



Research Article

Extensive ethnolinguistic diversity at the crossroads of North China and South Siberia reflects multiple sources of genetic diversity

Guang-Lin He^{1,2,3,4,5†*} , Meng-Ge Wang^{5,6,7†}, Xing Zou^{5,8}, Hui-Yuan Yeh⁴, Chang-Hui Liu⁶, Chao Liu^{6,7*}, Gang Chen^{9*}, and Chuan-Chao Wang^{1,2*} 

¹State Key Laboratory of Cellular Stress Biology, National Institute for Data Science in Health and Medicine, School of Life Sciences, Xiamen University, Xiamen 361005, China

²Department of Anthropology and Ethnology, Institute of Anthropology, School of Sociology and Anthropology, Xiamen University, Xiamen 361005, China

³Institute of Rare Diseases, West China Hospital of Sichuan University, Sichuan University, Chengdu 610044, China

⁴School of Humanities, Nanyang Technological University, Singapore 224050, Singapore

⁵Institute of Forensic Medicine, West China School of Basic Science and Forensic Medicine, Sichuan University, Chengdu 610041, China

⁶Guangzhou Forensic Science Institute, Guangzhou 510080, China

⁷Faculty of Forensic Medicine, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou 510080, China

⁸College of Medicine, Chongqing University, Chongqing 400016, China

⁹Hunan Key Laboratory of Bioinformatics, School of Computer Science and Engineering, Central South University, Changsha 410075, China

†These authors contributed equally to this work.

*Authors for correspondence. Guang-Lin He. E-mail: Guanglinhesu@163.com; Chao Liu. E-mail: liuchaogzf@163.com; Gang Chen. E-mail: chengangcs@gmail.com; Chuan-Chao Wang. E-mail: wang@xmu.edu.cn

Received 30 September 2021; Accepted 27 December 2021; Article first published online 9 January 2022

Abstract North China and South Siberia, populated by Altaic- and Sino-Tibetan-speaking populations, possess extensive ethnolinguistic diversity and serve as the crossroads for the initial peopling of America and western–eastern transcontinental communication. However, the population genetic structure and admixture history of northern East Asians remain poorly understood due to a lack of genome-wide data, especially for Mongolic-speaking people in China. We genotyped genome-wide single nucleotide polymorphisms for 510 individuals from 38 Mongolic, Tungusic, and Sinitic-speaking populations. We first explored the shared alleles and haplotypes within the studied groups. We then merged with 3508 published modern and ancient Eurasian individuals to reconstruct the deep evolutionary and natural selection history of northern East Asians. We identified genetic substructures within Altaic-speaking populations: Western Turkic people harbored more western Eurasian-related ancestry; Northern Mongolic people in Siberia and eastern Tungusic people in Amur River Basin (ARB) possessed a majority of Neolithic ARB related ancestry; Southern Mongolic people in China possessed apparent genetic influence from Neolithic Yellow River Basin (YRB) farmers. Additionally, we found the differentiated admixture history between western and eastern Mongolians and geographically close Northeast Hans: the former received a genetic impact from western Eurasians, and the latter retained the primary Neolithic YRB and ARB ancestry. Moreover, we demonstrated that Kalmyk people from the northern Caucasus Mountains possessed a strong genetic affinity with Neolithic Mongolian Plateau (MP) people, supporting the hypothesis of their eastern Eurasian origin and long-distance migration history. We also illuminated that historical pastoral empires in the MP contributed considerably to the gene pool of northern Mongolic people but rarely to the southern ones. We finally found natural selection signatures in Mongolians associated with alcohol metabolism. Our results demonstrated that the Neolithic ancestral sources from the MP or ARB played an important role in spreading Altaic populations and languages. The observed multisources of genetic diversity contributed significantly to the extensive ethnolinguistic diversity in northern East Asia.

Key words: demic diffusion, genetic admixture, genetic diversity, Northern East Asians, population substructure.

1 Introduction

North China and South Siberia (NCSS) have been inhabited by anatomically modern humans as early as 40 000 years ago (kya). They are now mainly peopled by Altaic-speaking groups and Sino-Tibetan-speaking Han Chinese (Yang et al., 2017). This region harbors massive ethnolinguistic, cultural, and archaeological diversity. With the advances in sequencing or array-based genotyping technologies and the innovation in analysis methods, many population genetic surveys have focused on the reconstruction of the deep population admixture of Eurasians, which provided a complex population mixed landscape of populations from Siberia and Mongolian Plateau (MP) (Yunusbayev et al., 2015; Bai et al., 2018; Tambets et al., 2018; Jeong et al., 2019). However, most previous genetic analyses have mainly focused on the Siberian Indo-European and Turkic-speaking people, especially those from the northern and central Siberia; therefore, dense sampling of modern groups from NCSS was needed to conduct further analysis.

Han Chinese are also the major group in this region, they have complex demographic history and the largest population size in the world. Genetic traces extracted from the ancient genomes supported Han Chinese being formed via the admixture of the dominant ancestry from Neolithic Yellow River Basin (YRB) farmers and geographically different indigenous peoples (Wang et al., 2021a). Zhang et al. (2019) recently analyzed the genome-wide single nucleotide variations (SNVs) of 114 783 Han Chinese (Han100K). They found six population subgroups: Northwest Han, Northeast Han, Central Han, Southwest Han, Southeast Han, and Lingnan Han (Zhang et al., 2019), which were recently evidenced via the ChinaMAP-data set (Cao et al., 2020). Other bioanthropological studies and medically derived cohort researches have also identified the admixture history between Han Chinese and their geographically close ethnolinguistically distinct indigenous groups (He et al., 2021a, 2021b; Li et al., 2021; Yao et al., 2021). However, the genetic origin and population structure of Northeast Han from Inner Mongolia, Jilin, Liaoning, and Heilongjiang remain uncharacterized. Although Northeast China is mainly populated by Mongolic and Tungusic people (Fig. S1), the recent northeastward migration of Han Chinese, such as historical migration events of Chuanguangdong and Zouxikou, as well as the establishment of Yuan and Ming dynasties, dramatically changed the genetic landscape of this region.

The MP and surrounding regions have played a crucial role in the formation of ancient eastern Eurasians at this crossroads. Population migrations between southern East Asia and north Siberia have been comprehensively characterized via ancient DNA studies. Archaeological and genetic evidence from the Tianyuan Cave (40 000-year-old Tianyuan) (Yang et al., 2017) and Siberian Paleolithic people (Yana, AG2-3, and Mal'ta) (Raghavan et al., 2014; Fu et al., 2016; Sikora et al., 2019) have suggested that at least two Paleolithic eastern Eurasian lineages widely existed in East Asia and Siberia. These two lineages have participated in the formation of Northeast Eurasians and Americans in the crossroads of the Baikal Lake and ARB regions. Furthermore, in the Neolithic period, there were at least five Neolithic genetic lineages discovered in East Asia, including inland/

coastal southern/northern East Asian lineages in Yangtze River and YRB, and the ancient northeastern East Asian lineages related to Neolithic DevilsGate and Boisman in MP and ARB (Jeong et al., 2016; Yang et al., 2020; Mao et al., 2021; Wang et al., 2021b). Neolithic Trans-Eurasian migration and subsequent admixture or transformation have played a pivotal and direct role in the formation of ancestral Eurasians (especially in Siberia) and also contributed substantially to the observed modern mosaic genetic structure (Mathieson et al., 2015; Damgaard et al., 2018; Jeong et al., 2020). A more dynamic population history has been observed in the eastern steppe during the Iron and historical periods. The directions of population dispersals had gradually changed from the previously dominant eastward to radial expansion from multiple centers (such as the Scythian federation) and were finally dominated by the westward migrations of historical pastoral tribes to Europe (Mathieson et al., 2015; Jeong et al., 2018). Historical empires of nomadic pastoralists combined and formed many elite dominance federation groups, such as Turkic, Xiongnu, Mongolian, and Tungusic. Population expansion centers have been shifted from West Eurasia to East Eurasia (Jeong et al., 2020; Wang et al., 2021a). The basic patterns of the genetic background of modern populations were gradually formed, with the eastern Eurasian-related genetic diversity replaced or mixed with the Proto-Indo-European's gene pool along with their languages (Damgaard et al., 2018; Sikora et al., 2019). Although the landscapes of three genetic clines in modern Siberians have been characterized, their genetic relationship with Chinese Altaic populations and the potentially existing differentiated population evolutionary history need to be comprehensively clarified and described.

The correlation between the genetic structure and dispersal of the Altaic language is another controversial topic that needs to be illuminated via more genetic data from geographically/ethnolinguistically different populations. The language/farming co-dispersal hypothesis has been evidenced in the spread of Sino-Tibetan and Austronesian language families in southern China and Southeast Asia (Wang et al., 2021a). Linguistic and archaeologic evidence consistently supported that the dissemination of Indo-European languages was mediated by the steppe pastoralists' expansion (Lazaridis et al., 2014; Narasimhan et al., 2019). Recent findings have revealed the influence of the co-spread of this language and the corresponding ancient Indo-European people extending to 2200-year-old Iron Age Shirezigou people in the northeastern Xinjiang (Ning et al., 2019). Patterns of genetic structure and the distribution of Indo-European languages were also influenced and characterized by large-scale eastward expansion and replacements of local populations (Sikora et al., 2019). Genomic evidence has found that these population migrations had no significant influence on the population demographic dynamics in Baikal Lake and ARB regions and North China (Zhang et al., 2019). After the Bronze Age migrations, population interaction between western and eastern Eurasian steppe had emerged among highly structured Scythian groups with possible Turkic languages, whose gene pool consisted of western sources (European farmer and Late Bronze Age herder), eastern source of southern Siberian hunter-gatherer, and Anatolian/Iran

farmer-related ancestry (Allentoft et al., 2015). Further westward dispersal of Xiongnu and Hun Khanates, and other historical migration events (expansion of Mongolian Empire) have promoted the domination of Altaic-speaking populations (Turkic, Mongolic, and Tungusic) in the Eurasian steppe and the replacement of previous Indo-Iranian-speaking groups (Wusun and Kangju) (Allentoft et al., 2015). However, the complex interaction between YRB farmers and Siberian hunter-gatherers and the extent of genetic contribution from these ancient populations to modern northern East Asians who mediated the dispersal of modern Altaic-speaking populations need to be further investigated. The geographical origins of Altaic-speaking populations have long been controversial based on the evidence from genetics and linguistics, including Pastoralist or Farming hypotheses that respectively supported that their originated centers were from the Altai Mountains, West Liao River, Baikal Lake, or ARB (Wang et al., 2021a), which needs to be tested via dense sampling of modern Altaic people and comprehensive comparison with spatiotemporally ancient sources.

Therefore, to comprehensively reconstruct the deep population history of Tungusic/Mongolic-speaking populations and their neighbors, we genotyped 510 modern individuals from Mongolic, Tungusic, and Sinitic-speaking populations using high-density single nucleotide polymorphisms (SNP) arrays and combined the primary data set with publicly available worldwide modern and ancient genome-wide data. This work mainly aimed to explore (i) how many ancestral sources contributed to modern Altaic-speaking populations; (ii) what was the association between the identified ancestries and the origin and mixture of modern Altaic macro-family; as well as (iii) what were the detailed admixture and natural selection histories of subgroups of Altaic-speaking populations; and finally (iv) to illuminate the genetic contribution from the historical pastoral empires from the MP to the southernmost Yunnan and Guangzhou Mongolians and westernmost Kalmyk people.

2 Material and Methods

2.1 Ethics statement

The Medical Ethics Committee of Xiamen University approved this study (Approval Number: XDYX2019009). Informed consent was obtained from all included individuals before the saliva or blood collection. The procedures of sample collection and experiment of this research were conducted by the recommendations provided by the revised Helsinki Declaration of 2000 (Association, 2001). All subjects were required to be self-declared indigenous ethnic groups.

2.2 Sample collection and genotyping

The reported data set consisted of 510 unrelated subjects from 31 Mongolic-speaking populations (Mongolian: 362, HulunBuir Daur: 9, and Tsitsihar Daur: 10), two Tungusic-speaking populations (Heihe Evenki: 8 and Jiamusi Hezhen: 2), and five Sinitic-speaking Han Chinese (119, Fig. 1A). PureLink Genomic DNA Mini Kit (Thermo Fisher Scientific) was used to isolate genomic DNA from saliva or anticoagulant-treated peripheral blood samples following the manufacturer's recommendations. Quantifiler Human

DNA Quantification Kit (Thermo Fisher Scientific) and Applied Biosystem 7500 Real-time PCR System (Thermo Fisher Scientific) were used to quantify the DNA concentration. Genotyping of 529 790 variants, including 18 711 parentally lineage informative single nucleotide polymorphisms (LISNPs), 4448 maternally LISNPs, and 506 616 autosomal and X chromosome-related SNPs, was carried out using the Affymetrix WeGene V1 Arrays.

2.3 Data purification

We used PLINK v1.90b6.13 (Chang et al., 2015) to perform quality control (QC) with the parameters of genotyping success more than 95% and the minor allele frequency more than 0.05%. We used PLINK and King to identify the close genetic relatives and ran Genome-wide Complex Trait Analysis (GCTA) v1.92.2 (Yang et al., 2011) to remove outliers. After quality filtering, we kept 478 individuals in the following population genetic analysis.

2.4 Reference population data sets

We merged our data with publicly available modern and ancient Eurasians genotyped via the 1240K-capture sequence platform, Human Origin (HO) arrays, and other recent public data into three data sets: Affymetrix data set, the merged HO data set, and the merged 1240K data set (Fig. S2). The reference populations in the Affymetrix data set consisted of two Hainan populations (Han and Hlai) (He et al., 2020), 14 Tai-Kadai-speaking populations (three Jing, four Zhuang, and seven Sui groups), and seven Hmong-Mien-speaking Miao populations from Guangxi (Huang et al., 2020), which were genotyped using the same arrays as in our study. The reference groups in the merged HO data set included 5081 individuals in the 1240K database and 7744 individuals from the 1240K + HO database obtained from the Allen Ancient DNA Resource (AADR) (Jeong et al., 2016; Yang et al., 2017; Lipson et al., 2018; McColl et al., 2018; Sikora et al., 2019; Changmai et al., 2021). These data sets included all publicly available ancient DNA data published before 2021 and 2054 individuals from the 1000 Genomes Project, 300 individuals from 142 worldwide populations from Simons Genome Diversity Project, and 928 sequenced genomes in the human genome diversity project (HGDP). Genotype data obtained using the Affymetrix HO arrays were included here (The International HapMap et al., 2007; Li et al., 2008; Huang et al., 2018). The reference populations in the merged 1240K data set included all publicly available ancient genomes and the sequenced genomes in the HGDP project (Bergstrom et al., 2020).

2.5 Principal component analysis (PCA) and model-based ADMIXTURE analysis

We used PLINK v1.9 (Chang et al., 2015) and the *smartpca* package in EIGENSOFT software (Patterson et al., 2006) to conduct Eurasian and regional PCA in the context of modern and ancient Eurasians or eastern Eurasians with two additional parameters (Project: YES and numoutlieriter: 0). Ancient individuals were projected onto the two-dimensional scaling plots. To characterize the individual and population ancestry compositions and reconstruct population genetic history, we used the genetic clustering algorithm implemented in ADMIXTURE to conduct the model-based

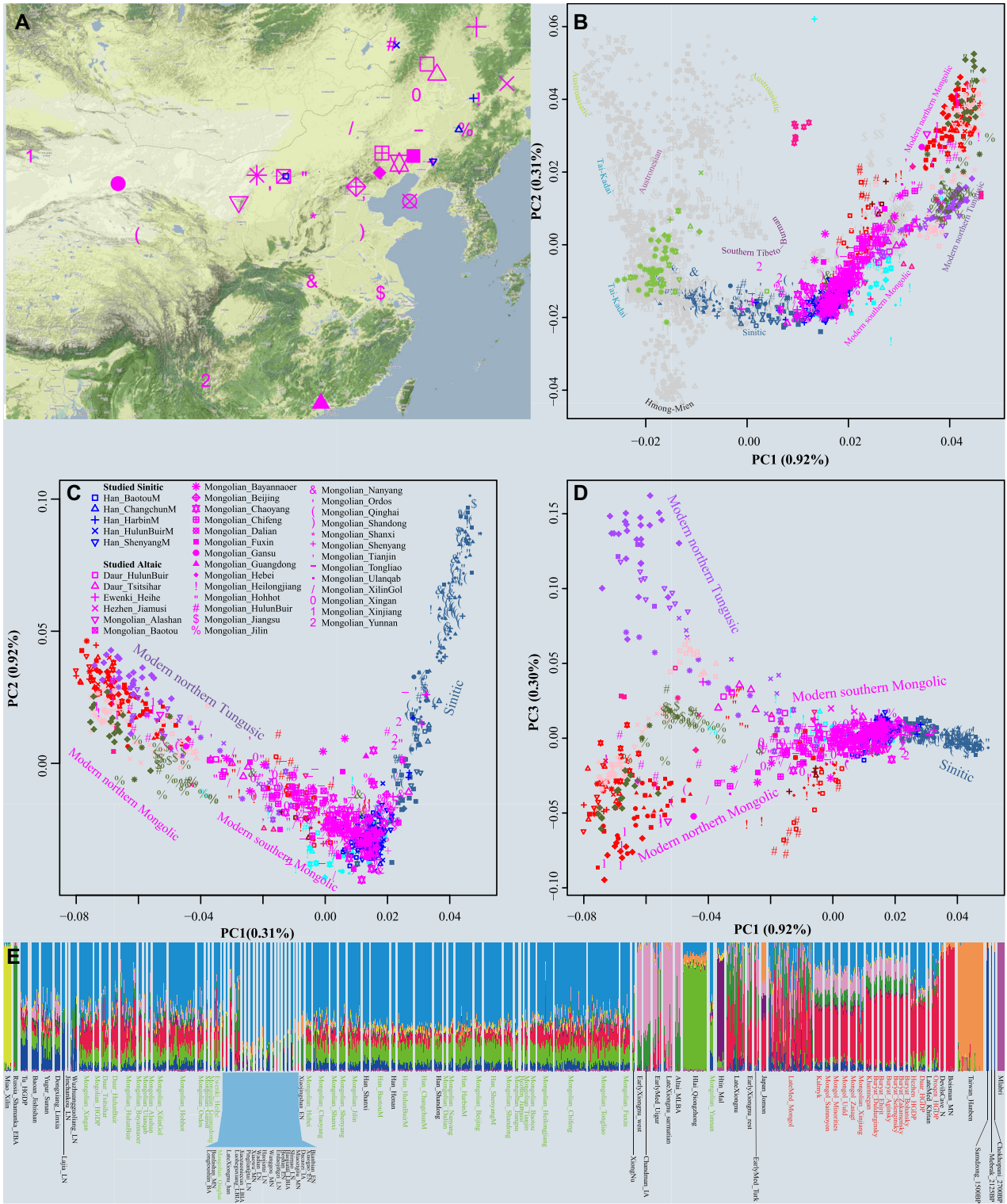


Fig. 1. Geographic position and genetic structure of studied Hans, Mongolic, and Tungusic people, as well as other modern and ancient East Asians based on the merged Human Origin (HO) data set (113 962 SNPs). **A**, Map of the newly sampled 40 Chinese populations. The topographic map of East Asia is drawn with the *ggmap* package in R software. **B**, Population genetic relationship patterns in the PCA analyses focus on all eastern Eurasian populations based on the top two components. Modern populations are colored-coded based on language categories, and the complete labels are presented in Fig. S5. Ancient samples were projected. **C**, **D**, Northern East Asian PCA analysis based on the top three components. **E**, ADMIXTURE analysis results for $K = 10$, which has the minimized cross-validation (CV) error and the complete results are presented in Fig. S6.

clustering analyses (Alexander et al., 2009). The analysis was performed based on the Eurasian data set and other subdata set comprising regional East Asians. Plink v1.9 was used to exclude linked-SNPs with the following parameter settings: pairwise SNP loci $r^2 > 0.4$; window: 200 SNPs; and sliding windows: 25 (–indep-pairwise 200 25 0.4). A total of 199 753 variants out of 346 634 passed filters and QC and remained in the merged 1240K data set. The number 90 965 out of 120 894 variants passed filters and QC in the merged HO. We assumed the number of predefined ancestry populations (K values) ranging from 2 to 20 and ran unsupervised ADMIXTURE with 10-fold cross-validation (CV) (–cv = 10). In total, 100 randomly seeded runs were used, and the log-likelihood scores and CV errors were used to find the most appropriate K values. Ancestry compositions between modern and ancient people were conducted using the settings above.

2.6 Three-population test (admixture- f_3 -statistics and outgroup- f_3 -statistics)

All possible source pairs in three data sets were used to conduct the admixture- f_3 (Source1, Source2; Studied populations) using the *qp3Pop* (Patterson et al., 2012). Formal admixture tests with statistically negative f_3 -statistic values ($Z \leq -3$) indicated that target populations were an admixture of two predefined ancestral populations. Genetic affinity was estimated using outgroup- f_3 (Reference, Studied populations; Mbuti).

2.7 Four-population test (f_4 -statistics)

We used the formal test of f_4 -statistics (Patterson et al., 2012) of the form $f_4(W, X; Y, \text{outgroup})$ to test whether our studied populations harbored more shared alleles with targeted reference populations than with other reference populations, for which W represented the Asian populations and X represented worldwide reference populations, and Y represented our studied populations. Generally, the statistical index was calculated as the differences between counts of BABA sites and counts of ABBA sites divided by the sum of counts of BABA sites and counts of ABBA sites. The null hypothesis was that W and X form a clade and descend from a homogeneous ancestral population separated from Y and outgroups. Therefore, there were no differences in the rate of allele sharing existing with Y , and the f_4 -statistic should be expected to be 0. A negative f_4 value will be produced if there is an excess of shared alleles between either or both pairs ($W, \text{outgroup}$) or (X, Y) since separation from the others. Similarly, a positive f_4 value can be obtained if a large number of shared alleles or gene flow exist between ($X, \text{outgroup}$) or (W, Y) since their separation from the others. Standard errors were estimated using the weighted block jackknife with the block size of 5 Mb in a single Hotelling's t^2 test. Z -scores were computed as the ratio between f_4 and standard error. $|Z| > 3$ was regarded as significantly different from 0, and the null hypothesis could be rejected.

2.8 TreeMix and qpGraph

We used the TreeMix v1.12 (Pickrell & Pritchard, 2012) to explore the topology of included populations with gene flow events ranging from 0 to 10, and we used Yoruba as the root.

We used ADMIXTOOLS2 to explore the best *qpGraph*-based models with different population compositions.

2.9 Mitochondrial and Y-chromosomal haplogroup assignment

In total, 18 435 phylogenetic informative Y-chromosomal SNPs and 4418 mitochondrial SNPs were used to classify the haplogroups using an in-house script based on the basal rules of the International Society of Genetic Genealogy (ISOGG; <http://www.isogg.org/>) and mtDNA tree Build 17 (<http://www.phylotree.org/>).

2.10 Pairwise qpWave and qpAdm-based admixture models

We merged geographically and ethnically defined populations into new genetically homogeneous populations (Mongolian_East, Daur, Mongolian_West, Ewenki_Heihe and Han_Northeast). We then performed *qpWave* analyses (Patterson et al., 2012) to explore the distribution of the f_4 -matrix via rank tests. Nine continental-representative populations were used as outgroups (Mbuti, Ust_Ishim, Kostenki14, Papuan, Australian, Mixe, MA1, Onge, and Atayal). We also conducted pairwise *qpWave* analysis focusing on the new-studied populations and previous sequenced HGDP Mongolic and Tungusic populations (Daur, Hezhen, Tu, Yakut, Mongolian, and Oroqen) as the targeted populations to explore their genetic homogeneity and conducted the *qpAdm* analysis to estimate the admixture proportion.

2.11 Identical by descent (IBD) segment inference and runs of homozygosity (ROH)

The different data sets were phased using ShapeIT. Then the Refined IBD (Browning & Browning, 2011) was used to characterize the IBD. Population-level shared IBD was corrected via the product of the sample size of the focused population pairs. We used PLINK v1.9 (Chang et al., 2015) to estimate the pairwise ROH. Effective population size changes were further estimated using the IBDNe (Browning & Browning, 2013).

2.12 Estimating admixture times with ALDER

We estimated admixture times using ALDER (Loh et al., 2013) with the default or two additional parameters (mindis: 0.005 and jackknife: YES). Admixture-mediated linkage disequilibrium can be decayed with the recombination that occurred after the initial admixture. We only captured the simple characteristics of the admixture process with the assumption of a single pulse-like admixture. We used ancestral sources from southern East Asians, eastern Siberians, and western Eurasians as the possible ancestral proxy sources.

2.13 Chromosome painting and fineSTRUCTURE

To infer finer scale scenarios of the admixture landscape, we used ChromoPainterv2 to paint the targeted chromosome using all other included donor chromosomes based on the phased haplotypes (Lawson et al., 2012). Chromosome painting could identify haplotypic distribution to further perform admixture dating and population identification. We conducted chromosome painting using the merged 1240K data set and Altaic people in the HO and Affymetrix data sets. Model-based Bayesian clustering instrumented in

fineSTRUCTURE was used to identify population substructure. FineSTRUCTURE v4 and fineSTRUCTURE R scripts (Lawson et al., 2012) based on the reconstructed co-ancestry matrix were used to dissect the fine-scale population structure via heatmap, clustering dendrogram, and PCA.

2.14 GLOBETROTTER admixture modeling

The R program of GLOBETROTTER (Hellenthal et al., 2014) was used to identify and date admixture events based on the merged 1240K data set, including ~366K genome-wide SNPs. Both complete and regional analyses were conducted based on the estimated copying vectors from ChromoPainterv2. GLOBETROTTER used the sampled surrogated populations to paint the haplotypic mixture of the truth unsampled source populations and then to identify whether the targeted Altaic populations had descended from any north-to-south or west-to-east admixture events (uncertain, no admixture, single complex admixture models), and – if so – it could calculate precisely when the admixture events occurred and the admixing sources involved.

2.15 Nature selection signatures identifying

We used selscan and Rehh to calculate the integrated haplotype score (iHS) and cross-population extended haplotype homozygosity (XP-EHH) as the selection indexes (Gautier et al., 2017). We then annotated the identified candidate loci using the PLINK. We highlighted the top selection SNPs using Manhattan. We further conducted the enrichment analysis using Metascape (Zhou et al., 2019).

3 Results

3.1 General population structure and genetic affiliations of northern East Asians

We successfully genotyped ~500K genome-wide SNPs from 510 individuals from 40 Mongolian (362), Daur (19), Ewenki (8), Hezhen (2), and Han (119) populations (Fig. 1A). We merged our data with modern and ancient East Asians included in the HO data set (110K). To explore the general patterns among studied populations and between other neighboring populations, we conducted PCA analysis among 3508 individuals. We found that our studied populations clustered with eastern Eurasian forming a north-to-south genetic cline (Figs. S3, S4). We also observed an apparent separation between northern and southern East Asians along PC1, in which northern East Asians consisted of the Altaic-speaking populations (Mongolic, Tungusic, and others; Figs. 1B, S5). Studied Mongolic and Tungusic people and previously published Chinese Altaic people from NCSS showed a separated localization with other northern Mongolic and Tungusic speakers from southern Siberia, Outer Mongolia and the ARB. Historic Xiongnu and Uigur were separated from all included modern Tungusic and Mongolic people along PC2, but other Neolithic Siberians overlapped with modern Tungusic and Mongolic people. This observed deviation from Bronze Age/historical Altai Mountain populations suggested that the primary ancestry of the contemporary Tungusic and Mongolic people originated from the eastern ancestral source from the MP or ARB. As

expected, our studied Mongolic and Tungusic populations were not only separated from southern and central Han Chinese (Fig. 1C) but also slightly separated from geographically close Northeast Hans. Neolithic YRB people were clustered closely with Hans compared with Mongolians. The cluster patterns inferred from the PCA showed more interesting population substructures among modern populations from NCSS, consistent with the categories of the subbranches of Altaic languages (Figs. 1C, 1D). PC1 separated Sinitic speakers from Mongolic and Tungusic people, and PC3 separated Tungusic-speaking populations from Mongolic people.

To further dissect the ancestry composition using model-based clustering analysis, we conducted ADMIXTURE with 10 predefined ancestral sources among eastern Eurasians (Figs. 1E, S6). We identified five homogeneous ancestry components in north groups (Tibetan [dark-blue], Han [light-blue], Neolithic ARB [red], and MP people [green], western Eurasian [light-pink]) and five components in southern ones (Austronesian [orange], Austroasiatic [purple and pink], Tai-Kadai [light-green], Hmong-Mien [yellow]). We found that most of the Mongolic and Tungusic people were shown as the mosaic form as a mixture of at least two ancestry components, in which the primary ancestry related to modern Han and Neolithic YRB farmers (light-blue) and the minor ancestral sources derived from sources related to Hlai, Tibetan and Neolithic ARB people (Fig. S6), suggesting that different ancestral sources contributed to the mixed gene pool of the studied populations. Mongolic-speaking Buryat in Russia and Mongolians in Mongolia harbored a marked amount of ancestries related to the YRB farmers and ARB ancients (Table S1). They also had ancestries related to western Eurasians and Shamanka_EBA, providing clues of the effect of western Eurasians and indigenous Baikal ancients on the genetic structure of northern Mongolic populations. Unlike northern Mongolic and Tungusic people, our newly studied Mongolians and Hans possessed higher YRB farmer-related ancestry. The identified ancestral components illuminated the differentiated demographic history between northern and southern Altaic people.

3.2 The differentiated demographic history between western and eastern Mongolians

From the PCA results in Fig. 1, we observed the western Eurasian affinity in Mongolians from Xinjiang, Qinghai, Gansu, and western Inner Mongolia. Ancestry composition inferred from ADMIXTURE modeling showed that Xinjiang Mongolians had a proportion of 18.6% of western Eurasian ancestry, 34.3% of Neolithic ARB ancestry, 28.3% of Neolithic YRB ancestry and some Neolithic Mongolian (7.0%) and Tibetan (4.6%) ancestry. Therefore, to further characterize the genetic history of Mongolians based on the higher-density variations (the 466K Affymetrix array), we combined our data with the previously published data set from Hainan Hans and Hlais (as the representatives of southern Chinese populations) to explore their genetic structure. Frequency-based PCA showed genetic differentiation between northern and southern East Asians (Fig. S7A), and haplotype-based PCA further showed a clear separation between Xinjiang Mongolians and other eastern Mongolians (Figs. S7B, S7D). More genetically homogeneous subgroups could also be

identified via the heatmap of pairwise coincidence and clustering dendrogram based on individual-level and population-level shared chunk counts (Figs. S7E, S7F). Southernmost Hans formed the same cluster with geographically close Hainan indigenous Hlais, and new-studied populations were separated into two major clusters and more subclusters. One cluster consisted of most of the Mongolic and Tungusic people, and the other comprised northern Hans and some Mongolians who possessed more modern southern East Asian ancestry. The identified longer IBD fragments, larger outgroup- f_3 values and smaller F_{ST} values also showed that the geographically close populations shared stronger genetic affinity (Figs. S7G, S7H; Table S2). However, we observed that Hainan people had the shortest shared IBD fragments and the largest F_{ST} values with our newly studied northern East Asians and different modeling-ancestry composition (Fig. S6I), suggesting that Mongolians had a distinct relationship with two southernmost Chinese representative populations. The identified shared IBD within geographically defined northern Mongolians was shorter than the identified patterns in the southern indigenous Hlais (137.456). These identified phenomena may be caused by more recent admixture events in isolated Hainan populations, supported by the observed longer ROH in Hlais. A larger effective population size could also produce relatively shorter IBD chunks. Therefore, we next estimated the effective population size changes among Mongolians, Hans, and Daur, and we observed recent population expansion within all three focused groups (Fig. S7).

To assess the detailed admixture landscape of Mongolians with more reference populations, we next conducted pairwise *qpWave* analysis and found multiple differentiated allele sharing between the targeted western Mongolians and reference populations, which suggested that western Mongolians did not form a clade with eastern Mongolians (Fig. 2A). The different patterns of genetic affinity between western and eastern Mongolians were further confirmed via the different affinities observed in outgroup- f_3 (Eurasians, eastern/western Mongolians; Mbuti) and the observed diverse admixture signatures in the admixture- f_3 (Source1, Source2; western/eastern Mongolians). Rank results of p_rank1 (0.0004) and p_rank2 (0.9567) showed that at least two admixture events were needed to explain the admixture landscape of western Mongolians. When we used Hezhen, Tu, Oroqen, Daur, and eastern Mongolians as the targeted populations, we found that they formed a clade with each other, but not with western Mongolians. Indeed, PCA analysis showed that eastern Mongolians overlapped with Daur, Ewenki, and some northern Hans, but western Mongolians were separated from them (Fig. 2B). Eastern Mongolians overlapped with the ancient YRB populations and Neolithic Siberian people, suggesting their genetic affinity. The observed significantly negative values in f_4 (Eurasians, eastern Mongolians; YRB farmers, Mbuti) and f_4 (MP/ARB hunter-gatherers, YRB farmers; eastern Mongolians, Mbuti) further showed that eastern Mongolians shared more alleles with Neolithic millet farmers (Tables S3, S4). We further explored the topologies among Chinese ancients, East Asians in HGDP, Tai-Kadai-speaking Sui and Zhuang, and Hmong-Mien-speaking Miao people using *TreeMix*-based phylogeny. We found that western Mongolians clustered

with Uyghur people, but eastern Mongolians clustered closely with southern East Asians and YRB ancient populations (Fig. 2C). *ADMIXTURE* results also revealed that western Mongolians harbored more western Eurasian ancestry related to Srubnaya, Sintashta, and Srubnaya. In addition, similar to the Yakut people, Western Mongolians also had more ancestry related to the Neolithic MP people. Eastern Mongolians carried more ancestry related to Neolithic YRB farmers and Tai-Kadai people (Sui and Hlai). Compared with northern Hans and geographically close northern East Asian ancient people, all Mongolians had more western Eurasian orange ancestry (Fig. 2D). Compared with historical Mongols, western Mongolians shared more alleles with ancient YRB farmers and modern East Asians related to Sino-Tibetan, Austronesian, Austroasiatic, and Tai-Kadai people, as the observed negative values in f_4 (Late_Mongol, western Mongolians; Eurasians, Mbuti). Differentiated sharing alleles between western and eastern Mongolians were also evidenced via the observed significant asymmetrical- f_4 values in f_4 (western Mongolians, eastern Mongolians; Eurasians, Mbuti).

Considering the complex admixture sources of northern East Asians identified in admixture- f_3 -statistics, *ALDER* and *GLOBETROTTER* (Table S5), we conducted *qpAdm* analysis to model their admixture processes via the three-way admixture models with ancestry from East Asians, Northeast Asians, and western Eurasians. As expected, western Mongolians were modeled as an admixture of more ancestry related to Alan, Saka, or Historical SaiduSharif (Fig. 2E). The remaining ancestry was related to Neolithic MP/ARB/YRB people or ancient Guangxi people. We also evaluated the ancestry composition of geographically separated populations. We found that the three-way admixture models with those above three ancestral sources could be well fitted for most studied populations with variable ancestry proportions (Fig. 3A). We reconstructed the deep demographic history using the *qpGraph*-based phylogenetic framework (Fig. 3B). We found that western Mongolians were fitted via 11% gene flow from steppe Sintashta pastoralists and the remained ancestry (89%) from eastern Eurasian ancients related to Neolithic millet farmers (Fig. 3B). The admixture modeling graph showed that the formation of eastern Mongolians involved three ancestral sources (Figs. 3C, 3D): western Eurasian (0.090), southern East Asian Hanben (0.291), and upper YRB farmers (0.180). Our results suggested that modern Chinese Mongolic and some of the Tungusic people were formed via the massive population movement and interaction of three ancestral sources related to East Asians, southern Siberians, and western Eurasians.

3.3 Genetic history of Northeast Han Chinese

To illuminate the genetic formation of Northeast Hans and explore how they interacted with adjoining Altaic-speaking people, we collected 119 Han Chinese individuals from Baotou (25), Changchun (24), Harbin (24), HulunBuir (21), and Shenyang (25; Fig. 1A). General patterns inferred from PCA showed that Northeast Hans was localized between Henan Hans and Mongolians and far away from indigenous Tungusic Ulchis and Nanais and shared a similar mixed genetic landscape with other northern Hans, who harbored major YRB ancestry (Figs. 1B–1D). No significant f_4 -values in

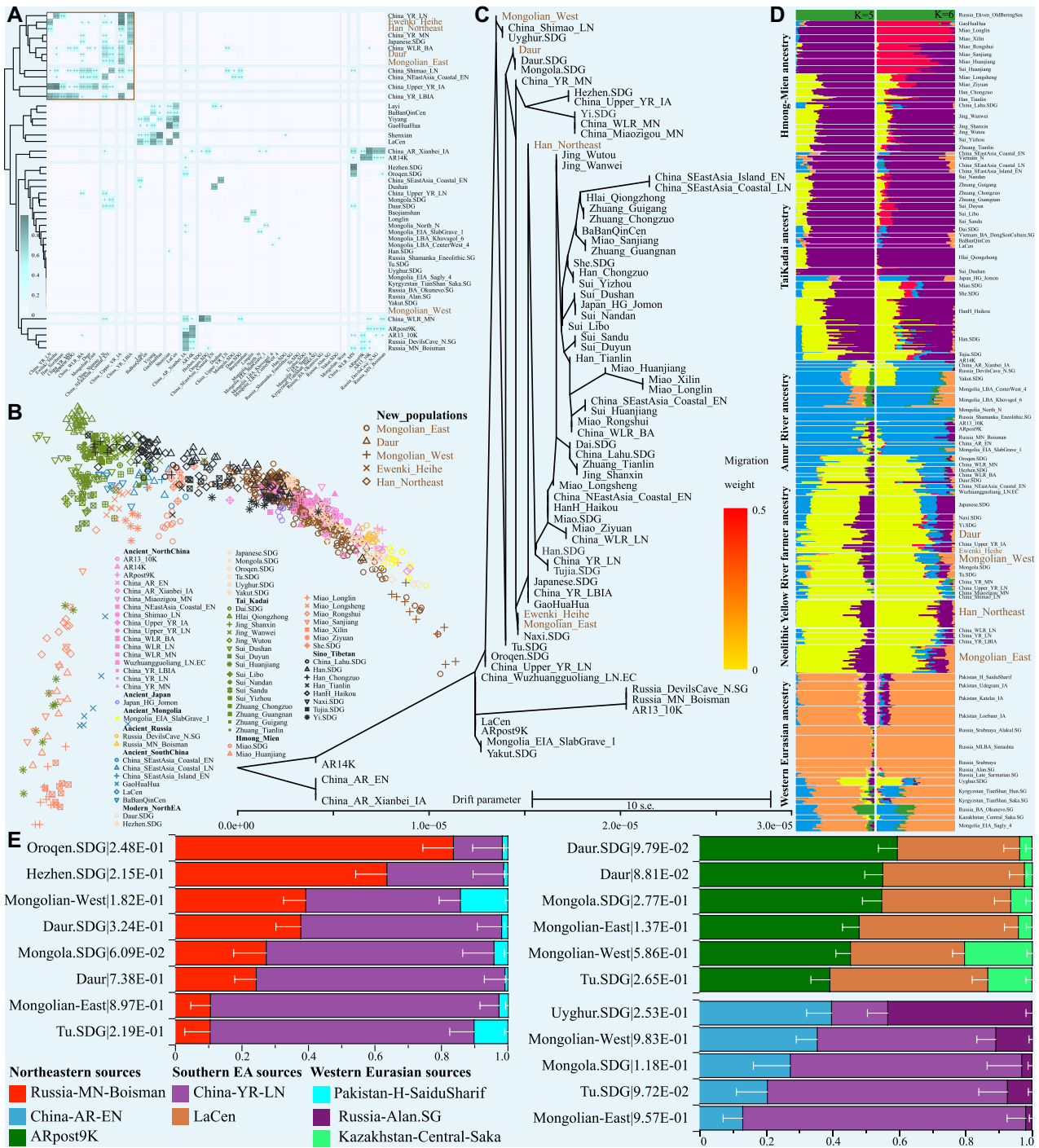


Fig. 2. Genetic sources and mixed landscape of eastern Asians based on the merged 1240K data set (369 519 SNPs). **A**, The p-values of the rank test (rank0) in the pairwise qpWave analysis. The color in the heatmap shows the detailed p-values. p-values more than 0.05 are marked with “++”, which denotes that two tagged populations are homogeneous; more than 0.01 are marked with “+”, which denotes a marginal significance; and others that denote two targeted populations are heterogeneous. **B**, PCA results among 1073 individuals from 71 modern and ancient populations. Ancient populations were projected. K = 5 had the lowest cross-validation (CV) error. **C**, TreeMix-based phylogeny without migration events among 71 populations. **D**, Model-based ADMIXTURE results show the possible ancestral sources and their corresponding admixture proportion. The breadth of populations is not correlated with the included population size, which is enlarged or reduced to be better visualization, especially for the eastern Mongolians. **E**, Three-way admixture qpAdm models with different ancestral sources show the mixed landscape of northern Mongolic and Tungusic people. The p-values (rank2) are presented following the population name and the white bar represents the standard error of the estimated ancestry proportion. SDG sample with diploid genotypes; SG: population with pseudo-haplotypes.

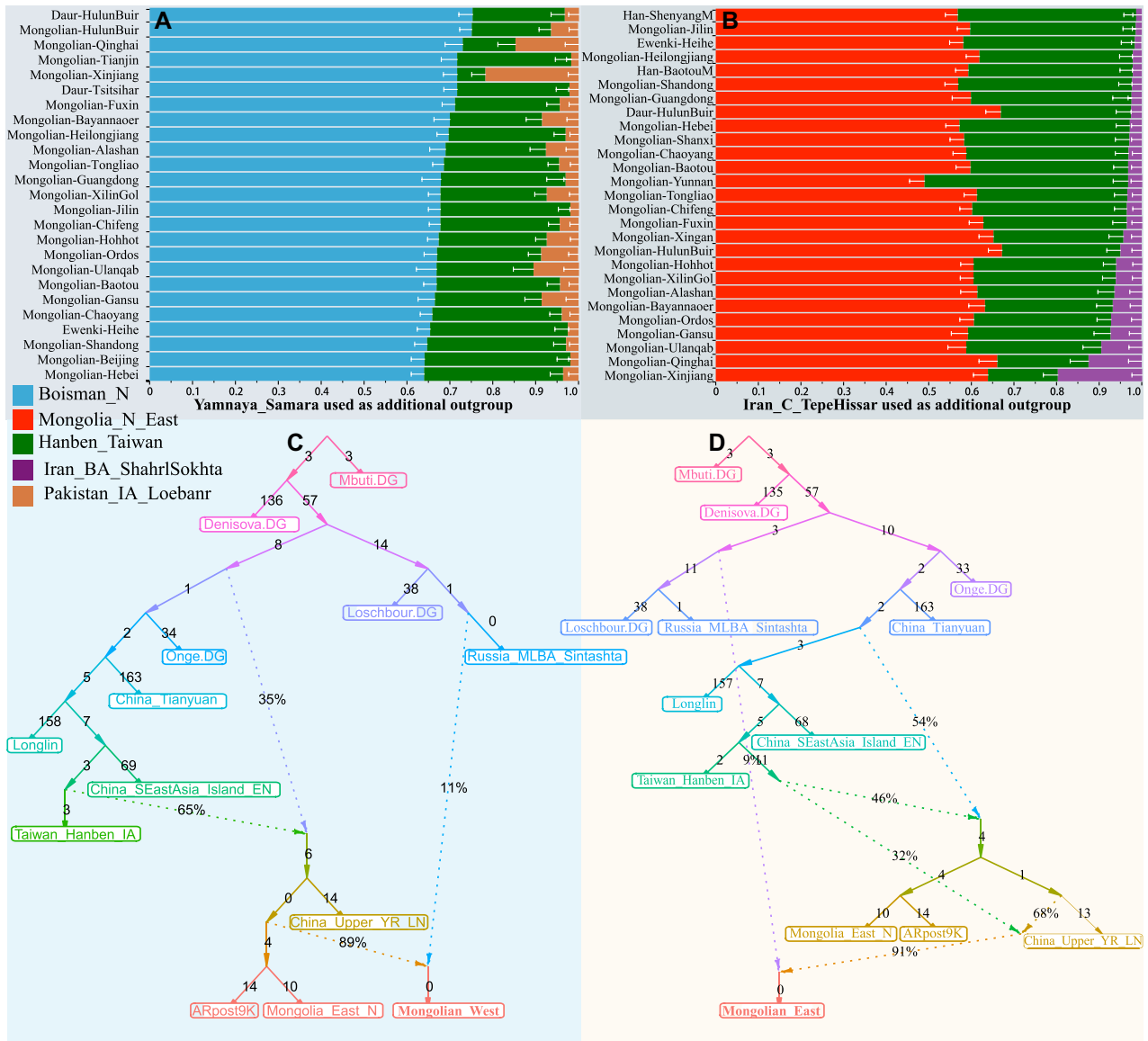


Fig. 3. Demographic history of western and eastern Mongolians estimated via qpAdm and qpGraph. **A, B,** The estimated ancestry coefficient of the representative sources in the three-way qpAdm models with Northeast East Asian sources from Mongolian Plateau and Amur River Basin, East Asian source related to Hanben people and western source related Iran and Pakistan ancients. The bar plots denote the standard error of the admixture proportion. The p -values of p_rank2 are more than 0.05. **C, D,** The best fitted qpGraph-based deep population admixture history of western Mongolians (C) and eastern Mongolians (D).

f_4 (Han1, Han2; Ancient Eurasians, Yoruba) further showed a close genetic affinity within Northeast Hans that was further confirmed by the co-ancestry matrix-based fineSTRUCTURE and shared-IBD-based heatmap (Fig. S7). To explore their genetic relationship with neighboring populations, we conducted pairwise qpWave analysis focused on the meta-Northeast Hans and one of modern and ancient northern East Asians. We found that Northeast Hans formed one clade with Henan Neolithic YRB farmers. Consistent clustering patterns were further identified via PCA results in Fig. S7B, as well as TreeMix and model-based ADMIXTURE results in the 1240K data set, in which Northeast Hans possessed more Tai-

Kadai ancestry compared with Henan ancients but lacked Neolithic MP ancestry (Figs. S7C, S7D), suggesting that modern Northeast Hans received more genetic influence from southern East Asians than from their ancient counterparts. Unlike the Northwest Hans, we could not identify apparent western ancestry in studied Hans based on the model-based ADMIXTURE modeling, which was confirmed by the limited western Eurasian ancestry estimated in three-way qpAdm admixture models (Figs. 3A, 3B). We directly evaluated the genetic affinity and ancestral composition of Northeast Hans using f_3 -statistics. We observed that Han shared the most genetic drift with each other, followed by

eastern Mongolians in the outgroup- f_3 -statistics. In addition, Northeast Hans had a similar admixture landscape as other Chinese Hans inferred from admixture- f_3 -statistics, in which one plausible source was from southern indigenes and the other from northern Altaic/Tibeto-Burman people. Asymmetrical and symmetrical four-population tests in f_4 (YRB farmers, Siberian ancients; Northeast Hans, Mbuti) $\geq 3 \times SE$ and f_4 (Hans1, Hans2; Eurasians, Mbuti) $\leq 3 \times SE$ also showed that Northeast Hans were one genetically homogeneous population, which showed a solid genetic affinity with Neolithic YRB farmers. Generally, Northeast Hans shared a stronger affinity with each other and a close relationship with ancient YRB farmers.

Additionally, we conducted qpAdm to assess the genomic formation of Northeast Hans, geographically close northern minorities (Oroqen, Hezhen, Daur, and Japanese), as well as southern Han Chinese and other ethnic groups (Miaos, Jings, Tujias, and Gelaos). We first modeled mixture proportions with early ARB ancients as the northern surrogate sources and Late Neolithic Coastal Southern East Asians as the southern surrogate sources (Figs. 4A, 4C) and found that five Northeast Hans were mixed via 0.506–0.586 ARB ancestry and 0.494–0.414 southern East Asian ancestry. Compared with Northeast Hans, Northeast minorities harbored more Neolithic northeastern East Asian ancestry ranging from 0.599 \pm 0.047 in Heihe Ewenki to 0.851 \pm 0.043 in Oroqen. Southern Hans, as expected, had more southern East Asian ancestry ranging from 0.565 \pm 0.039 in HGDP Hans to 0.583 \pm 0.039 in Zunyi Hans, and southern ethnically specific

minorities also harbored more southern East Asian ancestry ranging from 0.566 \pm 0.043 in Tai-Kadai-speaking Gelaos to 0.812 \pm 0.045 in Tai-Kadai-speaking Zhuang people. In addition, we used Late Neolithic YRB farmers as the northern representative ancestral source and found that only HulunBuir and Harbin Hans could be successfully modeled as a mixture of major (0.871–0.842) YRB farmer ancestry and minor (0.129–0.158) LaCen ancestry. All included populations sampled from the south of YRB could be fitted as 0.133–0.613 YRB farmer ancestry and 0.867–0.387 historical Guangxi ancestry, suggesting that other populations from north of YRB may be influenced by gene flow from Siberia. Therefore, we further used two ancestral sources respectively from YRB and ARB to model the genetic formation of included northeastern populations (Fig. 4C) and confirmed the significant influence of Neolithic ARB ancestry on the formation of the gene pool of Northeast Hans, but Mongolic and Tungusic people possessed more ARB ancestry ranging from 0.086 in Ewenki and 0.881 in Oroqen people. Finally, we explored the demographic history of Hans via the best fitted qpGraph models. We found that Northeast Hans were fitted via the 24% Hanben ancestry and 76% YRB farmer ancestry (Fig. 4D) with the limited gene flow from western Eurasians.

3.4 Primary Neolithic MP ancestry in Tungusic and Mongolic people in Siberia and Amur River Region

Except for the denser sampling of Mongolians, we also genotyped Mongolic Daur and Tungusic Ewenki and Hezhen people. PCA, ADMIXTURE, and fineSTRUCTURE-based

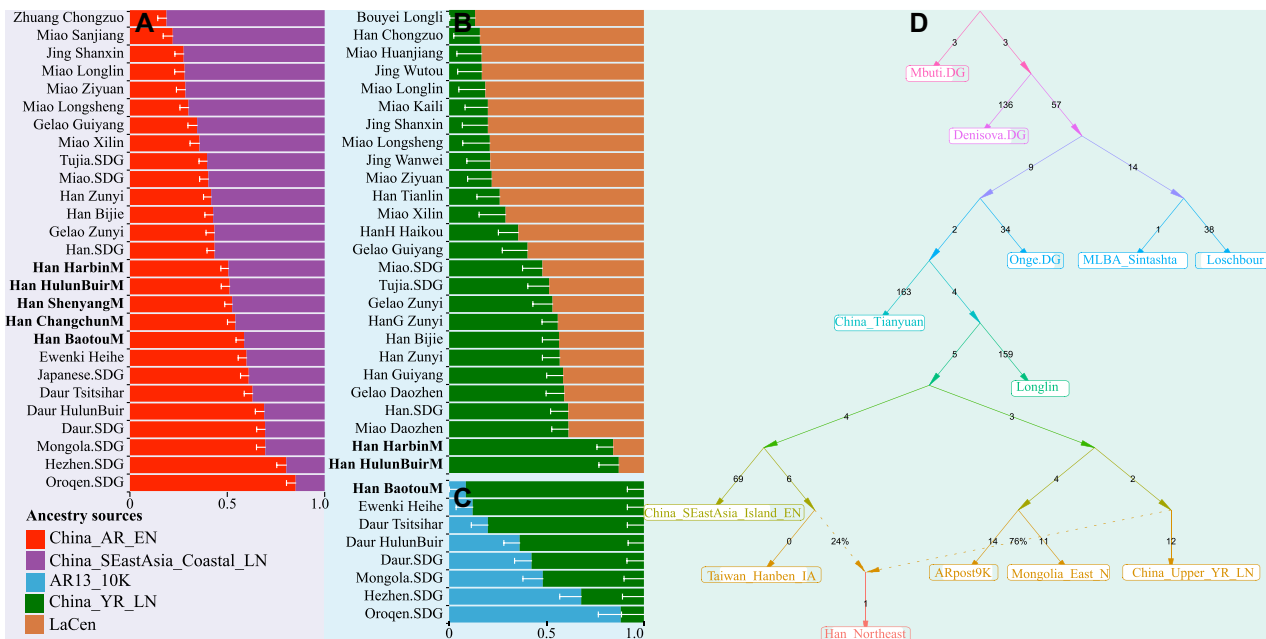


Fig. 4. Genetic formation of Northeastern Han Chinese inferred from the genome-wide SNPs data in the merged 1240K data set. **A**, Two-way admixture model (Amur Neolithic sources + Neolithic Coastal southern East Asian source). **B**, North-to-south admixture model with two sources related to YRB farmers and 2000-year-old LaCen inland southern East Asian shows the differentiated ancestry composition of Han Chinese and southern Chinese indigenous populations. **C**, Ancestry admixture landscape characterized via the two sources from the Yellow River farmers and Amur River hunter-gatherer. All fitted models with the p_{rank1} values are more than 0.05 and the bar plot represents the standard error. **D**, The deep population formation history of Northeast Hans via the best fitted qpGraph-based admixture graph.

population clustering patterns showed that HulunBuir and Tsitsihar Daur had a close genetic relationship with each other and clustered together. Hezhen and Ewenki people were clustered closely with Mongolians (Figs. 1, S5–S7). Similar mixed ancestral sources and landscape of ancestry components were confirmed via asymmetrical- f_4 -tests in the form $f(\text{pop}_1, \text{pop}_2; \text{Eurasians}, \text{Mbuti})$ based on the 1240K data set. Three-way qpAdm admixture models could well fitted most of our focused Hezhen, Daur, Mongolian, and Tu people (Fig. 2), which suggested their complex admixture history with multiple ancestral sources. We further confirmed the primary Neolithic ARB ancestry in Mongolic and Tungusic people via the three-way admixture models focused on non-meta populations in the merged HO data set (Fig. 3). However, we also found that our predefined models could not fit all our targeted populations, suggesting that the true scenarios of the admixture process were more complex than found when using simplified models.

Therefore, to comprehensively characterize this population landscape, we collected 235 Altaic-speaking individuals from 26 populations to conduct population structure analyses based on the shared alleles and haplotypes, except for Turkic people, whose population genetic history has been reconstructed recently (Yunusbayev et al., 2015). We could identify four-population clusters associated with the northern (pink circle) and southern (orange) Mongolic people, Japanese, and Korean (dark-green), as well as the Tungusic people (yellow) based on PCA patterns reconstructed from the allele frequency spectrum and co-ancestry matrix (Figs. S7J–S7N). Four ancestral sources were further confirmed via the ADMIXTURE results with four predefined ancestral sources. Here, we also confirmed the mixed genetic structure of Mongolians, Hezhens, and Xibos consisting of major southern Mongolian ancestry and some Japanese and Tungusic ancestry, while Daur and Oroqens were composed of four ancestries (Fig. 5A). Pairwise F_{ST} within 26 populations first revealed significant genetic differentiation between Tungusic-speaking Ulchis, Negidals, and Nanais from others, and then showed the differences between Japanese with other references (Table S6). Buryat people showed a close genetic relationship, followed by Mongolia Mongolians, and a distant relationship with southern Mongolic people (Fig. 5B). Longer sharing IBD observed in Nanai, Negidal, and Ulchi people compared with others showed their unique genetic structure and higher inbreeding phenomenon (Table S6). Similar patterns of longer sharing IBD fragments were further identified in Buryats and Mongols. Other Chinese Mongolic and Tungusic people had relatively shorter IBD chunks, suggesting their possible recent admixture history and higher effective population size (Figs. 5C, S8). Differentiated population genetic history of geographically and ethnolinguistically diverse Altaic populations and corresponding clustering patterns were further confirmed via the sharing number of IBD chunks in fineSTRUCTURE-based dendrogram and TreeMix-based phylogeny (Figs. 5D–5F).

We next used two-way and three-way admixture models to assess the genetic sources and admixture proportions of 42 Altaic-speaking populations. All included populations, except for four Tungusic-speaking populations (Negidal, Ulchi, Evenk-FarEast, and Nanai), could be well fitted via

the three-way admixture models with representative ancestral sources from Siberia, YRB, and western Eurasia. Altaic-speaking populations possessed major Neolithic YRB Xiaowu farmer ancestry ranging from 0.462 ± 0.013 in Evens to 0.907 ± 0.020 in Yugurs (Sunan) and non-ignorable Palaeo-Siberian ancestry (0.053 ± 0.021 in Yugurs to 0.339 ± 0.014 in Yukagirs) and western Andronovo pastoralist ancestry (0.033 ± 0.011 in Daur to 0.915 ± 0.009 in Veps) in the Palaeo-Siberians–YRB farmers–western Eurasians models (Fig. S9A). Mongolic and Tungusic people harbored more middle YRB Xiaowu farmer ancestry than Turkic people ($0.61–0.865$ vs. $0.069–0.597$). Most of the included Mongolic and Tungusic populations could be well fitted via the Neolithic Mongolians–YRB farmers–western Eurasians models (Figs. S9B, S9C) with primary Fofonovo ancestry in non-Chinese ethnic groups (ranging from 0.661 ± 0.070 in Mongol_Uuld to 0.880 ± 0.030 in Khamnegan) and primary Pingliangtai ancestry in Chinese groups (0.648 ± 0.057 in Dongxiangs to 0.806 ± 0.064 in Yugurs). Different from the ancestry composition of Tungusic and Mongolic people, we identified a significant effect of both western Eurasian (TepeHissar_C: 0.228 ± 0.013 in Kazakh_Aksay to 0.762 ± 0.014 in Nogai) and Fofonovo ancestries (0.168 ± 0.060 in Nogai to 0.638 ± 0.02 in Kazakh_Aksay) in the genetic formation of Turkic people. Similar patterns of qpAdm-based affinity were confirmed via the DevilsCave_N-Haojiatai_LN-Yamnaya three-way admixture model (Fig. S9C). We also confirmed that Chinese Mongolic and Tungusic people had primary Neolithic YRB farmer ancestry via the well fitted two-way admixture models (Figs. S9D–S9F). We could also successfully model the western Altaic people in the three-way qpAdm models with two western sources and one eastern source, and the four-way qpAdm models with two eastern sources and two western sources (Table S7), suggesting that both northern pastoralists and southern barley farmers participated in the formation of western Altaic people. Our findings supported the idea that geographically different Altaic-speaking populations harbored diverse genetic landscapes with differentiated demographic history: a northern one with primary Neolithic Mongolian gene flow, a southern one possessing significant influence from YRB farmers, and a western one harboring extensive genetic admixture with steppe pastoralists and Iran farmers.

3.5 Genetic relationships between northern Chinese populations and historical hierarchical and centrally organized empires in the eastern steppe

Increasing ancient genomes of historical pastoralists have been recently reported (Jeong et al., 2020), including Xiongnu, Uyghur, Khitan, Turkic, and Mongols. Jeong et al. (2020) used the ancient and modern genome-wide data from Siberia and Mongolia to perform individual-based qpWave analysis. They demonstrated that the genetic profiles of northern Mongolic-speaking populations have not changed since the Mongol Empire (Jeong et al., 2020). However, the genetic contribution from historical pastoralists in the MP into modern northern Chinese ethnic groups needed to be further characterized. To explore the shared ancestry between modern East Asians and historical ancient populations, we merged our data with modern Mongolic and Tungusic people in the HGDP data set, as well as the

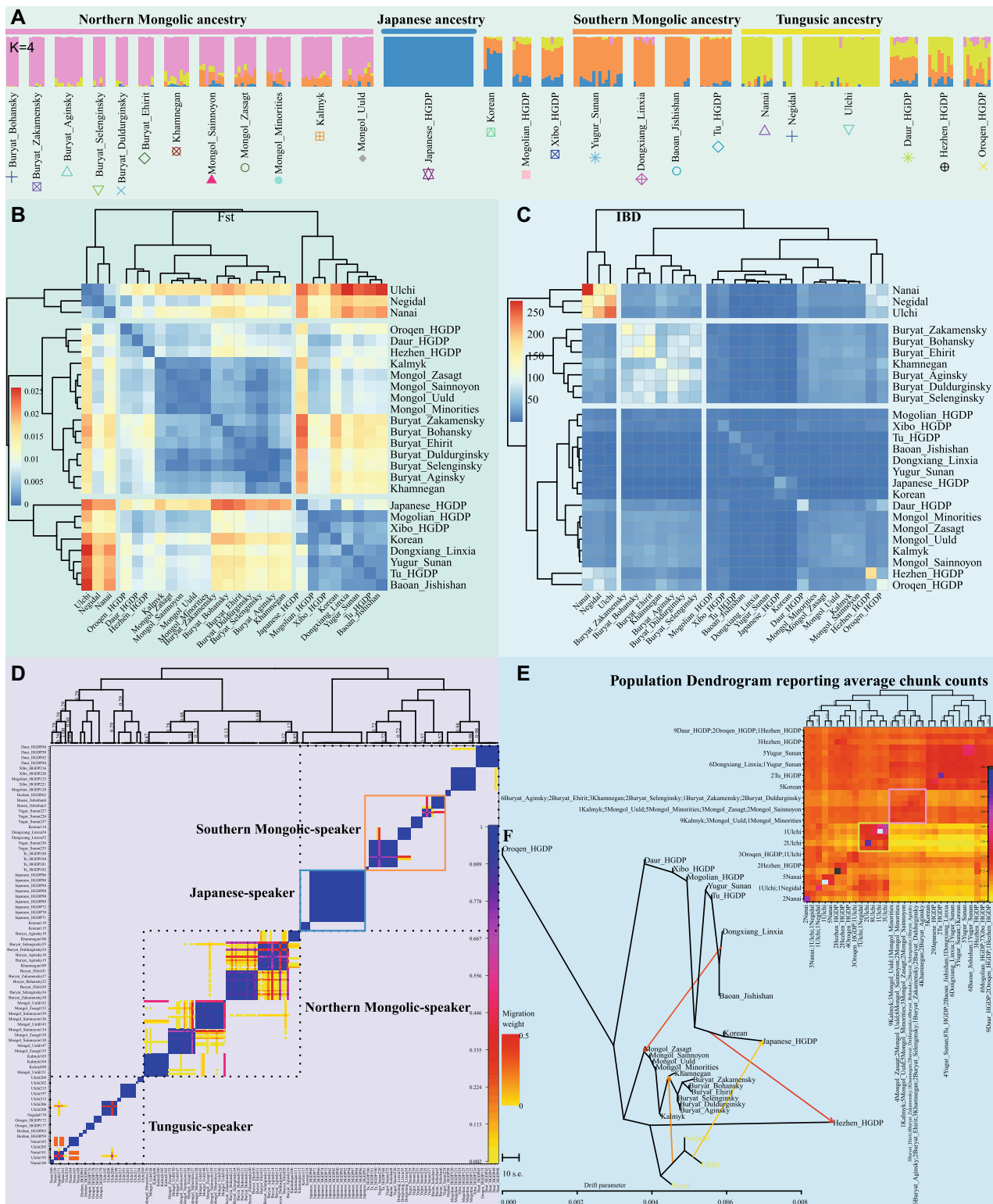


Fig. 5. Fine-scale population structure of 238 northern Mongolic and Tungusic people from 26 populations inferred from dense haplotype data of 600K SNPs in the Human Origin (HO) data set. **A**, ADMIXTURE results for four predefined ancestral sources. The colored-coded circle denotes different genetic homogenous population clusters, which is consistent with the following color-coded components in IBD-based PCA in Fig. S7. **B**, **C**, The pairwise F_{ST} genetic distance and share IBD fragments between 26 Mongolic and Tungusic populations. **D**, **E**, Pairwise coincidence matrix output using fineSTRUCTURE based on the chunk counts and the estimated average chunk counts. **F**, TreeMix-based phylogenetic relationship with four fitted admixture events.

ancient genomes from the MP, which included some Bronze Age populations and historical Xiongnu, Khitan, Turk, Uigur, and Mongol populations.

We identified the genetic differentiation between the studied populations and ancient Mongolian populations in

PCA analysis. Some late Medieval Mongols overlapped with eastern Mongolians, and some western Mongolians overlapped with Turk and Xiongnu people (Fig. 6A). The shared genetic drift estimated from f -statistics revealed that eastern Mongolians possessed stronger East Asian affinity with

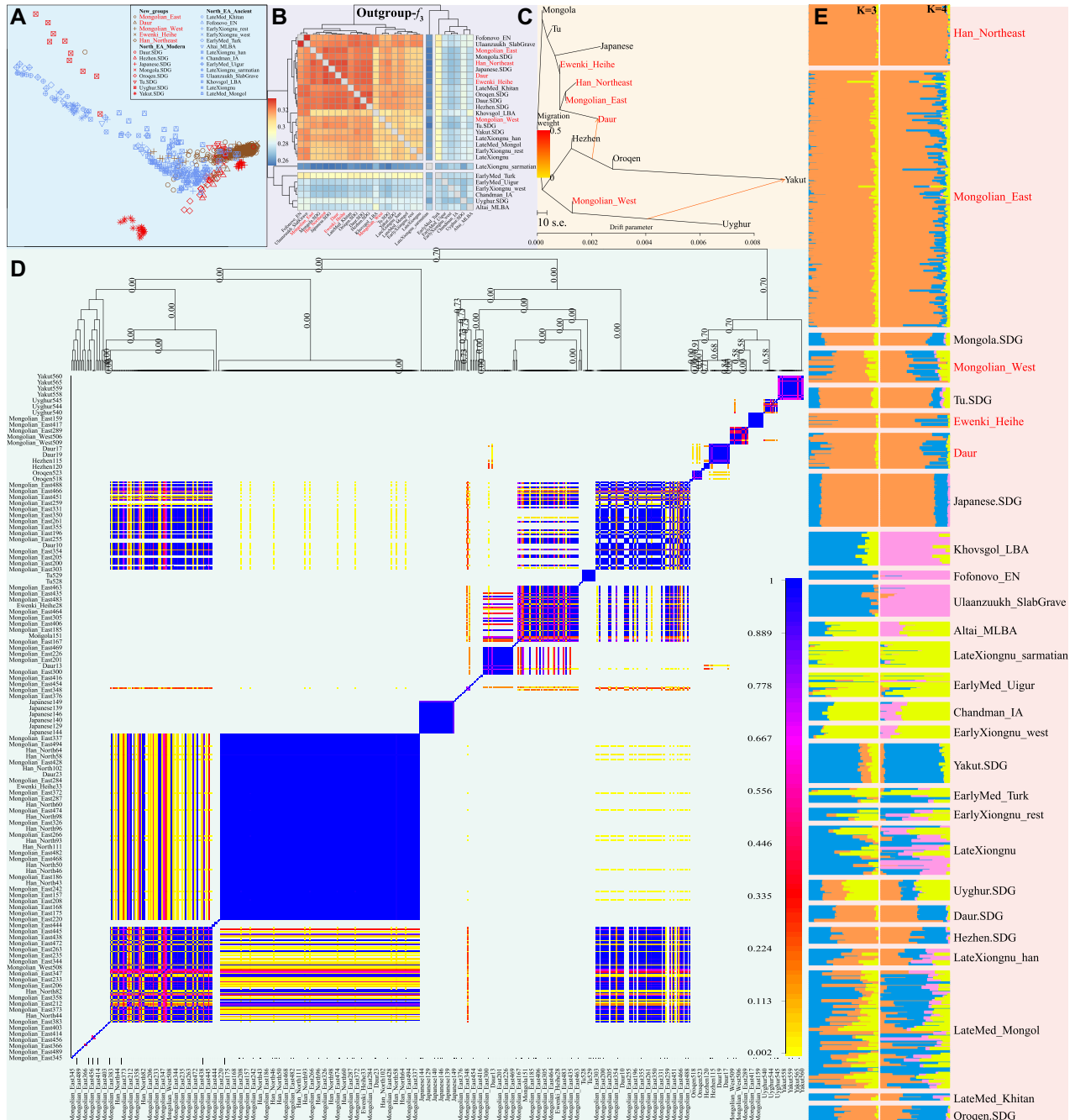


Fig. 6. Fine-scale population genetic structure shows the genetic relationship between newly studied Chinese populations. **A**, Smartpca-based PCA results among 760 individuals from 27 populations. Variations from 566 individuals from 13 populations were used to provide the genetic background. Ancient Mongolian Plateau historical people were projected. **B**, Pairwise shared genetic alleles estimated via outgroup- f_3 -statistics. **C**, Phylogenetic relationship between five studied populations and seven northern East Asian populations included in the HGDP with two admixture events. **D**, **E**, Pairwise coincidence matrix and ADMIXTURE-based results of 13 includes modern populations in the merged 1240K data set. $K=3$ had the lowest cross-validation (CV) value.

northern Hans, Daur, Japanese, and Hezhens, but they harbored a distant relationship with western Mongolians (0.3177). Our results were also confirmed via the inferred two major branches in the f_3 -based heatmap: one consisted of ancient populations harboring more western ancestry (Chandman, Turk, and Uigur) and the other comprised modern northern Asian populations and Ulaanzuukh_SlabGrave, Xiongnu, and Mongol populations (Fig. 6B). We also explored the genetic relationship using a TreeMix-based graph with admixture events and co-ancestry-based pairwise coincidence among 12 northern East Asians (Figs. 6C, 6D), confirming their two separated branches. Admixture signatures inferred from admixture- f_3 -statistics and GLOBETROTTER results revealed that the best ancestral sources for eastern Mongolians were from Hans and Mongol ancients. However, Eurasian-related populations (Uigur and Turk) in MP were evidenced as one of the best sources for western Mongolians. These identified admixture signatures also suggested that Daur and Ewenki people possessed the most ancestral components that were derived from the southern source related to Hlais and the northern source linked to Yakuts. In addition, Ulaanzuukh_SlabGrave people were evidenced as the ancestry surrogate for all studied northern East Asians, suggesting their apparent genetic influence on modern people. Generally, the identified genetic affinity and f_3 -based admixture signatures showed that ancient MP people related to Xiongnu and Mongols contributed more ancestry into modern northern Tungusic and Mongolic people but at different proportions.

We further modeled the ancestry composition using ADMIXTURE and found three ancestry components maximized in Han Chinese, ancient Mongolians and western Eurasian-related Xiongnu (Fig. 6E). Yakut ancestry was separated from Neolithic ancestry when we predefined four ancestral sources in the ADMIXTURE-based model fitting. We found a significant genetic difference between Mongolia's historical ancients and modern northern Chinese populations. We finally reconstructed the best fitted qpGraph-based admixture graph. We found different genetic lineages between Mongolia historical ancients and current northern Chinese groups with varying ancestry compositions of the populations (Figs. S9G–S9I). Here, eastern Mongolians shared the major lineages with YRB farmers, but the northern Yakuts and western Mongolians shared common lineages with historical Turkic and Xiongnu. We also found that western Mongolians shared a similar admixture history with late Medieval Mongols, who derived 44% ancestry from YRB farmers, 45.92% from ARB hunter-gatherers, and 10.08% from western Eurasian Andronovo (Figs. S4G, S4H). Finally, we confirmed that early Medieval Turkic derived 68% ancestry from Neolithic MP people, and ancient Turkic people also contributed 30% of genetic materials to western Mongolians whose remaining ancestry had derived from eastern Mongolian-related ancient sources (Fig. S9I).

3.6 East Eurasian origin of Mongolic-speaking Kalmyks in northern Caucasus Mountain

More genome-wide-scale population genetic analysis was needed to explore the genetic impacts of the westward expansion of the Mongol Empire in the Eurasian steppe on modern central Eurasians. Recent findings based on the

forensic STRs and SNPs identified the genetic link between geographically distinct Torghut Mongolians and 3000 kilometers away Jalaid Mongols (Wu et al., 2019). The identified shared alleles between Mongolians and Hazaras also supported the hypothesis that modern Hazaras were the descendants of ancient Mongols (He et al., 2019). Kalmyk is one of the Mongolic-speaking populations residing in Yashkul. Although the geographical origin of Kalmyks was in North Caucasus, the clustering patterns inferred from PCA and ADMIXTURE showed that Kalmyks were genetically located in East Eurasia (Fig. 1). The model-based ancestry composition revealed that Kalmyk people possessed major ancestry related to Neolithic ARB people (0.426), and some related to Neolithic YRB Longshan people (0.187), western Eurasians (0.196), and Neolithic MP people (0.100), and minor ancestry related to southern East Asian indigenes. These observed landscapes were consistent with the genetic structure observed in the northern Mongolic people (Fig. 1E). The slight allele frequency difference between Kalmyks and Xinjiang Mongolians, Siberia Mongols, and Buryats (0.0005–0.0034), suggested their genetic affinity and the common evolutionary history. Outgroup- f_3 values further showed that Kalmyks shared the most genetic drift with Tungusic-speaking populations and Neolithic ARB ancients ($f_3 > 0.2050$). Admixture- f_3 analysis showed that the ancestral sources of Kalmyks were eastern Tungusic people and western Indo-European groups, such as $f_3(\text{Nanai, French; Kalmyk}) = -38.165 \times \text{SE}$. Genetic relationships within Mongolic and Tungusic populations inferred from PCA, TreeMix, heatmaps, and fineSTRUCTURE (Fig. 5) grouped Kalmyks with Mongolians. Kalmyks also shared longer IBD chunks with eastern Eurasians than with others. Three-way qpAdm-based models confirmed that Kalmyks had an East Asian origin with an admixture of major YRB farmers or Neolithic Mongolian ancestry and some western Eurasians (Figs. S9A–S9C).

3.7 Paternal and maternal admixture history of northern East Asians

Finally, we explored the maternal and paternal history of studied northern East Asians (Table S8; Fig. S10). All 362 Mongolians were assigned into 204 terminal matrilineal haplogroups. Haplogroup D was the most predominant lineage (27.07%), followed by B, F, Z, G, C, A, N, and M7, other minor haplogroup lineages (HV, H, I, M8, M9, M10, M11, R, T, U, W, and Y) were sporadically distributed in studied Mongolians with frequencies of no more than 1.66%. Targeted Mongolic-speaking Daur people were assigned into 19 unique matrilineal lineages. Haplogroup B was the most dominant lineage (26.32%), followed by G, F, D, and A; in addition, haplogroups C, N, M8, M11, and R were observed once respectively. Haplogroups B, C, D, F, M8, N, and R were observed in studied Tungusic speakers. All 119 Han Chinese were assigned to 94 terminal matrilineal haplogroups. Haplogroup D was the most common lineage, followed by B, F, M7, N, Z, G, A, and C, other haplogroups (M8, M9, M10, M11, T, U, and Y) were distributed sporadically in Tungusic individuals with frequencies of no more than 3.36%. Among 175 male Mongolians, we identified 80 terminal paternal lineages with frequencies ranging from 0.0057 to 0.0629 (O2a1c1a1a1a1e: 11). The most frequent paternal haplogroup in

targeted Mongolians was O2, followed by C2, O1, and N1. Furthermore, haplogroups D1, E, I, G, Q, and R were sparsely distributed in studied Mongolian populations. We observed the distributions of haplogroups C2, N1, O1, and O2 in Mongolic-speaking Daur and haplogroups C2, O1, and R in studied Tungusic groups. Sixty Han Chinese men were assigned into 42 diverse paternal lineages. Haplogroup O2 was the most prevalent lineage (45.00%), followed by C2, N1, O1, Q, D1, and J.

3.8 Natural selection signatures in northern East Asian Mongolian and Tungusic people

The systematic landscape of the adaptive history of Mongolians has not been comprehensively analyzed. We next estimated the *iHS* and *XP-EHH* values among our newly genotyped Mongolians (Fig. S10). Central Hans were used as the reference populations. Many gene-coding candidate selection regions from chromosomes 1, 4, 6, 12, and 16 were identified via the *XP-EHH*. Here, we highlighted the top 803 SNPs, which were clustered and located in different candidate regions of the genomes. Nine SNPs (1: 77854238–77895554) in the *AK5* gene and six SNPs (1: 145444556–145524224) in *NBPF10* were the major candidate regions in chromosome 1, other signatures of natural selection from *ITGA10* (three SNPs) and *CAPN2* (three SNPs) were also identified in this chromosome. Linked haplotype block consisting of 51 SNPs located in *CYP19A1* (15: 51502844–51591204) was the longest genomic region under selection in Mongolian populations, which were associated with the etiology of breast cancer. Twenty-six SNPs were located in *HLA-C* (6: 31236998–31239912) with *XP-EHH* values ranging from -6.121 to -4.805 . Twenty-four SNPs located in *HLA-A* (6: 29910276–29913542) and 13 SNPs in *HLA-B* (6: 31321882–31324889) and 10 loci in *HLA-DPA1* (6: 33035771–32555987) also showed significant natural selection signatures. Twenty-three SNPs in *CPNE8* (12: 39050589–39261882) possessed the *XP-EHH* values ranging from -5.692 to -7.260 . In addition, at least five functional SNPs in these genes (*TRIM31*, *PHKB*, *TRIM40*, *HLA-DRB1*, *SCN4A*, *CTSB*, *GLDN*, *NBPF10*, *PGM5*, *NELL2*, *GALNT6*, *ITFG1*, *PLB1*, *EPHB1*, and *NREP*) also showed selection signatures, most of which were further confirmed via the high estimated *iHS* values. We following conducted enrichment analysis based on the top natural-selected loci inferred based on *iHS* and *XP-EHH* (Fig. S11). We identified that these candidate genes were associated with complex biological processes, mainly including allograft rejection, negative regulation of cell population proliferation, striated muscle tissue development, regulation of kinase activity, negative regulation of intracellular signal transduction, regulation of cell–cell adhesion, and regulation of cytokine production. The complex interaction inferred from the protein–protein and gene–gene interactions further showed that the naturally selected genes were polygenetic in their association with the occurrence of complex diseases and traits. Two alcohol metabolism genes (*ADH1B*-rs1229984, *ALDH2*-rs671) were proved to have undergone positive selection in East Asians (Taliun et al., 2021). Our study also identified 13 SNPs located in the gastric alcohol dehydrogenase gene (*ADH7*; 4: 100333932–100349135) with apparent selection signatures, suggesting that Mongolian people's alcohol metabolism underwent positive selection

during the pastoralist subsistence strategy. We further explored the spatiotemporal distribution of the allele frequency of *ADH7*-related SNPs in modern and ancient populations (Figs. S12, S13). No consistent and clear originated centers and periods were identified, which may be caused by the limited high-coverage ancient genomes with a larger sample size from China. Therefore, more spatiotemporally finer scale ancient DNA studies and the development of new methods that focus on the pseudo-haploid genotype data should be conducted in the next step to provide new insights into the spatiotemporal evolutionary history of selected candidate genes, such as the *EDAR V370A* (Mao et al., 2021).

4 Discussion

We have presented a comprehensive characterization of genetic variations and admixture history by merging 510 newly genotyped northern East Asian Mongolic, Tungusic, and Sinitic samples with all available modern and ancient eastern Eurasian genomes. This is the largest genome-wide study of Altaic-speaking populations in NCSS. We documented population stratification within northern East Asians, but all possessed the primary ancestry related to Neolithic MP/ARB people. This documented that the shared genetic legacy in Altaic-speaking populations was consistent with genetic continuity from 13 000-year-old ARB people to 7700-year-old DevilsGate/Boisman and then to modern Tungusic Ulchis. Shared common ancestry from eastern Eurasian lineage among modern Altaic people was also evidenced via the longer length of shared IBD fragments between modern Central Asian Turkic people and the southern Siberians (Yunusbayev et al., 2015). The previous Proto-Trans-Eurasian language origin hypothesis stated that the language subfamily of Proto-Mongolic, Proto-Tungusic, Proto-Turkic, and others shared cultural elements with the Hongshan culture in the West Liao River Basin (Robbeets et al., 2021). If the hypothesis is correct, we expected to observe dominant Neolithic Hongshan ancestry in modern Altaic people in the *ADMIXTURE*, *fineSTRUCTURE*, *qpAdm*, and *qpGraph*-based admixture models. However, we observed primary Neolithic hunter-gatherer-related ancestry from MP/ARB in modern northern East Asians, rejecting the farming-driven hypothesis of spreading Trans-Eurasian languages. Here, we also noted that we lacked more detailed robust evidence to confirm the exact geographical locations of the Proto-Altaic language in northern East Asia. More profound comprehensive studies based on cultural reconstruction, linguistic diversity, lexicostatistics, Bayesian phylogeography, and other interdisciplinary approaches need to be conducted to reconstruct a more accurate and complete picture of the evolutionary history of Proto-Altaic people and their languages.

The western Eurasian gene flow significantly shaped the genetic structure of western and northern Altaic-speaking populations. Three-way and four-way *qpAdm* models involving a western Eurasian admixture provided good fitness for Turkic and northern Mongolic people, which was also observed in our *qpGraph*-based phylogenetic framework. The *ALDER* and *GLOBTROTTER*-based data results

supported the idea that the admixture events occurred in historical times. We note that we only considered a single pulse-like admixture, but the real admixture may be continuous. The simplified model we used would result in a more recent admixture time in LD-based estimations. The population contact between western pastoralists and eastern Eurasians had indeed been attested in the early Bronze Age. Paleo-genomic studies have shown significant genetic differences in ancient Eurasian individuals at different times and geographic scales. The Yamnaya and Afanasievo populations spread eastward and westward across the Pontic–Caspian steppe between 3000 BC and 2100 BC, and was accompanied by the diffusion of the Early and Middle Bronze Age steppe population-related ancestral components (Allentoft et al., 2015) that had further reshaped the genetic landscape of Eurasians. The ancient Sintashta, Srubnaya, and Andronovo people further continued to spread westward and eastward to generate the Middle and Late Bronze Age steppe population-related ancestral components (Damgaard et al., 2018). From 100 BC to 200 BC, Scythians with different genetic compositions were active in the eastern, central and western of the Eurasian steppe (Unterlander et al., 2017). Their eastward movements may also leave genetic legacies in modern eastern Eurasians as identified in our qpAdm and qpGraph admixture models.

With this unprecedented data for Mongolians, we also comprehensively revealed the evolutionary genetic history of geographically diverse Chinese Mongolian populations. Our results showed that the Mongolian people were more closely related to East Asian populations than to other modern and ancient northern populations, suggesting their eastern Eurasian origin hypothesis. Moreover, eastern Mongolians shared ancestry with geographically close Hans and Neolithic YRB farmers. But the results from *f*-statistics and chromosome painting found marked differences in the genetic makeup between Mongolians and Hans. The eastern and western Mongolians were also genetically different, which could be attributed to their differentiated shared Siberian and western Eurasian ancestry and various proportions of YRB ancestry. We identified at least three major ancestral components in the Mongolic and Tungusic people, potentially corresponding to ancestral populations in western Eurasia, YRB, and MP/ARB. The precise source of eastern ancestry for Mongolians was challenging to determine, but our results showed that Neolithic YRB people contributed substantial genetic materials to the eastern Mongolic populations. The western Mongolians inherited more ancestry from western Eurasia and ARB/MP, which we proposed was induced by the intense population movement and gene flow from ancient western Eurasians as inferred from autosomal and mitochondrial DNA. In contrast, two major ancestral components from YRB and southern China were identified in Northeast Hans.

Our findings also demonstrated that the genetic landscape of the modern westernmost Mongolic Kalmyk people was the result of an admixture of multiple ancestral sources. Interestingly, our results suggested that Kalmyks were genetically closer to Tungusic people in ARB, although ARB and the Pontic–Caspian steppe were geographically apart. According to historical documents, the Kalmyk people were the Mongolian-related nomadic pastoralist regimes who

migrated from South Siberia into the North Caucasus during the Yuan and Qing dynasties. The ADMIXTURE analysis suggested no considerable gene flow between Kalmyks and surrounding populations after Kalmyk people had migrated from Siberia to West Asia, which could be attributed to speaking the Mongolic language. Focused on the relationship between historical pastoral empires and Chinese Mongolians, we found a significant effect of these historical and prehistorical admixture events on the ethnolinguistic diversity of modern northern and western Altaic people, but a limited contribution to eastern Chinese Mongolians compared with the influence from the Han expansion. The Xiongnu confederations grew strong in eastern Eurasia and moved westward in approximately the second or third century BC (Karafet et al., 2018; Jeong et al., 2019). Subsequently, Turkic, Mongolian and other people successively dominated inner Asia from AD 600 to AD 1500 (Karafet et al., 2018; Tambets et al., 2018; Jeong et al., 2019). The genetic impact of the intense east-to-west population movement on the genetic landscape of modern Altaic speakers was stronger than that caused by the north-to-south population migration. However, historical Yuan and Qing dynasties were established by the predecessors of modern Mongolic and Tungusic people. Finally, considering the unique and complex history of Mongolic and Tungusic populations, the genome-wide SNP data generated in this study is of great significance for the genetic studies of northern East Asians and serves as a useful control data set for genetic association studies.

5 Conclusions

Our population genomic results suggested a complex scenario of the admixture history of Altaic-speaking populations in northern East Asia. The results showed significant population substructures within Altaic-speaking populations (northern and southern Mongolic, Tungusic, and Turkic-speaking populations) and between eastern and western Mongolians. All Altaic-speaking populations were a mixture of primary Siberian Neolithic and non-negligible YRB ancestry, suggesting that Altaic people were more likely to have originated from Northeast Asia (the primary common ancestry was identified in the ARB and surrounding regions) and further experienced an influence from Neolithic YRB farmers. All Altaic people, except eastern and southern Mongolic-speaking populations, possessed a high proportion of West Eurasian-related ancestry, which was in accordance with the linguistically documented language borrowing in Turkic languages. Moreover, the genetic makeup of Mongolic-speaking populations, especially southern, central, and eastern Mongolic-speaking populations, harbored more Neolithic YRB farmer ancestry and a stronger genetic affinity with modern Northeast Han Chinese, suggesting the extensive genetic admixture between Chinese Mongolians and adjacent Han Chinese. Finally, we identified a close genetic connection between the Mongolic-speaking populations from North Caucasus and ARB, suggesting long-distance migration of Altaic-speaking people since the Mongolian Empire periods. These Mongolic groups kept relatively genetically isolated from the surrounding Indo-European-speaking populations.

Acknowledgements

This study was supported by the China Postdoctoral Science Foundation (2021M691879), the Major Project of National Social Science Foundation of China granted to Chuan-Chao Wang (21&ZD285) and Xiaohua Deng (20&ZD248), the National Natural Science Foundation of China (31801040), the “Double First-Class University Plan” key construction project of Xiamen University (the origin and evolution of East Asian populations and the spread of Chinese civilization, 0310/X2106027), European Research Council (ERC) grant to Dan Xu (ERC-2019-ADG-883700-TRAM), Nanqiang Outstanding Young Talents Program of Xiamen University (X2123302) and the Science and Technology Program of Guangzhou, China (2019030016). S. Fang and Z. Xu from the Information and Network Center of Xiamen University are acknowledged for their help with high-performance computing.

Conflict of Interest

The author declares no conflict of interest.

References

- Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 19: 1655–1664.
- Allentoft ME, Sikora M, Sjogren KG, Rasmussen S, Rasmussen M, Stenderup J, Damgaard PB, Schroeder H, Ahlstrom T, Vinner L, Malaspinas AS, Margaryan A, Higham T, Chivall D, Lynnerup N, Harvig L, Baron J, Della Casa P, Dabrowski P, Duffy PR, Ebel AV, Epimakhov A, Frei K, Furmanek M, Gralak T, Gromov A, Gronkiewicz S, Grupe G, Hajdu T, Jarysz R, Khartanovich V, Khokhlov A, Kiss V, Kolar J, Kriiska A, Lasak I, Longhi C, McGlynn G, Merkevicius A, Merkyte I, Metspalu M, Mkrtychyan R, Moiseyev V, Paja L, Palfi G, Pokutta D, Pospieszny L, Price TD, Saag L, Sablin M, Shishlina N, Smrcka V, Soenov VI, Szeverenyi V, Toth G, Trifanova SV, Varul L, Vicze M, Yepiskoposyan L, Zhitenev V, Orlando L, Sicheritz-Ponten T, Brunak S, Nielsen R, Kristiansen K, Willerslev E. 2015. Population genomics of Bronze Age Eurasia. *Nature* 522: 167–172.
- Association WM. 2001. World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects. *Bulletin of the World Health Organization* 79: 373.
- Bai H, Guo X, Narisu N, Lan T, Wu Q, Xing Y, Zhang Y, Bond SR, Pei Z, Zhang Y, Zhang D, Jirimutu J, Zhang D, Yang X, Morigenbatu M, Zhang L, Ding B, Guan B, Cao J, Lu H, Liu Y, Li W, Dang N, Jiang M, Wang S, Xu H, Wang D, Liu C, Luo X, Gao Y, Li X, Wu Z, Yang L, Meng F, Ning X, Hashenqimuge H, Wu K, Wang B, Suyalatu S, Liu Y, Ye C, Wu H, Leppala K, Li L, Fang L, Chen Y, Xu W, Li T, Liu X, Xu X, Gignoux CR, Yang H, Brody LC, Wang J, Kristiansen K, Burenbatu B, Zhou H, Yin Y. 2018. Whole-genome sequencing of 175 Mongolians uncovers population-specific genetic architecture and gene flow throughout North and East Asia. *Nature Genetics* 50: 1696–1704.
- Bergstrom A, McCarthy SA, Hui R, Almarri MA, Ayub Q, Danecek P, Chen Y, Felkel S, Hallast P, Kamm J, Blanche H, Deleuze JF, Cann H, Mallick S, Reich D, Sandhu MS, Skoglund P, Scally A, Xue Y, Durbin R, Tyler-Smith C. 2020. Insights into human genetic variation and population history from 929 diverse genomes. *Science* 367: eaay5012.
- Browning BL, Browning SR. 2011. A fast, powerful method for detecting identity by descent. *The American Journal of Human Genetics* 88: 173–182.
- Browning BL, Browning SR. 2013. Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics* 194: 459–471.
- Cao Y, Li L, Xu M, Feng Z, Sun X, Lu J, Xu Y, Du P, Wang T, Hu R, Ye Z, Shi L, Tang X, Yan L, Gao Z, Chen G, Zhang Y, Chen L, Ning G, Bi Y, Wang W, China MAPC. 2020. The ChinaMAP analytics of deep whole genome sequences in 10,588 individuals. *Cell Research* 30: 717–731.
- Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* 4: 7.
- Changmai P, Jaisamut K, Kampuansai J, Kutanan W, Altınışık NE, Flegontova O, Inta A, Yüncü E, Boonthai W, Pamjav H, Reich D, Flegontov P. 2021. Indian genetic heritage in Southeast Asian populations. *bioRxiv* 2001: 427591.
- Damgaard PB, Marchi N, Rasmussen S, Peyrot M, Renaud G, Korneliusen T, Moreno-Mayar JV, Pedersen MW, Goldberg A, Usmanova E, Baimukhanov N, Loman V, Hedeager L, Pedersen AG, Nielsen K, Afanasiev G, Akmatov K, Aldashev A, Alpaslan A, Baimbetov G, Bazaliiskii VI, Beisenov A, Boldbaatar B, Boldgiv B, Dorzhu C, Ellingvag S, Erdenebaatar D, Dajani R, Dmitriev E, Evdokimov V, Frei KM, Gromov A, Goryachev A, Hakonarson H, Hegay T, Khachatryan Z, Khaskhanov R, Kitov E, Kolbina A, Kubatbek T, Kukushkin A, Kukushkin I, Lau N, Margaryan A, Merkyte I, Mertz IV, Mertz VK, Mijiddorj E, Moiyesev V, Mukhtarova G, Nurmukhanbetov B, Orozbekova Z, Panyushkina I, Pieta K, Smrcka V, Shevnina I, Logvin A, Sjogren KG, Stolcova T, Taravella AM, Tashbaeva K, Tkachev A, Tulegenov T, Voyakin D, Yepiskoposyan L, Undrakhbold S, Varfolomeev V, Weber A, Wilson Sayres MA, Kradin N, Allentoft ME, Orlando L, Nielsen R, Sikora M, Heyer E, Kristiansen K, Willerslev E. 2018. 137 ancient human genomes from across the Eurasian steppes. *Nature* 557: 369–374.
- Fu Q, Posth C, Hajdinjak M, Petr M, Mallick S, Fernandes D, Furtwängler A, Haak W, Meyer M, Mittnik A. 2016. The genetic history of ice age Europe. *Nature* 534: 200–205.
- Gautier M, Klassmann A, Vitalis R. 2017. rehh 2.0: A reimplementation of the R package rehh to detect positive selection from haplotype structure. *Molecular Ecology Resources* 17: 78–90.
- He G, Adnan A, Rakha A, Yeh HY, Wang M, Zou X, Guo J, Rehman M, Fawad A, Chen P, Wang CC. 2019. A comprehensive exploration of the genetic legacy and forensic features of Afghanistan and Pakistan Mongolian-descent Hazara. *Forensic Science International: Genetics* 42: e1–e12.
- He G, Wang Z, Guo J, Wang M, Zou X, Tang R, Liu J, Zhang H, Li Y, Hu R, Wei LH, Chen G, Wang CC, Hou Y. 2020. Inferring the population history of Tai-Kadai-speaking people and southernmost Han Chinese on Hainan Island by genome-wide array genotyping. *European Journal of Human Genetics* 28: 1111–1123.
- He G, Li YX, Wang MG, Zou X, Yeh HY, Yang XM, Wang Z, Tang RK, Zhu SM, Guo JX, Luo T, Zhao J, Sun J, Xia ZY, Fan HL, Hu R, Wei LH, Chen G, Hou YP, Chuan-Chao W. 2021a. Fine-scale genetic structure of Tujia and central Han Chinese revealing massive genetic admixture under language borrowing. *Journal of Systematics and Evolution* 59: 1–20.
- He G, Wang MG, Li YX, Zou X, Yeh HY, Tang RK, Yang XM, Wang Z, Guo JX, Luo T, Zhao J, Sun J, Hu R, Wei LH, Chen G, Hou YP, Wang CC. 2021b. Fine-scale north-to-south genetic admixture profile in Shaanxi Han Chinese revealed by genome-wide

- demographic history reconstruction. *Journal of Systematics and Evolution*. doi:10.1111/jse.12715
- Hellenthal G, Busby GBJ, Band G, Wilson JF, Capelli C, Falush D, Myers S. 2014. A genetic atlas of human admixture history. *Science* 343: 747–751.
- Huang X, Zhou Q, Bin X, Lai S, Lin C, Hu R, Xiao J, Luo D, Li Y, Wei LH, Yeh HY, Chen G, Wang CC. 2018. The genetic assimilation in language borrowing inferred from Jing People. *American Journal of Physical Anthropology* 166: 638–648.
- Huang X, Xia Z-Y, Bin X, He G, Guo J, Lin C, Yin L, Zhao J, Ma Z, Ma F, Li Y, Hu R, Wei L-H, Wang C-C. 2020. Genomic insights into the demographic history of Southern Chinese. *bioRxiv* 2011(2008): 373225.
- Jeong C, Ozga AT, Witonsky DB, Malmstrom H, Edlund H, Hofman CA, Hagan RW, Jakobsson M, Lewis CM, Aldenderfer MS, Di Rienzo A, Warinner C. 2016. Long-term genetic stability and a high-altitude East Asian origin for the peoples of the high valleys of the Himalayan arc. *Proceedings of the National Academy of Sciences USA* 113: 7485–7490.
- Jeong C, Wilkin S, Amgalantugs T, Bouwman AS, Taylor WTT, Hagan RW, Bromage S, Tsolmon S, Trachsel C, Grossmann J, Littleton J, Makarewicz CA, Krigbaum J, Burri M, Scott A, Davaasambuu G, Wright J, Irmer F, Myagmar E, Boivin N, Robbeets M, Ruhli FJ, Krause J, Frohlich B, Hendy J, Warinner C. 2018. Bronze Age population dynamics and the rise of dairy pastoralism on the eastern Eurasian steppe. *Proceedings of the National Academy of Sciences USA* 115: E11248–E11255.
- Jeong C, Wang K, Wilkin S, Taylor WTT, Miller BK, Bemmman JH, Stahl R, Chiovelli C, Knolle F, Ulziibayar S, Khatanbaatar D, Erdenebaatar D, Erdenebat U, Ochir A, Ankhsanaa G, Vanchigdash C, Ochir B, Munkhbayar C, Tumen D, Kovalev A, Kradin N, Bazarov BA, Miyagashv DA, Konovalov PB, Zhambaltarova E, Miller AV, Haak W, Schiffels S, Krause J, Boivin N, Erdene M, Hendy J, Warinner C. 2020. A dynamic 6,000-year genetic history of Eurasia's Eastern Steppe. *Cell* 183(890–904): e829.
- Jeong C, Balanovsky O, Lukianova E, Kahbatkyzy N, Flegontov P, Zaporozhchenko V, Immel A, Wang CC, Ixan O, Khussainova E, Bekmanov B, Zaibert V, Lavryashina M, Pocheshkhova E, Yusupov Y, Agdzhoyn A, Koshel S, Bukin A, Nymadawa P, Turdikulova S, Dalimova D, Churnosov M, Skhalyakho R, Daragan D, Bogunov Y, Bogunova A, Shtrunov A, Dubova N, Zhabagin M, Yepiskoposyan L, Churakov V, Pislegin N, Damba L, Saroyants L, Dibirova K, Atramentova L, Utevska O, Idrisov E, Kamenshchikova E, Evseeva I, Metspalu M, Outram AK, Robbeets M, Djangugurova L, Balanovska E, Schiffels S, Haak W, Reich D, Krause J. 2019. The genetic history of admixture across inner Eurasia. *Nature Ecology & Evolution* 3: 966–976.
- Karafet TM, Osipova LP, Savina OV, Hallmark B, Hammer MF. 2018. Siberian genetic diversity reveals complex origins of the Samoyedic-speaking populations. *American Journal of Human Biology* 30: e23194.
- Lawson DJ, Hellenthal G, Myers S, Falush D. 2012. Inference of population structure using dense haplotype data. *PLoS Genetics* 8: e1002453.
- Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K, Sudmant PH, Schraiber JG, Castellano S, Lipson M, Berger B, Economou C, Bollongino R, Fu Q, Bos KI, Nordenfelt S, Li H, de Filippo C, Prufer K, Sawyer S, Posth C, Haak W, Hallgren F, Fornander E, Rohland N, Delsate D, Francken M, Guinet JM, Wahl J, Ayodo J, Babiker HA, Bailliet G, Balanovska E, Balanovsky O, Barrantes R, Bedoya G, Ben-Ami H, Bene J, Berrada F, Bravi CM, Brisighelli F, Busby GB, Cali F, Churnosov M, Cole DE, Corach D, Damba L, van Driem G, Dryomov S, Dugoujon JM, Fedorova SA, Gallego Romero I, Gubina M, Hammer M, Henn BM, Hervig T, Hodoglugil U, Jha AR, Karachanak-Yankova S, Khusainova R, Khusnutdinova E, Kittles R, Kivisild T, Klitz W, Kucinskis V, Kushniarevich A, Laredj L, Litvinov S, Loukidis T, Mahley RW, Melegh B, Metspalu E, Molina J, Mountain J, Nakkalajarvi K, Nesheva D, Nyambo T, Osipova L, Parik J, Platonov F, Posukh O, Romano V, Rothhammer F, Rudan I, Ruizbakiev R, Sahakyan H, Sajantila A, Salas A, Starikovskaya EB, Tarekegn A, Toncheva D, Turdikulova S, Uktveryte I, Utevska O, Vasquez R, Villena M, Voevoda M, Winkler CA, Yepiskoposyan L, Zalloua P, Zemunik T, Cooper A, Capelli C, Thomas MG, Ruiz-Linares A, Tishkoff SA, Singh L, Thangaraj K, Vilems R, Comas D, Sukernik R, Metspalu M, Meyer M, Eichler EE, Burger J, Slatkin M, Paabo S, Kelso J, Reich D, Krause J. 2014. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 513: 409–413.
- Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, Barsh GS, Feldman M, Cavalli-Sforza LL, Myers RM. 2008. Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 319: 1100–1104.
- Li L, Huang P, Sun X, Wang S, Xu M, Liu S, Feng Z, Zhang Q, Wang X, Zheng X, Dai M, Bi Y, Ning G, Cao Y, Wang W. 2021. The ChinaMAP reference panel for the accurate genotype imputation in Chinese populations. *Cell Research* 31: 1308–1310.
- Lipson M, Cheronet O, Mallick S, Rohland N, Oxenham M, Pietrusewsky M, Pryce TO, Willis A, Matsumura H, Buckley H, Domett K, Nguyen GH, Trinh HH, Kyaw AA, Win TT, Pradier B, Broomandkhoshbacht N, Candilio F, Changmai P, Fernandes D, Ferry M, Gamarra B, Harney E, Kampuansai J, Kutanan W, Michel M, Novak M, Oppenheimer J, Sirak K, Stewardson K, Zhang Z, Flegontov P, Pinhasi R, Reich D. 2018. Ancient genomes document multiple waves of migration in Southeast Asian prehistory. *Science* 361: 92–95.
- Loh PR, Lipson M, Patterson N, Moorjani P, Pickrell JK, Reich D, Berger B. 2013. Inferring admixture histories of human populations using linkage disequilibrium. *Genetics* 193: 1233–1254.
- Mao X, Zhang H, Qiao S, Liu Y, Chang F, Xie P, Zhang M, Wang T, Li M, Cao P, Yang R, Liu F, Dai Q, Feng X, Ping W, Lei C, Olsen JW, Bennett EA, Fu Q. 2021. The deep population history of northern East Asia from the Late Pleistocene to the Holocene. *Cell* 184(3256–3266): e3213.
- Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg SA, Harney E, Stewardson K, Fernandes D, Novak M, Sirak K, Gamba C, Jones ER, Llamas B, Dryomov S, Pickrell J, Arsuaga JL, de Castro JM, Carbonell E, Gerritsen F, Khokhlov A, Kuznetsov P, Lozano M, Meller H, Mochalov O, Moiseyev V, Guerra MA, Roodenberg J, Verges JM, Krause J, Cooper A, Alt KW, Brown D, Anthony D, Lalueza-Fox C, Haak W, Pinhasi R, Reich D. 2015. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* 528: 499–503.
- McColl H, Racimo F, Vinner L, Demeter F, Gakuhari T, Moreno-Mayar JV, van Driem G, Gram Wilken U, Seguin-Orlando A, de la Fuente Castro C, Wasef S, Shoocongdej R, Souksavatdy V, Sayavongkhamdy T, Saidin MM, Allentoft ME, Sato T, Malaspina AS, Aghakhanian FA, Korneliusson T, Prohaska A, Margaryan A, de Barros Damgaard P, Kaewsutthi S, Lertrit P, Nguyen TMH, Hung HC, Minh Tran T, Nghia Truong H, Nguyen GH, Shahidan S, Wiradnyana K, Matsumae H, Shigehara N, Yoneda M, Ishida H, Masuyama T, Yamada Y, Tajima A, Shibata H, Toyoda A, Hanihara T, Nakagome S, Deviese T, Bacon AM, Durringer P, Ponche JL, Shackelford L, Patole-Edoumba E, Nguyen AT, Bellina-Pryce B, Galipaud JC, Kinaston R, Buckley H, Pottier C, Rasmussen S, Higham T, Foley RA, Lahr MM, Orlando L, Sikora M, Phipps ME, Oota H, Higham C, Lambert DM, Willerslev E. 2018. The prehistoric peopling of Southeast Asia. *Science* 361: 88–92.

- Narasimhan VM, Patterson N, Moorjani P, Rohland N, Bernardos R, Mallick S, Lazaridis I, Nakatsuka N, Olalde I, Lipson M, Kim AM, Olivieri LM, Coppa A, Vidale M, Mallory J, Moiseyev V, Kitov E, Monge J, Adamski N, Alex N, Broomandkhoshbacht N, Candilio F, Callan K, Cheronet O, Culleton BJ, Ferry M, Fernandes D, Freilich S, Gamarra B, Gaudio D, Hajdinjak M, Harney E, Harper TK, Keating D, Lawson AM, Mah M, Mandl K, Michel M, Novak M, Oppenheimer J, Rai N, Sirak K, Slon V, Stewardson K, Zalzal F, Zhang Z, Akhatov G, Bagashev AN, Bagnera A, Baitanayev B, Bendezu-Sarmiento J, Bissembaev AA, Bonora GL, Charginov TT, Chikisheva T, Dashkovskiy PK, Derevianko A, Dobes M, Douka K, Dubova N, Duisengali MN, Enshin D, Epimakhov A, Fribus AV, Fuller D, Goryachev A, Gromov A, Grushin SP, Hanks B, Judd M, Kazizov E, Khokhlov A, Krygin AP, Kupriyanova E, Kuznetsov P, Luiselli D, Maksudov F, Mamedov AM, Mamirov TB, Meiklejohn C, Merrett DC, Micheli R, Mochalov O, Mustafokulov S, Nayak A, Pettener D, Potts R, Razhev D, Rykun M, Sarno S, Savenkova TM, Sikhymbaeva K, Slepchenko SM, Soltobaev OA, Stepanova N, Svyatko S, Tabaldiev K, Teschler-Nicola M, Tishkin AA, Tkachev VV, Vasilyev S, Veleminsky P, Voyakin D, Yermolayeva A, Zahir M, Zubkov VS, Zubova A, Shinde VS, Lalueza-Fox C, Meyer M, Anthony D, Boivin N, Thangaraj K, Kennett DJ, Frachetti M, Pinhasi R, Reich D. 2019. The formation of human populations in South and Central Asia. *Science* 365: eaat7487.
- Ning C, Wang CC, Gao S, Yang Y, Zhang X, Wu X, Zhang F, Nie Z, Tang Y, Robbeets M, Ma J, Krause J, Cui Y. 2019. Ancient genomes reveal Yamnaya-related ancestry and a potential source of Indo-European speakers in Iron Age Tianshan. *Current Biology* 29(2526–2532): e2524.
- Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. *PLoS Genet* 2: e190.
- Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. 2012. Ancient admixture in human history. *Genetics* 192: 1065–1093.
- Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics* 8: e1002967.
- Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, Moltke I, Rasmussen S, Stafford TW, Jr., Orlando L, Metspalu E, Karmin M, Tambets K, Rootsi S, Magi R, Campos PF, Balanovska E, Balanovsky O, Khusnutdinova E, Litvinov S, Osipova LP, Fedorova SA, Voevoda MI, DeGiorgio M, Sicheritz-Ponten T, Brunak S, Demeshchenko S, Kivisild T, Vilems R, Nielsen R, Jakobsson M, Willerslev E. 2014. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* 505: 87–91.
- Robbeets M, Bouckaert R, Conte M, Savelyev A, Li T, An D-I, Shinoda K-i, Cui Y, Kawashima T, Kim G, Uchiyama J, Dolińska J, Oskolskaya S, Yamano K-Y, Seguchi N, Tomita H, Takamiya H, Kanzawa-Kiriyama H, Oota H, Ishida H, Kimura R, Sato T, Kim J-H, Deng B, Bjørn R, Rhee S, Ahn K-D, Grunton I, Mazo O, Bentley JR, Fernandes R, Roberts P, Bausch IR, Gilaizeau L, Yoneda M, Kugai M, Bianco RA, Zhang F, Himmel M, Hudson MJ, Ning C. 2021. Triangulation supports agricultural spread of the Transeurasian languages. *Nature* 599: 616–621.
- Sikora M, Pitulko VV, Sousa VC, Allentoft ME, Vinner L, Rasmussen S, Margaryan A, de Barros Damgaard P, de la Fuente C, Renaud G, Yang MA, Fu Q, Dupanloup I, Giampoudakis K, Nogues-Bravo D, Rahbek C, Kroonen G, Peyrot M, McColl H, Vasilyev SV, Veselovskaya E, Gerasimova M, Pavlova EY, Chasnyk VG, Nikolskiy PA, Gromov AV, Khartanovich VI, Moiseyev V, Grebenyuk PS, Fedorchenko AY, Lebedintsev AI, Slobodin SB, Malyarchuk BA, Martiniano R, Meldgaard M, Arppe L, Palo JU, Sundell T, Mannerman K, Putkonen M, Alexandersen V, Primeau C, Baimukhanov N, Malhi RS, Sjogren KG, Kristiansen K, Wessman A, Sajantila A, Lahr MM, Durbin R, Nielsen R, Meltzer DJ, Excoffier L, Willerslev E. 2019. The population history of northeastern Siberia since the Pleistocene. *Nature* 570: 182–188.
- Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, Taliun SAG, Corvelo A, Gogarten SM, Kang HM, Pitsillides AN, LeFaive J, Lee SB, Tian X, Browning BL, Das S, Emde AK, Clarke WE, Loesch DP, Shetty AC, Blackwell TW, Smith AV, Wong Q, Liu X, Conomos MP, Bobo DM, Aguet F, Albert C, Alonso A, Ardlie KG, Arking DE, Aslibekyan S, Auer PL, Barnard J, Barr RG, Barwick L, Becker LC, Beer RL, Benjamin EJ, Bielak LF, Blangero J, Boehnke M, Bowden DW, Brody JA, Burchard EG, Cade BE, Casella JF, Chalazan B, Chasman DI, Chen YI, Cho MH, Choi SH, Chung MK, Clish CB, Correa A, Curran JE, Custer B, Darbar D, Daya M, de Andrade M, DeMeo DL, Dutcher SK, Ellinor PT, Emery LS, Eng C, Fatkin D, Fingerlin T, Forer L, Fornage M, Franceschini N, Fuchsberger C, Fullerton SM, Germer S, Gladwin MT, Gottlieb DJ, Guo X, Hall ME, He J, Heard-Costa NL, Heckbert SR, Irvin MR, Johnson JM, Johnson AD, Kaplan R, Kardina SLR, Kelly T, Kelly S, Kenny EE, Kiel DP, Klemmer R, Konkole BA, Kooperberg C, Kottgen A, Lange LA, Lasky-Su J, Levy D, Lin X, Lin KH, Liu C, Loos RJF, Garman L, Gerszten R, Lubitz SA, Lunetta KL, Mak ACY, Manichaikul A, Manning AK, Mathias RA, McManus DD, McCarvey ST, Meigs JB, Meyers DA, Mikulla JL, Minear MA, Mitchell BD, Mohanty S, Montasser ME, Montgomery C, Morrison AC, Murabito JM, Natale A, Natarajan P, Nelson SC, North KE, O'Connell JR, Palmer ND, Pankratz N, Peloso GM, Peyser PA, Pleiness J, Post WS, Psaty BM, Rao DC, Redline S, Reiner AP, Roden D, Rotter JI, Ruczinski I, Sarnowski C, Schoenherr S, Schwartz DA, Seo JS, Seshadri S, Sheehan VA, Sheu WH, Shoemaker MB, Smith NL, Smith SLR, Sotoodehnia N, Stilp AM, Tang W, Taylor KD, Telen M, Thornton TA, Tracy RP, Van Den Berg DJ, Vasan RS, Viaud-Martinez KA, Vrieze S, Weeks DE, Weir BS, Weiss ST, Weng LC, Willer CJ, Zhang Y, Zhao X, Arnett DK, Ashley-Koch AE, Barnes KC, Boerwinkle E, Gabriel S, Gibbs R, Rice KM, Rich SS, Silverman EK, Qasba P, Gan W, Consortium NT-OfPM, Papanicolaou GJ, Nickerson DA, Browning SR, Zody MC, Zollner S, Wilson JG, Cupples LA, Laurie CC, Jaquish CE, Hernandez RD, O'Connor TD, Abecasis GR. 2021. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* 590: 290–299.
- Tambets K, Yunusbayev B, Hudjashov G, Ilumae AM, Rootsi S, Honkola T, Vesakoski O, Atkinson Q, Skoglund P, Kushniarevich A, Litvinov S, Reidla M, Metspalu E, Saag L, Rantanen T, Karmin M, Parik J, Zhadanov SI, Gubina M, Damba LD, Bermisheva M, Reisberg T, Dibirova K, Evseeva I, Nelis M, Klovins J, Metspalu A, Esko T, Balanovsky O, Balanovska E, Khusnutdinova EK, Osipova LP, Voevoda M, Vilems R, Kivisild T, Metspalu M. 2018. Genes reveal traces of common recent demographic history for most of the Uralic-speaking populations. *Genome Biology* 19: 139.
- The International HapMap Consortium, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Shen Y, Sun W, Wang H, Wang Y, Wang Y, Xiong X, Xu L, Wayne MMY, Tsui SKW, Xue H, Wong JT-F, Galver LM, Fan J-B, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier J-F, Phillips MS, Roumy S, Sallée C, Verner A, Hudson TJ, Kwok P-Y, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui L-C, Mak W, Qiang Song Y, Tam PKH, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T,

- Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PIW, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Sham PC, Varilly P, Altshuler D, Stein LD, Krishnan L, Vernon Smith A, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Steve Qin Z, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Tsunoda T, Johnson T, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Zeng C, Zhao H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimma C, Royal CDM, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Wright Clayton E, Watkin J, Gibbs RA, Belmont JW, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Wheeler DA, Yakub I, Gabriel SB, Onofrio RC, Richter DJ, Ziaugra L, Birren BW, Daly MJ, Altshuler D, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archevêque P, Bellemare G, Saeki K, Wang H, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL, Brooks LD, McEwen JE, Guyer MS, Ota Wang V, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R, Stewart J. 2007. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449: 851.
- Unterlander M, Palstra F, Lazaridis I, Pilipenko A, Hofmanova Z, Gross M, Sell C, Blocher J, Kirsanow K, Rohland N, Rieger B, Kaiser E, Schier W, Pozdniakov D, Khokhlov A, Georges M, Wilde S, Powell A, Heyer E, Currat M, Reich D, Samashev Z, Parzinger H, Molodin VI, Burger J. 2017. Ancestry and demography and descendants of Iron Age nomads of the Eurasian Steppe. *Nature Communications* 8: 14615.
- Wang CC, Yeh HY, Popov AN, Zhang HQ, Matsumura H, Sirak K, Cheronet O, Kovalev A, Rohland N, Kim AM, Mallick S, Bernardos R, Tumen D, Zhao J, Liu YC, Liu JY, Mah M, Wang K, Zhang Z, Adamski N, Broomandkoshbacht N, Callan K, Candilio F, Carlson KSD, Culleton BJ, Eccles L, Freilich S, Keating D, Lawson AM, Mandl K, Michel M, Oppenheimer J, Ozdogan KT, Stewardson K, Wen S, Yan S, Zalzal F, Chuang R, Huang CJ, Loo H, Shiung CC, Nikitin YG, Tabarev AV, Tishkin AA, Lin S, Sun ZY, Wu XM, Yang TL, Hu X, Chen L, Du H, Bayarsaikhan J, Mijiddorj E, Erdenebaatar D, Iderkhangai TO, Myagmar E, Kanzawa-Kiriyama H, Nishino M, Shinoda KI, Shubina OA, Guo J, Cai W, Deng Q, Kang L, Li D, Li D, Lin R, Nini, Shrestha R, Wang LX, Wei L, Xie G, Yao H, Zhang M, He G, Yang X, Hu R, Robbeets M, Schiffels S, Kennett DJ, Jin L, Li H, Krause J, Pinhasi R, Reich D. 2021a. Genomic insights into the formation of human populations in East Asia. *Nature* 591: 413–419.
- Wang T, Wang W, Xie G, Li Z, Fan X, Yang Q, Wu X, Cao P, Liu Y, Yang R, Liu F, Dai Q, Feng X, Wu X, Qin L, Li F, Ping W, Zhang L, Zhang M, Liu Y, Chen X, Zhang D, Zhou Z, Wu Y, Shafey H, Gao X, Curnoe D, Mao X, Bennett EA, Ji X, Yang MA, Fu Q. 2021b. Human population history at the crossroads of East and Southeast Asia since 11,000 years ago. *Cell* 184(3829–3841): e3821.
- Wu R, Li R, Wang N, Peng D, Li H, Zhang Y, Zheng C, Sun H. 2019. Genetic polymorphism and population structure of Torghut Mongols and comparison with a Mongolian population 3000 kilometers away. *Forensic Science International: Genetics* 42: 235–243.
- Yang J, Lee SH, Goddard ME, Visscher PM. 2011. GCTA: A tool for genome-wide complex trait analysis. *The American Journal of Human Genetics* 88: 76–82.
- Yang MA, Gao X, Theunert C, Tong H, Aximu-Petri A, Nickel B, Slatkin M, Meyer M, Paabo S, Kelso J, Fu Q. 2017. 40,000-year-old individual from Asia provides insight into early population structure in Eurasia. *Current Biology* 27(3202–3208): e3209.
- Yang MA, Fan X, Sun B, Chen C, Lang J, Ko YC, Tsang CH, Chiu H, Wang T, Bao Q, Wu X, Hajdinjak M, Ko AM, Ding M, Cao P, Yang R, Liu F, Nickel B, Dai Q, Feng X, Zhang L, Sun C, Ning C, Zeng W, Zhao Y, Zhang M, Gao X, Cui Y, Reich D, Stoneking M, Fu Q. 2020. Ancient DNA indicates human population shifts and admixture in northern and southern China. *Science* 369: 282–288.
- Yao H, Wang M, Zou X, Li Y, Yang X, Li A, Yeh HY, Wang P, Wang Z, Bai J, Guo J, Chen J, Ding X, Zhang Y, Lin B, Wang CC, He G. 2021. New insights into the fine-scale history of western-eastern admixture of the northwestern Chinese population in the Hexi Corridor via genome-wide genetic legacy. *Molecular Genetics and Genomics* 296: 631–651.
- Yunusbayev B, Metspalu M, Metspalu E, Valeev A, Litvinov S, Valiev R, Akhmetova V, Balanovska E, Balanovsky O, Turdikulova S, Dalimova D, Nymadawa P, Bahmanimehr A, Sahakyan H, Tambets K, Fedorova S, Barashkov N, Khidiyatova I, Mihailov E, Khusainova R, Damba L, Derenko M, Malyarchuk B, Osipova L, Voevoda M, Yepiskoposyan L, Kivisild T, Khusnutdinova E, Villems R. 2015. The genetic legacy of the expansion of Turkic-speaking nomads across Eurasia. *PLoS Genetics* 11: e1005068.
- Zhang C, Gao Y, Ning Z, Lu Y, Zhang X, Liu J, Xie B, Xue Z, Wang X, Yuan K, Ge X, Pan Y, Liu C, Tian L, Wang Y, Lu D, Hoh BP, Xu S. 2019. PGG.SNV: Understanding the evolutionary and medical implications of human single nucleotide variations in diverse populations. *Genome Biology* 20: 215.
- Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. 2019. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications* 10: 1523.

Supplementary Material

The following supplementary material is available online for this article at <http://onlinelibrary.wiley.com/doi/10.1111/jse.12827/supinfo>:

Table S1. Admixture proportion among 2783 individuals estimated based on the model-based ADMIXTURE analysis.

Table S2. Shared IBD fragments and pairwise F_{ST} genetic indexes between 38 studied populations and two southmost Hainan.

Table S3. Differentiated shared derived alleles examined via f_4 -statistics in the form $f_4(\text{Eurasian}_1, \text{Eurasian}_2; \text{studied Altaic populations}, \text{Mbuti})$.

Table S4. Differentiated shared derived alleles examined via f_4 -statistics in the form $f_4(\text{Eurasian}_1, \text{studied Altaic populations}; \text{Eurasian}_2, \text{Mbuti})$.

Table S5. ALDER results focused on Altaic-speaking populations estimated based on the merged 1240K data set.

Table S6. Shared IBD fragments and pairwise F_{ST} genetic indexes among 26 Altaic-speaking populations included in the Human Origins data set.

Table S7. qpAdm-based admixture models with two western ancestral sources.

Table S8. Haplogroups of studied populations.

Fig. S1. The geographical position of our studied populations and Altaic-language distribution.

Fig. S2. The geographical position of our studied populations and other modern and ancient reference populations.

Fig. S3. Genetic affinity between newly studied populations and reference Eurasian populations revealed using principal component analysis.

Fig. S4. The genetic affinity between newly studied populations and extracted Tibeto-Burman and Trans-Eurasian-speaking populations revealed using principal component analysis.

Fig. S5. Principal component analysis among East Asians.

Fig. S6. Model-based ADMIXTURE results with 10 predefined ancestral sources.

Fig. S7. Fine-scale population genetic structure.

Fig. S8. Effective population size changes among three ethnolinguistically different populations with different estimated generation ranges.

Fig. S9. The estimated ancestry admixture coefficient of modern and ancient eastern Eurasians via qpAdm and qpGraph.

Fig. S10. Distribution of paternal and maternal lineage distribution among three studied populations.

Fig. S11. *p*-values of the estimated EXEHH focused on Mongolian populations with the Guizhou Hans as the reference population.

Fig. S12. Enrichment analysis results of Mongolian people.

Fig. S13. Allele frequency spectrum based on the major allele frequency of seven ADH7-related SNPs between modern and ancient worldwide populations.

Fig. S14. Spatiotemporal distribution of seven ADH7-related SNPs.