely related to Native Americans (Paleosiberians; table S1) as has been previously observed (*9*) and which we confirmed in f_4 analyses showing Paleosiberian 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 with more-inland Neolithic populations in Asia (northern East Asian, Siberian, and Tibetan; figs. S7 and S8). Coastal

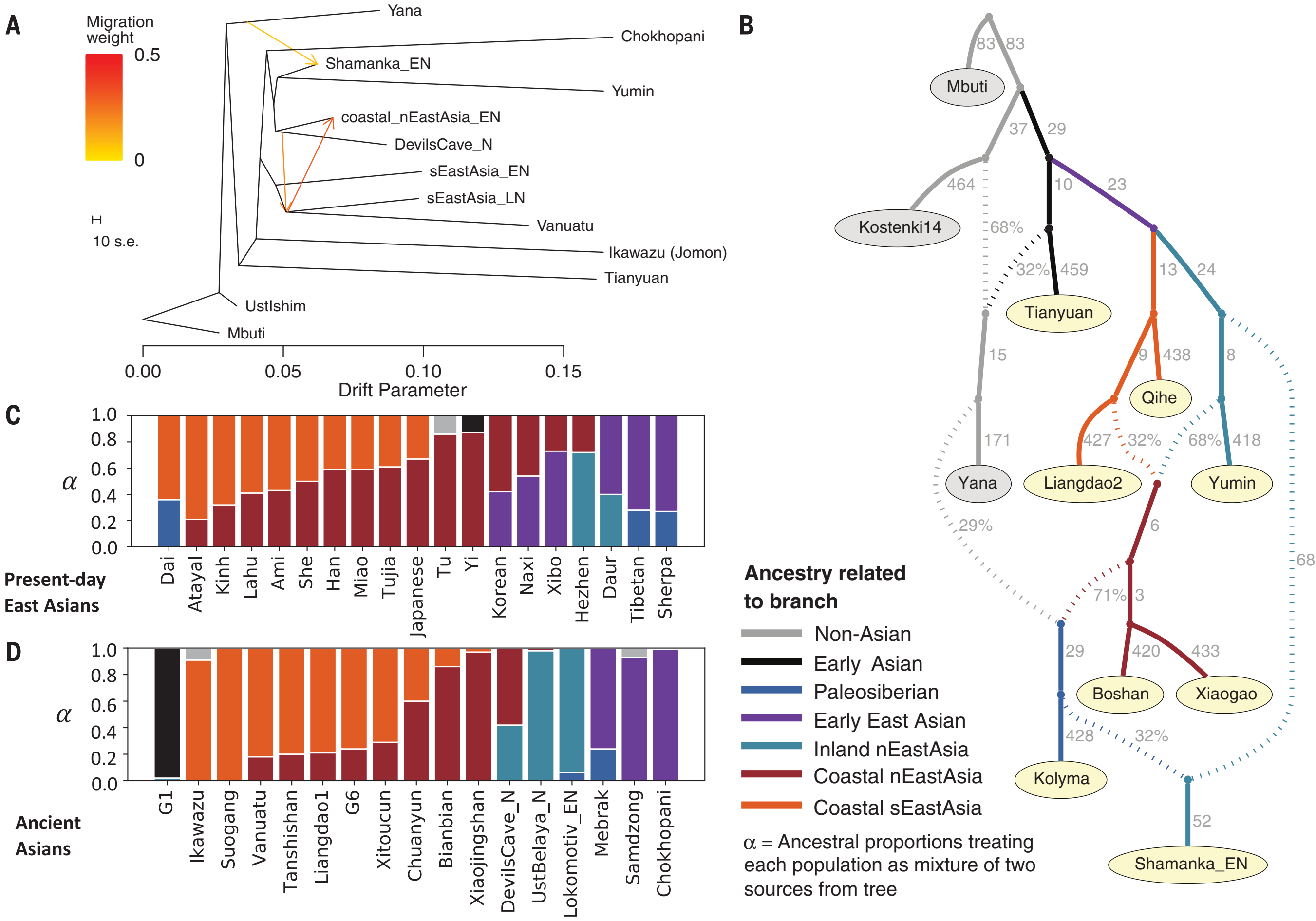

Fig. 2. Modeling tree relationships between ancient Asians. (**A**) A maximum-likelihood phylogeny allowing three migration events using the Treemix software (*23*). Aside from coastal nEastAsia_EN, sEastAsia_EN, and sEastAsia_LN (which refer to the labels in the Population columns in Table 1 or table S1 and group several ancient individuals from more than one site together), all other labels refer to the nomenclature used in the Name columns designated in Table 1 or table S1. s.e., average standard error of the entries in the sample covariance matrix. (**B**) An admixture graph that fits the data ($|\max Z| = 2.7$), where we include two Early Neolithic southern East Asians (sEastAsia: Qihe and Liangdao2), three Early Neolithic northern East Asians (inland and coastal nEastAsia: Yumin, Boshan, and Xiaogao), a Neolithic Siberian (Shamanka_EN), a Paleosiberian (Kolyma), an early Asian (Tianyuan), two individuals of non-Asian ancestry (Kostenki14 and 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 × 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. (**C** and **D**) The estimated mixture proportions observed when adding 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 with 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 (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.

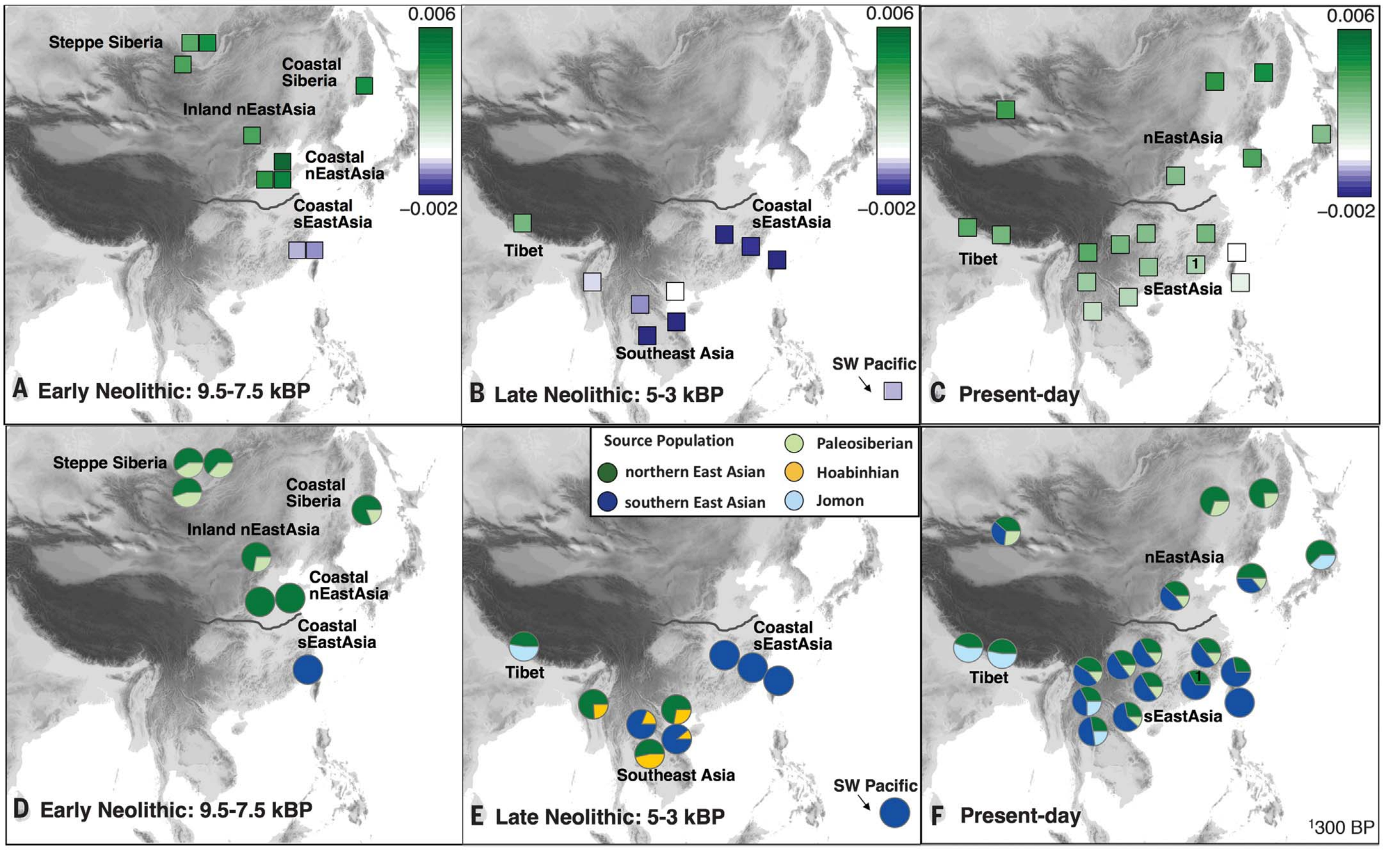

Fig. 3. Ancestry localized to the north and south in the Early Neolithic is found in admixed form across northern and southern East Asia today. (**A** to **C**) A heatmap showing f_4(*Mbuti*, *X*; *Qihe*, *Bianbian*), where Bianbian is the oldest northern East Asian sampled (~9500 cal yr B.P.) and Qihe is the oldest southern East Asian sampled (~8400 cal yr B.P.). *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. kBP, thousand calibrated years before the present. (**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), Hòabìnhian (orange), and Jōmon (light blue). Proportions were determined using *qpAdm* (*15*), 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. Labels on the map indicate the populations listed in table S1.

Early Neolithic northern East Asians show these connections to Early Neolithic southern East Asians as well (fig. S8B and fig. S9). These connections demonstrate 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 [fixation index (F_{ST})] shows that coastal Neolithic northern East Asians and coastal Neolithic southern East Asians are more highly differentiated from each other [F_{ST} = 0.042, standard error (SE) = 0.004] than are present-day northern and southern East Asians (F_{ST} = 0.023, SE = 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 those from the Neolithic, we used an f_4 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 f_4(*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 observed 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 pronounced change where all East Asians (including mainland southern East Asians) share an affinity with Neolithic northern East Asians, as shown by f_4(*Mbuti*, *present-day East Asian*; *Qihe*, *Bianbian*)>0 ($3 < Z < 8.8$; Fig. 3C and table S5A), rather than demonstrating affinities on the basis of location. Estimates of ancestry proportions in our mixture models show that this is because of increased northern East Asian–related ancestry in southern mainland East Asia, with estimates of northern East Asian–related ancestry ranging from 21 to 55% (Fig. 3F, fig. S13, and table S6B). Southern East Asian ancestry also extends north, as it is found in Han populations from northern China (36 to 41%; fig. S13 and table S6B) as well as in some northern East Asians (e.g., Xibo and Korean populations; 35 to 36%; Fig. 3F and table S6B). Paleosiberian-related ancestry from Siberia also greatly affects 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.

Although we are unable to estimate the timing of gene flow events, patterns in the f_4 comparisons suggest that population movement may have already started affecting East Asians by the Late Neolithic. In the test f_4(*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 further support this finding: The 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 f_4 test of symmetry, all Neolithic East Asians and coastal Neolithic Siberians share a closer relationship to Bianbian than to Yumin, whereas all inland Neolithic Siberians and Tibetans do not (fig. S15 and table S5D). 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, on the basis of geographic proximity across the Taiwan Strait and the occurrence of similar artifacts at archaeological sites from both regions (*11*, *25*, *26*). Furthermore, genetic modeling with mitochondrial DNA (mtDNA) genomes from present-day Southeast Asians (*27*) also suggests a southern Chinese origin. The mtDNA of the 8320- to 8060-year-old Liangdao1 individual belongs to haplogroup E1, which is common in Austronesian-speaking populations in Taiwan, the Philippines, and Indonesia today, and it 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 to 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., f_4(*Mbuti*, *Neolithic southern East Asians*; *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 3000-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 together with the ancient Vanuatu samples in a maximum-likelihood tree (bs = 82.3 to 100%; Fig. 2A and fig. S5) and share high genetic similarity with the Vanuatu in an outgroup f_3 analysis (Fig. 1D). Moreover, a direct f_4 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. Although most Neolithic Southeast Asians are a mixture of Hòabìnhian-related and southern East Asian–related ancestry, a 4000-year-old population from Vietnam shares an especially close relationship to coastal Late Neolithic southern East Asians (table S7B). Notably, these coastal connections extend further north. A 2700-year-old Jōmon individual from the Japanese archipelago that separated early from present-day East Asians was suggested to share genetic affinity to Hòabìnhians (*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 indicate that coastal regions were areas of interconnectivity and gene flow rather than of 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 Asians, it has been 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 (*13*, *30*). Although we did not find evidence of a first layer population in coastal southern East Asia by 8400 years ago, we did 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.

However, the spread of northern East Asian ancestry led 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. The fact 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 suggests 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 9500 to 300 cal yr B.P., we observe a close genetic relationship between East Asians, albeit with shifts in ancestry indicative of population movement and admixture during the Neolithic. Increased genetic sampling from the Paleolithic and from populations further inland in central China should help to further clarify the relationships among Paleolithic hunter-gatherers, Neolithic farmers, and present-day populations of East Asia.

REFERENCES AND NOTES

1. F. Zhang, B. Su, Y.-P. Zhang, L. Jin, *Phil. Trans. R. Soc. B* **362**, 987–996 (2007).
2. The HUGO Pan-Asian SNP Consortium, *Science* **326**, 1541–1545 (2009).
3. C. Tian *et al.*, *PLOS ONE* **3**, e3862 (2008).
4. J. Chen *et al.*, *Am. J. Hum. Genet.* **85**, 775–785 (2009).
5. A. Keinan, J. C. Mullikin, N. Patterson, D. Reich, *Nat. Genet.* **39**, 1251–1255 (2007).
6. M. Lipson *et al.*, *Science* **361**, 92–95 (2018).
7. H. McColl *et al.*, *Science* **361**, 88–92 (2018).
8. P. de Barros Damgaard *et al.*, *Science* **360**, eaar7711 (2018).
9. M. Sikora *et al.*, *Nature* **570**, 182–188 (2019).
10. P. Skoglund *et al.*, *Nature* **538**, 510–513 (2016).
11. L. Liu, X. Chen, *The Archaeology of China: From the Late Paleolithic to the Early Bronze Age* (Cambridge Univ. Press, 2012).
12. B. Sun, S. Cui, *Cultural Relics of Central China* **3**, 23–28 (2008).
13. H. Matsumura *et al.*, *Sci. Rep.* **9**, 1451 (2019).
14. See the supplementary materials for more details.
15. W. Haak *et al.*, *Nature* **522**, 207–211 (2015).
16. Q. Fu *et al.*, *Nature* **524**, 216–219 (2015).
17. A. W. Briggs *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 14616–14621 (2007).
18. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, *BMC Bioinformatics* **15**, 356 (2014).
19. Q. Fu *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 2223–2227 (2013).
20. M. A. Yang *et al.*, *Curr. Biol.* **27**, 3202–3208.e9 (2017).
21. C. Jeong *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 7485–7490 (2016).
22. M. Raghavan *et al.*, *Nature* **505**, 87–91 (2014).
23. J. K. Pickrell, J. K. Pritchard, *PLOS Genet.* **8**, e1002967 (2012).
24. W. Yan, *Wen Wu* **3**, 38–50 (1987).
25. T. Jiao, in *A Companion to Chinese Archaeology*, A. P. Underhill, Ed. (Wiley, 2013), pp. 599–611.
26. Fujian Museum, Longyan Municipal Bureau of Culture and Publishing, *Kaogu.* **5**, 7–19 (2013).
27. W. Kutanan *et al.*, *Eur. J. Hum. Genet.* **26**, 898–911 (2018).
28. A. M.-S. Ko *et al.*, *Am. J. Hum. Genet.* **94**, 426–436 (2014).
29. P. Bellwood, *Asian Perspect.* **26**, 107–117 (1984).
30. H. Matsumura *et al.*, *PLOS ONE* **13**, e0198689 (2018).
31. Y. Wang *et al.*, *Genomics Proteomics Bioinformatics* **15**, 14–18 (2017).
32. BIG Data Center Members, *Nucleic Acids Res.* **46**, D14–D20 (2018).

ACKNOWLEDGMENTS

We thank J. Kelso, M. Slatkin, and J. Gardner for comments, as well as S. Pääbo and his laboratory from the Max Planck Institute of Evolutionary Anthropology for their help in setting up the laboratory at the Institute of Vertebrate Paleontology and Paleoanthropology, CAS. We also thank our reviewers for providing insightful suggestions on the manuscript. **Funding:** This work was supported by the Chinese Academy of Sciences (XDB26000000), the National Natural Science Foundation of China (91731303, 41925009, and 41630102), CAS (XDA1905010 and QYZDB-SSW-DQC003), Zhengzhou University through the "Research on the roots of Chinese civilization" grant (XKZDJC202006), the Tencent Foundation through the XPLORER PRIZE, and the Howard Hughes Medical Institute (grant no. 55008731). **Author contributions:** Q.F. conceived the idea for the study. X.Fa., B.S., C.C., J.L., C.T., Y.-C.K., H.C., Q.B., X.W., A.M.-S.K., W.Z., Y.Z., C.S., M.Z., X.G., and Q.F. assembled archaeological materials and performed dating. Q.F., P.C., R.Y., F.L., X.Fe., Q.D., C.N., and Y.C. performed or supervised wet laboratory work. 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 supplementary materials. Y.-C.K., X.G., D.R., and M.S. helped revise the manuscript and supplementary materials. M.A.Y. 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 the Beijing Institute of Genomics (BIG) Data Center (*32*) (https://bigd.big.ac.cn/gsa-human; accession no. HRA000123). All software used is freely available online and is referenced in (*14*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/369/6501/282/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S18
Tables S1 to S9
References (*33*–*79*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

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

Science

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

Science **369** (6501), 282-288.
DOI: 10.1126/science.aba0909originally published online May 14, 2020

A genetic history of China
The history of human movements into and within China has been difficult to determine solely from archaeological investigations or genetic studies of contemporary peoples. Yang *et al.* sequenced DNA from 26 individuals from 9500 to 300 years ago from locations within China. Analyses of these individuals, along with previously sequenced ancient individuals and present-day genomes representing global populations, show a split between ancient humans in northern and southern China. Neolithic northern Chinese individuals are closest to modern-day East Asians, whereas ancient individuals from southern China are most closely related to modern-day Southeast Asians and show an affinity to modern-day Austronesian populations. These results indicate that there was a southward movement and admixture of peoples during the Neolithic that gave rise to modern-day populations in East Asia.

Science this issue p. 282

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.