

## 1 **Characterizing the genetic history of admixture across inner Eurasia**

2  
3 Choongwon Jeong<sup>1,2,31,\*</sup>, Oleg Balanovsky<sup>3,4,31</sup>, Elena Lukianova<sup>3</sup>, Nurzhibek Kahbatkyzy<sup>5,6</sup>, Pavel  
4 Flegontov<sup>7,8</sup>, Valery Zaporozhchenko<sup>3,4</sup>, Alexander Immel<sup>1</sup>, Chuan-Chao Wang<sup>1,9</sup>, Olzhas Ixan<sup>5</sup>, Elmira  
5 Khussainova<sup>5</sup>, Bakhytzhan Bekmanov<sup>5,6</sup>, Victor Zaibert<sup>10</sup>, Maria Lavryashina<sup>11</sup>, Elvira Pocheshkhova<sup>12</sup>,  
6 Yuldash Yusupov<sup>13</sup>, Anastasiya Agdzhoyan<sup>3,4</sup>, Koshel Sergey<sup>14</sup>, Andrei Bukin<sup>15</sup>, Pagbajabyn  
7 Nymadawa<sup>16</sup>, Michail Churnosov<sup>17</sup>, Roza Skhalyakho<sup>4</sup>, Denis Daragan<sup>4</sup>, Yuri Bogunov<sup>3,4</sup>, Anna  
8 Bogunova<sup>4</sup>, Alexandr Shtrunov<sup>4</sup>, Nadezda Dubova<sup>18</sup>, Maxat Zhabagin<sup>19,20</sup>, Levon Yepiskoposyan<sup>21</sup>,  
9 Vladimir Churakov<sup>22</sup>, Nikolay Pislegin<sup>22</sup>, Larissa Damba<sup>23</sup>, Ludmila Saroyants<sup>24</sup>, Khadizhat Dibirova<sup>3,4</sup>,  
10 Lubov Artamentova<sup>25</sup>, Olga Utevska<sup>25</sup>, Eldar Idrisov<sup>26</sup>, Evgeniya Kamenshchikova<sup>4</sup>, Irina Evseeva<sup>27</sup>, Mait  
11 Metspalu<sup>28</sup>, Martine Robbeets<sup>2</sup>, Leyla Djansugurova<sup>5,6</sup>, Elena Balanovska<sup>4</sup>, Stephan Schiffels<sup>1</sup>, Wolfgang  
12 Haak<sup>1</sup>, David Reich<sup>29,30</sup> & Johannes Krause<sup>1,\*</sup>

13

14 <sup>1</sup> Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany

15 <sup>2</sup> Eurasia3angle Research Group, Max Planck Institute for the Science of Human History, Jena, Germany

16 <sup>3</sup> Vavilov Institute of General Genetics Russian Academy of Sciences, Moscow, Russia

17 <sup>4</sup> Federal State Budgetary Institution «Research Centre for Medical Genetics», Moscow, Russia

18 <sup>5</sup> Department of Population Genetics, Institute of General Genetics and Cytology, SC MES RK, Almaty, Kazakhstan

19 <sup>6</sup> Department of Molecular Biology and Genetics, Kazakh National University by al-Farabi, Almaty, Kazakhstan

20 <sup>7</sup> Department of Biology and Ecology, Faculty of Science, University of Ostrava, Ostrava, Czech Republic

21 <sup>8</sup> Biology Centre, Czech Academy of Sciences and Faculty of Science, University of South Bohemia, České Budějovice, Czech  
22 Republic

23 <sup>9</sup> Department of Anthropology and Ethnology, Xiamen University, Xiamen 361005, China

24 <sup>10</sup> Institute of Archeology and Steppe Civilization, Kazakh National University by al-Farabi, Almaty, Kazakhstan

25 <sup>11</sup> Kemerovo State University, Krasnaya 3, Kemerovo, Russia

26 <sup>12</sup> Kuban State Medical University, Krasnodar, Russia

27 <sup>13</sup> Institute of Strategic Research of the Republic of Bashkortostan, Ufa, Russia

28 <sup>14</sup> Faculty of Geography, Lomonosov Moscow State University, Moscow, Russia

29 <sup>15</sup> Transbaikalian State University, Chita, Russia

30 <sup>16</sup> Mongolian Academy of Medical Sciences, Ulaanbaatar, Mongolia

31 <sup>17</sup> Belgorod State University, Belgorod, Russia

32 <sup>18</sup> The Institute of Ethnology and Anthropology of the Russian Academy of Sciences, Moscow, Russia

33 <sup>19</sup> National Laboratory Astana, Nazarbayev University, Astana, Kazakhstan

34 <sup>20</sup> National Center for Biotechnology, Astana, Kazakhstan

35 <sup>21</sup> Laboratory of Ethnogenomics, Institute of Molecular Biology of National Academy of Sciences, Yerevan, Armenia

36 <sup>22</sup> Udmurt Institute of History, Language and Literature of Udmurt Federal Research Center of the Ural Branch of the Russian  
37 Academy of Sciences, Russia

38 <sup>23</sup> Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

39 <sup>24</sup> Leprosy Research Institute, Astrakhan, Russia

40 <sup>25</sup> V. N. Karazin Kharkiv National University, Kharkiv, Ukraine

41 <sup>26</sup> Astrakhan branch of the Russian Academy of National Economy and Public Administration under the President of the Russian  
42 Federation, Astrakhan, Russia

43 <sup>27</sup> Northern State Medical University, Arkhangelsk, Russia

44 <sup>28</sup> Estonian Biocenter, Tartu, Estonia

45 <sup>29</sup> Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA

46 <sup>30</sup> Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts 02115, USA

47 <sup>31</sup> These authors contributed equally to this work

48 \* Correspondence: [jeong@shh.mpg.de](mailto:jeong@shh.mpg.de) (C.J.), [krause@shh.mpg.de](mailto:krause@shh.mpg.de) (J.K.)

49

50 **Abstract**

51           The indigenous populations of inner Eurasia, a huge geographic region covering the central  
52 Eurasian steppe and the northern Eurasian taiga and tundra, harbor tremendous diversity in their genes,  
53 cultures and languages. In this study, we report novel genome-wide data for 763 individuals from  
54 Armenia, Georgia, Kazakhstan, Moldova, Mongolia, Russia, Tajikistan, Ukraine, and Uzbekistan. We  
55 furthermore report genome-wide data of two Eneolithic individuals (~5,400 years before present)  
56 associated with the Botai culture in northern Kazakhstan. We find that inner Eurasian populations are  
57 structured into three distinct admixture clines stretching between various western and eastern Eurasian  
58 ancestries. This genetic separation is well mirrored by geography. The ancient Botai genomes suggest yet  
59 another layer of admixture in inner Eurasia that involves Mesolithic hunter-gatherers in Europe, the  
60 Upper Paleolithic southern Siberians and East Asians. Admixture modeling of ancient and modern  
61 populations suggests an overwriting of this ancient structure in the Altai-Sayan region by migrations of  
62 western steppe herders, but partial retaining of this ancient North Eurasian-related cline further to the  
63 North. Finally, the genetic structure of Caucasus populations highlights a role of the Caucasus Mountains  
64 as a barrier to gene flow and suggests a post-Neolithic gene flow into North Caucasus populations from  
65 the steppe.

66

67

68 **Introduction**

69           Present-day human population structure is often marked by a correlation between geographic and  
70 genetic distances,<sup>1;2</sup> reflecting continuous gene flow among neighboring groups, a process known as  
71 “isolation by distance”. However, there are also striking failures of this model, whereby geographically  
72 proximate populations can be quite distantly related. Such barriers to gene flow often correspond to major  
73 geographic features, such as the Himalayas<sup>3</sup> or the Caucasus Mountains.<sup>4</sup> Many cases also suggest the  
74 presence of social barriers to gene flow. For example, early Neolithic farming populations in Europe  
75 show a remarkable genetic homogeneity suggesting minimal genetic exchange with local hunter-gatherer

76 populations through the initial expansion; genetic mixing of these two gene pools became evident only  
77 after thousands of years in the middle Neolithic.<sup>5</sup> Modern Lebanese populations provide another example  
78 by showing a population stratification reflecting their religious community.<sup>6</sup> There are also examples of  
79 geographically very distant populations that are closely related: for example, people buried in association  
80 with artifacts of the Yamnaya horizon in the Pontic-Caspian steppe and the contemporaneous Afanasievo  
81 culture 3,000 km east in the Altai-Sayan Mountains.<sup>7; 8</sup>

82         The vast region of the Eurasian inland (“inner Eurasia” herein) is split into distinct ecoregions,  
83 such as the Eurasian steppe in central Eurasia, boreal forests (taiga) in northern Eurasia, and the Arctic  
84 tundra at the periphery of the Arctic Ocean. These ecoregions stretch in an east-west direction within  
85 relatively narrow north-south bands. Various cultural features show a distribution that broadly mirrors the  
86 eco-geographic distinction in inner Eurasia. For example, indigenous peoples of the Eurasian steppe  
87 traditionally practice nomadic pastoralism,<sup>9; 10</sup> while northern Eurasian peoples in the taiga mainly rely on  
88 reindeer herding and hunting<sup>11</sup>. The subsistence strategies in each of these ecoregions are often considered  
89 to be adaptations to the local environments.<sup>12</sup>

90         At present there is limited information about how environmental and cultural influences are  
91 mirrored in the genetic structure of inner Eurasians. Recent genome-wide studies of inner Eurasians  
92 mostly focused on detecting and dating genetic admixture in individual populations.<sup>13-16</sup> So far only two  
93 studies have reported recent genetic sharing between geographically distant populations based on the  
94 analysis of “identity-by-descent” segments.<sup>13; 17</sup> One study reports a long-distance extra genetic sharing  
95 between Turkic populations based on a detailed comparison between Turkic-speaking groups and their  
96 non-Turkic neighbors.<sup>13</sup> Another study extends this approach to some Uralic and Yeniseian-speaking  
97 populations.<sup>17</sup> However, a comprehensive spatial genetic analysis of inner Eurasian populations is still  
98 lacking.

99         Ancient DNA studies have already shown that human populations of this region have  
100 dramatically transformed over time. For example, the Upper Paleolithic genomes from the Mal'ta and  
101 Afontova Gora archaeological sites in southern Siberia revealed a genetic profile, often referred to as

102 “Ancient North Eurasians (ANE)”, which is deeply related to Paleolithic/Mesolithic hunter-gatherers in  
103 Europe and also substantially contributed to the gene pools of modern-day Native Americans, Siberians,  
104 Europeans and South Asians.<sup>18; 19</sup> Studies of Bronze Age steppe populations found the appearance of  
105 additional Western Eurasian-related ancestries across the steppe from the Pontic-Caspian region in the  
106 West to the Altai-Sayan region in the East, here we collectively refer to as “Western Steppe Herders  
107 (WSH)”: the earlier populations associated with the Yamnaya and Afanasievo cultures (often referred to  
108 “steppe Early and Middle Bronze Age”; “steppe\_EMBA”) and the later ones associated with many  
109 cultures such as Potapovka, Sintashta, Srubnaya and Andronovo to name a few (often referred to “steppe  
110 Middle and Late Bronze Age”; “steppe\_MLBA”).<sup>8</sup> Still, important questions remain unanswered due to  
111 limited availability of ancient genomes, including the identity of the eastern Eurasian gene pools that  
112 interacted with Pleistocene ANE or Bronze Age WSH populations and the genetic profile of pre-Bronze  
113 Age inner Eurasians. An example of the latter is the Eneolithic Botai culture in northern Kazakhstan in  
114 the 4<sup>th</sup> millennium BCE.<sup>20</sup> In addition to their role in the earliest horse domestication so far known,<sup>21</sup>  
115 Botai is at the crossroads, both in time and in space, connecting various earlier hunter-gatherer and later  
116 WSH populations in inner Eurasia.

117 In this study, we analyzed newly produced genome-wide genetic variation data for 763  
118 individuals belonging to 60 self-reported ethnic groups to provide a dense portrait of the genetic structure  
119 of indigenous populations in inner Eurasia. We also produced genome-wide data of two individuals  
120 associated with the Eneolithic Botai culture in Kazakhstan to explore the genetic structure of pre-Bronze  
121 Age populations in inner Eurasia. We aimed at characterizing the genetic composition of inner Eurasians  
122 in fine resolution by applying both allele frequency- and haplotype-based methods. Based on the fine-  
123 scale genetic profile, we further explored if and where the barriers and conduits of gene flow exist in  
124 inner Eurasia.

125

126

127 **Materials and Methods**

128

129 ***Study participants and genotyping***

130 We collected samples from 763 participants from nine countries (Armenia, Georgia, Kazakhstan,  
131 Moldova, Mongolia, Russia, Tajikistan, Ukraine, and Uzbekistan). The sampling strategy included  
132 sampling a majority of large ethnic groups in the studied countries. Within groups, we sampled subgroups  
133 if they were known to speak different dialects; for ethnic groups with large area, we sampled within  
134 several districts across the area. We sampled individuals whose grandparents were all self-identified  
135 members of the given ethnic groups and were born within the studied district(s). All individuals provided  
136 a written informed consent approved by the Ethic Committee of the Research Centre for Medical  
137 Genetics, Moscow, Russia. Most of the ethnic Russian samples were collected from indigenous Russian  
138 areas (present-day Central Russia) and had been stored for years in the Estonian Biocenter; samples from  
139 Mongolia, Tajikistan, Uzbekistan, and Ukraine were collected partially in the framework of the  
140 Genographic project. Most DNA samples were extracted from venous blood via the phenol-chloroform  
141 method. For this study we identified 112 subgroups (belonging to 60 ethnic group labels) which were not  
142 previously genotyped on the Affymetrix Axiom® Genome-wide Human Origins 1 (“HumanOrigins”)  
143 array platform<sup>22</sup> and selected on average 7 individuals per subgroup (Figure 1 and Table S1). Genome-  
144 wide genotyping experiments were performed on the HumanOrigins array platform. We removed 18  
145 individuals from further analysis either due to high genotype missing rate ( $> 0.05$ ;  $n=2$ ) or due to being  
146 outliers in principal component analysis (PCA) relative to other individuals from the same group ( $n=16$ ).  
147 The remaining 745 individuals assigned to 60 group labels were merged to published HumanOrigins data  
148 sets of world-wide contemporary populations<sup>19</sup> and of four Siberian ethnic groups (Enets, Kets,  
149 Nganasans and Selkups).<sup>23</sup> Diploid genotype data of six contemporary individuals (two Saami, two  
150 Sherpa and two Tibetans) were obtained from the Simons Genome Diversity Panel data set.<sup>24</sup> We also  
151 added ancient individuals from published studies,<sup>3; 8; 18; 19; 25-40</sup> by randomly sampling a single allele for  
152 581,230 autosomal single nucleotide polymorphisms (SNPs) in the HumanOrigins array (Table S2).

153

154 ***Sequencing of the ancient Botai genomes***

155 We extracted genomic DNA from four skeletal remains belonging to two individuals and built  
156 sequencing libraries either with no uracil-DNA glycosylase (UDG) treatment or with partial treatment  
157 following published protocols ([Table 1](#)).<sup>41; 42</sup> Radiocarbon dating of BKZ001 was conducted by the CEZ  
158 Archaeometry gGmbH (Mannheim, Germany) for one of two bone samples used for DNA extraction. All  
159 libraries were barcoded with two library-specific 8-mer indices.<sup>43</sup> The samples were manipulated in  
160 dedicated clean room facilities at the University of Tübingen or at the Max Planck Institute for the  
161 Science of Human History (MPI-SHH). Indexed libraries were enriched for about 1.24 million  
162 informative nuclear SNPs using the in-solution capture method (“1240K capture”).<sup>5; 31</sup>

163 Libraries were sequenced on the Illumina HiSeq 4000 platform with either single-end 75 bp  
164 (SE75) or paired-end 50 bp (PE50) cycles following manufacturer’s protocols. Output reads were  
165 demultiplexed by allowing up to 1 mismatch in each of two 8-mer indices. FASTQ files were processed  
166 using EAGER v1.92.<sup>44</sup> Specifically, Illumina adapter sequences were trimmed using AdapterRemoval  
167 v2.2.0,<sup>45</sup> aligned reads (30 base pairs or longer) onto the human reference genome (hg19) using BWA  
168 aln/samse v0.7.12<sup>46</sup> with relaxed edit distance parameter (“-n 0.01”). Seeding was disabled for reads from  
169 non-UDG libraries by adding an additional parameter (“-1 9999”). PCR duplicates were then removed  
170 using DeDup v0.12.2<sup>44</sup> and reads with Phred-scaled mapping quality score < 30 were filtered out using  
171 Samtools v1.3.<sup>47</sup> We did several measurements to check data authenticity. First, patterns of chemical  
172 damages typical to ancient DNA were tabulated using mapDamage v2.0.6.<sup>48</sup> Second, mitochondrial  
173 contamination for all libraries was estimated by Schmutzi.<sup>49</sup> Third, nuclear contamination for libraries  
174 derived from males was estimated by the contamination module in ANGSD v0.910.<sup>50</sup> Prior to genotyping,  
175 the first and last 3 bases of each read were masked for libraries with partial UDG treatment using the  
176 trimBam module in bamUtil v1.0.13.<sup>51</sup> To obtain haploid genotypes, we randomly chose one high-quality  
177 base (Phred-scaled base quality score  $\geq 30$ ) for each of the 1.24 million target sites using pileupCaller  
178 (<https://github.com/stschiff/sequenceTools>). We used masked reads from libraries with partial UDG  
179 treatment for transition (Ts) SNPs and used unmasked reads from all libraries for transversions (Tv).

180 Mitochondrial consensus sequences were obtained by the log2fasta program in Schmutzi with the quality  
181 cutoff 10 and subsequently assigned to haplogroups using HaploGrep2.<sup>52</sup> Y haplogroup R1b was assigned  
182 using the yHaplo program.<sup>53</sup> To estimate the phylogenetic position of the Botai Y haplogroup more  
183 precisely, Y chromosomal SNPs were called with Samtools mpileup using bases with quality score  $\geq 30$ :  
184 a total of 2,481 SNPs out of ~30,000 markers included in the 1240K capture panel were called with mean  
185 read depth of 1.2. Twenty-two SNP positions relevant to the up-to-date haplogroup R1b tree  
186 ([www.isogg.org](http://www.isogg.org); [www.yfull.com](http://www.yfull.com)) confirmed that the sample was positive for the markers of R1b-P297  
187 branch but negative for its R1b-M269 sub-branch.

188 The frequency distribution map of this Y chromosomal clade was created by the GeneGeo  
189 software<sup>54;55</sup> using the average weighed interpolation procedure with the weight function of degree 3 and  
190 radius 1, 200 km. The initial frequencies were calculated as proportion of samples positive for “root” R1b  
191 marker M343 but negative for M269; these proportions were calculated for the 577 populations from the  
192 in-home *Y-base* database, which was compiled mainly from the published datasets.

193

#### 194 ***Analysis of population structure***

195 We performed principal component analysis (PCA) of various groups using smartpca v13050 in  
196 the EIGENSOFT v6.0.1 package.<sup>56</sup> We used the “*lsqproject: YES*” option to project individuals not used  
197 for calculating PCs (this procedure avoids bias due to missing genotypes). We performed unsupervised  
198 model-based genetic clustering as implemented in ADMIXTURE v1.3.0.<sup>57</sup> For that purpose, we used  
199 116,468 SNPs with minor allele frequency (maf) 1% or higher in 3,332 individuals after pruning out  
200 linked SNPs ( $r^2 > 0.2$ ) using the “--indep-pairwise 200 25 0.2” command in PLINK v1.90.<sup>58</sup> We then  
201 converted diploid genotypes to haploid data by randomly choosing one of the two alleles to minimize a  
202 bias due to artificial genetic drift in haploid genotype calls of most low coverage ancient individuals. For  
203 each value of K ranging from 2 to 20, we ran 5 replicates with different random seeds and took one with  
204 the highest log likelihood value.

205

206 ***F-statistics analysis***

207 We computed various  $f_3$  and  $f_4$  statistics using the qp3Pop (v400) and qpDstat (v711) programs in  
208 the ADMIXTOOLS package.<sup>22</sup> We computed  $f_4$ -statistics with the “*f4mode: YES*” option. For these  
209 analyses, we studied a total of 301 groups, including 73 inner Eurasian target groups and 167  
210 contemporary and 61 ancient reference groups (Table S2). We included two groups from the Aleutian  
211 Islands (“Aleut” and “Aleut\_Tlingit”; Table S2) as positive control targets with known recent admixture.  
212 Aleut\_Tlingits are Aleut individuals whose mitochondrial haplogroup lineages are related to Tlingits.<sup>29</sup>  
213 For each target, we calculated outgroup  $f_3$  statistic of the form  $f_3(\text{Target}, X; \text{Mbuti})$  against all targets and  
214 references to quantify overall allele sharing and performed admixture  $f_3$  test of the form  $f_3(\text{Ref}_1, \text{Ref}_2;$   
215 Target) for all pairs of references to explore the admixture signal in targets. We estimated standard error  
216 (SE) using a block jackknife with 5 centiMorgan (cM) block.<sup>56</sup>

217 We performed  $f_4$  statistic-based admixture modeling using the qpAdm (v632) program<sup>19</sup> in the  
218 ADMIXTOOLS package. We used a basic set of 7 outgroups, unless specified otherwise, to provide high  
219 enough resolution to distinguish various western and eastern Eurasian ancestries: Mbuti (n=10; central  
220 African), Natufian (n=6; early Holocene Levantine),<sup>19</sup> Onge (n=11; from the Andaman Islands), Iran\_N  
221 (n=5; Neolithic Iranian),<sup>19</sup> Villabruna (n=1; Paleolithic European),<sup>26</sup> Ami (n=10; Taiwanese aborigine)  
222 and Mixe (n=10; Central American). Prior to qpAdm modeling, we checked if the reference groups are  
223 well distinguished by their relationship with the outgroups using the qpWave (v400) program.<sup>59</sup>

224 We used the qpGraph (v6065) program in the ADMIXTOOLS package for graph-based  
225 admixture modeling. Starting with a graph of (Mbuti, Ami, WHG), we iteratively added AG3 (n=1;  
226 Paleolithic Siberian),<sup>26</sup> EHG (n=3; Mesolithic hunter-gatherers from Karelia or Samara),<sup>5;26</sup> and Botai  
227 onto the graph by testing all possible topologies allowing up to one additional gene flow. After obtaining  
228 the best two-way admixture model for Botai, we tested additional three-way admixture models.

229

230 ***GLOBETROTTER analysis***



231 We performed a GLOBETROTTER analysis of admixture for 73 inner Eurasian target  
232 populations to obtain haplotype sharing based evidence of admixture, independent of the allele frequency  
233 based  $f$ -statistics, as well as estimates of admixture dates and a fine-scale profile of their admixture  
234 sources.<sup>14</sup> We followed the “regional” approach described in Hellenthal et al.,<sup>14</sup> in which target  
235 haplotypes can only be copied from the haplotypes of 167 contemporary reference groups, but not from  
236 those of the other target groups. This approach is recommended when multiple target groups share a  
237 similar admixture history,<sup>14</sup> which is likely to be the case for our inner Eurasian populations.

238 We jointly phased the contemporary genome data without a pre-phased set of reference  
239 haplotypes, using SHAPEIT2 v2.837 in its default setting.<sup>60</sup> We used a genetic map for the 1000  
240 Genomes Project phase 3 data, downloaded from:  
241 [https://mathgen.stats.ox.ac.uk/impute/1000GP\\_Phase3.html](https://mathgen.stats.ox.ac.uk/impute/1000GP_Phase3.html). We used haplotypes from a total of 2,615  
242 individuals belonging to 240 groups (73 recipients and 167 donors; [Table S2](#)) for the GLOBETROTTER  
243 analysis. To reduce computational burden and to provide more balanced set of donor populations, we  
244 randomly sampled 20 individuals if a group contained more than 20 individuals. Using these haplotypes,  
245 we performed GLOBETROTTER analysis following the recommended workflow.<sup>14</sup> We first ran 10  
246 rounds of the expectation-maximization (EM) algorithm for chromosomes 4, 10, 15 and 22 in  
247 ChromoPainter v2 with “-in” and “-iM” switches to estimate chunk size and switch error rate  
248 parameters.<sup>61</sup> Both recipient and donor haplotypes were modeled as a patchwork of donor haplotypes. The  
249 “chunk length” output was obtained by running ChromoPainter v2 across all chromosomes with the  
250 estimated parameters averaged over both recipient and donor individuals (“-n 238.05 -M 0.000617341”).  
251 We also generated 10 painting samples for each recipient group by running ChromoPainter with the  
252 parameters averaged over all recipient individuals (“-n 248.455 -M 0.000535236”). Using the  
253 chunklength output and painting samples, we ran GLOBETROTTER with the “prop.ind: 1” and “null.ind:  
254 1” options. We estimated significance of estimated admixture date by running 100 bootstrap replicates  
255 using the “prop.ind: 0” and “bootstrap.date.ind: 1” options; we considered date estimates between 1 and  
256 400 generations as evidence of admixture.<sup>14</sup> For populations that gave evidence of admixture by this

257 procedure, we repeated GLOBETROTTER analysis with the “null:ind: 0” option.<sup>14</sup> We also compared  
258 admixture dates from GLOBETROTTER analysis with those based on weighted admixture linkage  
259 disequilibrium (LD) decay, as implemented in ALDER v1.3.<sup>62</sup> As the reference pair, we used (French,  
260 Eskimo\_Naukan), (French, Nganasan), (Georgian, Ulchi), (French, Ulchi) and (Georgian, Ulchi) for the  
261 target group categories 1 to 5, respectively, based on their genetic profile (Table S2). We used a minimum  
262 inter-marker distance of 1.0 cM to account for LD in the references.

263

### 264 *EEMS analysis*

265 To visualize the heterogeneity in the rate of gene flow across inner Eurasia, we performed the  
266 EEMS (“estimated effective migration surface”) analysis.<sup>63</sup> We included a total of 1,180 individuals from  
267 94 groups in the analysis (Table S2). In this dataset, we kept 101,320 SNPs with  $\text{maf} \geq 0.01$  after LD  
268 pruning ( $r^2 \leq 0.2$ ). We computed the mean squared genetic difference matrix between all pairs of  
269 individuals using the “bed2diffs\_v1” program in the EEMS package. To reduce distortion in northern  
270 latitudes due to map projection, we used geographic coordinates in the Albers equal area conic projection  
271 (“+proj=aea +lat\_1=50 +lat\_2=70 +lat\_0=56 +lon\_0=100 +x\_0=0 +y\_0=0 +ellps=WGS84  
272 +datum=WGS84 +units=m +no\_defs”). We converted geographic coordinates of each sample and the  
273 boundary using the “spTransform” function in the R package rgdal v1.2-5. We ran five initial MCMC  
274 runs of 2 million burn-ins and 4 million iterations with different random seeds and took a run with the  
275 highest likelihood. Starting from the best initial run, we set up another five MCMC runs of 2 million  
276 burn-ins and 4 million iterations as our final analysis. We used the following proposal variance  
277 parameters to keep the acceptance rate around 30-40%, as recommended by the developers<sup>63</sup>:  
278  $\text{qSeedsProposalS2} = 5000$ ,  $\text{mSeedsProposalS2} = 1000$ ,  $\text{qEffctProposalS2} = 0.0001$ ,  $\text{mrateMuProposalS2}$   
279  $= 0.00005$ . We set up a total of 532 demes automatically with the “nDemes = 600” parameter. We  
280 visualized the merged output from all five runs using the “eems.plots” function in the R package  
281 rEEMSplots.<sup>63</sup>

282 We performed the EEMS analysis for Caucasus populations in a similar manner, including a total  
283 of 237 individuals from 21 groups (Table S2). In this dataset, we kept 95,442 SNPs with  $\text{maf} \geq 0.01$  after  
284 LD pruning ( $r^2 \leq 0.2$ ). We applied the Mercator projection of geographic coordinates to the map of  
285 Eurasia (“+proj=merc +datum=WGS84”). We ran five initial MCMC runs of 2 million burn-ins and 4  
286 million iterations with different random seeds and took a run with the highest likelihood. Starting from  
287 the best initial run, we set up another five MCMC runs of 1 million burn-in and 4 million iterations as our  
288 final analysis. We used the default following proposal variance parameters:  $\text{qSeedsProposalS2} = 0.1$ ,  
289  $\text{mSeedsProposalS2} = 0.01$ ,  $\text{qEffctProposalS2} = 0.001$ ,  $\text{mrateMuProposalS2} = 0.01$ . A total of 171 demes  
290 were automatically set up with the “nDemes = 200” parameter.

291

292

## 293 **Results**

294

### 295 *Inner Eurasians form distinct east-west genetic clines mirroring geography*

296 In a PCA of Eurasian individuals, we find that PC1 separates eastern and western Eurasian  
297 populations, PC2 splits eastern Eurasians along a north-south cline, and PC3 captures variation in western  
298 Eurasians with Caucasus and northeastern European populations at opposite ends (Figure 2A and Figures  
299 S1-S2). Inner Eurasians are scattered across PC1 in between, largely reflecting their geographic locations.  
300 Strikingly, inner Eurasian populations seem to be structured into three distinct west-east genetic clines  
301 running between different western and eastern Eurasian groups, instead of being evenly spaced in PC  
302 space. Individuals from northern Eurasia, speaking Uralic or Yeniseian languages, form a cline  
303 connecting northeast Europeans and the Uralic (Samoyedic) speaking Nganasans from northern Siberia  
304 (“forest-tundra” cline). Individuals from the Eurasian steppe, mostly speaking Turkic and Mongolic  
305 languages, are scattered along two clines below the forest-tundra cline. Both clines run into Turkic- and  
306 Mongolic-speaking populations in southern Siberia and Mongolia, and further into Tungusic-speaking  
307 populations in Manchuria and the Russian Far East in the East; however, they diverge in the west, one

308 heading to the Caucasus and the other heading to populations of the Volga-Ural area (the “southern steppe”  
309 and “steppe-forest” clines, respectively; [Figure 2](#) and [Figure S2](#)).

310 A model-based clustering analysis using ADMIXTURE shows a similar pattern ([Figure 2B](#) and  
311 [Figure S3](#)). Overall, the proportions of ancestry components associated with eastern or western Eurasians  
312 are well correlated with longitude in inner Eurasians ([Figure 3A](#)). Notable outliers from this trend include  
313 known historical migrants such as Kalmyks, Nogais and Dungans. The forest-tundra cline populations  
314 derive most of their eastern Eurasian ancestry from a component most enriched in Nganasans, while those  
315 on the steppe-forest and southern steppe clines have this component together with another component  
316 most enriched in populations from the Russian Far East, such as Ulchi and Nivkh. The southern steppe  
317 cline groups are distinct from the others in their western Eurasian ancestry profile, in the sense that they  
318 have a high proportion of a component most enriched in Mesolithic Caucasus hunter-gatherers (“CHG”)<sup>28</sup>  
319 and Neolithic Iranians (“Iran\_N”)<sup>19</sup> and frequently harbor another component enriched in South Asians  
320 ([Figure S4](#)).

321 The genetic barriers splitting the inner Eurasian clines are also evidenced in the EEMS  
322 (“estimated effective migration surface”) analysis ([Figure 3B](#)). A strong genetic barrier is detected  
323 between the Caucasus and the Pontic-Caspian steppe regions, separating the southern steppe and steppe-  
324 forest clines. On the eastern side, another barrier north of Lake Baikal separates southern Siberians from  
325 the forest-tundra cline groups in the North. These two barriers are partially connected by a weaker barrier  
326 north of the Altai-Sayan region, likely reflecting both the east-west connection within the steppe-forest  
327 cline and the north-south connection along the Yenisei River.

328

### 329 *High-resolution tests of admixture distinguish the genetic profile of source populations in the inner* 330 *Eurasian clines*

331 We performed both allele frequency-based three-population ( $f_3$ ) tests and a haplotype-sharing-  
332 based GLOBETROTTER analysis to characterize the admixed gene pools of inner Eurasian groups. For  
333 these group-based analyses, we manually removed 87 outliers from our contemporary individuals based

334 on PCA results (Table S1). We also split a few inner Eurasian groups showing genetic heterogeneity into  
335 subgroups based on PCA results and their sampling locations (Table S1). This was done to minimize false  
336 positive admixture signals. We chose 73 groups as the targets of admixture tests and another 228 groups  
337 (167 contemporary and 61 ancient groups) as the “sources” to represent world-wide genetic diversity  
338 (Table S2).

339 Testing all possible pairs of 167 contemporary “source” groups as references, we detect highly  
340 significant  $f_3$  statistics for 66 of 73 targets ( $< -3$  SE; standard error; Table S3). Negative  $f_3$  values mean  
341 that allele frequencies of the target group are on average intermediate between the allele frequencies of  
342 the reference populations, providing unambiguous evidence that the target population is a mixture of  
343 groups related, perhaps deeply, to the source populations.<sup>22</sup> Extending the references to include 61 ancient  
344 groups, we find that the seven non-significant groups also have small  $f_3$  statistics around zero ( $-5.1$  SE to  
345  $+2.7$  SE). Reference pairs with the most negative  $f_3$  statistics for the most part involve one eastern  
346 Eurasian and one western Eurasian group supporting the qualitative impression of east-west admixture  
347 from PCA and ADMIXTURE analysis. To highlight the difference between the distinct inner Eurasian  
348 clines, we looked into  $f_3$  results with representative reference pairs comprising two western Eurasian  
349 (French to represent Europeans and Georgian to represent Caucasus populations) and three eastern  
350 Eurasian groups (Nganasan, Ulchi and Korean). In the populations of the southern steppe cline, reference  
351 pairs with Georgians tend to produce more negative  $f_3$  statistics than those with French while the opposite  
352 pattern is observed for the steppe-forest and forest-tundra populations (Figure 4A). Reference pairs with  
353 Nganasans mostly result in more negative  $f_3$  statistic than those with Ulchi in the forest-tundra  
354 populations, but the opposite pattern is dominant in the southern steppe populations. Populations of the  
355 steppe-forest cline show an intermediate pattern: the northern ones tend to have more negative  $f_3$  statistics  
356 with Nganasans while the southern ones tend to have more negative  $f_3$  statistics with Ulchi.

357 To perform a higher resolution characterization of the admixture landscape, we performed a  
358 haplotype-based GLOBETROTTER analysis. We took a “regional” approach, meaning that all 73  
359 recipient groups were modeled as a patchwork of haplotypes from the 167 donor groups but not those

360 from any recipient group. The goal of this approach was to minimize false negative results due to sharing  
361 of admixture history between recipient groups. All of 73 recipient groups show a robust signal of  
362 admixture: i.e. a correlation of ancestry status shows a distinct pattern of decay over genetic distance in  
363 all bootstrap replicates (bootstrap  $p < 0.01$  for all 73 targets; [Table S4](#)). When the relative contribution of  
364 donors, categorized to 12 groups ([Table S2](#)), into the two main sources of the admixture signal (“date 1  
365 PC 1”) is considered, we observe a pattern comparable to PCA, ADMIXTURE and  $f_3$  results ([Figure 4B](#)).  
366 The European donors provide a major contribution for the western Eurasian-related source in the forest-  
367 tundra and steppe-forest recipients while the Caucasus/Iranian donors do so in the southern steppe  
368 recipients. Similarly, Siberian donors make the highest contribution to the eastern Eurasian-related source  
369 in the forest-tundra recipients, followed by the steppe-forest and southern steppe ones.

370 The GLOBETROTTER analysis also provides an estimate of admixture dates, either one- or two-  
371 date estimates, depending on the best model of admixture ([Figure S5](#) and [Table S4](#)). We obtain a mean  
372 admixture date estimate of 24.3 generations for the steppe-forest and southern steppe cline populations,  
373 ranging from 10.7 to 38.1 generations (309 to 1104 years ago, using 29 years per generation<sup>64</sup>). These  
374 young dates do not change much even when taking the older dates from the two-date model, as here we  
375 obtain a mean of 29.8 generations ranging from 10.7 to 68.1 generations (310 to 1975 years ago). The  
376 forest-tundra cline groups have older estimates with a mean of 40.1 generations and a range of 6.8-55.2  
377 generations (197 to 1601 years ago). All but two groups have an estimate older than the steppe mean of  
378 29.8 generations. Estimates of admixture dates using ALDER result in similar values ([Figure S5](#)). The  
379 admixture dates of the steppe populations are consistent with previous estimates using similar  
380 methodologies,<sup>13</sup> but much younger than expected if they had been driven by admixtures in the Late  
381 Bronze and Iron Ages.<sup>8; 38</sup>

382

### 383 ***The Eneolithic Botai gene pool provides a glimpse of a lost prehistoric cline***

384 The Eneolithic Botai individuals are closer to each other in the PC space than to any other ancient  
385 or present-day individual, and are in proximity to the upper Paleolithic Siberians from the Mal'ta (MA-1)

386 or Afontova Gora (AG3) archaeological sites (Figure 2). Consistent with this, Botai has the highest  
387 outgroup  $f_3$  statistic with AG3 and other upper Paleolithic Siberians, as well as with the Mesolithic eastern  
388 European hunter-gathers from Karelia and Samara (“EHG”) (Figure S6A). East Asians (EAS) are more  
389 closely related to Botai than to AG3 as shown by significantly positive  $f_4$  symmetry statistics in the form  
390 of  $f_4(\text{Mbuti}, \text{EAS}; \text{AG3}, \text{Botai})$ , suggesting East Asian gene flow into Botai (Figure S6B).

391 We estimated the proportion of East Asian ancestry in Botai using qpAdm. The two-way  
392 admixture model of AG3+Korean provides a good fit to Botai with 17.3% East Asian contribution ( $\chi^2 p =$   
393 0.286; Table S5), while the models EHG+EAS do not fit ( $\chi^2 p \leq 1.44 \times 10^{-7}$ ). However, we find that Botai  
394 harbors an extra affinity with Mesolithic western European hunter-gatherers (“WHG”) unexplained by  
395 this model:  $f_4(\text{Mbuti}, \text{WHG}; \text{AG3}+\text{Korean}, \text{Botai})$  is significantly positive in a plausible range of the  
396 ancestry proportions (+3.0 to +4.2 SE for 77.7-87.7% AG3 ancestry, mean  $\pm 2$  SE; Figure S7). We still  
397 obtain a reasonable fit for the same model when we add WHG to the outgroups ( $\chi^2 p = 0.089$ ; Table S5),  
398 but adding EHG as an additional source slightly increases model fit with a similar amount of contribution  
399 from the East Asian source ( $\chi^2 p = 0.016$ ; 17.3 $\pm$ 2.2% East Asian contribution; Table S5).

400 A graph-based admixture modeling using qpGraph provides similar results: the best two-way  
401 admixture model for Botai added to a scaffold graph composed of Mbuti, Onge, Ami, AG3, WHG, and  
402 EHG still shows an unexplained affinity between WHG and Botai, and adding an admixture edge from  
403 EHG-related branches substantially improves the model fit (Figure S8). Thus, we conclude that the ANE-  
404 related ancestry in Botai is intermediate between EHG and AG3, which corresponds to its intermediate  
405 geographic position. This suggests a genetic cline of decreasing ANE-related ancestry stretching from  
406 AG3 in Siberia to WHG in Western Europe. A substantial East Asian contribution into Botai make them  
407 offset from the WHG-ANE cline. A strong genetic affinity between Botai and the Middle Bronze Age  
408 Okunevo individuals in the Altai-Sayan region also suggests a wide geographic and temporal distribution  
409 of Botai-related ancestry in central Eurasia (Figure S6C).

410 The Y-chromosome of the male Botai individual (TU45) belongs to the haplogroup R1b (Table  
411 S6). However, it falls into neither a predominant European branch R1b-L51<sup>65</sup> nor into a R1b-GG400

412 branch found in Yamnaya individuals.<sup>66</sup> Thus, phylogenetically this Botai individual should belong to the  
413 R1b-M73 branch which is frequent in the Eurasian steppe (Figure S9). This branch was also found in  
414 Mesolithic samples from Latvia<sup>67</sup> as well as in numerous modern southern Siberian and Central Asian  
415 groups.

416

417 *Admixture modeling of contemporary inner Eurasians shows multiple gene flows producing new*  
418 *genetic clines overwriting the ancient ones*

419 Our results show that contemporary inner Eurasians form genetic clines distinct from the ancient  
420 WHG-ANE cline, from which a majority of the Botai ancestry is derived. To see if this ancient cline of  
421 “ANE” ancestry left any legacy in the genetic structure of inner Eurasians, we performed admixture  
422 modeling of populations from the Altai-Sayan region and those belonging to the forest-tundra cline.  
423 Specifically, we investigated if an additional contribution from ANE-related ancestry is required to  
424 explain their gene pools beyond a simple mixture model of contemporary eastern Eurasians and ancient  
425 western Eurasian populations.

426 Contemporary Altai-Sayan populations are effectively modeled as a two-way mixture of ancient  
427 populations from the region with WSH ancestry and contemporary eastern Eurasians, either  
428 Afanasievo+Ulchi or Sintashta+Nganasan ( $\chi^2 p \geq 0.05$  for 8/12 and 5/12 Altai-Sayan groups, respectively;  
429 Table S7). Among the ancient groups, Sintashta+EAS generally fits Andronovo individuals well with a  
430 small eastern Eurasian contribution (6.4±1.4% for estimate ± 1 SE with Nganasans), while later Karasuk  
431 or Iron Age individuals from the Altai are modeled better with the older Afanasievo as their WSH-related  
432 source (Table S7). If the pre-Bronze Age populations of the Altai-Sayan region were related to either  
433 Botai in the west or the Upper Paleolithic Siberians in the east, these results suggest that these pre-Bronze  
434 Age populations in southern Siberia did not leave a substantial genetic legacy in the present-day  
435 populations in the region. The Okunevo individuals are the only case that WSH+EAS mixture cannot  
436 explain ( $\chi^2 p \leq 3.85 \times 10^{-4}$ ); similar to Botai, a model of AG3+EAS provides a good fit ( $\chi^2 p = 0.396$  for  
437 AG3+Korean; Table S5).



438 For the forest-tundra cline populations, for which currently no relevant Holocene ancient  
439 genomes are available, we took a more generalized approach of using proxies for contemporary  
440 Europeans: WHG, WSH (represented by “Yamnaya\_Samara”), and early Neolithic European farmers  
441 (EEF; represented by “LBK\_EN”; [Table S2](#)). Adding Nganasans as the fourth reference, we find that  
442 most Uralic-speaking populations in Europe (i.e. west of the Urals) and Russians are well modeled by this  
443 four-way admixture model ( $\chi^2 p \geq 0.05$  for all but three groups; [Figure 5](#) and [Table S8](#)). Nganasan-related  
444 ancestry substantially contributes to their gene pools and cannot be removed from the model without a  
445 significant decrease in model fit (4.7% to 29.1% contribution;  $\chi^2 p \leq 1.12 \times 10^{-8}$ ; [Table S8](#)). The ratio of  
446 contributions from three European references varies from group to group, probably reflecting genetic  
447 exchange with neighboring non-Uralic groups. For example, Saami from northern Fennoscandia contain a  
448 higher WHG and lower WSH contribution (16.1% and 41.3%, respectively) than Udmurts or Besermyans  
449 from the Volga river region do (4.9-6.6% and 50.7-53.2%, respectively), while the three groups have  
450 similar amounts of Nganasan-related ancestry (25.5-29.1%).

451 For the four forest-tundra cline groups east of the Urals (Enets, Selkups, Kets and Mansi), the  
452 above four-way model estimates negative contribution from EEF ( $< -1.6\%$ ). Replacing EEF with EHG,  
453 one of the top  $f_3$  references for these groups, we obtain well-fitted models with a small WHG contribution  
454 ( $\chi^2 p \geq 0.253$ ; -1.0% to 5.5% WHG contributions). The three-way model excluding WHG shows a good  
455 fit for Enets, Selkups and Kets ( $\chi^2 p \geq 0.098$ ; [Figure 5](#)). Simpler models without either EHG or WSH  
456 ancestry do not fit ( $\chi^2 p \leq 0.019$  and 0.003, respectively), suggesting a legacy of the ancient WHG-ANE  
457 cline.

458

### 459 *The Caucasus Mountains form a barrier to gene flow*

460 When the Altai-Sayan Mountains are often considered as a crossroad of migrations and mark the  
461 eastern boundary of the western Eurasian steppe, the Caucasus area plays a similar role for the western  
462 end of the steppe. To explore the genetic structure of populations of the Caucasus region, we first  
463 performed a PCA of western Eurasians including Caucasus populations ([Figure S10](#)). Consistent with

464 previous studies,<sup>4</sup> Caucasus populations are clustered on the PC space in the vicinity of West Asians  
465 further in the south but far from eastern Europeans. The genetic structure within the Caucasus is less  
466 pronounced but still evident: populations from the North and South Caucasus, geographically divided by  
467 the Greater Caucasus ridge, also show a genetic differentiation. North Caucasus populations show a  
468 further subdivision into northwest and northeast groups.

469 By applying EEMS to the Caucasus region, we identify a strong barrier to gene flow separating  
470 North and South Caucasus populations (Figure 6). This genetic barrier coincides with the Greater  
471 Caucasus mountain ridge even to small scale: a weaker barrier in the middle, overlapping with Ossetia,  
472 matches well with the region where the ridge also becomes narrow. We also observe weak barriers  
473 running in the north-south direction that separate northeastern populations from northwestern ones.  
474 Together with PCA, EEMS results suggest that the Caucasus Mountains have posed a strong barrier to  
475 human migration.

476 We quantified the genetic difference within Caucasus populations using  $f_4$  statistics of the form  
477  $f_4(\text{Mbuti}, X; \text{Caucasus}_1, \text{Caucasus}_2)$  against world-wide populations outside the Caucasus (“X”). We find  
478 many significant  $f_4$  statistics suggesting that gene flows from exogenous gene pools have been involved in  
479 the development of the population structure of the Caucasus (Figure S11). Compared to both northwest  
480 and northeast groups, South Caucasians show extra affinity to Near Eastern populations, such as Neolithic  
481 Levantines and Anatolians (“Levant\_N” and “Anatolia\_N”, respectively; Table S2). In turn, North  
482 Caucasus populations have extra affinity with populations of the steppe and broadly of eastern Eurasia.  
483 Northeast Caucasians, for example Laks and Lezgins, show the strongest signals with ANE- and WSH-  
484 related ancient groups, with MA-1, AG3, Botai and EHG at the top. Northwest Caucasians (e.g. Adygei  
485 and Ossetians) are closer to East Asians than Northeast or South Caucasians are. We speculate that these  
486 results may suggest at least two layers of gene flow into the North Caucasus region: an older layer related  
487 to the ANE- or WSH-related ancestries and the younger layer related to East Asians. The former may  
488 have involved an interaction with Iron Age nomads, such as Scythians or Sarmatians. The latter most  
489 strongly affected Northwest Caucasians and might be related to historical movements of Turkic

490 populations with some East Asian ancestry into the Caucasus. The genetic legacy of this movement is  
491 obvious for the Nogais that are scattered along PC1 between the rest of Caucasus populations and Central  
492 Asians ([Figures S1-S2](#)).

493 To explicitly model and quantify the steppe-related gene flows into the Caucasus, we performed a  
494 qpAdm-based admixture modeling of 22 Caucasus populations. For 7 of 22 Caucasus populations, a two-  
495 way admixture model using Armenians and an ancient Scythian individual<sup>31</sup> is sufficient ( $\chi^2 p \geq 0.05$ ;  
496 [Table S9](#)). Except for Georgians from the South Caucasus (6.8% contribution from Scythians), all the  
497 other groups have a substantial contribution from Scythians (38.0-50.6%). When we add Nanais as the  
498 third reference to model potential gene flow from Eastern Eurasians, most of the Caucasus populations  
499 are consistent with the model: 15 of 22 Caucasus populations with  $\chi^2 p \geq 0.05$  and another three with  $\chi^2 p$   
500  $\geq 0.01$  ([Table S9](#)). 9 of the 15 groups are adequately modeled by the three references but not by the two:  
501 they indeed have positive admixture coefficients for Nanais. Except for Nogais (19.8% for Nogai1 and  
502 48.0% for Nogai2), the other seven groups have only a small amount of East Asian ancestry that is  
503 prominent neither in PCA nor in ADMIXTURE (2.7-5.1%; [Table S9](#)).

504

505

## 506 **Discussion**

507 In this study, we analyzed newly reported genome-wide variation data of indigenous people from  
508 inner Eurasia, providing a dense representation for human genetic diversity in this vast region. Our  
509 finding of inner Eurasian populations being structured into three distinct clines shows a striking  
510 correlation between genes and geography ([Figures 1-2](#)). The genetic grouping of samples into three clines  
511 with gaps in between ([Figure 2](#)) corresponds with the fact that samples tend to group into the same clines  
512 on the geographic map ([Figure 1](#)) with lower density of studied populations in between. However, this  
513 non-uniformity of sampling results from the non-uniformity in the density of (language-defined) ethnic  
514 groups. Moreover, the reality of the clines was confirmed by the barrier and  $f_4$  analyses. The steppe cline  
515 populations derive their eastern Eurasian ancestry from a gene pool similar to contemporary Tungusic

516 speakers from the Amur river basin (Figures 2 and 4), thus suggesting a genetic connection among the  
517 speakers of languages belonging to the Altaic macrofamily (Turkic, Mongolic and Tungusic families).  
518 Based on our results as well as early Neolithic genomes from the Russian Far East,<sup>37</sup> we speculate that  
519 such a gene pool may represent the genetic profile of prehistoric hunter-gatherers in the Amur river basin.  
520 On the other hand, a distinct Nganasan-related eastern Eurasian ancestry in the forest-tundra cline  
521 suggests a substantial separation between these two eastern ancestries. Nganasans have high genetic  
522 affinity with prehistoric individuals with the “ANE” ancestry in North Eurasia, such as the Upper  
523 Paleolithic Siberians or the Mesolithic EHG, which is exceeded only by Native Americans and by  
524 Beringians among eastern Eurasians (Figure S12). Also, Northeast Asians are closer to Nganasans than  
525 they are to either Beringians or Native Americans, and the ANE affinity in East Asians is correlated well  
526 with their affinity with Nganasans (Figure S13). We hypothesize that Nganasans may be relatively  
527 isolated descendants of a prehistoric Siberian gene pool, which formed modern Northeast Asians by  
528 mixing with populations related to the Neolithic Northeast Asians.<sup>37</sup>

529         The Botai genomes provide a critical snapshot of the genetic profile of pre-Bronze Age steppe  
530 populations. Our admixture modeling positions Botai primarily on an ancient genetic cline of the pre-  
531 Neolithic western Eurasian hunter-gatherers: stretching from the post-Ice Age western European hunter-  
532 gatherers (e.g. WHG) to EHG in Karelia and Samara to the Upper Paleolithic southern Siberians (e.g.  
533 AG3). Botai’s position on this cline, between EHG and AG3, fits well with their geographic location and  
534 suggests that ANE-related ancestry in the East did have a lingering genetic impact on Holocene Siberian  
535 and Central Asian populations at least till the time of Botai. A recent study reports 6,000 to 8,000 year old  
536 genomes from a region slightly north of Botai, whose genetic profiles are similar to our Botai  
537 individuals.<sup>68</sup> This ancient cline in Altai-Sayan region has now largely been overwritten by waves of  
538 genetic admixtures. Starting from the Eneolithic Afanasievo culture, multiple migrations from the Pontic-  
539 Caspian steppe to the east have significantly changed the western Eurasian ancestry during the Bronze  
540 Age.<sup>7; 8</sup> Our admixture modeling finds that no contemporary population in the Altai-Sayan region is  
541 required to have additional ANE ancestry beyond what the mixture model of Bronze Age steppe plus

542 modern Eastern Eurasians can explain (Table S7). The most recent clear connection with the Botai  
543 ancestry can be found in the Middle Bronze Age Okunevo individuals (Figure S6C). In contrast,  
544 additional EHG-related ancestry is required to explain the forest-tundra populations to the east of the  
545 Urals (Figure 5 and Table S8). Their multi-way mixture model may in fact portrait a prehistoric two-way  
546 mixture of a WSH population and a hypothetical eastern Eurasian one that has an ANE-related  
547 contribution higher than that in Nganasans. Botai and Okunevo individuals prove the existence of such  
548 ANE ancestry-rich populations. Pre-Bronze Age genomes from Siberia will be critical for testing this  
549 hypothesis.

550         The study of ancient genomes from inner Eurasia will be extremely important for going forward.  
551 Inner Eurasia has functioned as a conduit for human migration and cultural transfer since the first  
552 appearance of modern humans in this region. As a result, we observe deep sharing of genes between  
553 western and eastern Eurasian populations in multiple layers: the Pleistocene ANE ancestry in Mesolithic  
554 EHG and contemporary Native Americans, Bronze Age steppe ancestry from Europe to Mongolia, and  
555 Nganasan-related ancestry extending from western Siberia into Eastern Europe. More recent historical  
556 migrations, such as the westward expansions of Turkic and Mongolic groups, further complicate genomic  
557 signatures of admixture and have overwritten those from older events. Ancient genomes of Iron Age  
558 steppe individuals, already showing signatures of west-east admixture in the 5<sup>th</sup> to 2<sup>nd</sup> century BCE,<sup>38</sup>  
559 provide further direct evidence for the hidden old layers of admixture, which is often difficult to  
560 appreciate from present-day populations as shown in our finding of a discrepancy between the estimates  
561 of admixture dates from contemporary individuals and those from ancient genomes.

562

563

#### 564 **Supplemental Data**

565 Supplemental Data include 13 figures and 9 tables.

566

#### 567 **Declaration of Interests**

568 The authors declare no competing interests.

569

## 570 **Acknowledgements**

571 We thank Iain Mathieson and Iosif Lazaridis for their helpful comments. The research leading to these  
572 results has received funding from the Max Planck Society, the Max Planck Society Donation Award and  
573 the European Research Council (ERC) under the European Union's Horizon 2020 research and  
574 innovation programme (grant agreement No 646612 granted to M.R.). Analysis of the Caucasus dataset  
575 was supported by RFBR grant 16-06-00364 and analysis of the Far East dataset was supported by Russian  
576 Scientific Fund project 17-14-01345. D.R. was supported by the U.S. National Science Foundation  
577 HOMINID grant BCS-1032255, the U.S. National Institutes of Health grant GM100233, by an Allen  
578 Discovery Center grant, and is an investigator of the Howard Hughes Medical Institute. P.F. was  
579 supported by IRP projects of the University of Ostrava and by the Czech Ministry of Education, Youth  
580 and Sports (project OPVVV 16\_019/0000759). C.C.W. was funded by Nanqiang Outstanding Young  
581 Talents Program of Xiamen University and the Fundamental Research Funds for the Central Universities.  
582 M.Z. has been funded by research grants from the MES RK No. AP05134955 and No. 0114RK00492.

583

## 584 **Web Resources**

585 IMPUTE version 2 (IMPUTE2), [https://mathgen.stats.ox.ac.uk/impute/1000GP\\_Phase3.html](https://mathgen.stats.ox.ac.uk/impute/1000GP_Phase3.html)

586 International Society of Genetic Genealogy (ISOGG), <http://www.isogg.org>

587 pileupCaller, <https://github.com/stschiff/sequenceTools>

588 Sequence Read Archive (SRA), <https://www.ncbi.nlm.nih.gov/sra>

589 YFULL<sup>TM</sup>.com, <http://www.yfull.com>

590

## 591 **Accession Numbers**

592 Genome-wide sequence data of two Botai individuals (BAM format) are available at the Sequence Read  
593 Archive under the accession number PRJNA470593. Array genotype data will be made available through  
594 the Reich Lab and MPI-SHH webpages upon the publication of the manuscript.

595

596

## 597 **References**

598

- 599 1. Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., Cann, H.M.,  
600 Barsh, G.S., Feldman, M., and Cavalli-Sforza, L.L. (2008). Worldwide human relationships  
601 inferred from genome-wide patterns of variation. *Science* 319, 1100-1104.
- 602 2. Wang, C., Zöllner, S., and Rosenberg, N.A. (2012). A quantitative comparison of the similarity  
603 between genes and geography in worldwide human populations. *PLoS Genet.* 8, e1002886.
- 604 3. Jeong, C., Ozga, A.T., Witonsky, D.B., Malmström, H., Edlund, H., Hofman, C.A., Hagan, R.,  
605 Jakobsson, M., Lewis, C.M., Aldenderfer, M., et al. (2016). Long-term genetic stability and a  
606 high altitude East Asian origin for the peoples of the high valleys of the Himalayan arc. *Proc.*  
607 *Natl. Acad. Sci. USA* 113, 7485-7490.
- 608 4. Yunusbayev, B., Metspalu, M., Järve, M., Kutuev, I., Rootsi, S., Metspalu, E., Behar, D.M., Varendi,  
609 K., Sahakyan, H., Khusainova, R., et al. (2012). The Caucasus as an asymmetric semipermeable  
610 barrier to ancient human migrations. *Mol. Biol. Evol.* 29, 359-365.
- 611 5. Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S.,  
612 Harney, E., Stewardson, K., et al. (2015). Massive migration from the steppe was a source for  
613 Indo-European languages in Europe. *Nature* 522, 207-211.
- 614 6. Haber, M., Gauguier, D., Youhanna, S., Patterson, N., Moorjani, P., Botigué, L.R., Platt, D.E.,  
615 Matisoo-Smith, E., Soria-Hernanz, D.F., Wells, R.S., et al. (2013). Genome-wide diversity in the  
616 Levant reveals recent structuring by culture. *PLoS Genet.* 9, e1003316.
- 617 7. Martiniano, R., Cassidy, L.M., Ó'Maoldúin, R., McLaughlin, R., Silva, N.M., Manco, L., Fidalgo, D.,  
618 Pereira, T., Coelho, M.J., Serra, M., et al. (2017). The population genomics of archaeological  
619 transition in west Iberia: Investigation of ancient substructure using imputation and haplotype-  
620 based methods. *PLoS Genet.* 13, e1006852.
- 621 8. Allentoft, M.E., Sikora, M., Sjogren, K.-G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard,  
622 P.B., Schroeder, H., Ahlstrom, T., Vinner, L., et al. (2015). Population genomics of Bronze Age  
623 Eurasia. *Nature* 522, 167-172.
- 624 9. Barfield, T.J. (1993). *The nomadic alternative*. (Englewood Cliffs, NJ: Prentice Hall).
- 625 10. Frachetti, M.D. (2009). *Pastoralist landscapes and social interaction in Bronze Age Eurasia*. (Berkeley,  
626 CA: Univ of California Press).
- 627 11. Burch, E.S. (1972). The caribou/wild reindeer as a human resource. *Am. Antiquity* 37, 339-368.
- 628 12. Sherratt, A. (1983). The secondary exploitation of animals in the Old World. *World Archaeol.* 15, 90-  
629 104.
- 630 13. Yunusbayev, B., Metspalu, M., Metspalu, E., Valeev, A., Litvinov, S., Valiev, R., Akhmetova, V.,  
631 Balanovska, E., Balanovsky, O., Turdikulova, S., et al. (2015). The genetic legacy of the  
632 expansion of Turkic-speaking nomads across Eurasia. *PLoS Genet.* 11, e1005068.
- 633 14. Hellenthal, G., Busby, G.B., Band, G., Wilson, J.F., Capelli, C., Falush, D., and Myers, S. (2014). A  
634 genetic atlas of human admixture history. *Science* 343, 747-751.

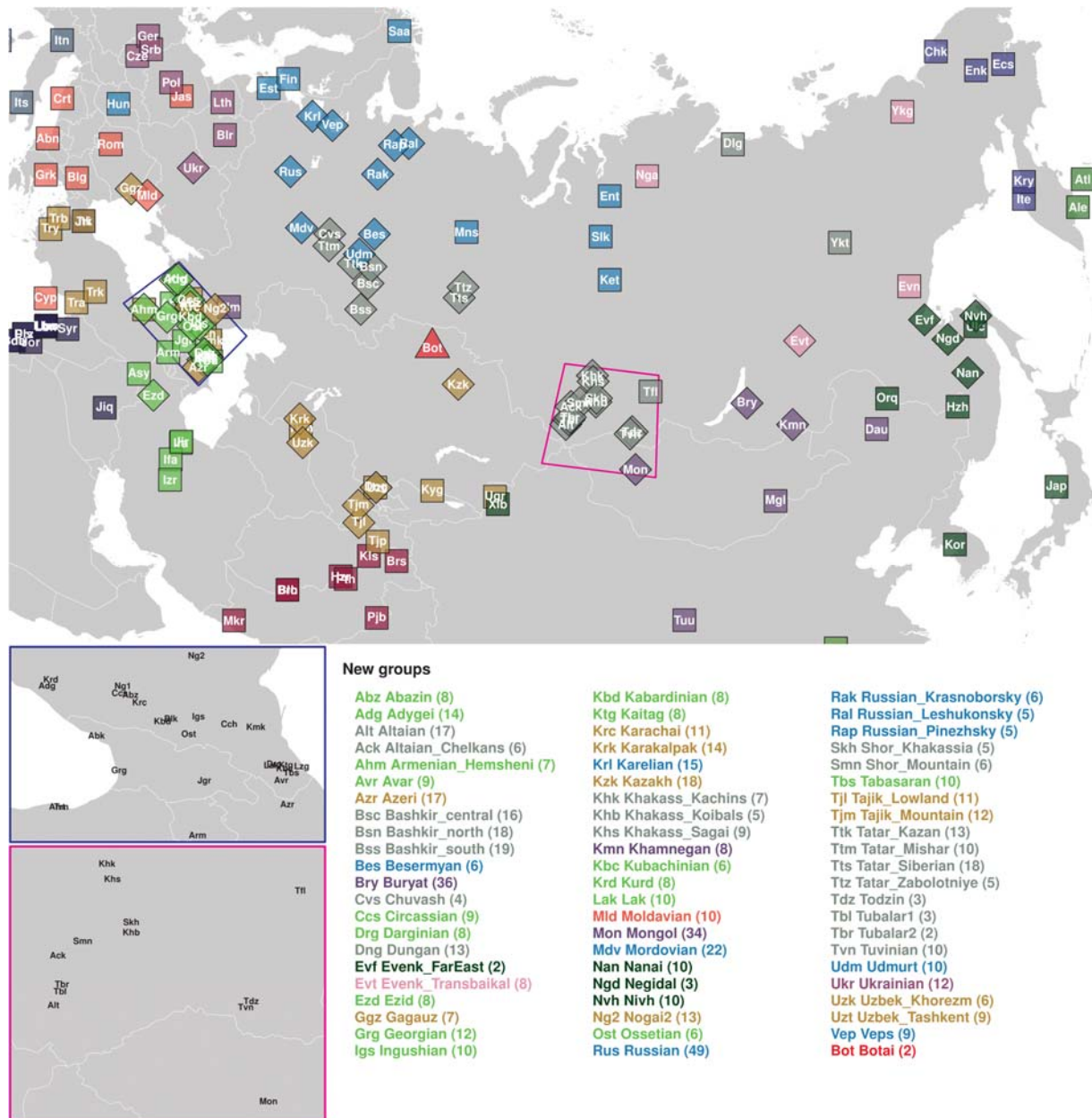
- 635 15. Flegontov, P., Changmai, P., Zidkova, A., Logacheva, M.D., Altınışık, N.E., Flegontova, O., Gelfand,  
636 M.S., Gerasimov, E.S., Khrameeva, E.E., Konovalova, O.P., et al. (2016). Genomic study of the  
637 Ket: a Paleo-Eskimo-related ethnic group with significant ancient North Eurasian ancestry. *Sci.*  
638 *Rep.* 6, 20768.
- 639 16. Pugach, I., Matveev, R., Spitsyn, V., Makarov, S., Novgorodov, I., Osakovsky, V., Stoneking, M., and  
640 Pakendorf, B. (2016). The complex admixture history and recent southern origins of Siberian  
641 populations. *Mol. Biol. Evol.* 33, 1777-1795.
- 642 17. Triska, P., Chekanov, N., Stepanov, V., Khusnutdinova, E.K., Kumar, G.P.A., Akhmetova, V.,  
643 Babalyan, K., Boulygina, E., Kharkov, V., Gubina, M., et al. (2017). Between Lake Baikal and  
644 the Baltic Sea: genomic history of the gateway to Europe. *BMC Genet.* 18, 110.
- 645 18. Raghavan, M., Skoglund, P., Graf, K.E., Metspalu, M., Albrechtsen, A., Moltke, I., Rasmussen, S.,  
646 Stafford Jr, T.W., Orlando, L., Metspalu, E., et al. (2014). Upper Palaeolithic Siberian genome  
647 reveals dual ancestry of Native Americans. *Nature* 505, 87-91.
- 648 19. Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D.C., Rohland, N., Mallick, S., Fernandes, D., Novak,  
649 M., Gamarra, B., Sirak, K., et al. (2016). Genomic insights into the origin of farming in the  
650 ancient Near East. *Nature* 536, 419-424.
- 651 20. Levine, M., and Kislenco, A. (1997). New Eneolithic and early Bronze Age radiocarbon dates for  
652 north Kazakhstan and south Siberia. *Camb. Archaeol. J.* 7, 297-300.
- 653 21. Outram, A.K., Stear, N.A., Bendrey, R., Olsen, S., Kasparov, A., Zaibert, V., Thorpe, N., and  
654 Evershed, R.P. (2009). The earliest horse harnessing and milking. *Science* 323, 1332-1335.
- 655 22. Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T.,  
656 and Reich, D. (2012). Ancient admixture in human history. *Genetics* 192, 1065-1093.
- 657 23. Flegontov, P., Altinisik, N.E., Changmai, P., Rohland, N., Mallick, S., Bolnick, D.A., Candilio, F.,  
658 Flegontova, O., Jeong, C., Harper, T.K., et al. (2017). Paleo-Eskimo genetic legacy across North  
659 America. *bioRxiv*, 203018.
- 660 24. Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N.,  
661 Nordenfelt, S., Tandon, A., et al. (2016). The Simons Genome Diversity Project: 300 genomes  
662 from 142 diverse populations. *Nature* 538, 201-206.
- 663 25. Fu, Q., Li, H., Moorjani, P., Jay, F., Slepchenko, S.M., Bondarev, A.A., Johnson, P.L.F., Aximu-Petri,  
664 A., Prüfer, K., de Filippo, C., et al. (2014). Genome sequence of a 45,000-year-old modern  
665 human from western Siberia. *Nature* 514, 445-449.
- 666 26. Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., Furtwängler, A., Haak, W.,  
667 Meyer, M., Mittnik, A., et al. (2016). The genetic history of Ice Age Europe. *Nature* 534, 200-205.
- 668 27. Haber, M., Doumet-Serhal, C., Scheib, C., Xue, Y., Danecek, P., Mezzavilla, M., Youhanna, S.,  
669 Martiniano, R., Prado-Martinez, J., Szpak, M., et al. (2017). Continuity and admixture in the last  
670 five millennia of Levantine history from ancient Canaanite and present-day Lebanese genome  
671 sequences. *Am. J. Hum. Genet.* 101, 274-282.
- 672 28. Jones, E.R., Gonzalez-Fortes, G., Connell, S., Siska, V., Eriksson, A., Martiniano, R., McLaughlin,  
673 R.L., Gallego Llorente, M., Cassidy, L.M., Gamba, C., et al. (2015). Upper Palaeolithic genomes  
674 reveal deep roots of modern Eurasians. *Nat. Commun.* 6, 8912.
- 675 29. Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P.H.,  
676 Schraiber, J.G., Castellano, S., Lipson, M., et al. (2014). Ancient human genomes suggest three  
677 ancestral populations for present-day Europeans. *Nature* 513, 409-413.
- 678 30. Lazaridis, I., Mittnik, A., Patterson, N., Mallick, S., Rohland, N., Pfrengle, S., Furtwängler, A.,  
679 Peltzer, A., Posth, C., Vasilakis, A., et al. (2017). Genetic origins of the Minoans and  
680 Mycenaean. *Nature* 548, 214-218.
- 681 31. Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S.A., Harney, E.,  
682 Stewardson, K., Fernandes, D., Novak, M., et al. (2015). Genome-wide patterns of selection in  
683 230 ancient Eurasians. *Nature* 528, 499-503.



- 684 32. Raghavan, M., DeGiorgio, M., Albrechtsen, A., Moltke, I., Skoglund, P., Korneliussen, T.S.,  
685 Grønnow, B., Appelt, M., Gulløv, H.C., Friesen, T.M., et al. (2014). The genetic prehistory of the  
686 New World Arctic. *Science* 345, 1255832.
- 687 33. Rasmussen, M., Anzick, S.L., Waters, M.R., Skoglund, P., DeGiorgio, M., Stafford Jr, T.W.,  
688 Rasmussen, S., Moltke, I., Albrechtsen, A., Doyle, S.M., et al. (2014). The genome of a Late  
689 Pleistocene human from a Clovis burial site in western Montana. *Nature* 506, 225-229.
- 690 34. Rasmussen, M., Li, Y., Lindgreen, S., Pedersen, J.S., Albrechtsen, A., Moltke, I., Metspalu, M.,  
691 Metspalu, E., Kivisild, T., Gupta, R., et al. (2010). Ancient human genome sequence of an extinct  
692 Palaeo-Eskimo. *Nature* 463, 757-762.
- 693 35. Rasmussen, M., Sikora, M., Albrechtsen, A., Korneliussen, T.S., Moreno-Mayar, J.V., Poznik, G.D.,  
694 Zollikofer, C.P.E., Ponce de León, M.S., Allentoft, M.E., Moltke, I., et al. (2015). The ancestry  
695 and affiliations of Kennewick Man. *Nature* 523, 455-458.
- 696 36. Saag, L., Varul, L., Scheib, C.L., Stenderup, J., Allentoft, M.E., Saag, L., Pagani, L., Reidla, M.,  
697 Tambets, K., and Metspalu, E. (2017). Extensive farming in Estonia started through a sex-biased  
698 migration from the Steppe. *Curr. Biol.* 27, 2185-2193. e2186.
- 699 37. Siska, V., Jones, E.R., Jeon, S., Bhak, Y., Kim, H.-M., Cho, Y.S., Kim, H., Lee, K., Veselovskaya, E.,  
700 Balueva, T., et al. (2017). Genome-wide data from two early Neolithic East Asian individuals  
701 dating to 7700 years ago. *Sci. Adv.* 3, e1601877.
- 702 38. Unterländer, M., Palstra, F., Lazaridis, I., Pilipenko, A., Hofmanová, Z., Groß, M., Sell, C., Blöcher,  
703 J., Kirsanow, K., Rohland, N., et al. (2017). Ancestry and demography and descendants of Iron  
704 Age nomads of the Eurasian Steppe. *Nat. Commun.* 8, 14615.
- 705 39. Yang, M.A., Gao, X., Theunert, C., Tong, H., Aximu-Petri, A., Nickel, B., Slatkin, M., Meyer, M.,  
706 Pääbo, S., Kelso, J., et al. (2017). 40,000-Year-Old Individual from Asia Provides Insight into  
707 Early Population Structure in Eurasia. *Curr. Biol.* 27, 3202-3208.e3209.
- 708 40. Kılınç, Gülşah M., Omrak, A., Özer, F., Günther, T., Büyükkarakaya, Ali M., Bıçakçı, E., Baird, D.,  
709 Dönertaş, Handan M., Ghalichi, A., Yaka, R., et al. (2016). The demographic development of the  
710 first farmers in Anatolia. *Curr. Biol.* 26, 2659-2666.
- 711 41. Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C.,  
712 García, N., Pääbo, S., Arsuaga, J.-L., et al. (2013). Complete mitochondrial genome sequence of a  
713 Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci.*  
714 *USA* 110, 15758-15763.
- 715 42. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., and Reich, D. (2015). Partial uracil–DNA–  
716 glycosylase treatment for screening of ancient DNA. *Phil. Trans. R. Soc. B* 370, 20130624.
- 717 43. Kircher, M. (2012). Analysis of high-throughput ancient DNA sequencing data. In *Ancient DNA:*  
718 *methods and protocols*, B. Shapiro and M. Hofreiter, eds. (New York, NY, USA, Humana Press),  
719 pp 197-228.
- 720 44. Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J., and Nieselt, K. (2016). EAGER:  
721 efficient ancient genome reconstruction. *Genome Biol.* 17, 60.
- 722 45. Schubert, M., Lindgreen, S., and Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming,  
723 identification, and read merging. *BMC Res. Notes* 9, 88.
- 724 46. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler  
725 transform. *Bioinformatics* 25, 1754-1760.
- 726 47. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and  
727 Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078-  
728 2079.
- 729 48. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L., and Orlando, L. (2013). mapDamage2.0: fast  
730 approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29, 1682-  
731 1684.
- 732 49. Renaud, G., Slon, V., Duggan, A.T., and Kelso, J. (2015). Schmutzi: estimation of contamination and  
733 endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* 16, 224.

- 734 50. Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: analysis of next generation  
735 sequencing data. *BMC Bioinformatics* *15*, 356.
- 736 51. Jun, G., Wing, M.K., Abecasis, G.R., and Kang, H.M. (2015). An efficient and scalable analysis  
737 framework for variant extraction and refinement from population-scale DNA sequence data.  
738 *Genome Res.* *25*, 918-925.
- 739 52. Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.-J.,  
740 Kronenberg, F., Salas, A., and Schönherr, S. (2016). HaploGrep 2: mitochondrial haplogroup  
741 classification in the era of high-throughput sequencing. *Nucleic Acids Res.* *44*, W58-W63.
- 742 53. Poznik, G.D. (2016). Identifying Y-chromosome haplogroups in arbitrarily large samples of  
743 sequenced or genotyped men. *bioRxiv*, 088716.
- 744 54. Balanovsky, O., Dibirova, K., Dybo, A., Mudrak, O., Frolova, S., Pocheshkhova, E., Haber, M., Platt,  
745 D., Schurr, T., Haak, W., et al. (2011). Parallel Evolution of Genes and Languages in the  
746 Caucasus Region. *Mol. Biol. Evol.* *28*, 2905-2920.
- 747 55. Koshel, S. (2012). Geoinformation technologies in genegeography. In *Sovremennaya*  
748 *geograficheskaya kartografiya (Modern Geographic Cartography)*, I. Lourie and V. Kravtsova,  
749 eds. (Moscow, Data+), pp 158-166.
- 750 56. Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genet.*  
751 *2*, e190.
- 752 57. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in  
753 unrelated individuals. *Genome Res.* *19*, 1655-1664.
- 754 58. Chang, C.C., Chow, C.C., Tellier, L., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-  
755 generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* *4*, 7.
- 756 59. Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N., Parra, M.V., Rojas, W.,  
757 Duque, C., Mesa, N., et al. (2012). Reconstructing native American population history. *Nature*  
758 *488*, 370-374.
- 759 60. Delaneau, O., Zagury, J.-F., and Marchini, J. (2013). Improved whole-chromosome phasing for  
760 disease and population genetic studies. *Nat. Methods* *10*, 5-6.
- 761 61. Lawson, D.J., Hellenthal, G., Myers, S., and Falush, D. (2012). Inference of population structure  
762 using dense haplotype data. *PLoS Genet.* *8*, e1002453.
- 763 62. Loh, P.-R., Lipson, M., Patterson, N., Moorjani, P., Pickrell, J.K., Reich, D., and Berger, B. (2013).  
764 Inferring admixture histories of human populations using linkage disequilibrium. *Genetics* *193*,  
765 1233-1254.
- 766 63. Petkova, D., Novembre, J., and Stephens, M. (2016). Visualizing spatial population structure with  
767 estimated effective migration surfaces. *Nat. Genet.* *48*, 94-100.
- 768 64. Fenner, J.N. (2005). Cross-cultural estimation of the human generation interval for use in  
769 genetics-based population divergence studies. *Am. J. Phys. Anthropol.* *128*, 415-423.
- 770 65. Myres, N.M., Rootsi, S., Lin, A.A., Järve, M., King, R.J., Kutuev, I., Cabrera, V.M., Khusnutdinova,  
771 E.K., Pshenichnov, A., Yunusbayev, B., et al. (2010). A major Y-chromosome haplogroup R1b  
772 Holocene era founder effect in Central and Western Europe. *Eur. J. Hum. Genet.* *19*, 95.
- 773 66. Balanovsky, O., Chukhryaeva, M., Zaporozhchenko, V., Urasin, V., Zhabagin, M., Hovhannisyan, A.,  
774 Agdzhoyan, A., Dibirova, K., Kuznetsova, M., Koshel, S., et al. (2017). Genetic differentiation  
775 between upland and lowland populations shapes the Y-chromosomal landscape of West Asia.  
776 *Hum. Genet.* *136*, 437-450.
- 777 67. Jones, E.R., Zarina, G., Moiseyev, V., Lightfoot, E., Nigst, P.R., Manica, A., Pinhasi, R., and Bradley,  
778 D.G. (2017). The Neolithic Transition in the Baltic Was Not Driven by Admixture with Early  
779 European Farmers. *Curr. Biol.* *27*, 576-582.
- 780 68. Narasimhan, V.M., Patterson, N.J., Moorjani, P., Lazaridis, I., Mark, L., Mallick, S., Rohland, N.,  
781 Bernardos, R., Kim, A.M., Nakatsuka, N., et al. (2018). The genomic formation of South and  
782 Central Asia. *bioRxiv*, 292581.
- 783 69. Sedghifar, A., Brandvain, Y., Ralph, P., and Coop, G. (2015). The spatial mixing of genomes in  
784 secondary contact zones. *Genetics* *201*, 243-261.

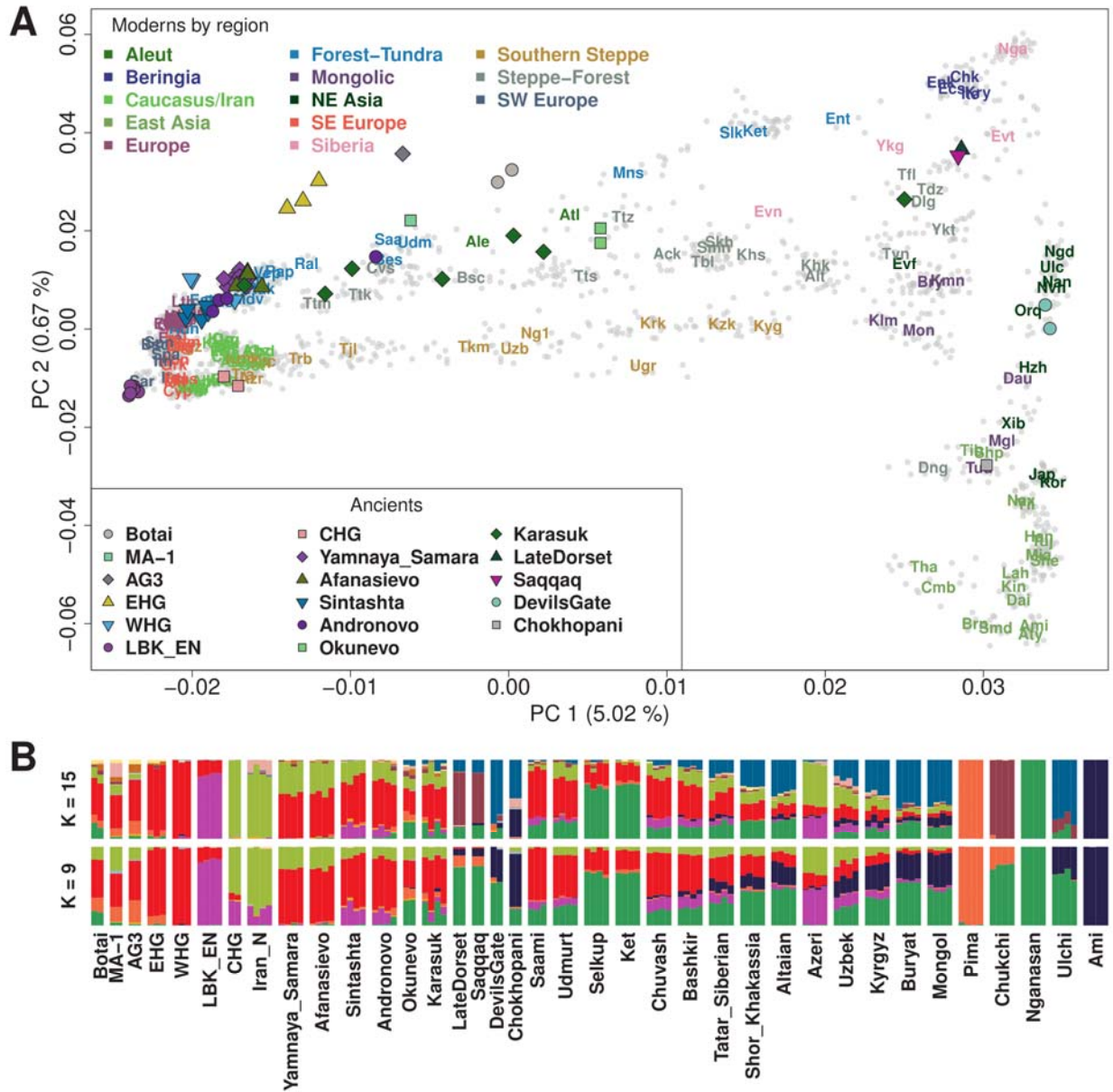
- 785 70. Levine, M. (1999). Botai and the origins of horse domestication. *J. Anthropol. Archaeol.* 18, 29-78.  
786 71. Bronk Ramsey, C. (2009). Bayesian analysis of radiocarbon dates. *Radiocarbon* 51, 337-360.  
787 72. Reimer, P.J., Bard, E., Bayliss, A., Beck, J.W., Blackwell, P.G., Ramsey, C.B., Buck, C.E., Cheng, H.,  
788 Edwards, R.L., Friedrich, M., et al. (2016). IntCal13 and Marine13 radiocarbon age calibration  
789 curves 0–50,000 years cal BP. *Radiocarbon* 55, 1869-1887.  
790  
791  
792  
793



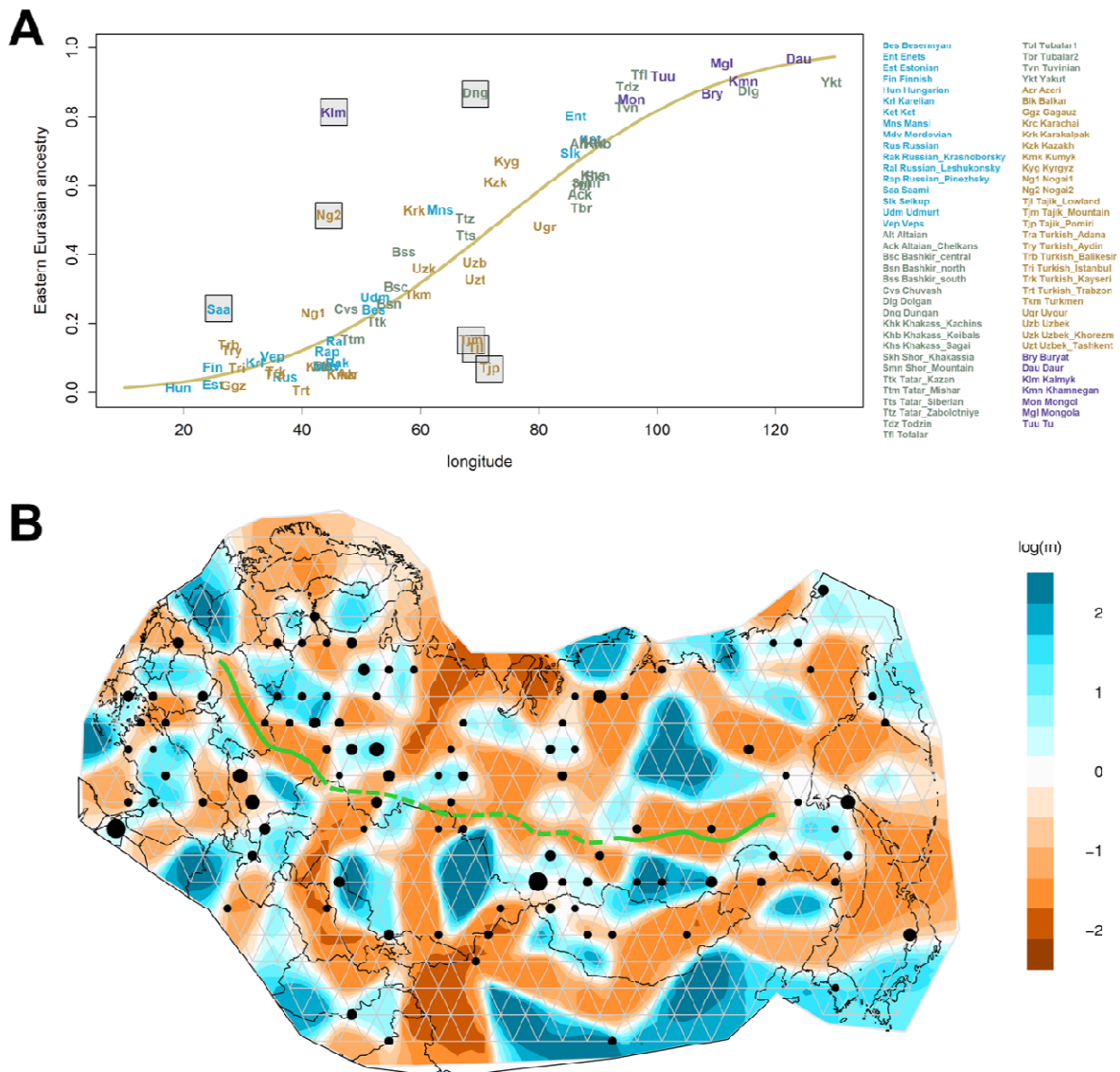
794  
795

**Figure 1. Geographic locations of the Eneolithic Botai site (red triangle), 65 groups including newly sampled individuals (filled diamonds) and nearby groups with published data (filled squares).** Mean latitude and longitude values across all individuals under each group label were used. Two zoom-in plots for the Caucasus (blue) and the Altai-Sayan (magenta) regions are presented in the lower left corner. A list of new groups, their three-letter codes, and the number of new individuals (in parenthesis) are provided at the bottom. Corresponding information for the previously published groups is provided in [Table S2](#). The main inner Eurasian map is on the Albers equal area projection and was produced using the spTransform function in the R package rgdal v1.2-5.

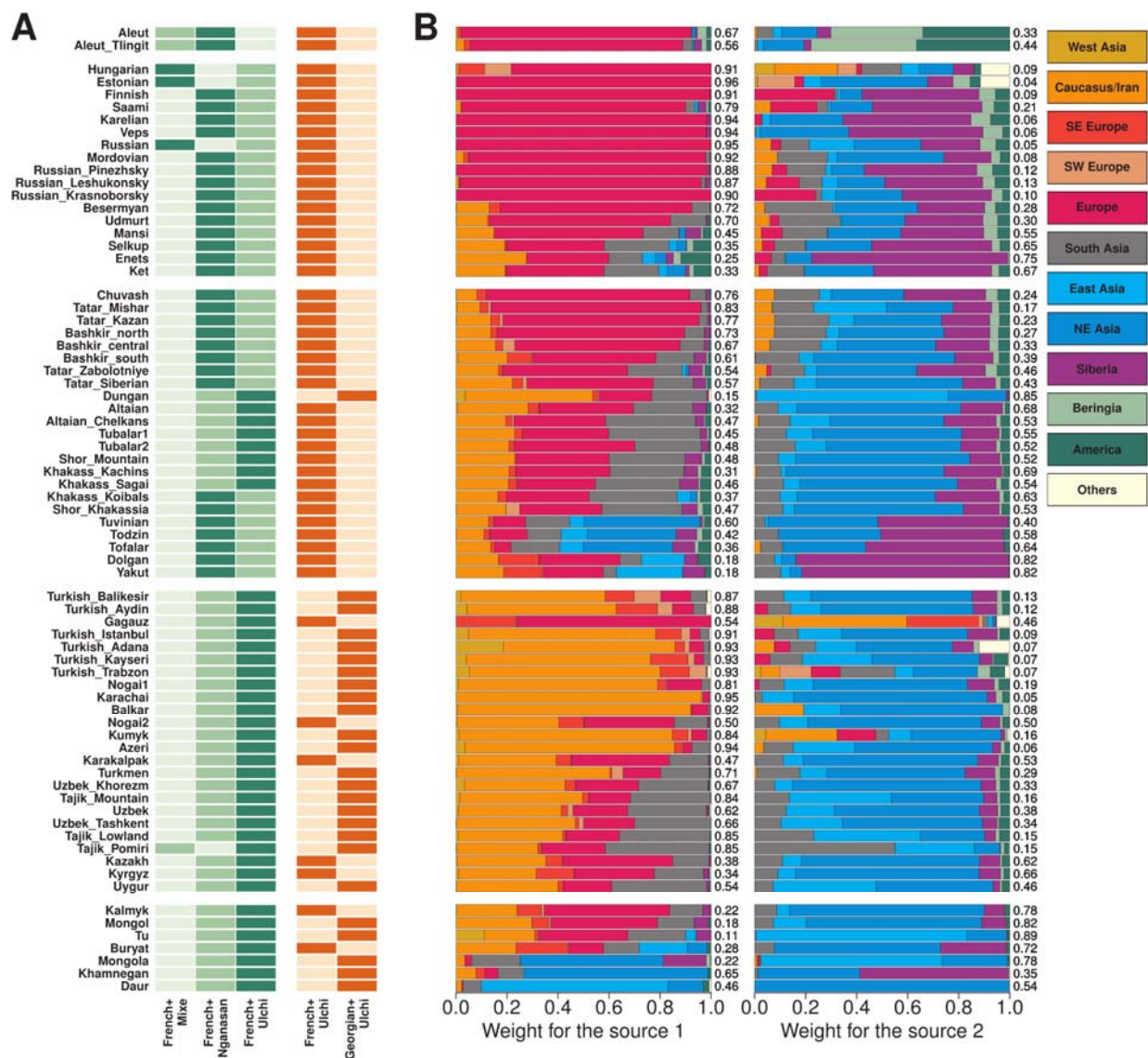
804  
805



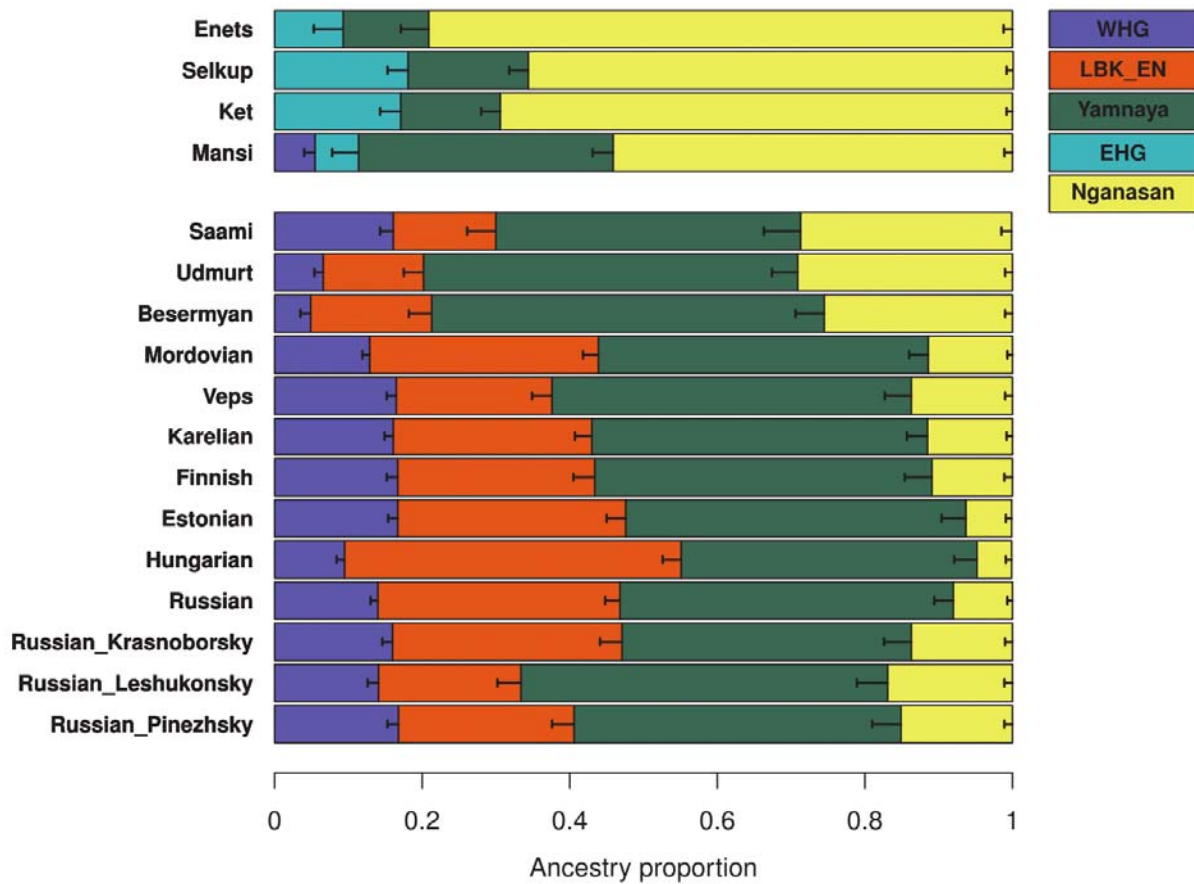
806  
 807  
 808 **Figure 2. The genetic structure of inner Eurasian populations.** (A) The first two PCs of 2,077  
 809 Eurasian individuals separate western and eastern Eurasians (PC1) and Northeast and Southeast Asians  
 810 (PC2). Most inner Eurasians are located between western and eastern Eurasians on PC1. Ancient  
 811 individuals (color-filled shapes) are projected onto PCs calculated based on contemporary individuals.  
 812 Modern individuals are marked by grey dots, with their per-group mean coordinates marked by three-  
 813 letter codes listed in [Table S2](#). (B) ADMIXTURE results for a chosen set of ancient and modern groups  
 814 (K = 9 and 15). Most inner Eurasians are modeled as a mixture of components primarily found in eastern  
 815 or western Eurasians. Results for the full set of individuals are provided in [Figure S3](#).  
 816



817  
818  
819 **Figure 3. Inner Eurasian admixture in geographical context.** (A) A comparison of mean longitudinal  
820 coordinates (x-axis) and mean eastern Eurasian ancestry proportions (y-axis) of inner Eurasians. Eastern  
821 Eurasian ancestry proportions are estimated from ADMIXTURE results with K=15 by summing up six  
822 components maximized in Karitiana, Pima, Chukchi, Nganasan, Ulchi and Ami, respectively (Figure S3).  
823 The yellow curve shows a probit regression fit following the model in Sedghifar et al.<sup>69</sup> Seven groups  
824 substantially deviating from the curve, including known historical migrants, are marked with grey  
825 background. (B) Barriers (brown) and conduits (blue) of gene flow across inner Eurasia estimated by the  
826 EEMS program. Black dots show the location of vertices to which individuals are assigned, with sizes  
827 correlated with the number of individuals. Solid green curves highlight strong barriers to gene flow  
828 separating the steppe-forest cline and the southern steppe cline populations (the western curve) or the  
829 steppe-forest cline and the forest-tundra cline populations (the eastern curve). The dotted green curve  
830 marks a region between the two curves where this barrier seems to be weaker than in the flanking regions.  
831  
832



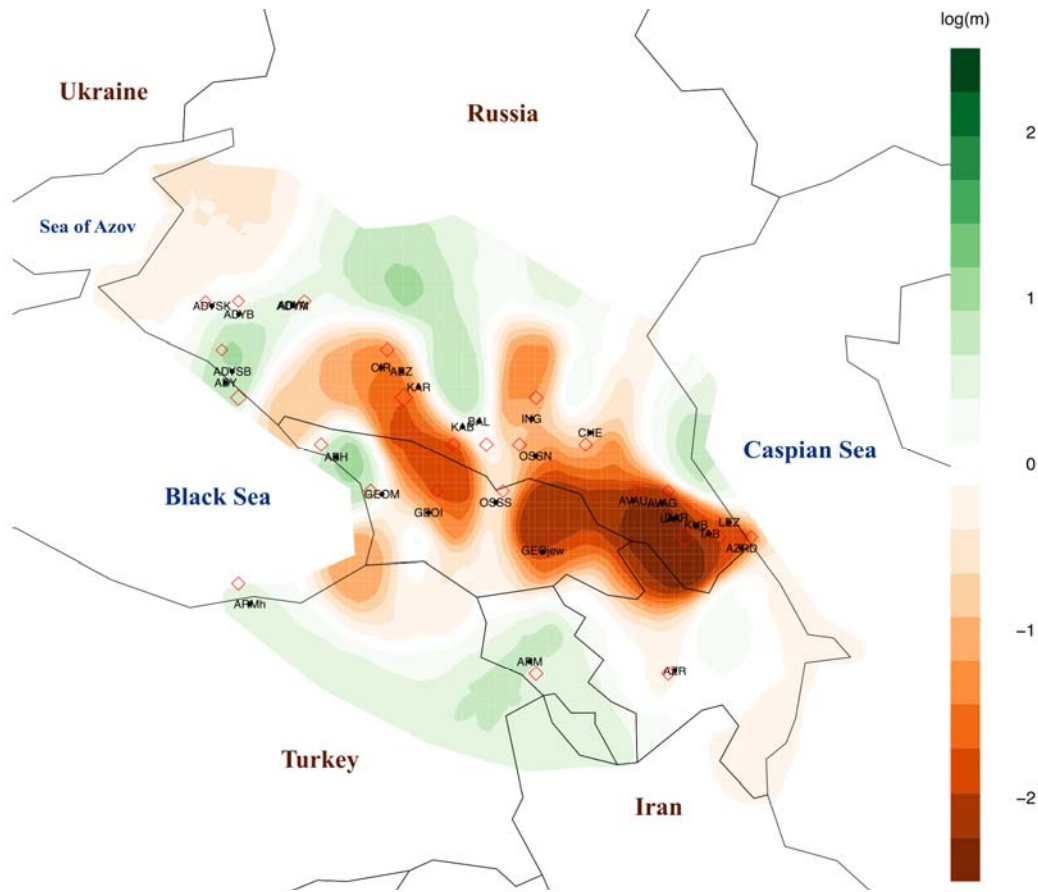
833  
834  
835 **Figure 4. Characterization of the western and eastern Eurasian source ancestries in inner Eurasian**  
836 **populations.** (A) Admixture  $f_3$  values are compared for different eastern Eurasian references (Mixe,  
837 Nganasan, Ulchi; left) or western Eurasian ones (French, Georgian; right). For each target group, darker  
838 shades mark more negative  $f_3$  values. (B) Weights of donor populations in two sources characterizing the  
839 main admixture signal (“date 1 PC 1”) in the GLOBETROTTER analysis. We merged 167 donor  
840 populations into 12 groups, as listed on the top right side. Target populations are split into five groups:  
841 Aleuts, the forest-tundra cline populations, the steppe-forest cline populations, the southern steppe cline  
842 populations and the Mongolic-speaking populations, from the top to bottom.  
843  
844



845  
846  
847  
848  
849  
850  
851  
852  
853

**Figure 5. qpAdm-based admixture models for the forest-tundra cline populations.** For populations to the east of the Urals (Enets, Selkups, Kets, and Mansi), EHG+Yamnaya+Nganasan provides a good fit, except for Mansi, for which adding WHG significantly increases the model fit. For the rest of the groups, WHG+LBK\_EN+Yamnaya+Nganasan in general provides a good fit. 5 cM jackknifing standard errors are marked by the horizontal bar. Details of the model information are presented in [Table S8](#).





854  
855  
856  
857  
858  
859  
860  
861  
862  
863

**Figure 6. The Greater Caucasus mountain ridge as a barrier to genetic exchange.** Barriers (brown) and conduits (green) of gene flow around the Caucasus region are estimated by the EEMS program. Red diamonds show the location of vertices to which groups are assigned. A strong barrier to gene flow overlaps with the Greater Caucasus mountain ridge reflecting the genetic differentiation between populations of the north and south of the Caucasus. The barrier becomes considerably weaker in the middle where present-day Ossetians live.

864 **Table 1. Sequencing statistics and radiocarbon dates of two Eneolithic Botai individuals analyzed in**  
 865 **this study.**  
 866

ID	Genetic Sex	Uncal. <sup>14</sup> C Date	Cal. <sup>14</sup> C Date (2-sigma) <sup>b</sup>	# of reads sequenced	# of SNPs covered <sup>c</sup>	MT / Y haplogroup	MT.cont <sup>d</sup>	X.cont <sup>e</sup>
TU45	M	4620 ± 80 <sup>a</sup>	3632-3100 cal. BCE	84,170,835	77,363	K1b2 / R1b1a1	0.02 (0.01-0.03)	0.0122 (0.0050)
BKZ001	F	4660 ± 25	3517-3367 cal. BCE	69,678,735	432,078	Z1 / NA	0.01 (0.00-0.02)	NA

867 <sup>a</sup> The uncalibrated date of TU45 was published in Levine (1999) under the ID OxA-4316.<sup>70</sup>

868 <sup>b</sup> The calibrated <sup>14</sup>C dates are calculated based on uncalibrated dates, by the OxCal v4.3.2 program<sup>71</sup> using the  
 869 INTCAL13 atmospheric curve.<sup>72</sup>

870 <sup>c</sup> The number of autosomal SNPs in the HumanOrigins array (out of 581,230) covered at least by one read. Only  
 871 transversion SNPs are considered for the non-UDG libraries (both of the TU45 libraries, one of two BKZ001  
 872 libraries).

873 <sup>d</sup> The contamination rate of mitochondrial reads estimated by the Schmutzi program (95% confidence interval in  
 874 parentheses)

875 <sup>e</sup> The nuclear contamination rate for the male (TU45) estimated based on X chromosome data by ANGSD software  
 876 (standard error in parentheses)

877

878

879