

1 **The submission is intended as an original article**

2 **Genomic insights into the differentiated population admixture structure and demographic history**  
3 **of North East Asians**

4  
5 **Guanglin He<sup>1,2,3\*,#</sup>, Mengge Wang<sup>2,4,5,\*</sup>, Xing Zou<sup>2,6</sup>, Renkuan Tang<sup>7</sup>, Hui-Yuan Yeh<sup>3</sup>, Zheng Wang<sup>2</sup>, Xiaomin Yang<sup>1</sup>, Ziyang Xia<sup>1</sup>, Yingxiang Li<sup>1,8</sup>, Jianxin Guo<sup>1</sup>, Rui Wang<sup>1</sup>, Jing Liu<sup>2</sup>, Kongyang Zhu<sup>1</sup>, Jing Chen<sup>9</sup>, Meiqing Yang<sup>9</sup>, Qu Shen<sup>1</sup>, Jinwen Chen<sup>1</sup>, Jing Zhao<sup>1</sup>, Hao Ma<sup>1</sup>, Lan-Hai Wei<sup>1</sup>, Ling Chen<sup>10</sup>, Changhui Liu<sup>4</sup>, Chao Liu<sup>4,5,10,#</sup>, Gang Chen<sup>11,#</sup>, Yiping Hou<sup>2,#</sup>, Chuan-Chao Wang<sup>1,12,13,#</sup>**

9 <sup>1</sup>Department of Anthropology and Ethnology, Institute of Anthropology, National Institute for Data Science in Health and Medicine, State Key Laboratory of Cellular Stress Biology, School of Life Sciences, State Key Laboratory of Marine Environmental Science, Xiamen University, Xiamen, 361005, China

12 <sup>2</sup>Institute of Forensic Medicine, West China School of Basic Science and Forensic Medicine, Sichuan University, Chengdu, 610041, China

14 <sup>3</sup>School of Humanities, Nanyang Technological University, Nanyang Avenue, 639798, Singapore

15 <sup>4</sup>Guangzhou Forensic Science Institute, Guangzhou, 510080, China

16 <sup>5</sup>Faculty of Forensic Medicine, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, 510080, China

17 <sup>6</sup>College of Basic Medicine, Chongqing University, Chongqing, 400016, China

18 <sup>7</sup>Department of Forensic Medicine, College of Basic Medicine, Chongqing Medical University, Chongqing, 400016, China

20 <sup>8</sup>AnLan AI, Shenzhen, 518000, China

21 <sup>9</sup>Department of Forensic Medicine, Guizhou Medical University, Guiyang, 550000, China

22 <sup>10</sup>Department of Forensic Genetics, School of Forensic Medicine, Southern Medical University, Guangzhou, 510515, China

24 <sup>11</sup>Hunan Key Lab of Bioinformatics, School of Computer Science and Engineering, Central South University, Changsha, 410075, China.

26 <sup>12</sup>School of Basic Medical Sciences, Zhejiang University School of Medicine, Hangzhou, 310000, China

27 <sup>13</sup>Institute of Asian Civilizations, Zhejiang University, Hangzhou, 310000, China

28 \*These authors contributed equally to this work and should be considered co-first authors.

29 #Corresponding author: [Guanglinhescu@163.com](mailto:Guanglinhescu@163.com) (G.L. H.), [liuchaogzf@163.com](mailto:liuchaogzf@163.com) (C.L.), [chengangcs@gmail.com](mailto:chengangcs@gmail.com) (G.C.), [forensic@scu.edu.cn](mailto:forensic@scu.edu.cn) (Y.P. H.) and [wang@xmu.edu.cn](mailto:wang@xmu.edu.cn) (C.C. W.)

32 **ABSTRACT**

33 **North China and South Siberia, mainly populated by Altaic-speaking populations, possess extensive ethnolinguistic diversity and serve as the crossroad for the initial peopling of America and western-eastern trans-continental communication. Yet, the complex scenarios of genetic origin, population structure, and admixture history of North-East Asia remain to be fully characterized, especially for Mongolic people in China with a genome-wide perspective. Thus, we genotyped genome-wide SNPs for 510 individuals from 38 Chinese Mongolic, Tungusic, and Sinitic populations to explore the sharing alleles and haplotypes within the studied groups and following merged it with 3508 modern and ancient Eurasian individuals to reconstruct the deep evolutionary and natural selection history of northern East Asians. We identified significant substructures within Altaic-speaking populations with the primary common ancestry linked to the Neolithic northern East Asians: Western Turkic people harbored more western Eurasian ancestry; Northern Mongolic people in Siberia and eastern Tungusic people in Amur River Basin (ARB) possessed dominant Neolithic Mongolian Plateau (MP) or ARB ancestry; Southern Mongolic people in China owned obvious genetic impact 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 dominant YRB and ARB Neolithic ancestry. Moreover, we demonstrated that Kalmyk people from the northern Caucasus Mountain possessed a strong genetic affinity with Neolithic MP people, supporting the hypothesis of their eastern Eurasian origin and long-distance migration history. We also illuminated that historic pastoral empires in the MP contributed considerably to the gene pool of northern Mongolic people but rarely to southern ones. We finally found natural signatures in Mongolians associated with alcohol metabolism. Generally, our results not only illuminated that complex population migration and admixture of Neolithic ancestral sources from the MP or ARB played an important role in the spread of Altaic-speaking populations and Proto-Altaic language, which partly supported the Northeast Asia-origin hypothesis, but also demonstrated that the observed multi-sources of genetic diversity contributed significantly to the modern existing extensive ethnolinguistic diversity in North-East Asia.**

61 **KEYWORDS:** North East Asians, ancient DNA, genetic admixture, demic diffusion, population  
62 substructure, genetic diversity  
63

64 **INTRODUCTION**

65 North China and South Siberia have been inhabited by anatomically modern humans as early as 40  
66 thousand years ago (kya)(1). Two hypotheses, referred to as the southern migration hypothesis(2) and  
67 northern migration hypothesis(3), have been put forward to elucidate the mechanism and process of  
68 peopling of East Asia and Siberia. The southern migration hypothesis states that the ancestor of East  
69 Asians diverged with the ancestor of Europeans at approximately 49~54 kya in South Asia or the Arabian  
70 Peninsula and then split with the ancestor of Oceanians in Southeast Asia at approximately 44~49 kya  
71 and finally formed ancient East Asians via this southern route(4). The northern migration hypothesis  
72 states that one wave of ancestral population diverged from the ancestor of Europeans, Australians, and  
73 East Asians and migrated via Central Asia into North Siberia, forming Ancient North Eurasian and  
74 Ancient North Siberian, and finally mixed with incoming ancestors from East Asia to form the founding  
75 populations of Americans(5, 6). Crossroads of Mongolian Plateau (**MP**) and surrounding regions  
76 between South East Asia and North Siberia played a key role in the formation of ancient and modern  
77 Eastern Eurasians. Indeed, archeological and genetic evidence from the Chinese Tianyuan Cave (40,000-  
78 year-old Tianyuan)(1) and Siberian Paleolithic ancient people (31,600-year-old Yana\_RHS, 17,000-year-  
79 old AG2, 16,500-year-old AG3, and 24,000-year-old Mal'ta)(3, 7, 8) have suggested that at least two  
80 Paleolithic eastern Eurasian lineages widely existed in East Asia and Siberia, and these two lineages have  
81 participated in the formation of Northeast Eurasians and Americans in the crossroad of Baikal Lake and  
82 Amur River Basin (**ARB**) regions.

83  
84 Neolithic Trans-Eurasian migration and subsequent admixture or transformation played a pivotal role in  
85 the formation of ancestral Eurasians (especially in Siberia) and also contributed to the observed modern  
86 mosaic genetic structure(9-12). Similarly, Neolithic genetic evidence in East Asia not only documented  
87 two deeply diverged eastern Eurasian lineages(13), which composed of Hoabinhian-related clade in  
88 Southeast Asia and Jomon-related lineage in the Japanese archipelago, but also identified at least five  
89 Neolithic East Asian lineages (inland/coastal southern/northern East Asian lineages respectively  
90 associated with Liangdao/Qihe in Fujian, Longlin in Guangxi, Bianbian and Yumin in **Yellow River**  
91 **Basin (YRB)**, and Chokhopani in Tibet, as well as the ancient northeast East Asian lineages related to  
92 Neolithic DevilsGate and Boisman in **MP** and **ARB**)(14-17). These observed patterns of the admixed  
93 ancestry among northern and southern Neolithic East Asians showed that both southward and northward  
94 gene flow events and coastal population connections from Vietnam and coastal Siberia have reshaped  
95 the subsequent genetic landscape of East Asia. Further intense interplays between the coastal and inland  
96 East Asian pre-Neolithic Hunter-Gatherers and Neolithic farmers in Fujian and Guangxi, as well as lower  
97 or middle or upper YRB, were illuminated in the perspectives of both archeology and ancient genomes.  
98 Besides, Neolithic ancient DNA evidence from the Baikal Lake region(7, 18) and ARB(7, 19) not only  
99 suggested the different extent of genetic continuity that existed in eastern Siberia but also demonstrated  
100 extensive genetic interaction and connection between YRB farmers and other Siberian ancient Hunter-  
101 Gatherers(20).

102  
103 Differentiated western Eurasian ancestral sources contributed different genetic components into the gene  
104 pool of North-East Asia. In the Copper or early and middle Bronze Age, the archeologically documented  
105 material culture showed that Yamnaya pastoralists, formed in the Pontic-Caspian steppe (~3000 BCE),  
106 revolutionized the perception of personhood, property, and family(10, 11). These early herders shaped  
107 the genetic structure of later populations in Europe and South Asia(21, 22). Subsequent cultural shifts,  
108 such as Afanasievo culture in the Minusinsk Basin and the Altaic Mountain, Sintashta culture emerged  
109 in the Urals, Andronovo culture in Central Asia and even in Xinjiang and Sayan Mountain, were  
110 significantly associated with the demic diffusion of early eastward migration of steppe pastoralists(10,  
111 11, 22). European or Anatolian farmers following admixed with local Pontic-Caspian herders, which then  
112 formed the Late Bronze Age pastoralists with different cultural backgrounds (Koryakova, Mezhovskaya,  
113 Srubnaya, Okunevo, and Karasuk)(9, 12). Although multiple genetic turnovers and admixture events  
114 occurred in western Eurasia, the Bronze Age population history of eastern Eurasians kept a genetic  
115 continuity as the Neo-Siberians, which has been evidenced via sporadic studies of ancient DNA(7, 12).  
116 Subsistence strategies of these people kept primary hunting-gathering/fishing or reindeer different from  
117 the pastoral lifestyles in the west Eurasia. Recent 6000-year genomic dynamics in Mongolia showed the  
118 instantaneous admixture signatures between western and eastern Eurasians in central Mongolia, as the  
119 observed tripartite population structures starting to vanish(9).

120  
121 More dynamic population history occurred in the eastern steppe during the Iron and historic periods.  
122 Population expansion directions have gradually changed from the previously dominant eastward  
123 expansion to radial expansion from multiple dominant centers (such as the Scythian federation) and

124 finally to westward dispersal via historic pastoral tribes(11, 23). Historic empires of nomadic pastoralists  
125 combined and formed many elite dominance federation groups, such as Turkic, Xiongnu, Mongolia, and  
126 Tungusic et al., and the population expansion centers have been shifted from West Eurasia to East  
127 Eurasia(9, 20). And the basic patterns of the genetic background of modern populations were gradually  
128 formed, with the eastern Eurasian-related genetic diversity replaced or mixed with the Proto-Indo-  
129 European's gene pool as well as their languages(7, 10). Although the landscapes of three genetic clines  
130 in modern Siberians have been characterized, how their genetic relationship with Chinese Altaic  
131 populations and the potentially existing differentiated evolutionary population history needed to be  
132 comprehensively clarified and characterized.

133  
134 The language/farming co-dispersal hypothesis has been evidenced in the spread of Sino-Tibetan, and  
135 micro-family in South China and Southeast Asia(20), as well as the possible association between the  
136 spread of Anatolian farmer ancestry and Indo-European people(12). Linguistic and archaeologic  
137 evidence also consistently supported that early dissemination of Indo-European languages was mediated  
138 by the steppe pastoralists' expansion(21, 22). Recent findings have revealed the influence of the co-spread  
139 of this language and the corresponding ancient Indo-European people extending to 2200-year-old Iron  
140 Age Shirenzigou people in the northeastern Xinjiang(24). Patterns of genetic structure and Indo-  
141 European distribution were also influenced and characterized by large-scale eastward expansion and  
142 replacements of local populations(7). Genomic evidence has found that these population migrations have  
143 no significant influence on the population demographic dynamics in Baikal Lake and ARB regions and  
144 North China(25). After the Bronze Age migrations, population interaction between western and eastern  
145 Eurasian steppe has emerged among highly structured Scythian groups with possible Turkic language,  
146 whose gene pool was consisted of western sources (European farmer and Late Bronze Age herder),  
147 eastern source of southern Siberian Hunter-Gatherer and Anatolian or Iran farmer-related ancestry(26).  
148 Further westward dispersal of Xiongnu and Hun Khanates, and other historical migration events  
149 (expansion of Mongolian Empire) have promoted Trans-Eurasian-speaking populations (Turkic,  
150 Mongolic and Tungusic) dominated in the Eurasian steppe and replaced previous Indo-Iranian-speaking  
151 groups (Wusun, Kangju et al.)(26). However, the detailed interaction between YRB farmers and Siberian  
152 Hunter-Gatherers, and what the extent of genetic contribution from these ancient populations to modern  
153 northern East Asians, as well as who mediated the dispersal of modern Altaic-speaking populations  
154 needed to be further explored. The geographical origins of Altaic-speaking populations have been stated  
155 to be controversial based on evidence from genetics and linguistics, including Pastoralist Hypothesis or  
156 Farming Hypothesis supported from the Altaic Mountain, West Liao River, Baikal Lake and ARB(20).  
157 These controversial hypotheses needed to be tested via dense sampling of modern Altaic people and  
158 comprehensive comparison with spatiotemporally ancient sources. As the demographic and evolutionary  
159 history of the eastern regions (Baikal, ARB, MP and YRB) is still in its infancy and unclear compared  
160 with the well-documented complex admixture history in the western Eurasian steppe. These regions have  
161 been inhabited mainly by Mongolic and Tungusic-speaking populations today and there are also some  
162 remnants of early Turkic-speaking populations and historic northward expanded Han Chinese.

163  
164 To comprehensively reconstruct the deep population history, explore the ancestral origin and population  
165 dynamics of Tungusic/Mongolic-speaking populations and their neighbors, we genotyped 510 modern  
166 individuals from Mongolic, Tungusic, and Sinitic-speaking populations using a high-density SNP array  
167 and combined the basic dataset with publicly available worldwide modern and ancient genome-wide  
168 data(3, 4, 7, 9-16, 20-22, 26, 27). This work mainly aimed to explore **I**) how many ancestral sources  
169 contributed to modern Altaic-speaking populations, **II**) what's the association between the identified  
170 ancestries and the origin and mixture of modern Altaic language macro-family, as well as **III**) what's the  
171 detailed admixture and natural selection histories of subgroups of Altaic-speaking populations and  
172 geographically close Northeast Hans, and finally **IV**) illuminate the genetic contribution from the historic  
173 pastoral empires from the MP to the southernmost Yunnan and Guangzhou Mongolians and westernmost  
174 Kalmyk people. We reconstructed the deep population genetic history of North-East Asia using multiple  
175 computational statistical methods (including sharing-allele  $f$ -statistics, haplotype-based chromosome  
176 painting skills, and sharing paternal and maternal founding lineages). We found a shared Neolithic  
177 ancestry maximized in Neolithic populations from the MP and ARB, which was widely distributed  
178 among all northern Mongolic/Tungusic-speaking populations from as far east as coastal Siberia (Ulchi)  
179 to as far west as the northern Caucasus (Kalmyk). We also identified population substructures within the  
180 Altaic-speaking populations, suggesting that the geographically diverse sampling populations were the  
181 mixture results of different ancestral sources or proportions. We also identified biogeographically  
182 structured Mongolian populations: western Mongolian with substantial ancestry related to western  
183 Eurasians, and eastern Mongolian with excessive shared alleles associated with Neolithic southern East

184 Asians. We finally provided one case for a better understanding of deep population history via the shared  
185 alleles, haplotypes and paternal lineages.

186

## 187 RESULTS

### 188 General population structure

189 We successfully genotyped around 500K genome-wide SNPs from 510 individuals from Mongolian,  
190 Daur, Ewenki, Hezhen and Han in 40 populations (**Fig. 1A**). We merged our data with population data  
191 of modern and ancient East Asians included in the Human Origin dataset (110K). Eurasian PCA among  
192 3508 individuals clustered studied populations with Eastern Eurasian North-to-South cline  
193 (**Supplementary Figs. 1~3**). The final eastern Eurasian dataset included 2826 individuals from 291  
194 populations. Patterns of population genetic relationship in eastern Eurasian PCA showed an obvious  
195 separation between northern and southern East Asians along PC1, in which the southern East Asians  
196 consisting of Austronesian, Austroasiatic, Tai-Kadai and Hmong-Mien people and the counterpart  
197 consisting of the Altaic-speaking populations (Mongolic, Tungusic and others, **Fig. 1B**). Hmong-Mien-  
198 speaking populations from southwestern China and Mainland Southeast Asia formed one separated clade,  
199 which was different from the Tai-Kadai, Austronesian and Austroasiatic people from South China and  
200 Southeast Asia (**Fig. 1C**). Studied Mongolic and Tungusic people and the previously published ones from  
201 North China showed a separated localization with other northern Mongolic and Tungusic speakers from  
202 southern Siberia, Outer Mongolia and the Amur River. Ancient projected populations  
203 (LateXiongnu\_sarmatian, EarlyMed\_Uigur and EarlyXiongnu\_west) also separated from all included  
204 modern Tungusic and Mongolic people along PC2, but other ancient populations with high Neolithic  
205 northeastern ancestry overlapped with modern Tungusic and Mongolic people. This observed deviation  
206 from Bronze Age/historic Altaic Mountain populations suggested that the primary ancestry of the modern  
207 Tungusic and Mongolic people originated from the eastern ancestral source from the MP or ARB. As  
208 expected, our studied Mongolic and Tungusic populations not only separated with southern and central  
209 Han Chinese but also slightly separated with geographically close northeastern Hans. Neolithic people  
210 from YRB were clustered closely with Hans than with Mongolians. Focused on the genetic variations of  
211 modern populations from North China and southern Siberia, cluster patterns inferred from the PCA  
212 showed more interesting population substructures among them, consistent with the categories of the  
213 subbranches of Altaic languages. PC1 separated Sinitic speakers and Mongolic and Tungusic people, and  
214 PC3 separated Tungusic and Mongolic people (**Fig. 1D~E**)

215

216 Admixture results based on the modern and ancient Eurasians further showed the mixture nature of our  
217 studied populations and other Mongolic and Tungusic people. Different from the homogeneous genetic  
218 structure observed in some ancient or geographically/ethnically representative source populations, most  
219 of the modern East Asians were showed as the mosaic form as a mixture of at least two ancestry  
220 components. Red ancestry was maximized in Boisman\_MN and DevilsCave\_N Neolithic people, which  
221 was previously referred to as Neo-Siberian ancestry or northeastern Asian ancestry in ARB. Light-blue  
222 ancestry was maximized in Neolithic YRB millet farmers from Henan, Shandong and Gansu provinces  
223 (Wanggou\_MN, Lajia\_LN and Erdaojingzi\_LN). Mongolic-speaking Buryat in Russia and Mongols in  
224 Mongolia harbored remarkable red and light-blue ancestries, which showed a strong genetic affinity with  
225 indigenous Siberian ancients. They also had pink ancestry related to western Eurasians and light-green  
226 ancestry related to Russia-Shamanka\_EBA, providing clues of the effect of western Eurasians and  
227 indigenous Baikal ancients on the genetic structure of northern Mongolic populations. Different from  
228 northern Mongolic and Tungusic people, new studied Mongolians and Hans in China possessed much  
229 light-blue ancestry, and limited dark-blue, red and light-green ancestries. Light-green ancestry dominant  
230 in Qiongzong Hlai was one of the representative ancestral sources of southern East Asians, and dark-  
231 blue ancestry related to Nepal ancients was the representative source of indigenous Tibetan ancients.  
232 Southern Altaic-speaking populations in China harbored more Neolithic YRB farmer-related ancestry  
233 and little Siberian-related ancestry (**Fig. 1F**). And more western Eurasian ancestry was visualized in  
234 Turkic people. Similar patterns of predefined ancestry sources and ancestry composition were also  
235 identified with different included reference populations in the model-based ADMIXTURE results  
236 (**Supplementary Figs. 4~7**)

237

### 238 The differentiated demographic history between western and eastern Mongolians

239 From the PCA results in **Fig. 1**, we observed the western Eurasian affinity of Mongolians from Xinjiang,  
240 Qinghai, Gansu and western Inner Mongolia. Ancestry composition inferred from ADMIXTURE  
241 modeling showed that Xinjiang Mongolians had 0.186 western Eurasian pink ancestry, 0.343 Neolithic  
242 Amur red ancestry, 0.283 Neolithic millet farmer light-blue ancestry and some of Neolithic Mongolia  
243 (0.070) and Tibeto-Burman (0.046) ancestry. Thus, we first explored the genetic structure of our newly

244 genotyped 510 individuals and two previously published Hainan Hans and Hlai genotyped using the  
245 same Affymetrix array (**466K**). Plink-based PCA results showed the genetic differentiation between  
246 northern and southern East Asians (**Fig. 2A**). The number and length of the shared ancestry fragment  
247 were more powerful to dissect the finer-scale population structure. We obtained the shared haplotype via  
248 phasing and painting our generated genome-wide SNP data and then explored the coancestry matrix and  
249 clustering patterns. PCA results along the top four components extracted from the coancestry matrix  
250 showed the clear separation between Xinjiang Mongolians and other eastern Mongolians (**Fig. 2B~D**).  
251 Heatmap of pairwise coincidence and clustering dendrogram based on individual-level and population-  
252 level shared chunk counts showed more genetically homogeneous subgroups in accordance (**Fig. 2E~F**).  
253 Southernmost Hans formed the same branch with geographically close indigenous Hlai. New studied  
254 populations were separated into two major branches and more subbranches. One branch consisted of  
255 most of the Mongolic and Tungusic people and the other comprised of northern Hans and some  
256 Mongolians possessing more modern southern East Asian ancestry. We found geographically close  
257 populations shared longer IBD fragments, larger outgroup- $f_3$  values and smaller  $F_{st}$  values (**Fig. 2G~H**).  
258 Two southernmost Chinese Hainan populations had the shortest shared IBD fragments and the largest  
259  $F_{st}$  values with our newly studied northern East Asians. Different from the southern indigenous Hlai  
260 (sharing IBD length within the population is 137.456), the identified shared IBD within geographically  
261 defined northern Mongolians was shorter, which suggested more recent admixture events occurred in the  
262 isolated Hainan populations consistent with the observed longer ROH in Hlai people. We also observed  
263 the Mongolians separated from Northeast Han in the clustering patterns, suggesting their differentiated  
264 demographic history. Modeling of ancestry composition based on the ADMIXTURE showed that the  
265 differentiated ancestry compositions were observed in Xinjiang and other eastern Mongolians (**Fig. 2I**).  
266 Pairwise  $F_{st}$  and inverse outgroup- $f_3$ -values among populations from the merged HO dataset also showed  
267 different genetic affinity between western and eastern Mongolians with their Eurasian references. Indeed,  
268 clustering patterns in the multidimensional scaling plots (MDS), N-J tree and heatmap based on these  
269 two-type genetic distances further confirmed that western Mongolians showed a close relationship with  
270 western Eurasians and Northern Altaic people, but eastern Mongolians were clustered closely with Sinitic  
271 people (**Supplementary Figs. 8~22**).

272  
273 We following assessed the detailed admixture landscape of Mongolians with more reference populations  
274 included in the merged 1240K dataset (**369K**). Results from the pairwise qpWave analysis focused on  
275 our western Mongolians showed multiple differentiated allele-sharing between the left and right  
276 populations, suggesting western Mongolians did not form a clade with eastern Mongolians and other  
277 reference populations relative to these used right populations (**Fig. 3A**). The different patterns of genetic  
278 affinity between western and eastern Mongolians were further confirmed via the different affinities  
279 observed in outgroup- $f_3$  in the form  $f_3(\text{Eurasians, eastern/western Mongolians; Mbuti})$ , diverse admixture  
280 signatures via the observed admixture signatures in the admixture- $f_3(\text{Source1, Source2; western/eastern}$   
281  $\text{Mongolians})$ . Rank results of  $p\_rank1$  (0.0004) and  $p\_rank2$  (0.9567) showed that at least two admixture  
282 events were needed to explain the admixture landscape of western Mongolians. When we focused on the  
283 genetic diversity of the Hezhen, Tu, Oroqen, Daur, and eastern Mongolians in the right populations in  
284 the qpWave analysis, we found they formed a clade with each other, but not with western Mongolians.  
285 Indeed, PCA analysis based on the merged 1240K-dataset showed that eastern Mongolians overlapped  
286 with Daur, Ewenki, and some of northern Hans, but western Mongolians were separated with them (**Fig.**  
287 **3B**). Eastern Mongolians were overlapped with the ancient YRB populations from North China and  
288 Neolithic people in southern Siberia, suggesting their close genetic affinity. The observed significantly  
289 negative values in  $f_4(\text{Eurasians, Eastern Mongolians; YRB farmers, Mbuti})$  and  $f_4(\text{MP/ARB Hunter-}$   
290  $\text{Gatherers; YRB farmers; Eastern Mongolians, Mbuti})$  further showed that eastern Mongolians shared  
291 more alleles with Neolithic millet farmers. We further explored the topologies among Chinese ancients,  
292 East Asians in HGDP, Tai-Kadai-speaking Sui and Zhuang, and Hmong-Mien-speaking Miao using  
293 TreeMix-based phylogeny. Western Mongolians clustered with Uyghur people, however, eastern  
294 Mongolians clustered closely to southern East Asians and YRB ancient populations (**Fig. 3C**). Admixture  
295 results in this merged 1240K-dataset also revealed that western Mongolians harbored more western  
296 Eurasian ancestry (orange ancestry maximized in Srubnaya, Sintashta and Srubnaya). Besides, Western  
297 Mongolians, similar to Yakut people in Siberia, also harbored more ancestry related to the Neolithic MP  
298 people (light-blue). Eastern Mongolians harbored more yellow ancestry related to Neolithic YRB farmers  
299 and purple ancestry related to Tai-Kadai people (Sui and Hlai). Compared with northern Hans and  
300 geographically close northern East Asian ancient people, Mongolians had more western Eurasian orange  
301 ancestry (**Fig. 3D**). Compared to historic Mongols in Mongolia, western Mongolians shared more alleles  
302 with ancient YRB farmers and modern East Asians related to Sino-Tibetans, Austronesian, Austroasiatic  
303 and Tai-Kadai people, as the observed negative values in  $f_4(\text{Late\_Mongol, western Mongolians;}$

304 Eurasians, Mbuti). Differentiated sharing alleles between western and eastern Mongolians were also  
305 evidenced via the observed significant asymmetrical- $f_4$  values in  $f_4$ (western Mongolians, eastern  
306 Mongolians; Eurasians, Mbuti). Differentiated allele sharing patterns were also evidenced in the merged  
307 HO dataset (**Supplementary Figs. 23~38**).

308  
309 Considering the complex admixture sources of northern East Asians identified in admixture- $f_3$ -statistics,  
310 ALDER and GLOBETROTTER, we could successfully fit the admixture model via the three-way  
311 admixture models with ancestry from East Asians, Northeast Asians and western Eurasians. As we  
312 expected, western Mongolians were modeled as an admixture of more ancestry related to Alan, Saka or  
313 Historic SaiduSharif in Pakistan, and the remained ancestry related to Neolithic MP/ARB/YRB people  
314 or southern East Asian Guangxi ancient people. We also evaluated the ancestry composition of  
315 geographically separated populations using qpAdm-based models (**Fig. 4B**). We found that the three-  
316 way admixture models with the aforementioned three ancestral sources could be well-fitted for most of  
317 the included studied populations with variable ancestry proportions. We further reconstructed the deed  
318 demographic history using qpGraph. Here, we included Mbuti, Denisovan, Onge, Loschbour, Tianyuan  
319 and Longshan people to explore the basal model (**Fig. 4B**). We also used the Sintashta, island Neolithic-  
320 to-Iron-Age southern East Asians (Liangdao and Hanben), ARpost9k and Mongolia\_East\_N, as well as  
321 upper YRB farmers (Jinchankou and Lajia) as the ancestral source proxies from western Eurasia, Fujian,  
322 YRB, MP and ARB. We could find that western Mongolians were fitted via 11% gene flow from steppe  
323 Sintashta pastoralists and remained ancestry (89%) from eastern Eurasian ancients related to Neolithic  
324 millet farmers (**Fig. 4B**). Focused on the genomic formation of eastern Mongolians, the admixture  
325 modeling graph showed that they were produced via three ancestral sources: western Eurasian (0.090),  
326 Southern East Asian Hanben (0.291) and upper YRB farmers (0.180). Our results suggested that modern  
327 Chinese Mongolic and some of the Tungusic people were formed via the massive population movement  
328 and interaction of three ancestral sources related to East Asians, southern Siberians, and western  
329 Eurasians.

330

### 331 **Genetic history of Northeast Han Chinese populations**

332 Han Chinese with the largest population size in the world have complex demographic history. Genetic  
333 traces extracted from the ancient genomes supported that Han Chinese were formed via admixture of the  
334 major ancestry from Neolithic YRB farmer ancestry and geographically different indigenes(20). Xu et  
335 al. analyzed the genome-wide SNVs of 114,783 Han Chinese (Han100K) individuals and found six  
336 population subgroups: Northwest Hans, Northeast Hans, Central Hans, Southwest Hans, Southeast Hans  
337 and Lingnan Hans(25). Our recent genetic survey focused on geographically distinct Hans also identified  
338 their differentiated demographic history: Northwest Hans from Gansu harboring more western Eurasian  
339 ancestry; Southwest Hans from Sichuan possessing more genetic influence from Tibeto-Burman speakers;  
340 Southmost Hans from Hainan having more ancestry related to Tai-Kadai people; Southeast Hans from  
341 Fujian possessing more Austronesian-related ancestry; Central Hans from Chongqing, Guizhou, Hubei,  
342 and Hunan harboring equal ancestry related to Neolithic YRB millet farmers and southern indigenes, and  
343 North Hans in Shaanxi and Shanxi possessing dominant local millet farmer ancestry. Whole-genome  
344 sequence data from Taiwan Hans, Singapore and Peranakan Chinese also showed that Hans outside of  
345 Mainland China possessed additional genetic admixture from indigenous Austronesian Ami or Malay(28,  
346 29). However, the genetic origin and population structure of Northeast Hans from Inner Mongolia, Jilin,  
347 Liaoning and Heilongjiang keep uncharacterized. Northeast China was majorly populated by Mongolic  
348 and Tungusic people, but the recent northeastward migration of Hans, such as historic migration events  
349 of Chuanguandong and Zouhukou, as well as the established Yuan and Ming dynasties, changed the  
350 genetic landscape of this region. Thus, we collected 119 Han Chinese individuals from Baotou (25),  
351 Changchun (24), Harbin (24), HulunBuir (21) and Shenyang (25, **Fig. 1A**) to illuminate the genetic  
352 formation of Northeast Hans and explore how they interacted with adjoining Altaic-speaking people.

353

354 Northeast Hans was localized between North Hans from Henan, Shandong and Shaanxi provinces and  
355 Mongolians based on the merged HO dataset and was far away from indigenous Tungusic Ulchi and  
356 Nanai (**Fig. 1B~E**). Five studied Hans shared a similar mixed genetic landscape with North Hans, which  
357 harbored major light-blue and light-green ancestry and minor red and dark-blue ancestry. No significant  
358  $f_4$ -values in  $f_4$ (Studied Han Chinese1, Studied Han Chinese2; Ancient Eurasian populations, Yoruba)  
359 further showed a close genetic affinity within Northeast Hans (**Supplementary Figs. 39**). Han people  
360 were also clustered together based on the co-ancestry matrix-based FineSTRUCTURE and shared-IBD-  
361 based heatmap (**Fig. 2**). Pairwise qpWave focused on the meta-Northeast Hans and one of modern and  
362 ancient northern East Asians showed that Northeast Hans formed one clade with Late Neolithic-to-Iron  
363 Age YRB farmers in Henan province, suggesting a close genetic connection with YRB farmers. Genetic

364 variations extracted from the first two components in the merged 1240K dataset further confirmed the  
365 close relationship between Northeast Hans and middle YRB farmers, as some of them were overlapped  
366 in **Fig. 2B**. Studied Hans were also clustered together in the TreeMix tree and model-based  
367 ADMIXTURE results in the 1240K dataset, in which Northeast Hans possessed more purple Tai-Kadai  
368 ancestry than Henan ancients but lack light-blue Neolithic MP ancestry (**Fig. 2C~D**), suggesting they  
369 received more genetic influence from southern East Asians than from their ancient counterparts. Different  
370 from the Northwest Hans, we could not identify obvious western ancestry in studied Hans based on the  
371 model-based ADMIXTURE modeling, which was following confirmed by the limited western Eurasian  
372 ancestry estimated in three-way qpAdm admixture models (**Fig. 4A~B**). The most shared genetic drift  
373 with one of geographically Northeast Han is another Han population, following by eastern Mongolian in  
374 the outgroup- $f_3$ -statistics. We evaluated the ancestral source composition of Hans using admixture- $f_3$ -  
375 statistics and found that included Hans had a similar admixture landscape as other Chinese Hans, in  
376 which one plausible source was from southern indigenous and the other from northern Altaic or Tibeto-  
377 Burman people. Results from  $f_4(\text{YRB farmers, Siberian ancients; northeastern Hans, Mbuti}) \geq 3 * SE$  and  
378  $f_4(\text{Hans1, Hans2; Eurasians, Mbuti}) \leq 3 * SE$  showed that northeastern Hans were one genetically  
379 homogeneous population, which showed a strong genetic affinity with Neolithic YRB farmers.

380  
381 Additionally, we assessed the genomic formation of Northeast Hans, as well as geographically close  
382 northern minorities (Oroqen, Hezhen, Daur and Japanese), as well as southern Han Chinese and other  
383 ethnic groups (Miao, Jing, Tujia and Gelao) from South China using the qpAdm analysis with different  
384 representative ancestral sources in the merged 1240K dataset. We first modeled mixture proportion with  
385 early ARB ancients as the northern surrogate sources and Late Neolithic Coastal South East Asians  
386 (Tanshishan and Xitoucun) as the southern surrogate sources (**Fig. 5A**). Five Northeast Hans were mixed  
387 via 0.506~0.586 ARB ancestry and 0.494~0.414 southern East Asian ancestry. Compared with Northeast  
388 Hans, Northeast minorities harbored more Neolithic northeastern East Asian ancestry ranging from  
389 0.599±0.047 in Heihe Ewenki to 0.851±0.043 in Oroqen. Southern Hans, as expected, had more southern  
390 East Asian ancestry ranging from 0.565±0.039 in HGDP Han to 0.583±0.039 in Zunyi Han, and southern  
391 ethnically specific minorities also harbored more southern East Asian ancestry ranging from 0.566±0.043  
392 in Tai-Kadai-speaking Gelao to 0.812±0.045 in Tai-Kadai-speaking Zhuang. We following used late  
393 Neolithic YRB farmers as the northern representative ancestral source, as China\_YR\_LN shared the most  
394 ancestry with Northeast Hans in the PCA, ADMIXTURE, outgroup- $f_3$ -based allele sharing. Only  
395 HulunBuir and Harbin Hans could be successfully modeled as major (0.871~0.842) ancestry related to  
396 YRB farmers and minor (0.129 ~0.158) ancestry related to Guangxi LaCen people. All included  
397 populations sampled from the south of YRB could be fitted as 0.133~0.613 YRB millet farmer ancestry  
398 and 0.867~0.387 southern historic Guangxi ancestry. We further used two ancestral sources respectively  
399 from YRB and ARB to model the genetic formation of included northeastern populations (**Fig. 5C**). We  
400 confirmed the significant influence of Neolithic ARB ancestry on the formation of the gene pool of  
401 Northeast Hans, but Mongolic and Tungusic people possessed more ARB ancestry ranging from 0.086  
402 in Ewenki and 0.881 in Oroqen people. The best-fitted qpGraph models focused on Hans also supported  
403 the limited gene flow from western Eurasian into Northeast Hans, in which Northeast Hans were fitted  
404 as the 24% ancestry related to Southeast coastal Iron Age Hanben people and 76% ancestry from  
405 Neolithic upper YRB farmer ancestry (**Fig. 5D**).

#### 406 407 **Dominant Neolithic Mongolian Plateau ancestry in Tungusic and Mongolic people in Siberia and** 408 **Amur River Region**

409 Ancient DNA researches from the MP, ARB, Baikal Lake regions and other eastern Eurasian steppe have  
410 demonstrated dynamic population admixture history of ancient eastern Eurasian populations. Sikora et  
411 al. recovered 34 31,000~600-year-old ancient genomes and identified three ancient representative  
412 populations localized in northeastern Siberia, including Ancient North Siberians, Ancient Palaeo-  
413 Siberians, and Neo-Siberians(7). Jeong et al obtained large-scale ancient genomes from Outer Mongolia  
414 and Baikal Region and found the initial indigenous MP ancestry, the first genetic communication between  
415 western and eastern Eurasian, as well as the repeated and complicated mixed history of multi-ancestral  
416 sources since Bronze Age to historic empire periods(9). Mao et al. recently sequenced upper Paleolithic  
417 to the Iron Age genomes in the ARB and identified the population migration and replacement in this  
418 region 9000 years ago and observed a regional genetic continuity since then(14). Ancient findings from  
419 the easternmost Eurasian steppe also identified the genetic stability from 7700-year-old ancient  
420 DevilsGate people to modern Tungusic Ulchi and Nanai people. Recent linguistic and genetic evidence  
421 also evoked the controversial opinion of the origin of the modern Trans-Eurasian (Altaic) language and  
422 corresponding people. Linguists proposed modern Altaic people were originated from West Liao River  
423 regions associated with the dissemination of millet agriculture (farming-and-language-dispersal

424 hypothesis). However, recent ancient DNA traces provided another possibility that the population  
425 expansion of the Hunter-Gatherer people in the MP dispersed the Proto-Altai language. Although  
426 previous genome-wide SNPs have been used to solve this problem, modern comprehensive and deep  
427 population history focused on the genetic variations of modern people should be further conducted to  
428 illuminate how previously documented representative sources contributed to modern northern East  
429 Asians.

430  
431 Except for the denser sampling of Mongolians, we also genotyped Mongolic Daur and Tungusic Ewenki  
432 and Hezhen people. We following combined our data with 25 previously published Tungusic and Mongolic  
433 populations in China, Mongolia and Russia to comprehensively characterize the potentially existed  
434 differentiated population history and their relationship to the ancient people, as well as the possible  
435 association between population migration and the spread of Trans-Eurasian language. HulunBuir and  
436 Tsitsihar Daur had a close genetic relationship with each other and clustered together in PCA,  
437 ADMIXTURE and FineStructure-based population clustering patterns (**Fig. 1~2**) in the merged HO  
438 dataset. Hezhen and Ewenki were clustered closely with Mongolian people. Similar mixed ancestral  
439 sources and landscape of ancestry components were confirmed based on the 1240k dataset via  
440 asymmetrical- $f_4$ -tests in the form  $f_4(\text{pop1, pop2; Eurasians, Mbuti})$ . Our studied and previously sequenced  
441 Hezhen and Daur could be fitted via three-way admixture qpAdm models, in which Hezhen people  
442 harbored more ancestry related to Neolithic ARB ancients (**Fig. 3**). Dominant Neolithic ARB ancestry in  
443 Mongolic and Tungusic people was further confirmed via the three-way admixture models focused on  
444 non-Meta populations in the merged HO dataset (**Fig. 4**).

445  
446 Eastern Eurasian PCA has revealed the separation between northern and southern Mongolic people as  
447 well as between the Mongolic and Tungusic people (**Fig. 1**). Thus, to further comprehensively  
448 characterize the landscape of these populations, we collected 235 Altaic-speaking individuals from 26  
449 populations to conduct population structure analyses based on the shared alleles and haplotypes except  
450 for Turkic people whose population genetic history has been recently explored(30). We could identify  
451 four population clusters respectively associated with the northern (pink circle) and southern (orange)  
452 Mongolic people, Japanese and Korean (dark-green), as well as the Tungusic people (yellow) based on  
453 PCA patterns reconstructed from the allele frequency spectrum and co-ancestry matrix (**Fig. 6A~E**). Four  
454 ancestral sources were further confirmed via the ADMIXTURE results with four predefined ancestral  
455 sources. Here, we also confirmed the mixed genetic structure of Mongolian, Hezhen and Xibo consisting  
456 of major southern Mongolian ancestry and some of the Japanese and Tungusic ancestry, as well as Daur  
457 and Oroqen composing of four ancestries (**Fig. 6F**). Pairwise  $F_{st}$  within 26 populations first revealed  
458 significant genetic differentiation between Tungusic-speaking Ulchi, Negidal and Nanai with others,  
459 following showed the differences between Japanese with other references. Siberian Buryat people  
460 showed a close genetic relationship with each other, following with Mongols in Mongolia, and distant  
461 with southern Mongolic people (**Fig. 6G**). Longer sharing IBD observed in Nanai, Negidal and Ulchi  
462 than others showed their unique genetic structure and higher inbreeding phenomenon. Similar patterns  
463 of longer sharing IBD fragments were further identified in Siberian Buryat and Mongols. Other Chinese  
464 Mongolic and Tungusic people had relatively shorter IBD chunks, suggesting more frequent population  
465 movement and admixture (**Fig. 6H**). Differentiated population genetic history of  
466 geographically/ethnolinguistically diverse Altaic populations and corresponding clustering patterns were  
467 further confirmed via the sharing number of IBD chunks in FineStructure-based dendrogram and  
468 TreeMix-based phylogeny (**Fig. 6I~K**).

469  
470 We following used three-way and two-way admixture models to assess the genetic sources and admixture  
471 proportion of 42 Altaic-speaking populations (included Turkic people). All included populations except  
472 for four Tungusic-speaking populations (Negidal, Ulchi, Evenk-FarEast and Nanai) could be well-fitted  
473 via the three-way admixture models with three representative ancestral sources from northeastern Asia  
474 (Russia\_OldBeringSea\_Ekven, Fofonovo\_EN, DevilsCave\_N and Boisman\_MN), YRB (Xiaowu\_MN,  
475 Pingliangtai\_LN and Haojiatai\_LN) and western Eurasia (Russia\_Andronovo, Iran\_C\_TepeHissar and  
476 Russia\_EBA\_Yamnaya\_Samara). In the Palaeo-Siberians-YRB-farmers-western-Eurasians models (**Fig.**  
477 **7A**), Altaic-speaking populations possessed major Neolithic YRB Xiaowu farmer ancestry ranging from  
478  $0.462 \pm 0.013$  in Even to  $0.907 \pm 0.020$  in Yugur (Sunan) and non-ignorable Palaeo-Siberian ancestry  
479 ( $0.053 \pm 0.021$  in Yugur to  $0.339 \pm 0.014$  in Yukagir) and western Andronovo pastoralist ancestry  
480 ( $0.033 \pm 0.011$  in Daur to  $0.915 \pm 0.009$  in Veps). Mongolic and Tungusic people harbored more Middle  
481 YRB Xiaowu farmer ancestry than Turkic people ( $0.61 \sim 0.865$  versus  $0.069 \sim 0.597$ ). Most of included  
482 Mongolic and Tungusic populations could be well-fitted via the Neolithic-Mongolians-YRB-farmers-  
483 western-Eurasian models (**Fig. 7B~C**) with dominant Neolithic Mongolia Fofonovo ancestry in non-



484 Chinese ethnic groups (ranging from  $0.661\pm 0.070$  in Mongol\_Uuld to  $0.880\pm 0.030$  in Khamnegan) and  
485 dominant Late Neolithic YRB Pingliangtai ancestry in Chinese groups ( $0.648\pm 0.057$  in Dongxiang to  
486  $0.806\pm 0.064$  in Yugur). Different from the ancestry composition of Tungusic and Mongolic people, we  
487 identified a significant effect of both western Eurasian (Iran\_C\_TepeHissar:  $0.228\pm 0.013$  in  
488 Kazakh\_Aksay to  $0.762\pm 0.014$  in Nogai) and Neolithic Fofonovo ancestry ( $0.168\pm 0.060$  in Nogai to  
489  $0.638\pm 0.02$  in Kazakh\_Aksay) on the genetic formation of Turkic people. Similar patterns of qpAdm-  
490 based genomic affinity were confirmed via the DevilsCave\_N-Haojiatai\_LN-Yamnaya three-way  
491 admixture model (**Fig. 7C**). we also confirmed that Chinese Mongolic and Tungusic people had dominant  
492 Neolithic Yellow farmer ancestry via the well-fitted two-way admixture models (**Fig. 7D-F**). Our  
493 findings supported that geographically different Altaic-speaking populations harboring different genetic  
494 landscapes with differentiated demographic history: Northern one with dominant Neolithic Mongolia  
495 gene flow, southern one possessing significant influence from YRB farmers, and western one harboring  
496 extensive genetic admixture with steppe pastoralists and Iran farmers.

#### 497 498 **Genetic relationships between northern Chinese populations and historic hierarchical and** 499 **centrally organized empires in the Eastern steppe**

500 More and more ancient genomes of historic pastoralists were recently reported(9), including Xiongnu,  
501 Uyghur, Khitan, Turkic and Mongols. Jeong et al. used the ancient and modern genome-wide data from  
502 Siberia and Mongolia to perform the individual-based qpWave analysis and demonstrated that genetic  
503 profiles of northern Mongolic-speaking populations have not changed since the Mongol Empire(9).  
504 However, the genetic contribution from historic pastoralists in the MP into modern northern Chinese  
505 ethnic groups needed to be further characterized.

506  
507 To explore the shared ancestry between modern East Asians and historic ancient populations, we merged  
508 our data with modern Mongolic and Tungusic people in the HGDP dataset, as well as the ancient genomes  
509 from the MP, which included some Bronze Age populations and historic Xiongnu, Khitan, Turk, Uigur  
510 and Mongol. PCA analysis showed the genetic differentiation between studied populations and ancient  
511 Mongolian populations, but some late Medieval Mongols overlapped with eastern Mongolians, and some  
512 western Mongolians overlapped with Turk and Xiongnu people (**Fig. 8A**). Shared genetic drift estimated  
513 using outgroup- $f_3$ -statistics revealed that eastern Mongolians possessed stronger East Asian affinity with  
514 northern Hans, Heihe, Daur, Japanese and Hezhen, but distant with western Mongolians ( $0.3177$ ).  
515 Western Mongolians, Daur, Han and Ewenki also shared more ancestry with each other and then with  
516 modern northern East Asians not with ancient populations. Heatmap based on the outgroup- $f_3$  matrix  
517 showed two major branches, one consisted of ancient populations harboring more western ancestry  
518 (Chandman, Turk and Uigur), and the other comprised modern northern Asian populations and  
519 Ulaanzuukh\_SlabGrave, Xiongnu and Mongol populations (**Fig. 8B**). Admixture signatures inferred  
520 from admixture- $f_3$ -statistics focused on the eastern Mongolians revealed that the best ancestral sources  
521 were from Hans and Mongol ancients. Different from the eastern Mongolians, western Eurasian-related  
522 populations (LateXiongnu\_sarmatian, Chandman\_IA, EarlyMed\_Uigur, Altai\_MLBA and  
523 EarlyMed\_Turk) in MP were evidenced as one of the best sources for western Mongolians. Daur and  
524 Ewenki possessed the most ancestral sources deriving from the southern source related to Hlai and the  
525 northern source linked to Yakut. Ulaanzuukh\_SlabGrave people were evidenced as the ancestry surrogate  
526 for all studied northern East Asians. Our results from the shared genetic ancestry showed that ancient  
527 Xiongnu and Mongols contributed more ancestry into modern northern Tungusic and Mongolic people.  
528 But the proportion of ancestry contribution was different, as the differentiated population genetic  
529 structure revealed via the TreeMix-based phylogeny and fine-scale substructure inferred from the  
530 FineSTRUCTURE (**Fig. 8C-D**). Under the genetic variations of these included modern and ancient  
531 populations, we conducted the model-based ADMIXTURE analysis to explore the ancestry composition.  
532 We found three ancestry components maximized in Han Chinese, Neolithic Mongolian ancestry  
533 (Fofonovo, Ulaanzuukh and SlabGrave), and western Eurasian-related Xiongnu ancestry respectively.  
534 And Yakut ancestry was separated from Neolithic ancestry when we predefined four ancestral sources in  
535 the ADMIXTURE-based model fitting. We found a significant genetic difference between Mongolia  
536 historic ancients and modern northern Chinese populations. We finally reconstructed the best-fitted  
537 qpGraph-based admixture graph and found different genetic lineages between Mongolia historic ancients  
538 and modern northern Chinese groups with different population compositions (**Fig. 7G-I**). We found  
539 eastern Mongolians shared the major lineages with YRB farmers and northern Yakut, and western  
540 Mongolians shared common lineages with historic Turkic and Xiongnu. Here, we found that western  
541 Mongolians shared a similar ancestry history with late Medieval Mongols, which was descended 0.440  
542 ancestry from YRB farmers, 0.4592 from ARB Hunter-Gatherers, and 0.1008 from western Eurasian  
543 Andronovo (**Fig. 7G-H**). We also confirmed that early Medieval Turkic derived 68% ancestry from

544 Neolithic MP people. Ancient Turkic people also contributed 30% of genetic materials to western  
545 Mongolians whose remaining ancestry derived from eastern Mongolian-related ancient sources (**Fig. 7I**).  
546 **East Eurasian origin of Mongolic-speaking Kalmyk in northern Caucasus Mountain**  
547 More genome-wide-scale population genetic analysis was needed to explore the genetic impacts of the  
548 westward expansion of the Mongol empire in the Eurasian steppe on modern central Eurasians.  
549 GLOBETROTTER-based admixture sources and dates showed that eastern Mongolian-related haplotype  
550 was directly observed in modern central Uyghur, Hazara and others with different proportions, such as  
551 8% in the Turkish(31). Recent findings based on the forensic STRs and SNPs identified the genetic link  
552 between geographically distinct Torglut Mongolians and 3000 kilometers away Jalaid Mongols(32). The  
553 identified shared alleles between Chinese Mongolians and Hazara also supported the hypothesis of  
554 modern Hazara were the descendants of ancient Mongols(33). Kalmyk is one of the Mongolic-speaking  
555 populations residing in the Yashkul, Republic of Kalmykia Russia. The genetic origin, population  
556 structure and admixture history were formally explored here.

557  
558 Although the geographical origin of our focused Kalmyk was in North Caucasus, the clustering patterns  
559 inferred from PCA and ADMIXTURE showed that the genetic localization of Kalmyk was East Eurasia  
560 (**Fig. 1**). The model-based ancestry composition revealed that Kalmyk people possessed major ancestry  
561 related to Neolithic ARB people (0.426), and some related to Neolithic YRB Longshan (0.187), western  
562 Eurasians (0.196) and Neolithic MP (0.100), and minor related to southern East Asian indigenes. These  
563 observed landscapes were consistent with the genetic structure observed in the northern Mongolic people  
564 (**Fig. 1F**). The small allele frequency difference between Kalmyk and Xinjiang Mongolian, Siberia  
565 Mongols and Buryats (0.0005~0.0034), suggested their genetic affinity and the common evolutionary  
566 history. Outgroup- $f_3$  values further showed that Kalmyk shared the most genetic drift with Tungusic-  
567 speaking populations and Neolithic ARB ancient of DevilsCave\_N ( $f_3 > 0.2050$ ). Admixture- $f_3$  analysis  
568 showed the ancestral sources of Kalmyk were eastern Tungusic people and western Indo-European  
569 groups, such as  $f_3(\text{Nanai, French; Kalmyk}) = -38.165 * SE$ . Genetic relationships within Mongolic and  
570 Tungusic populations inferred from PCA, TreeMix, heatmap and FineSTRUCTURE (**Fig. 6**) grouped  
571 Kalmyk with Mongolians. Kalmyk also shared longer IBD chunks with eastern Eurasians than with  
572 others. Formal testing of ancestry proportion estimation confirmed that Kalmyk had an East Asian origin  
573 with an admixture of major YRB millet farmers or Neolithic Mongolian ancestry and some western  
574 Eurasians (**Fig. 7A~C**).

#### 575 576 **Natural selection signatures in northern East Asian Mongolian and Tungusic people**

577 The cross-population extended haplotype homozygosity (XP-EHH) was calculated using our newly  
578 genotyped Mongolians as the target and Guizhou Hans as the reference populations (**Fig. 1**). Many gene-  
579 coding candidate selection regions from Chromosomes 1, 4, 6 12 and 16 were identified via the XP-EHH.  
580 Here, we highlighted the top 803 SNPs which were clustered and located in different candidate regions  
581 of the genomes. Nine SNPs (1: 77854238-77895554) located in the AK5 gene and six SNPs (1:  
582 145444556-145524224) in NBP10 were the major candidate regions in Chromosome 1, other signatures  
583 of natural selection from ITGA10 (3 SNPs) and CAPN2 (3 SNPs) were also identified in this  
584 chromosome. Linked haplotype block consisting of fifty-one SNPs located in the CYP19A1 (15:  
585 51502844-51591204) was the longest genomic region under selection in Mongolian populations, which  
586 were associated with the etiology of breast cancer. Twenty-six SNPs were located in HLA-C (6:  
587 31236998-31239912) with XPEHH values ranging from -6.121 to -4.805. Twenty-four SNPs located in  
588 HLA-A (6: 29910276-29913542) and thirteen SNPs in HLA-B (6: 31321882-31324889) and ten loci in  
589 HLA-DPA1 (6: 33035771-32555987) also showed significant natural selection signatures. Twenty-three  
590 SNPs in CPNE8 (12: 39050589-39261882) possessed the XPEHH values ranging from -5.692 to -7.260.  
591 Thirteen SNPs located in the gastric alcohol dehydrogenase gene (ADH7, 4: 100333932-100349135)  
592 showed obvious selection signatures, suggesting that alcohol metabolism in Mongolian people  
593 underwent selection during the pastoralist subsistence strategy, which was also evidenced via our  
594 estimated iHS results. Besides, at least five functional SNPs in these genes (TRIM31, PHKB, TRIM40,  
595 HLA-DRB1, SCN4A, CTSB, GLDN, NBP10, PGM5, NELL2, GALNT6, ITFG1, PLB1, EPHB1 and  
596 NREP) also showed selection signatures, most of which were further confirmed via the estimated high  
597 iHS values.

#### 598 599 **Paternal and maternal admixture history of northern East Asians**

600 Finally, we explored gendered dimensions of the population history of studied northern East Asians. All  
601 362 Mongolians were assigned into 204 terminal matrilineal haplogroups. Haplogroup D was the  
602 most predominant lineage (27.07%), followed by B (11.60%), F (10.77%), Z (8.01%), G (7.73%), C  
603 (6.91%), A (6.08%), N (5.25%), and M7 (5.25%), other haplogroups (HV, H, I, M8, M9, M10, M11, R,

604 T, U, W, and Y) were sporadically distributed in studied Mongolians with frequencies of no more than  
605 1.66%. Target Mongolic-speaking Daur people were assigned into 19 unique matrilineal lineages.  
606 Haplogroup B was the most dominant lineage (26.32%), followed by G (15.79%), F (10.53%), D  
607 (10.53%), and A (10.53%), in addition, haplogroups C, N, M8, M11, and R were observed once  
608 respectively. Haplogroups B (2/10), C (1/10), D (3/10), F (1/10), M8 (1/10), N (1/10), and R (1/10) were  
609 observed in studied Tungusic speakers. All 119 Han Chinese were assigned to 94 terminal matrilineal  
610 haplogroups. Haplogroup D was the most common lineage (23.53%), followed by B (12.61%), F  
611 (10.08%), M7 (9.24%), N (7.56%), Z (6.72%), G (6.72%), A (6.72%), and C (5.88%), other haplogroups  
612 (M8, M9, M10, M11, T, U, and Y) were sporadically distributed in Tungusic individuals with frequencies  
613 of no more than 3.36%. Among 175 male Mongolians, we identified 80 terminal paternal lineages with  
614 frequencies ranging from 0.0057 to 0.0629 (O2a1c1a1a1a1e: 11). The most frequent paternal haplogroup  
615 in target Mongolians was O2 (49.14%), followed by C2 (22.86%), O1 (12.00%), and N1 (6.29%).  
616 Furthermore, haplogroups D1, E, I, G, Q, and R were sparsely distributed in studied Mongolian  
617 populations. We observed the distributions of haplogroups C2, N1, O1, and O2 in Mongolic-speaking  
618 Daur individuals and haplogroups C2, O1, and R in studied Tungusic groups. Sixty Han Chinese men  
619 were assigned into 42 diverse paternal lineages. Haplogroup O2 was the most prevalent lineage (45.00%),  
620 followed by C2 (26.67%), N1 (10.00%), O1 (10.00%), Q (5.00%), D1 (1.67%), and J (1.67%).

621

## 622 Discussion

623 The genetic, archaeological and anthropological findings have demonstrated that the Eastern Eurasian  
624 populations had experienced several waves of population turnover. Ancient North Siberians (Yana)(7),  
625 ancient northern Eurasians (Mal'ta)(3) and 330,00-year old ARB served as the first batch of pre-LGM  
626 Paleo-Neolithic populations, which possessed higher East Asian Tianyuan-related ancestry. Ancient  
627 Palaeo-Siberians, Neo-Siberians(7) and other East Asians from YRB, Guangxi and Fujian served as the  
628 second batch of founding lineages that worked as the key components to participate in the formation of  
629 Holocene to modern northern East Asians. Furthermore, the Yamnaya and Afanasievo people in the  
630 Bronze Age had dominated the eastern Eurasian steppe, and these steppe herders gradually expanded  
631 eastward to the Altaic Mountains(11, 22) and even reached Shirengigou of Xinjiang in northwestern  
632 China(24) in the Iron Age and central region of Mongolia in the Bronze Age(9). Besides, the ancient  
633 genomes from MP also documented the genetic legacy of Iran farmer-related ancestry. Thus, ancient  
634 western Eurasians related to northern steppe pastoralists and southern agriculturalists served as the third  
635 batch of ancestral sources. The reconstructed deep population history in North-East Asia was dynamic.  
636 Recent findings from ARB supported over 13,000-years of genetic stability in ARB, and complex  
637 interactions and admixture history between local ancestry and surrounding incoming ancestry were  
638 spatiotemporally documented in YRB, MP and Baikal Lake regions. Damgaard et al. provided the first  
639 batch of ancient genomes from 74 ancient individuals across Inner Asia and Anatolia and found that there  
640 was a distinct East-West cline in the ancient Eurasian steppe populations, namely, eastern European  
641 Hunter-Gatherers harbored the highest proportion of western ancestral components, and Neolithic  
642 populations (such as Shamanka) of Lake Baikal harbored the highest proportion of eastern ancestral  
643 components(12). The genetic findings from 101 newly recovered ancient genomes that dated to between  
644 3000 and 1000 years ago showed that the Bronze Age was a highly dynamic period involving large-scale  
645 population migrations and replacements, which also provided supporting evidence for the spread of Indo-  
646 European languages during the Early Bronze Age(26). In 2018, Damgaard et al. further sequenced the  
647 genomes of 137 ancient humans in Eurasia and found that population substructures existed in the  
648 Scythian groups that dominated the Eurasian steppe throughout the Iron Age(10). However, the early  
649 population migrations in Eurasia had not significantly affected eastern Eurasia. A recent paleogenomic  
650 study of 22 Khovsgol individuals dated to between 5300 to 4700 years ago indicated that dairy  
651 pastoralism was introduced through a process of cultural transmission rather than population  
652 replacement(23). Highly spatiotemporally sampling ancient DNA from Mongolia further characterized  
653 the dynamic history of the Eastern Steppe, which documented the tripartite population structures in the  
654 Bronze Age and multiple population admixture in historic periods(9). How these ancient admixture  
655 events influenced the modern genetic landscape in North-East Asia needed to be comprehensively  
656 explored.

657

658 We have presented a comprehensive characterization of genetic variations in 510 northern East Asian  
659 Mongolic, Tungusic and Sinitic samples by merging with all available Eastern Eurasian modern and  
660 ancient genomes, which is the largest genome-wide genetic study of the Altaic-speaking populations in  
661 the boundary regions between South China and Siberia. We documented population stratification within  
662 northern East Asians, but all of them possessed the dominant ancestry related to Neolithic MP/ARB  
663 people. This documented common genetic legacy in Altaic-speaking populations was consistent with the

664 ancient findings of genetic continuity from 13,000-year-old ARB people to 7700-year-old  
665 DevilsGate/Boisman and then to modern Tungusic Ulchi. Shared common ancestry from eastern  
666 Eurasian lineage among modern Altaic people was also evidenced via the longer length shared IBD  
667 fragments among modern central Asian Turkic people from the southern Siberians(30). Previous Trans-  
668 Eurasian language origin hypothesis stated that the language subfamily of Mongolic, Tungusic, Turkic  
669 and others shared cultural elements with the Hongshan culture in the West Liao River Basin. If the  
670 hypothesis is correct, we expect to observed dominant Neolithic Hongshan ancestry in modern Altaic  
671 people in the ADMIXTURE, FineSTRUCTURE, qpAdm and qpGraph-based admixture models.  
672 However, we observed dominant Neolithic Hunter-Gatherer-related ancestry from MP/ARB in modern  
673 northern East Asians. Considering the co-dispersal of language, gene and culture, as well as the  
674 established genetic continuity of deep population history in ARB and surrounding regions, our  
675 established genetic landscape in North-East Asia supported Altaic language originated from Northeast  
676 Asia. Moreover, archeologically documented stability in the material cultures and anthropologically  
677 attested similarity of morphology in MP/ARB further supported our observed dominant local ancestry in  
678 northern Altaic people in ARB and neighboring regions. Here, we also noted that we lack more detailed  
679 robust evidence to confirm the exact geographical locations of the Proto-Altaic language in Northern  
680 East Asia. More complex originated landscapes maybe existed in East Eurasia, such as Robbees et al.  
681 provided evidence supporting for new ‘farming Hypothesis’ and against the traditional ‘Pastoralis  
682 Hypothesis’ by integrating linguistics, archaeological and genetic phylogenetic findings(34). Deeper and  
683 more comprehensive studies based on cultural reconstruction, linguistic diversity, lexicostatistics,  
684 Bayesian phylogeography and other interdisciplinary approaches should be conducted to reconstruct a  
685 more accurate and complete picture of the evolutionary history of Proto-Altaic people and languages.

686  
687 Western Eurasian gene flow significantly shaped the genetic structure of western and northern Altaic-  
688 speaking populations. Three-way qpAdm models provided good fitness for Turkic and northern  
689 Mongolic people, which was also observed in our qpGraph-based phylogenetic framework. ALDER-  
690 based and GLOBTTOTER-based date results supported the admixture events that occurred in the historic  
691 times. We should pay attention that only single simple or complex models with double dates and sources  
692 were considered, which provided the simplest framework of evolutionary history. The truth admixture  
693 scenarios may be continuous admixture, which could make a more recent admixture times in the ancient  
694 admixture event dating. Ancient population contact between western pastoralists and eastern Eurasians  
695 has indeed been attested in the early ancient populations. Early paleo-genomic studies have shown that  
696 significant genetic differences existed in ancient Eurasian individuals at different times and geographic  
697 scales. But the Yamnaya and Afanasievo populations spread eastward and westward across the Pontic  
698 Caspian steppe between 3000 BC to 2100 BC, which was accompanied by the diffusion of the early and  
699 middle Bronze Age steppe population-related ancestral components(26). The ancient individuals related  
700 to Sintashta, Srubnaya and Andronovo cultures further continued to spread westward and eastward to  
701 generate the middle and late Bronze Age steppe population-related ancestral components(10). From 100  
702 BC to 200 BC, Scythians with significantly different genetic structures were active in the eastern, central  
703 and western of the Eurasian steppe, respectively(35). These westward movements may also leave genetic  
704 legacies in modern eastern Eurasians as identified in our qpAdm and qpGraph admixture models.

705  
706 With this unprecedented data of Mongolians, we also comprehensively revealed the genetic origins,  
707 admixture history, and population structure of geographically diverse Chinese Mongolian populations.  
708 Our results showed that the Mongolian people were most closely related to East Asian populations  
709 compared with other modern and ancient global populations, especially the northern Tungusic and  
710 Mongolic people, suggesting their eastern Eurasian origin hypothesis. Moreover, eastern Mongolians  
711 shared most of the genetic makeup with the geographically close Han Chinese populations and Neolithic  
712 YRB farmers. Interestingly, eastern Mongolians and Han Chinese were genetically distinguishable from  
713 each other, as well as there was a different genetic structure between eastern and western Mongolians,  
714 which could attribute to the differentiated Siberian and western Eurasian ancestry in eastern and western  
715 Mongolians, and different proportions of YRB ancestry in Northeast Hans. Indeed, all results from allele-  
716 shared  $f$ -statistics and haplotype-shared chromosome painting found remarkable differences in the  
717 genetic makeup between Mongolians and Hans, as well as eastern and western Mongolians. Specifically,  
718 at least three major ancestral components were identified in the Mongolic and Tungusic people, which  
719 were potentially derived from ancestral populations in Western Eurasia, YRB, and MP/ARB. In contrast,  
720 two major ancestral components from YRB and southern China were identified in Northeast Hans.  
721 QpAdm/qpGraph-based modeling admixture history of northern Altaic and western Mongolians  
722 indicated that these three identified ancestral components were derived from Neolithic ancestors or their  
723 admixed descendants. The precise source of eastern ancestry for Mongolians was difficult to constrain

724 but our results showed that Neolithic YRB people contributed substantial genetic materials to the eastern  
725 Chinese Mongolic populations. The intense population movement and gene flow between ancient  
726 Xinjiang people and western Eurasians, as well as the deep connection between Xinjiang local Xiaohu  
727 people and South Siberians, have been reported via autosomal and mitochondrial DNA, which provided  
728 an explanation for our observed differentiated genetic structure in western Mongolians possessing more  
729 ancestry from western Eurasia and ARB/MP.

730  
731 Our findings also demonstrated that the genetic landscape of modern westernmost Mongolic Kalmyk  
732 people was a result of the genetic origin, migration, and admixture history of multiple ancestral sources.  
733 Interestingly, among the Mongolic populations in Eurasia, the genetic differentiation between Chinese  
734 eastern and western Mongolians and others was even larger than that between Kalmyk and other eastern  
735 Mongolic populations, though ARB and Pontic-Caspian steppe were geographically apart. Our results  
736 suggested that Kalmyk was genetically closer related to Tungusic people in ARB. According to historical  
737 documents, the Kalmyk people were the Mongolian-related nomadic pastoralist regimes who migrated  
738 from South Siberia into North Caucasus during the Yuan and Qing dynasties. ADMIXTURE analysis  
739 suggested that there was no considerable gene flow between Kalmyk and surrounding populations after  
740 Kalmyk people migrated from Siberia into West Asia, which could be attributed to that they speak the  
741 Mongolic language, whereas most of the surrounding populations speak Indo-European language.  
742 Focused on the relationship between historic pastoral empires and Chinese Mongolians, we found a  
743 significant effect of these historic and prehistoric admixture events on the ethnolinguistic diversity of  
744 northern and western modern Altaic people, but the limited contribution to Chinese eastern Mongolians  
745 compared to the influence from the Han expansion. The Xiongnu confederations grew strong in eastern  
746 Eurasia and moved westward about the second or third century BC(27, 36). Subsequently, Turkic,  
747 Mongolian and other people successively dominated inner Asia from 600 AD to 1500 AD(27, 36, 37).  
748 The genetic impact of the intense east-to-west population movement on the genetic landscape of modern  
749 Altaic speakers was stronger than that caused by the north-to-south population migration although  
750 historic Chinese Yuan and Qing dynasties were established via the predecessors of modern Mongolic and  
751 Tungusic people. Finally, considering the unique and complex history of Mongolic and Tungusic  
752 populations, the genome-wide SNP data generated in this study is of great significance for the genetic  
753 studies of northern East Asians and serves as a useful control data set for genetic association studies.

754

## 755 **Conclusions**

756 Our population genomic results suggested a complex scenario of admixture history of Altaic-speaking  
757 populations in North-East Asia via comprehensively analyzing the newly generated genome-wide SNP  
758 data of 31 Mongolic-speaking, two Tungusic-speaking, and five Sinitic-speaking populations in China.  
759 The results showed that there were significant population substructures within Altaic-speaking  
760 populations (northern and southern Mongolic, Tungusic, and Turkic-speaking populations) and between  
761 eastern and western Mongolians. All Altaic-speaking populations were a mixture of dominant Siberian  
762 Neolithic ancestry and non-negligible YRB ancestry, suggesting that Altaic-people and their language  
763 were more likely to originate from the Northeast Asia (mostly likely the ARB and surrounding regions  
764 as the primary common ancestry identified here) and further experienced influence from Neolithic YRB  
765 farmers. All Altaic people but eastern and southern Mongolic-speaking populations possessed a high  
766 proportion of West Eurasian-related ancestry, in accordance with the linguistically documented language  
767 borrowing in Turkic language. Our findings supported the “ARB-origin hypothesis” over the “Baikal-  
768 origin hypothesis” for the origin of Mongolic and Tungusic-speaking populations, as they possessed  
769 documented dominant ARB ancestry. Moreover, the genetic makeup of Mongolic-speaking populations,  
770 especially southern, central and eastern Mongolic-speaking populations, harbored more YRB Neolithic  
771 farmer ancestry and a stronger genetic affinity with modern Northeast Hans, suggesting the extensive  
772 genetic admixture between Chinese Mongolians and adjacent Hans. Finally, we identified a close genetic  
773 connection between the Mongolic-speaking populations from North Caucasus and ARB, suggesting  
774 long-distance migration of Altaic-speaking people since the Mongolian Empire periods, and these  
775 Mongolic groups kept relatively genetically isolated with surrounding Indo-European-speaking  
776 populations.

777

## 778 **Materials and methods**

### 779 **Ethics statement**

780 This study was approved by the Medical Ethics Committee of Xiamen University. Informed consent was  
781 obtained from all included individuals before the saliva or blood collection. The procedures of sample  
782 collection and experiment of this research were in accordance with the recommendations provided by  
783 the revised Helsinki Declaration of 2000(38). All subjects were required to be indigenous self-declared

784 ethnic groups.

### 785 **Sample collection and genotyping**

786 The reported dataset consisted of 510 unrelated subjects from 31 Mongolic-speaking populations  
787 (Mongolian: 362, HulunBuir Daur: 9, and Tsitsihar Daur: 10), two Tungusic-speaking populations (Heihe  
788 Evenki: 8 and Jiamusi Hezhen: 2), and five Sinitic-speaking Han Chinese (119, **Fig. 1**). Large-scale  
789 geographically different Mongolians were obtained from 29 groups: Jilin (18), HulunBuir (22),  
790 Chaoyang (21), XilinGol (19), Xinjiang (8), Chifeng (43), Heilongjiang (21), Hohhot (38), Tongliao (42),  
791 Fuxin (13), Xingan (18), Shenyang (9), Ordos (8), Nanyang (6), Beijing (9), Dalian (3), Qinghai (3),  
792 Alashan (3), Shanxi (8), Hebei (7), Tianjin (4), Gansu (3), Yunnan (4), Jiangsu (4), Shandong (12),  
793 Bayannaer (5), Guangdong (2), Ulancharu (2), and Baotou (7). Han individuals were collected from  
794 five northeastern Chinese populations: Baotou (25), Shenyang (25), Changchun (24), Harbin (24), and  
795 HulunBuir (21). PureLink Genomic DNA Mini Kit (Thermo Fisher Scientific) was used to isolate  
796 genomic DNA from saliva or anticoagulant-treated peripheral blood samples following the  
797 manufacturer's recommendations. Quantifiler Human DNA Quantification Kit (Thermo Fisher Scientific)  
798 and Applied Biosystem 7500 Real-time PCR System (Thermo Fisher Scientific) were used to quantify  
799 the DNA concentration. Genotyping of 529,790 variants, including 18,711 parentally lineage informative  
800 single nucleotide polymorphisms (LISNPs), 4,448 maternally LISNPs, and 506,616 autosomal and X-  
801 related SNPs, was carried out using the Affymetrix WeGene V1 Arrays.

### 802 **Data purification**

803 We used PLINK v1.90b6.13 with the parameters of genotyping success large than 95% and the minor  
804 allele frequency large than 0.05% to perform quality control. We used PLINK and King to identify the  
805 close genetic relationships within degree relatives and ran Genome-wide Complex Trait Analysis (GCTA)  
806 version 1.92.2 to remove outliers. After quality filtering, we kept 478 individuals in the following  
807 population genetic analysis.

### 808 **Reference population datasets**

809 For comparative purposes, we merged our data with publicly available modern and ancient Eurasians  
810 included in the 1240K, HO, and other recent public databases into three datasets: Affymetrix dataset, the  
811 merged HO dataset, and the merged 1240K dataset. The reference populations in the Affymetrix dataset  
812 consisting of Chinese populations which were genotyped using the same array in our groups, which  
813 included two Hainan populations (Han and Hlai)(39), 14 Tai-Kadai-speaking populations (three Jing,  
814 four Zhuang, and seven Sui groups) and seven Hmong-Mien-speaking Miao from Guangxi province(40).  
815 The reference groups in the merged HO dataset included 5081 individuals included in the 1240K database  
816 and 7744 individuals from the 1240k+HumanOrigin database obtained from David Reich Lab(1, 7, 13,  
817 17, 41). These datasets included all publicly available ancient DNA data published before 2021 and 2054  
818 individuals from the 1000 Genomes Project, 300 individuals from 142 worldwide populations from  
819 Simons Genome Diversity Project, and 928 sequence genomes in the sequenced HGDP. Genotype data  
820 obtained using the Affymetrix Human Origins were also included here(42-44). The used reference  
821 populations in the merged 1240K dataset included all publicly available ancient genomes and the  
822 sequenced genomes in the HGDP project(45).

### 823 **PCA and Model-based ADMIXTURE analysis**

824 We used PLINK v1.90b6.13(46) and SmartPCA(47) to conduct Eurasian and regional PCA in the context  
825 of modern and ancient Eurasians or eastern Eurasians with one additional parameter (Project: YES).  
826 Ancient individuals were projected onto the two-dimensional scaling plots. To characterize the individual  
827 and population ancestry compositions and reconstruct population genetic history, we used genetic  
828 clustering algorithm implemented in ADMIXTURE, one STRUCTURE-like but applying the maximum  
829 likelihood-based approach to conduct the model-based clustering analyses in the population genetic and  
830 evolutionary genetic studies (48). The analysis was performed based on the Eurasian dataset and other  
831 sub-dataset comprising regional East Asians. Plink v1.9 (46) was used to exclude SNPs in strong Linkage  
832 Disequilibrium (LD) with the following parameters settings: pairwise SNP loci  $r^2 > 0.4$ ; window: 200  
833 SNPs; and sliding windows: 25 (--indep-pairwise 200 25 0.4). A total of 199,753 variants out of 346,634  
834 variants passed filters and quality control (QC) and remained in the merged 1240K. 90,965 out of 120,894  
835 variants passed filters and QC in the merged HO. We assumed the number of predefined ancestry  
836 populations (K values) ranging from 2 to 20 and ran unsupervised ADMIXTURE with the 10-fold cross-  
837 validation (--cv=10). 100 randomly seeded runs were employed, and the log-likelihood scores and cross-  
838 validation errors were used to find the most appropriate K values. Ancestry compositions among modern  
839 and ancient people were conducted using the aforementioned settings.

### 840 **Three-population test (admixture $f_3$ -statistics and outgroup- $f_3$ -statistics)**

841 All possible pairs of source populations in three datasets were used to conduct the admixture- $f_3$ -statistics  
842 in the form  $f_3(\text{Source1}, \text{Source2}; \text{Studied populations})$  using the qp3Pop in the AdmixTools (49). Formal  
843 tests of admixture with statistically negative  $f_3$ -statistic values ( $Z \leq -3$ ) indicated that target populations

844 were an admixture of two predefined ancestral populations. Genetic affinity was estimated using  
845 outgroup- $f_3$ -statistics in the form  $f_3(\text{reference, studied populations; Mbuti})$ .

#### 846 **Four-population test ( $f_4$ -statistics)**

847 We employed the formal test of the four-population test (also called *D-statistics* or  $f_4$ -statistics)  
848 implemented in the ADMIXTOOLS(49) to test whether our studied populations harbor more shared  
849 alleles or genetic drift with targeted reference populations than other worldwide reference populations.  
850 We used the statistics in the form  $f_4(W, X; Y, \text{outgroup})$ , where W represents the Asian populations  
851 belonging to the Tungusic-, Mongolian-, Turkic-, Hmong-Mien-, Tai-Kadai-, Sinitic-, Tibeto-Burman-,  
852 Austronesian-, and Austroasiatic-speaking populations, X represents worldwide reference populations,  
853 and Y represents studied populations. The central African of Mbuti was used as the outgroup. Generally,  
854 the statistical index was calculated as the differences between counts of BABA sites and counts of ABBA  
855 sites divided by the sum of counts of BABA sites and counts of ABBA sites. The null hypothesis is that  
856 W and X form a clade and descend from a homogeneous ancestral population separated from Y and  
857 Outgroups. Thus, there are no differences in the rate of allele sharing existing with Y, and the D-statistic  
858 should be expected to be 0. If there is an excess of shared alleles or gene flow between either or both  
859 pairs (W, outgroup) or (X, Y) since separation from the others, a negative D value will be produced.  
860 Similarly, a positive D value can be obtained if a large number of the shared allele or gene flow exist  
861 between (X, Outgroup) or (W, Y) since separation from the others. Standard errors were estimated using  
862 the weighted block jackknife with the block size of 5Mb in a single Hotelling's  $T^2$  test. Z-scores were  
863 computed as the ratio between D and standard error.  $|Z| > 3$  was regarded as significantly different from  
864 0 and the null hypothesis could be rejected.

#### 865 **TreeMix and qpGraph**

866 We used the TreeMix version 1.12(50) to explore the topology of included populations with gene flow  
867 events ranging from 0 to 10. Phylogenetic trees with Yoruba as the root and un-root trees were established.  
868 We used ADIXTOOLS2 to explore the best qpGraph-based models with different population  
869 compositions. We also reconstruct the neighboring trees used Mega 7.0(51) based on the pairwise Fst  
870 and inverse outgroup- $f_3$ . PCA and MDS based on the genetic distance matrixes were conducted via the  
871 IBM SPSS Statistics 21(52).

#### 872 **Mitochondrial DNA haplogroup and Y-chromosome haplogroup assignment**

873 A total of 18,435 phylogenetic informative Y-chromosomal SNPs and 4,418 phylogenetic mitochondrial-  
874 related SNPs were used to classify the haplogroups using an in-house script.

#### 875 **Pairwise qpWave and qpAdm-based admixture models**

876 We merged geographically and ethnically defined populations into new genetically homogeneous  
877 populations (Mongolian\_East, Daur, Mongolian\_West, Ewenki\_Heihe, and Han\_Northeast) and  
878 performed qpWave analyses(49) to explore the distribution of the  $f_4$ -matrix via rank tests. Nine  
879 continental-representative populations were used as outgroups (Mbuti, Ust\_Ishim, Kostenki14, Papuan,  
880 Australian, Mixe, MA1, Onge, and Atayal). We also conducted pairwise qpWave analysis used the new-  
881 studied populations and previous sequenced HGDP Mongolic and Tungusic populations (Daur, Hezhen,  
882 Tu., Yakut, Mongolian, and Oroqen) as the right populations.

#### 883 **IBD segments inference and ROH**

884 The different datasets were phased using ShapeIT and then we used Refined IBD(53) to characterize the  
885 identical by descent (IBD), population-level shared IBD were corrected via the product of sample size  
886 of the focused population pairs. We used Plink to calculate the consecutive stretches of homozygous  
887 SNPs (ROH).

#### 888 **Estimating admixture times with MALDER**

889 We estimated admixture times using MALDER(54) with the default parameters or with two additional  
890 parameters (mindis: 0.0005 and jackknife: YES). Admixture mediated linkage disequilibrium can be  
891 decay with the recombination that occurred after the initial admixture. We only captured the simple  
892 characteristics of the admixture process with the assumption of a single pulse-like admixture. We used  
893 ancestral sources from southern East Asians, eastern Siberians, and western Eurasians as the possible  
894 ancestral proxy sources.

#### 895 **Chromosome painting and FineSTRUCTURE**

896 To infer finer-scale scenarios of the admixture landscape, we used ChromoPainterv2 to paint the targeted  
897 chromosome using all other included donor chromosomes based on the phased haplotype data(55).  
898 Chromosome painting could identify haplotypic distribution to further perform admixture dating and  
899 population identification. We conducted chromosome painting using the merged 1240K dataset, Altaic  
900 people in the HO dataset, and the Affymetrix dataset. Model-based Bayesian clustering instrumented in  
901 the FineSTRUCTURE was used to identify population substructure. FineSTRUCTURE version4 and  
902 FineSTRUCTURE R scripts based on the reconstructed coancestry matrix were conducted to dissect the  
903 fine-scale population structure via heatmap, clustering dendrogram, and PCA.

904 **ChromoPainterv2 and GLOBETROTTER admixture modeling**

905 R program of GLOBETROTTER(31) was used to identify and date admixture events based on the  
906 merged 1240K dataset, which included around 366K genome-wide SNPs. Both full analysis and regional  
907 analysis were conducted based on the estimated copying vectors from ChromoPainterv2. GLOBETROTTER  
908 used the sampled surrogated populations to paint the haplotypic mixture of the truth unsampled source  
909 populations and then to identify whether the targeted Altaic populations descend from any north-to-south  
910 or west-to-east admixture events (uncertain, no admixture, single admixture models, or complex  
911 admixture models with multiple dates or sources), and - if so - it can statistically calculate precisely when  
912 the admixture events occurred and the admixing sources involved.

913 **Nature selections signatures identifying**

914 We used to selscan and Rehh to calculate the iHS and XP-EHH as the selection indexes and then  
915 annotated the identified selection candidate loci using the PLINK. We highlighted the top selection SNPs  
916 using the Manhattan.

917 **Data availability**

918 The variation data reported in this paper can be shared via personal communication with corresponding  
919 authors. We make the data available upon request by asking the person requesting the data to agree in  
920 writing to the following restrictions: 1, the data can be only used for studying population history; 2, the  
921 data cannot be used for commercial purposes; 3, the data cannot be used to identify the sample donors;  
922 4, the data cannot be used for studying natural/cultural selections, medical or other related studies.

923 **Acknowledgments**

924 This study was supported by the National Natural Science Foundation of China (31801040), Nanqiang  
925 Outstanding Young Talents Program of Xiamen University (X2123302), China Postdoctoral Science  
926 Foundation (2021M691879), Fundamental Research Funds for the Central Universities (ZK1144), and  
927 Science and Technology Program of Guangzhou, China (2019030016).

928 **Disclosure of potential conflict of interest**

929 The author declares no conflict of interest.

930

931 **Reference**

- 932 1. M. A. Yang *et al.*, 40,000-Year-Old Individual from Asia Provides Insight into Early Population  
933 Structure in Eurasia. *Curr Biol* **27**, 3202-3208 e3209 (2017).
- 934 2. H. P.-A. S. Consortium *et al.*, Mapping human genetic diversity in Asia. *Science* **326**, 1541-  
935 1545 (2009).
- 936 3. M. Raghavan *et al.*, Upper Palaeolithic Siberian genome reveals dual ancestry of Native  
937 Americans. *Nature* **505**, 87-91 (2014).
- 938 4. S. Mallick *et al.*, The Simons Genome Diversity Project: 300 genomes from 142 diverse  
939 populations. *Nature* **538**, 201-206 (2016).
- 940 5. M. A. Yang, Q. Fu, Insights into Modern Human Prehistory Using Ancient Genomes. *Trends in*  
941 *genetics : TIG* **34**, 184-196 (2018).
- 942 6. P. Skoglund *et al.*, Genetic evidence for two founding populations of the Americas. *Nature* **525**,  
943 104-108 (2015).
- 944 7. M. Sikora *et al.*, The population history of northeastern Siberia since the Pleistocene. *Nature*  
945 **570**, 182-188 (2019).
- 946 8. Q. Fu *et al.*, The genetic history of ice age Europe. *Nature* **534**, 200-205 (2016).
- 947 9. C. Jeong *et al.*, A Dynamic 6,000-Year Genetic History of Eurasia's Eastern Steppe. *Cell* **183**,  
948 890-904 e829 (2020).
- 949 10. P. B. Damgaard *et al.*, 137 ancient human genomes from across the Eurasian steppes. *Nature*  
950 **557**, 369-374 (2018).
- 951 11. I. Mathieson *et al.*, Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**,  
952 499-503 (2015).
- 953 12. P. de Barros Damgaard *et al.*, The first horse herders and the impact of early Bronze Age steppe  
954 expansions into Asia. *Science* **360**, eaar7711 (2018).
- 955 13. H. McColl *et al.*, The prehistoric peopling of Southeast Asia. *Science* **361**, 88-92 (2018).
- 956 14. X. Mao *et al.*, The deep population history of northern East Asia from the Late Pleistocene to  
957 the Holocene. *Cell* **184**, 3256-3266 e3213 (2021).
- 958 15. T. Wang *et al.*, Human population history at the crossroads of East and Southeast Asia since  
959 11,000 years ago. *Cell* **184**, 3829-3841 e3821 (2021).
- 960 16. M. A. Yang *et al.*, Ancient DNA indicates human population shifts and admixture in northern  
961 and southern China. *Science* **369**, 282-288 (2020).
- 962 17. C. Jeong *et al.*, Long-term genetic stability and a high-altitude East Asian origin for the peoples  
963 of the high valleys of the Himalayan arc. *Proc Natl Acad Sci U S A* **113**, 7485-7490 (2016).



- 964 18. P. de Barros Damgaard *et al.*, The first horse herders and the impact of early Bronze Age steppe  
965 expansions into Asia. *Science* **360**, 1422-+ (2018).
- 966 19. V. Siska *et al.*, Genome-wide data from two early Neolithic East Asian individuals dating to  
967 7700 years ago. *Science advances* **3**, e1601877 (2017).
- 968 20. C. C. Wang *et al.*, Genomic insights into the formation of human populations in East Asia.  
969 *Nature* **591**, 413-419 (2021).
- 970 21. V. M. Narasimhan *et al.*, The formation of human populations in South and Central Asia. *Science*  
971 **365**, eaat7487 (2019).
- 972 22. I. Lazaridis *et al.*, Ancient human genomes suggest three ancestral populations for present-day  
973 Europeans. *Nature* **513**, 409-413 (2014).
- 974 23. C. Jeong *et al.*, Bronze Age population dynamics and the rise of dairy pastoralism on the eastern  
975 Eurasian steppe. *Proc Natl Acad Sci U S A* **115**, E11248-E11255 (2018).
- 976 24. C. Ning *et al.*, Ancient Genomes Reveal Yamnaya-Related Ancestry and a Potential Source of  
977 Indo-European Speakers in Iron Age Tianshan. *Curr Biol* **29**, 2526-2532 e2524 (2019).
- 978 25. C. Zhang *et al.*, PGG.SNV: understanding the evolutionary and medical implications of human  
979 single nucleotide variations in diverse populations. *Genome Biol* **20**, 215 (2019).
- 980 26. M. E. Allentoft *et al.*, Population genomics of Bronze Age Eurasia. *Nature* **522**, 167-172 (2015).
- 981 27. C. Jeong *et al.*, The genetic history of admixture across inner Eurasia. *Nat Ecol Evol* **3**, 966-976  
982 (2019).
- 983 28. D. Wu *et al.*, Genetic admixture in the culturally unique Peranakan Chinese population in  
984 Southeast Asia. *Mol Biol Evol* 10.1093/molbev/msab187 (2021).
- 985 29. C. Y. Wei *et al.*, Genetic profiles of 103,106 individuals in the Taiwan Biobank provide insights  
986 into the health and history of Han Chinese. *NPJ Genom Med* **6**, 10 (2021).
- 987 30. B. Yunusbayev *et al.*, The genetic legacy of the expansion of Turkic-speaking nomads across  
988 Eurasia. *PLoS Genet* **11**, e1005068 (2015).
- 989 31. G. Hellenthal *et al.*, A genetic atlas of human admixture history. *Science* **343**, 747-751 (2014).
- 990 32. R. Wu *et al.*, Genetic polymorphism and population structure of Torghut Mongols and  
991 comparison with a Mongolian population 3000 kilometers away. *Forensic Sci Int Genet* **42**, 235-  
992 243 (2019).
- 993 33. G. He *et al.*, A comprehensive exploration of the genetic legacy and forensic features of  
994 Afghanistan and Pakistan Mongolian-descent Hazara. *Forensic Sci Int Genet* **42**, e1-e12 (2019).
- 995 34. M. Robbeets *et al.*, Triangulation supports agricultural spread of the Transeurasian languages  
996 (2021), 10.21203/rs.3.rs-255765/v1.
- 997 35. M. Unterlander *et al.*, Ancestry and demography and descendants of Iron Age nomads of the  
998 Eurasian Steppe. *Nat Commun* **8**, 14615 (2017).
- 999 36. T. M. Karafet, L. P. Osipova, O. V. Savina, B. Hallmark, M. F. Hammer, Siberian genetic  
1000 diversity reveals complex origins of the Samoyedic-speaking populations. *American journal of*  
1001 *human biology : the official journal of the Human Biology Council* **30**, e23194 (2018).
- 1002 37. K. Tambets *et al.*, Genes reveal traces of common recent demographic history for most of the  
1003 Uralic-speaking populations. *Genome Biol* **19**, 139 (2018).
- 1004 38. W. M. Association, World Medical Association Declaration of Helsinki. Ethical principles for  
1005 medical research involving human subjects. *Bulletin of the World Health Organization* **79**, 373  
1006 (2001).
- 1007 39. G. He *et al.*, Inferring the population history of Tai-Kadai-speaking people and southernmost  
1008 Han Chinese on Hainan Island by genome-wide array genotyping. *Eur J Hum Genet* **28**, 1111-  
1009 1123 (2020).
- 1010 40. X. Huang *et al.*, Genomic Insights into the Demographic History of Southern Chinese. *bioRxiv*  
1011 10.1101/2020.11.08.373225, 2020.2011.2008.373225 (2020).
- 1012 41. M. Lipson *et al.*, Ancient genomes document multiple waves of migration in Southeast Asian  
1013 prehistory. *Science* **361**, 92-95 (2018).
- 1014 42. C. The International HapMap *et al.*, A second generation human haplotype map of over 3.1  
1015 million SNPs. *Nature* **449**, 851 (2007).
- 1016 43. J. Z. Li *et al.*, Worldwide human relationships inferred from genome-wide patterns of variation.  
1017 *Science* **319**, 1100-1104 (2008).
- 1018 44. X. Huang *et al.*, The genetic assimilation in language borrowing inferred from Jing People. *Am*  
1019 *J Phys Anthropol* **166**, 638-648 (2018).
- 1020 45. A. Bergstrom *et al.*, Insights into human genetic variation and population history from 929  
1021 diverse genomes. *Science* **367** (2020).
- 1022 46. S. Purcell *et al.*, PLINK: a tool set for whole-genome association and population-based linkage  
1023 analyses. *Am J Hum Genet* **81**, 559-575 (2007).

- 1024 47. N. Patterson, A. L. Price, D. Reich, Population structure and eigenanalysis. *PLoS Genet* **2**, e190  
1025 (2006).
- 1026 48. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated  
1027 individuals. *Genome Res* **19**, 1655-1664 (2009).
- 1028 49. N. Patterson *et al.*, Ancient admixture in human history. *Genetics* **192**, 1065-1093 (2012).
- 1029 50. J. K. Pickrell, J. K. Pritchard, Inference of population splits and mixtures from genome-wide  
1030 allele frequency data. *PLoS Genet* **8**, e1002967 (2012).
- 1031 51. S. Kumar, G. Stecher, K. Tamura, MEGA7: Molecular Evolutionary Genetics Analysis Version  
1032 7.0 for Bigger Datasets. *Mol Biol Evol* **33**, 1870-1874 (2016).
- 1033 52. J. Hansen, Using SPSS for Windows and Macintosh: Analyzing and Understanding Data. *Amer.*  
1034 *Statistician* **59**, 113-113 (2005).
- 1035 53. B. L. Browning, S. R. Browning, A fast, powerful method for detecting identity by descent. *Am*  
1036 *J Hum Genet* **88**, 173-182 (2011).
- 1037 54. P. R. Loh *et al.*, Inferring admixture histories of human populations using linkage disequilibrium.  
1038 *Genetics* **193**, 1233-1254 (2013).
- 1039 55. D. J. Lawson, G. Hellenthal, S. Myers, D. Falush, Inference of population structure using dense  
1040 haplotype data. *PLoS Genet* **8**, e1002453 (2012).
- 1041

1042 **Legends of Figures**

1043 **Fig. 1 Geographic position and genetic structure of studied Hans, Mongolic and Tungusic people,**  
1044 **as well as other modern and ancient East Asians based on the merged HO dataset (113,962 SNPs).**

1045 (A) Map of the newly sampled 40 Chinese populations and other 251 modern and ancient reference  
1046 populations. (B-C) population genetic relationship patterns in the PCA analyses focused on all eastern  
1047 Eurasian populations based on the top three components. Modern populations were colored-coded based  
1048 on language categories (D-E) Northern East Asian PCA analysis based on the top three components.  
1049 Ancient samples were projected. (F) ADMIXTURE analysis result for K=10 which has the minimized  
1050 cross-validation error.

1051 **Fig. 2. Fine-scale population genetic structure of studied populations and two Hainan populations**  
1052 **based on 465,941SNPs genotyped using the Affymetrix chip.** (A) Plink-based PCA results based on  
1053 the allele frequency spectrum showed two clines within 573 individuals from 40 populations. (B-D) PCA  
1054 patterns based on the coancestry matrix of linked SNP markers. (E-F) Pairwise coincidence matrix  
1055 estimated based on the chunk counts using the linked SNPs and population-level average chunk counts.  
1056 The coloring in E represented the posterior coincidence probability. (G-H) Heatmap showed the pairwise  
1057 sharing IBD length and pairwise Fst genetic distance within 40 populations.

1058 **Fig. 3. Genetic sources and mixed landscape of eastern Asians based on the merged 1240K dataset**  
1059 **(369519 SNPs).** (A) The p values of rank1 in the pairwise qpWave analysis. P values larger than 0.05  
1060 were marked with “++” and larger than 0.01 were marked with “+”. The color in the heatmap showed  
1061 the detailed p values. (B). PCA results among 1073 individuals from 71 modern and ancient populations.  
1062 Ancient populations were projected. (C) TreeMix-based phylogeny without migration events among 71  
1063 populations. (D) Model-based ADMIXTURE results showed the possible ancestral sources and their  
1064 corresponding admixture proportion. The breadth of populations is not correlated with the included  
1065 population size, which was be enlarged or reduced to be better visualization, especially for the eastern  
1066 Mongolians. (F) Three-way admixture qpAdm models with different ancestral sources showed the mixed  
1067 landscape of northern Mongolic and Tungusic people. The p\_rank2 values were presented following the  
1068 population name and the white bar represented the standard error of the estimated ancestry proportion.

1069 **Fig. 4 Demographic history of western and eastern Mongolians estimated via the qpAdm and**  
1070 **qpGraph.** (A-B) The estimated ancestry coefficient of the representative sources in the three-way  
1071 ADMIXTURE models with Northeast East Asian sources from Mongolian Plateau and Amur River Basin,  
1072 East Asian source related to Hanben people and western source related Iran and Pakistan ancients. The  
1073 bar plot denoted the standard error of the admixture proportion. The p values of p\_rank2 were larger than  
1074 0.05. (C-D) The best fitted qpGraph-based deep population admixture history of western Mongolians  
1075 (C) and eastern Mongolians (D)

1076 **Fig. 5 Genetic formation of Northeastern Han Chinese inferred from the genome-wide SNPs data**  
1077 **in the merged 1240K dataset.** (A) Two-way admixture model (Amur Neolithic sources+ Neolithic  
1078 Coastal southern East Asian source). (B) North-to-south admixture model with two sources related to  
1079 YRB farmers and 2000-year-old LaCen inland southern East Asian showed the differentiated ancestry  
1080 composition of Han Chinese and southern Chinese indigenous populations. (C) Ancestry admixture  
1081 landscape characterized via the two sources from the Yellow River farmers and Amur River Hunter-  
1082 Gatherer. All fitted models with the p\_rank1 values were larger than 0.05 and the bar plot represented  
1083 the standard error. (D) The deep population formation history of Northeast Hans via the best-fitted  
1084 qpGraph-based admixture graph.

1085 **Fig. 6 Fine-scale population structure of 238 northern Mongolic and Tungusic people from 26**  
1086 **populations inferred from dense haplotype data of 600K SNPs in the HO dataset.** (A-E) Two-  
1087 dimensional plots based on the allele frequency spectrum (A) and coancestry matrix (B-E) showed four  
1088 genetic clusters within the studied populations. (F) ADMIXTURE results for four predefined ancestral  
1089 sources. (G-H) The pairwise Fst genetic distance and share IBD fragments among 26 Mongolic and  
1090 Tungusic populations. (I-J) Pairwise coincidence matrix output by FineSTRUCTURE based on the  
1091 chunk counts and the estimated average chunk counts. (K) TreeMix-based phylogenetic relationship with  
1092 four fitted admixture events.

1093 **Fig. 7 The estimated ancestry admixture coefficient of modern and ancient eastern Eurasians via**  
1094 **qpAdm and qpGraph.** (A-C) Three-way ADMIXTURE models focused on the Altaic-speaking  
1095 populations with ancestry from northeastern East Asia, Yellow River Basin and western Eurasia. (D-F)  
1096 Two-way admixture models with one source from Mongolian Plateau or Amur River Basin and Yellow  
1097 River Basin. (G-I) The best fitted qpGraph models with different population compositions showed the  
1098 genetic relationship between Neolithic, historic and modern northern East Asians. The estimated genetic  
1099 drift ( $1000 \times f_2$  values) was marked along each branch. The dot line represented the admixture events and  
1100 the admixture proportion was also marked along the dot line.

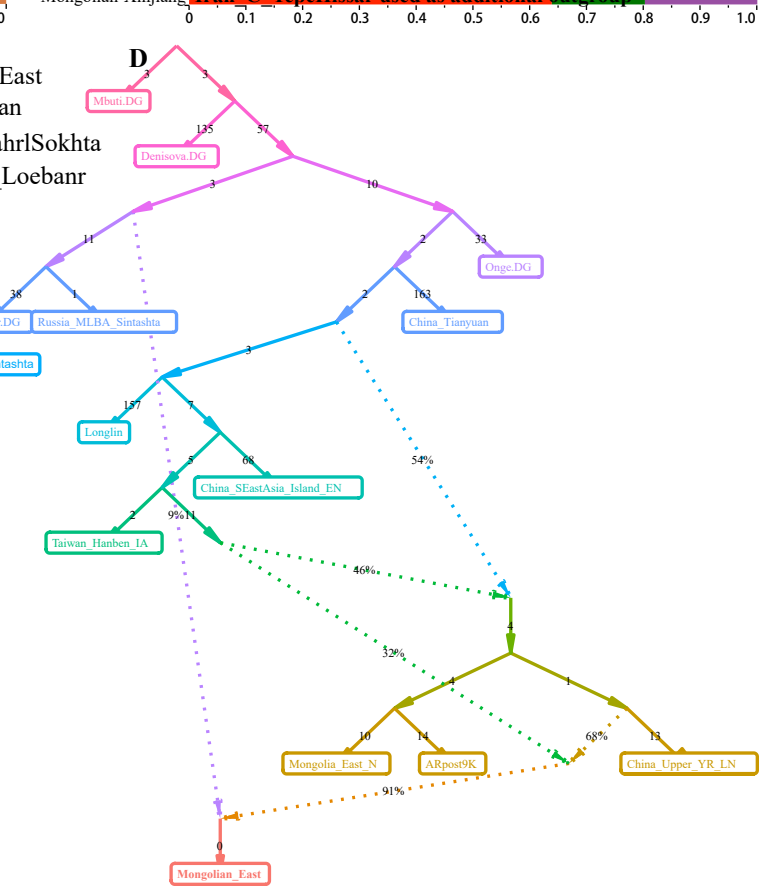
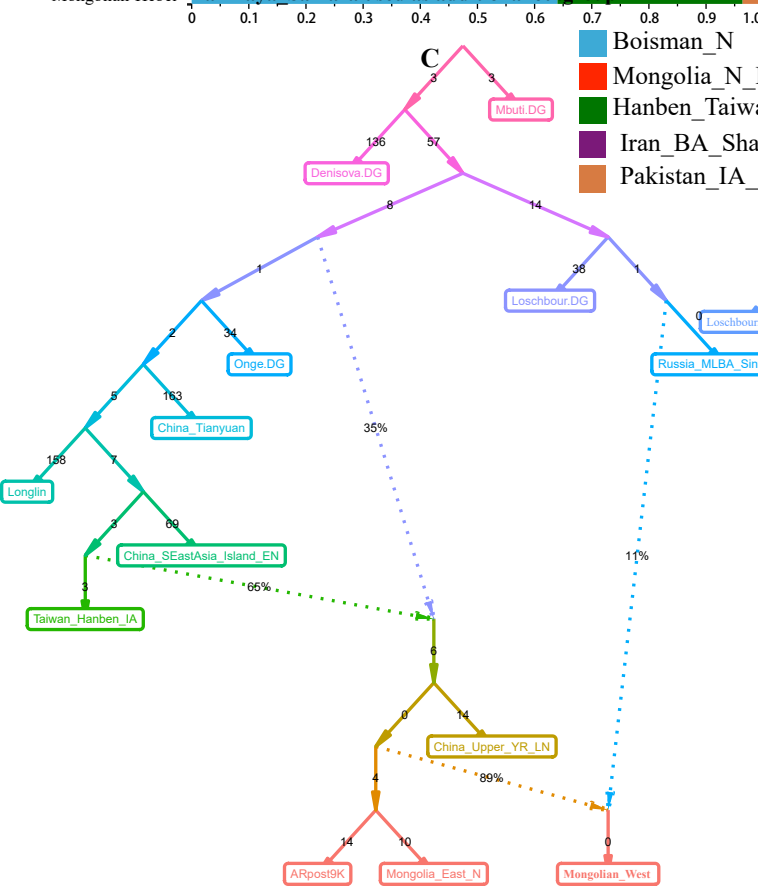
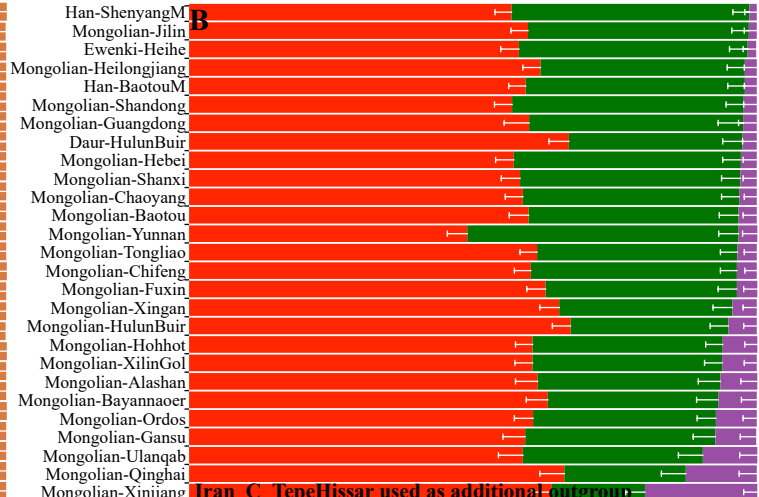
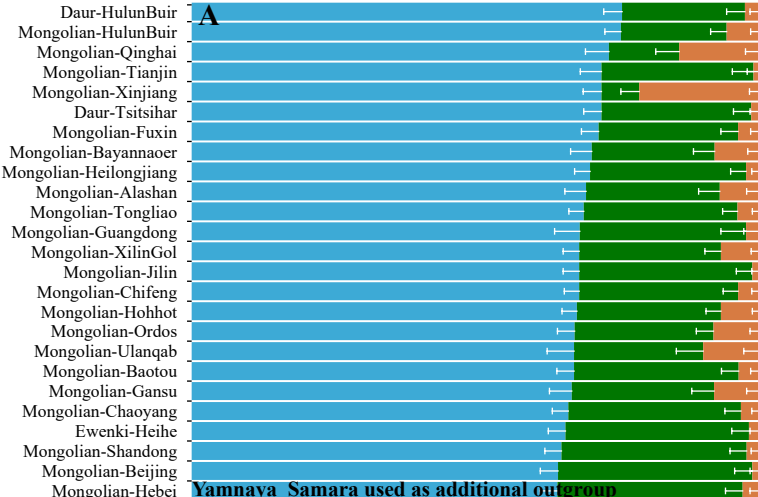
1101

1102 **Fig. 8. Fine-scale population genetic structure showed the genetic relationship between newly**  
1103 **studied Chinese populations. (A)** Smartpca-based PCA results among 760 individuals from 27  
1104 populations. Variations from 566 individuals from 13 populations were used to provide the genetic  
1105 background. Ancient Mongolian Plateau historic people were projected. **(B)** Pairwise shared genetic  
1106 alleles estimated via outgroup- $f_3$ -statistics. **(C)** Phylogenetic relationship between five studied  
1107 populations and seven northern East Asian populations included in the HGDP with two admixture events.  
1108 **(D~K)** Pairwise coincidence matrix and ADMIXTURE-based results of 13 included modern populations  
1109 in the merged 1240K dataset.  
1110 **Fig. 9 P values of the estimated EXEHH focused on Mongolian populations with the Guizhou Hans**  
1111 **as the reference population. (A~D)** Circle plots of chromosomes 1, 4, 6 12 and 16 showed the selection  
1112 signatures. **(E)** Manhattan plots showed all tested candidate loci under natural selection.

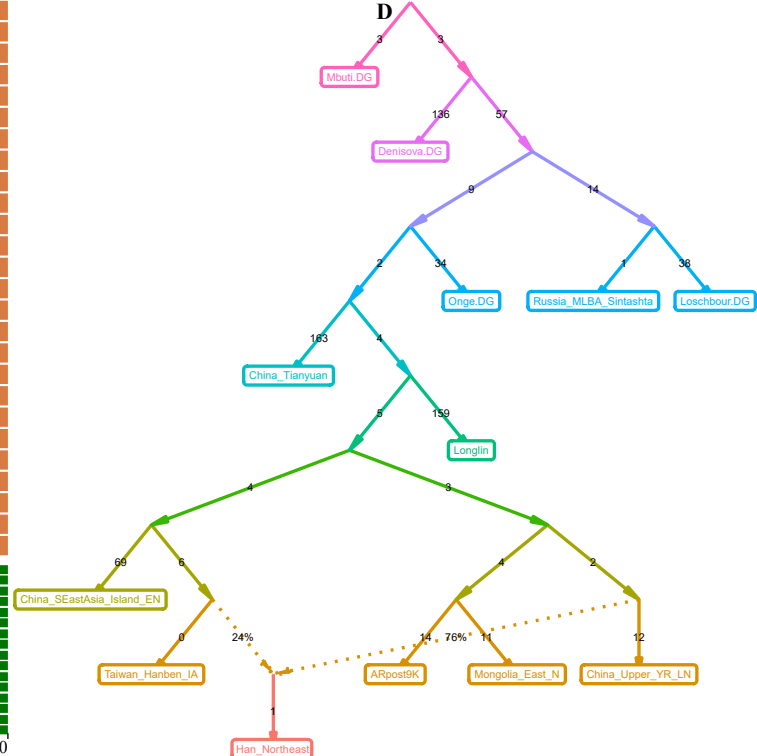
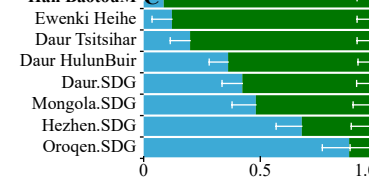
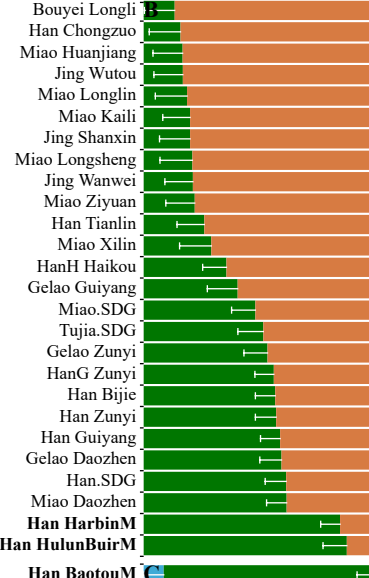
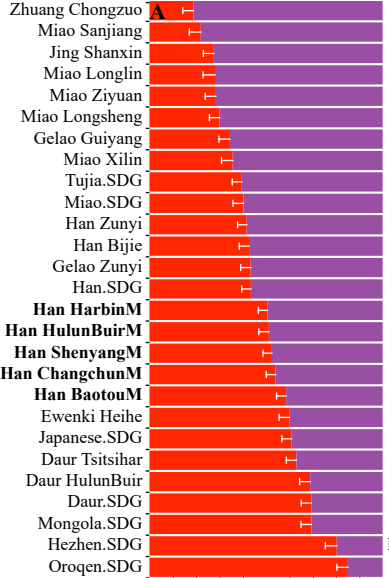






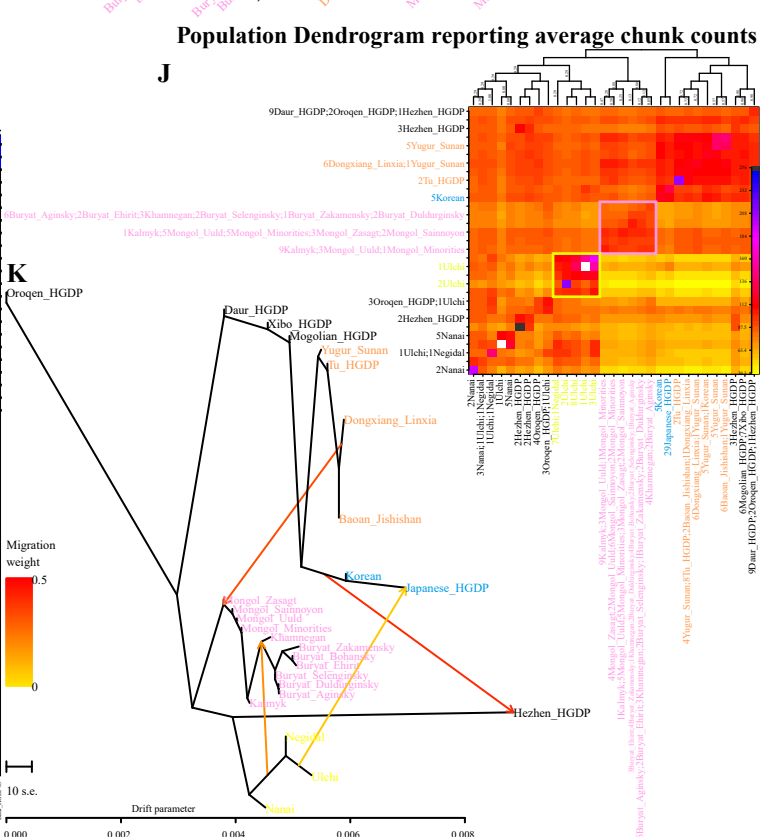
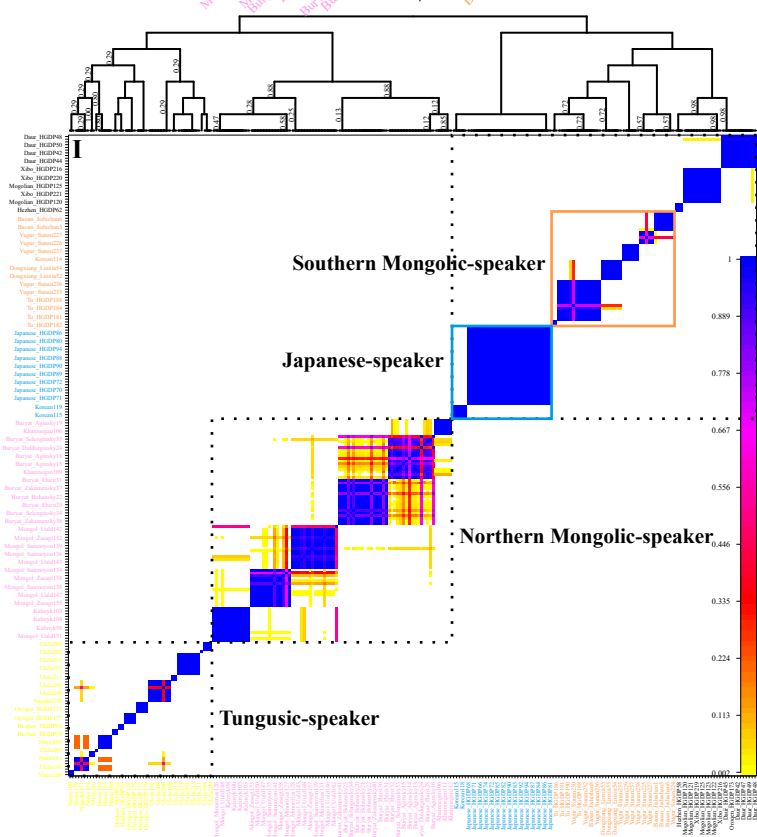
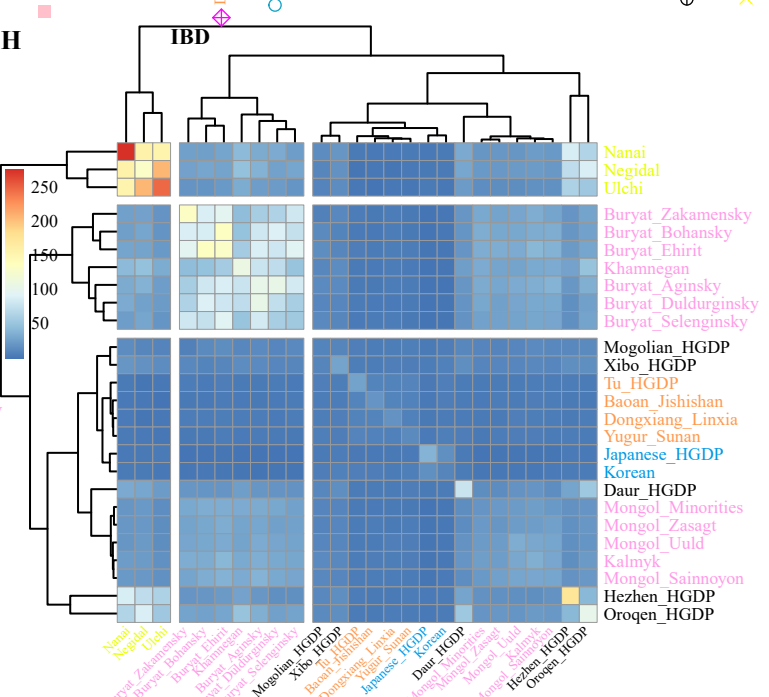
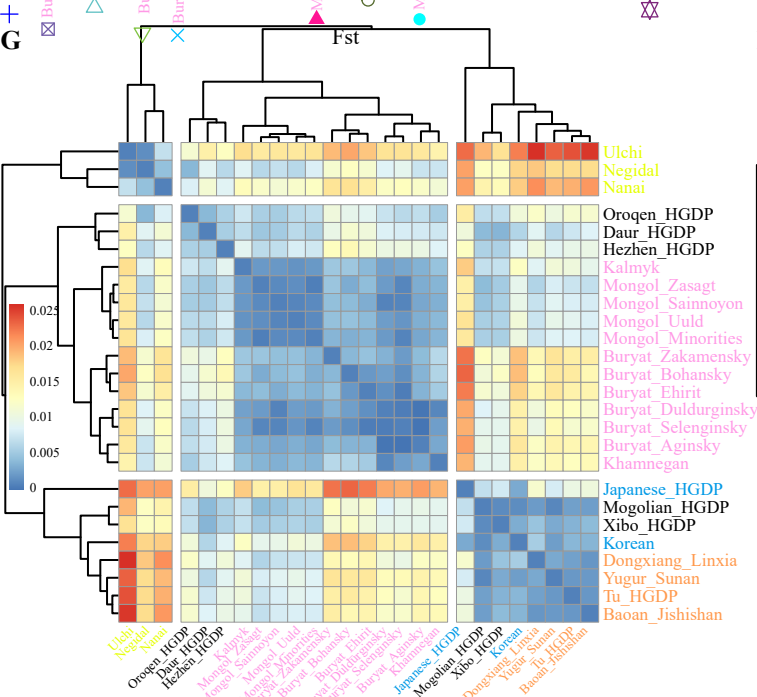
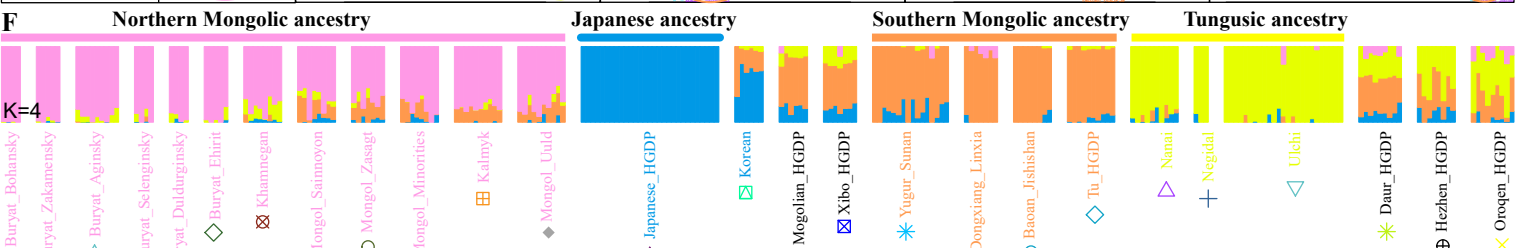
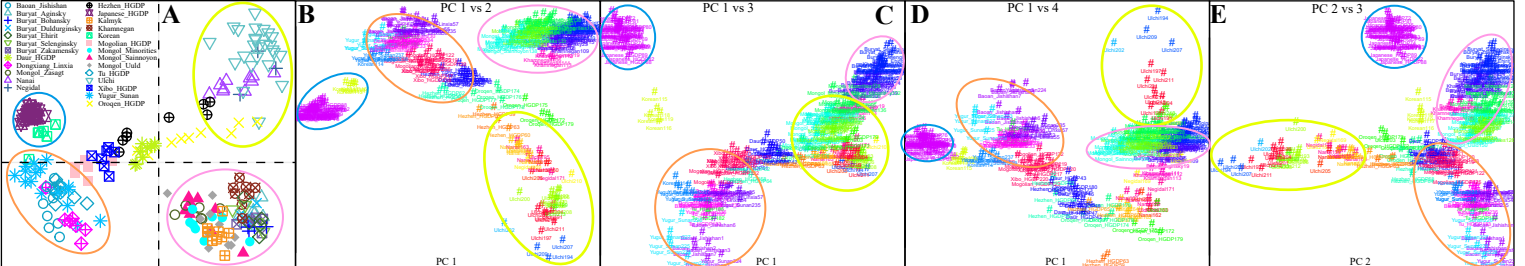


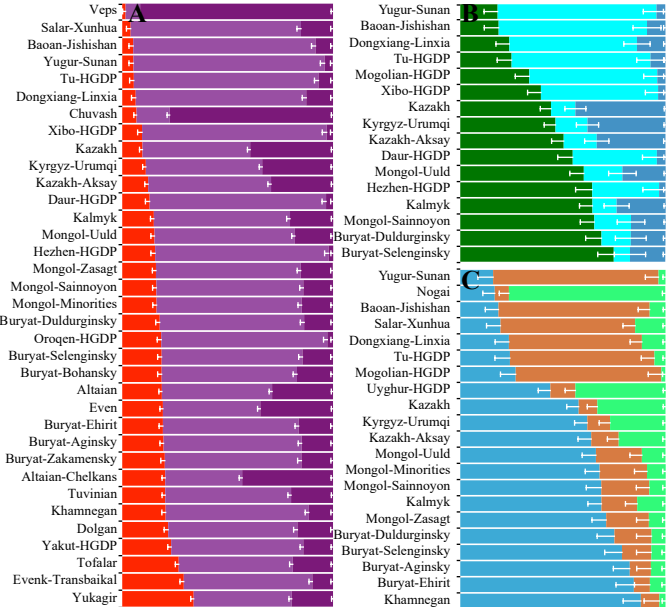




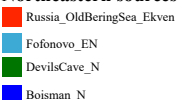
**Ancestry sources**

- China\_AR\_EN
- China\_SEastAsia\_Coastal\_LN
- AR13\_10K
- China\_YR\_LN
- LaCen

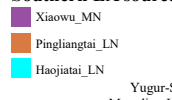




**Northeastern sources**



**Southern EA sources**



**Western Eurasian sources**

