

This is the author's accepted manuscript of the following article. The final publication is available in *Heredity*: <https://doi.org/10.1038/s41437-017-0024-3>.

The full details of the published version of the article are as follows:

TITLE: Genome-wide genotyping uncovers genetic profiles and history of the Russian cattle breeds

AUTHORS: Yurchenko, A; Yudin, N; Aitnazarov, R; Plyusnina, A; Brukhin, V; Soloshenko, V; Lhasaranov, B; Popov, R; Paronyan, I A; Plemyashov, K V; **Larkin, D M**

JOURNAL TITLE: *Heredity*

PUBLICATION DATE: 8 December 2017 (online)

PUBLISHER: Nature Publishing Group

DOI: 10.1038/s41437-017-0024-3

1 **Genome-wide genotyping uncovers genetic profiles and history of the**  
2 **Russian cattle breeds**

3

4 Andrey Yurchenko<sup>1,2</sup>, Nikolay Yudin<sup>1,3</sup>, Ruslan Aitnazarov<sup>1,3</sup>, Alexandra Plyusnina<sup>1</sup>,  
5 Vladimir Brukhin<sup>4</sup>, Vladimir Soloshenko<sup>5</sup>, Bulat Lhasaranov<sup>6</sup>, Ruslan Popov<sup>7</sup>, Ivan A.  
6 Paronyan<sup>8</sup>, Kirill V. Plemyashov<sup>8</sup>, Denis M. Larkin<sup>1,9\*</sup>

7

8 <sup>1</sup>*The Federal Research Center Institute of Cytology and Genetics, The Siberian Branch of the*  
9 *Russian Academy of Sciences (ICG SB RAS), 630090 Novosibirsk, Russia;*

10 <sup>2</sup>*Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow,*  
11 *Glasgow, G12 8QQ, UK;*

12 <sup>3</sup>*Novosibirsk State University, 630090 Novosibirsk, Russia;*

13 <sup>4</sup>*Theodosius Dobzhansky Center for Genome Bioinformatics, St. Petersburg State University,*  
14 *St. Petersburg, 199034, Russia;*

15 <sup>5</sup>*Siberian Research Institute of Animal Husbandry, 630501 Krasnoobsk, Russia;*

16 <sup>6</sup>*Shuluuma IAPC, Kizhinga, 671450 Buryatia, Russia;*

17 <sup>7</sup>*Yakutian Research Institute of Agriculture, 677001 Yakutsk, Russia;*

18 <sup>8</sup>*Russian Research Institute of Farm Animal Genetics and Breeding, 196601 St. Petersburg,*  
19 *Russia;*

20 <sup>9</sup>*Royal Veterinary College, University of London, NW01 0TU London, UK*

21

22 \*Corresponding author: Denis M. Larkin, Royal Veterinary College, University of London,  
23 Royal College Street, NW01 0TU London, UK, Phone: +44(0)7577057645, Fax: +44 (0)20  
24 73881027, email: dmlarkin@gmail.com

25

26 Running title: Genetics of the Russian cattle

27

28 Word count: 6,616

29 **ABSTRACT**

30

31 One of the most economically important areas within the Russian agricultural sector is dairy  
32 and beef cattle farming contributing about \$11 billion to the Russian economy annually. Trade  
33 connections, selection and breeding have resulted in the establishment of a number of breeds  
34 that are presumably adapted to local climatic conditions. Little however is known about the  
35 ancestry and history of Russian native cattle. To address this question, we genotyped 274  
36 individuals from 18 breeds bred in Russia and compared them to 135 additional breeds from  
37 around the world that had been genotyped previously. Our results suggest a shared ancestry  
38 between most of the Russian cattle and European taurine breeds, apart from a few breeds that  
39 shared ancestry with the Asian taurines. The Yakut cattle, belonging to the latter group, was  
40 found to be the most diverged breed in the whole combined dataset according to structure  
41 results. Haplotype sharing further suggests that the Russian cattle can be divided into four  
42 major clusters reflecting ancestral relations with other breeds. Herein, we therefore shed light  
43 on to the history of Russian cattle and identified closely related breeds to those from Russia.  
44 Our results will facilitate future research on detecting signatures of selection in cattle genomes  
45 and eventually inform future genetics-assisted livestock breeding programs in Russia and in  
46 other countries.

47

48 **Keywords:** Russian native cattle breeds, adaptation, breed formation, SNP genotyping,  
49 admixture, European cattle, Asian cattle

## INTRODUCTION

50

51

52 Thousands of years of artificial selection coupled with human-driven migration and adaptation  
53 to diverse environmental conditions resulted in ~1000 cattle breeds worldwide, which are  
54 tailored to local economic needs, aesthetic demands and possess unique genetic profiles  
55 (Mason, 1969). During the last two centuries, some cattle populations were further improved  
56 resulting in several commercial breeds demonstrating outstanding productivity when properly  
57 handled (Boichard and Brochard, 2012). Currently there is a tendency to replace or ‘improve’  
58 local breeds with the genetic material from superior commercial ones, meaning that genetic  
59 diversity, signatures of adaptations to local conditions, and the history of formation encoded in  
60 native breed genomes often diminish before being recorded and properly studied (Gaouar *et*  
61 *al*, 2015). On the other hand, genomes of native breeds could be mined for combinations of the  
62 genetic variants invaluable in the development of a new generation of commercial breeds that  
63 would better fit into a range of environmental conditions (Gao *et al*, 2017). The first step  
64 towards uncovering this information is to understand the origin, structure and admixture events  
65 involving the native breed populations and to place them into the context of a wider set of  
66 world breeds (Beynon *et al*, 2015; Bovine HapMap Consortium, 2009; Matukumalli *et al*,  
67 2009).

68 The genetic diversity of domestic cattle stems from the two main sources of  
69 domestication of the ancient *Bos* subspecies: *B. taurus* and *B. indicus* originating from the  
70 Fertile Crescent and the Indus Valley respectively, and adapted to distinct environments  
71 (Loftus *et al*, 1994). Some extant breeds originate from old and/or recent interbreeding between  
72 the *B. taurus* and *B. indicus* resulting in a wide geo-climatic adaptation of the hybrids (Larkin  
73 and Yudin, 2016).

74 According to a recent study involving the whole-genome genotyping of 129 bovine  
75 breeds (Decker *et al*, 2014), the European cattle breed pool consists mainly of animals of *B.*

76 *taurus* ancestry without a great deal of contribution from *B. indicus* genes, with the exception  
77 of Turkish breeds. In addition, the Iberian populations of cattle also have a significant genetic  
78 component tracing back to the African taurines (Decker *et al*, 2014). This comprehensive study,  
79 however, did not include breeds from Russia, despite some of them expressing unique  
80 adaptations (e.g. the ability to live above the polar circle expressed by the Yakut cattle). Other  
81 recent studies of native European cattle did however include a limited number of samples from  
82 several Russian native breeds (Iso-Touru *et al*, 2016; Upadhyay *et al*, 2017; Zinovieva *et al*,  
83 2016) but did not carry out a comprehensive comparison between the Russian cattle and the  
84 world breeds. A high divergence of the Yakut cattle (Iso-Touru *et al*, 2016) was suggested as  
85 well as distinct genetic profiles of several Russian breeds placing some of them apart from the  
86 European Holstein-Friesian population (Zinovieva *et al*, 2016).

87         Due to Russia's unique geographic position in both Europe and Asia, its large territory,  
88 diverse climate conditions and its rich history, it is expected that Russian native cattle will  
89 demonstrate a variety of adaptations and are likely to form a link between the European and  
90 Asian cattle populations. According to historical records, the extant Russian cattle breeds  
91 originate from the ancient Eurasian cattle, including the steppe cattle (Li and Kantanen 2010)  
92 and later (starting from the early 18th century) were affected by 'uncontrolled' interbreeding  
93 with multiple European cattle populations (Dmitriev and Ernst, 1989). Currently there are 16  
94 native breeds recognised in Russia (Dunin and Dankvert, 2013) with even more being extinct  
95 (DAD-IS, 2017). The Russian cattle breeds can be classified as the breeds of Eastern European  
96 origin (e.g. Kholmogory and Yaroslavl), crossbred Eastern European breeds (e.g. Istoben, and  
97 Kazakh Whiteheaded), and Asian/Siberian/Turano-Mongolian breeds (e.g. Yakut, Buryat)  
98 (Buchanan and Lenstra, 2015). A comprehensive molecular genetic study of the Russian cattle  
99 is missing or limited to the studies based on mitochondrial DNA and a small number of

100 autosomal (Li and Kantanen, 2010), and Y-chromosome microsatellite markers (Edwards *et*  
101 *al*, 2011).

102         The aim of this study therefore was to analyse of a dataset composed mostly of Russian  
103 and native breeds from neighbouring countries in the context of the dataset of world breeds.  
104 We used the GGP HD150K and Illumina Bovine 50K arrays to genotype individuals from 18  
105 breeds bred in Russia, combined our data with the dataset containing additional 129 cattle  
106 breeds collected from around the world (Decker *et al*, 2014) and samples from ten breeds from  
107 Russia and Europe genotyped previously (Iso-Touru *et al*, 2016). We aimed at building on  
108 these established resources to use them as a reference to reveal the genetic structure and history  
109 of Russian native cattle and to develop hypotheses about their relationships with breeds  
110 worldwide. To reveal the complex history of Russian cattle breeds, multiple complementary  
111 methods of population genetics were applied to the datasets, and hypotheses pertaining to the  
112 origin and structure of the extant breeds were built based on integration of the results.

## MATERIALS AND METHODS

113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137

### **Sample collection**

We used breed society and herdbook information to locate the herds of nine native cattle breeds bred in Russia and the Siberian population of Herefords. Collection of blood (maximum volume = 10 ml) was carried out by superficial venepuncture using sterile 10-ml BDK2EDTA Vacutainers® (Wellkang Ltd, London, UK). In addition, sperm samples from bulls of seven breeds were purchased from breeding companies, and sperm samples from six breeds were obtained from Russian Research Institute of Farm Animal Genetics and Breeding (St. Petersburg, Russia). Additional DNA samples for three breeds were identified from the Russian Cattle Genomic Diversity Panel v.1.0 (Yudin *et al*, 2015). Where pedigree details were available, we attempted to avoid sampling of individuals known to be closely related (e.g. siblings, parent and offspring). Additionally, a balanced combination of the same breed samples from different sources/locations was selected for genotyping (Table 1), however for seven breeds the number of samples collected was <10 with as few as two for the Red Pied cattle suggesting that sampling may not account completely for the breed's genetics. Whole blood and sperm were both stored at -80°C until further use.

### **DNA extraction and genotyping of single nucleotide polymorphisms (SNPs)**

DNA from blood samples was extracted using cell lysis followed by phenol-chloroform extraction (Sambrook *et al*, 2006). The semen samples were pretreated with guanidinium thiocyanate (AppliChem, Darmstadt, Germany) and DNA was extracted using a salting out method (Miller *et al*, 1988). DNA quality and quantity were determined using a NanoDrop 2000c (Thermo Scientific, Wilmington, DE, USA). High quality samples (i.e. having DNA concentrations of at least 50 ng/µl and A260/280 ratios of ca. 1.8) were then subjected to array genotyping. When the number of DNA samples from purebred unrelated animals of the same

138 or different source/location (Table 1) exceeded ten for a breed, genotyping was performed on  
139 the GeneSeek Genomic Profiler High-Density (GGP HD150K) array containing ~139,000 SNP  
140 markers with plans to include this dataset into the follow-up studies on detecting signatures of  
141 selection in bovine genomes which would benefit from a higher number/density of SNPs being  
142 genotyped. Otherwise, samples were genotyped on the BovineSNP50 Analysis BeadChip  
143 (BovineSNP50K) array containing ~54,000 SNP markers compatible with many previously  
144 published datasets, which is a number sufficient for the present study. Each genotyping set  
145 contained several duplicated DNA samples (three for GGP HD150K and two for  
146 BovineSNP50K) to control for the quality of genotyping and to identify potentially problematic  
147 SNP markers.

148 Genotypes were called using the GenomeStudio 2 software (Illumina, San Diego, USA), and  
149 samples with call rates of < 95% were excluded from the further analyses. A pedigree (.ped)  
150 file containing the genotype calls, sample and family identifiers and a map (.map) file  
151 containing the chromosomal location and identifier for each SNP were generated using  
152 GenomeStudio 2 and imported into the PLINK whole genome analysis toolkit (Purcell *et al*,  
153 2007) for further processing.

#### 154 **Data merging and filtering**

155 To identify relationships between the Russian cattle breeds and worldwide breed collections  
156 our GGP HD150K and BovineSNP50K genotyping sets were combined with a set of 48  
157 samples originating from the Ukrainian Grey cattle (Boussaha *et al*, 2015) applying the PLINK  
158 *--merge* command and a common set of ~43,000 SNP markers shared between the GGP  
159 HD150K and BovineSNP50K arrays. To the merged set we added the genotyping sets  
160 generated by Decker *et al.* (2014) (128 additional breeds) and Eurasian breeds from Iso-Touru  
161 *et al.* (2016) (10 breeds). The latter two datasets contained a total of 1,836 individuals. The  
162 datasets were combined with the PLINK *--merge* command using only SNP with unique IDs

163 and chromosomal positions as identified by the SNPchiMp v.3 software (Nicolazzi et al, 2015)  
164 and custom Python scripts. The combined dataset was further filtered to exclude duplicate  
165 samples, poorly genotyped individuals (< 95% of SNPs), loci genotyped in < 99% of  
166 individuals and rare alleles (MAF < 0.001) in PLINK: *--geno 0.01 --mind 0.05 --maf 0.001*  
167 resulting in a subset of 26,740 SNP that were used for the analyses described below.

### 168 **Population structure and phylogenetic analyses**

169 Population structure was characterised using: 1) individual distance-based phylogenetic  
170 analysis, 2) model-based clustering and 3) assumption-free Principal Component Analysis  
171 (PCA). To ensure that analyses would not be distorted by the presence of SNPs in a strong  
172 linkage disequilibrium (LD), the *--indep* command in PLINK was used to prune the SNPs that  
173 passed the initial filtering step. This was achieved by removing one locus from each pair for  
174 which LD ( $r^2$ ) exceeded 0.1 within 50 SNP blocks resulting in 16,645 remaining SNPs. To  
175 estimate and test the phylogenetic relationship of different breeds we constructed a neighbour-  
176 joining (NJ) tree (Saitou and Nei, 1987) based on individual genotypes in FastNJ software (Li,  
177 2015). Tree topology was tested with 1,000 bootstrap replications. Nodes with < 70% support  
178 were collapsed and the resulting tree was visualised using FigTree software (Rambaut and  
179 Drummond, 2012). To evaluate the fractions of putative ancient populations in the modern  
180 genetic pool we used the fastSTRUCTURE (v1.0) clustering and stratification program (Raj *et*  
181 *al*, 2014). The program runs were carried out assuming between one and 40 groups (K) for  
182 both global set and Russian breed (including also closely related world breeds with more than  
183 five sampled individuals identified from the global NJ tree) sets. The cluster membership  
184 matrixes of the fastSTRUCTURE outputs were visualised using PONG software (Behr *et al*,  
185 2016). We used model complexity that maximises marginal likelihood to infer the putative  
186 optimal number of genetic clusters. As an assumption-free illustration of the differentiation

187 between breeds, PCA, was performed using the SNPrelate Bioconductor package (Zheng *et al*,  
188 2012).

### 189 **Single nucleotide polymorphism diversity, linkage disequilibrium and haplotype sharing**

190 An estimate of expected heterozygosity ( $H_e$ ) at each locus was calculated using the *--hardy*  
191 command in PLINK and the mean value was calculated for each breed. The proportion of  
192 polymorphic loci ( $P_n$ ) in each breed and the mean inbreeding coefficient ( $F$ ) values were  
193 calculated using the PLINK commands *--freq* and *--het*, respectively. To calculate pairwise  
194 differentiation ( $F_{ST}$ ) between different breeds we used smartpca software from the Eigensoft  
195 package (v 6.1.4) (Patterson *et al*, 2006).

196 Runs of homozygosity (ROH) represent long stretches of haplotypes identical by  
197 descent (IBD) and provide valuable information about past and recent demographic events  
198 which accompanied the history of populations. To calculate ROH we used the methodology of  
199 (Purfield *et al*, 2012) with stringent settings suitable for low-density genotype samples:  $\geq 1$   
200 SNP per 80 Kbp region,  $> 30$  SNPs per region, with no more than one heterozygous SNP  
201 (PLINK commands: *--homozyg-density 80 --homozyg-snp 30 --homozyg-het 1*). To investigate  
202 the relationship and to infer signatures of recent gene flows between pairs of populations we  
203 used the method based on the detection of IBD-shared haplotypes according to (Ralph and  
204 Coop, 2013). Briefly, the genotypes of the global dataset were split by chromosome and phased  
205 using SHAPEIT 2 software (Delaneau *et al*, 2013) with 400 conditioning states (*--states 400*)  
206 and the effective population size ( $N_e$ ) equal 15,000 as a safe provisional estimate for our diverse  
207 dataset. We used a high-density genetic map of the cattle genome (Ma *et al*, 2015) to correct  
208 for local variations in recombination rate during the haplotype inference. The haplotype sharing  
209 analysis was conducted using BEAGLE 4.1 software (Browning and Browning, 2013) based  
210 on phased haplotypes with LOD score  $\geq 2.5$  (*ibdlod=2.5*), the length of shared haplotypes  $\geq$   
211 100 Kbp (*ibdcm=0.01*) and the number of markers trimmed from the end of the shared

212 haplotypes when testing for IBD equalled three (*ibdtrim*=3). The inferred shared haplotypes  
213 were binned into three categories according to the size (< 3 Mbp, 3-5 Mbp, > 7 Mbp) and  
214 plotted using the R libraries *igraph* and *ggcorrplot* (R Development Core Team, 2008).

215 To further reveal traces of genetic admixtures and their directions between the Russian  
216 and closely related breeds, and between the Russian and breeds of *B. indicus* origin, we applied  
217 the maximum-likelihood algorithm implemented in the Treemix software (Pickrell and  
218 Pritchard, 2012) which models migration events on the phylogenetic tree. Two datasets  
219 analysed separately were: 1) Russian and closely related world breeds as defined from the NJ  
220 analysis, 2) Russian breeds and breeds of known *B. indicus* and *B. javanicus* origin from  
221 Decker et al. 2014 with at least five sampled individuals per breed. The Treemix analysis was  
222 performed with 1 SNP per block for estimation of the covariance matrix ( $k=1$ ) and gradual  
223 addition from one to 15 migration events with the step equal to one for the first dataset and  
224 from three to 18 migration events with the step equal to three for the second dataset. We rooted  
225 the trees on the Yakut cattle and *B. javanicus* for the first and second datasets, respectively.  
226 The optimal number of migration events was determined after examining the difference  
227 between the likelihoods of the tree after each migration step being added and the tree's previous  
228 step likelihood ( $\Delta$  Likelihood).

229 To estimate the historical and recent effective population sizes ( $N_e$ ) in Russian breeds  
230 we applied a method based on the relationship between the extent of LD,  $N_e$  and the  
231 recombination rate within the populations implemented in the SNeP software (Barbato *et al*,  
232 2015). The calculations were performed on SNPs with  $MAF \geq 0.05$ , with sample size  
233 correction (*-samplesize*), and with minimum and maximum distances equal to 5,000 bp and  
234 2,000,000 bp respectively. The recombination correction was applied according to (Sved and  
235 Feldman, 1973). LD values for size bins in the range from 28 Kbp to 600 Kbp were extracted  
236 from the SNeP output and plotted to estimate the LD decay for Russian cattle breeds.

## RESULTS

237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253

### **Single nucleotide polymorphism, diversity, inbreeding and linkage disequilibrium within the Russian breeds**

Both the GGP HD150K and BovineSNP50K SNP arrays were found highly informative for the Russian cattle breeds (Suppl. Table 1). The proportion of loci polymorphic ( $P_n$ ) in at least one breed for the overlapping set of 26,701 SNPs shared between the arrays varied from 0.650 for the Red Pied to 0.977 for the Black Pied breeds with a mean of 0.891 (Table 1). The mean MAF was found highly consistent among the breeds ranging from 0.205 (Yakut) to 0.269 (Black Pied and Kalmyk). Similarly, the expected heterozygosity ( $H_e$ ) was relatively high in all Russian breeds (range 0.271- 0.352, mean 0.324) with the lowest values observed in the Red Pied (0.271) and Yakut (0.273) and the highest in the Black Pied (0.352) and Tagil (0.350). The inbreeding coefficient ( $F$ ) demonstrated negative values for all the breeds, but the largest deviations from zero ( $>0.1$ ) should be taken with caution because they were observed for the breeds with the lowest number of samples analysed (i.e. Istoben, Red Pied, Red Steppe, and Yurino; Table 1) suggesting that the genetic composition was likely not covered in full for these breeds.

254  
255  
256  
257  
258  
259  
260

We estimated the recent and past effective population ( $N_e$ ) sizes for the native breeds and plotted the results (Suppl. Figure 1). All of the Russian breeds demonstrated a highly similar pattern of  $N_e$  decay with an increased rate starting  $\sim 200$  generations ago (Suppl. Figure 1b) likely being caused by bottle necks associated with contemporary breed formation. The highest historical  $N_e$  sizes were observed for the Buryat and Kalmyk while the lowest  $N_e$  for the Yakut (Suppl. Figure 1a,b). The LD decay plot (Figure 1) suggested the presence of long haplotypes usually associated with low  $N_e$  size (e.g. Yakut, Kostroma and Kholmogory); the

261 most pronounced effect was observed for the Yakut cattle. The Buryat and Kalmyk cattle  
262 demonstrated a rapid LD decay consistent with the historically larger  $N_e$  sizes of these breeds.

263 Consistent with the expectation of high inbreeding within Yakut, Kostroma, Kazakh  
264 Whiteheaded and Ukrainian Gray breeds, the presence of the longest and most frequent ROHs  
265 (>500 Kbp/>4 per animal) were observed within these breeds (Suppl. Figure 2). Tagil and  
266 Buryat demonstrated the shortest and the least frequent average number of ROHs in their  
267 genomes suggesting that these breeds could have been managed effectively to avoid excessive  
268 inbreeding. All other breeds expressed an intermediate level of ROHs consistent with the  
269 higher  $N_e$  and expected moderate level of inbreeding.

#### 270 **Ancestry of Russian cattle breeds**

271 To identify ancestral relationships between native breeds from Russia and the cattle breeds  
272 distributed worldwide, we analysed our datasets with that of world breeds (Decker *et al*, 2014)  
273 and additional Eurasian breeds (Iso-Touru *et al*, 2016). As expected, the first two components  
274 of PCA differentiated the main clusters of breeds from Africa, Asia, and Europe representing  
275 mainly African taurine, cattle of Eurasian taurine origin and cattle of Asian indicine origin,  
276 with breeds expressing various levels of hybridisation found in between (Suppl. Figure 3).  
277 Breeds from the Americas clustered with the European and Asian breeds. The majority of  
278 Russian breeds followed the European taurine breed cluster with additional breeds found in the  
279 cluster of taurine Asian breeds (Suppl. Figure 3). These results were highly consistent with the  
280 fastSTRUCTURE analysis which suggested a close relationship between the breeds from  
281 Russia and other taurine breeds of European and several of Asian origin (Figure 2). However,  
282 at K=4 a separate cluster was formed by the Yakut cattle. The first other breed that formed a  
283 separate cluster was the British Shorthorn (K=5).

284 The collapsed NJ tree grouped samples into the well-supported breed-specific nodes  
285 confirming the expected phylogenetic relationships within the breed populations (Figure 3).

286 The three major well-resolved branches of the tree separated the breeds of Asian, African and  
287 European/American origins consistent with the PCA and fastSTRUCTURE results. The  
288 majority of the breeds from Russia (N = 17) were distributed along the branch of the  
289 European/American taurine breeds with some of them forming well-supported clusters with  
290 other breeds indicating close relationships. The Yakut cattle was found in the same cluster with  
291 Hanwoo and Wagyu cattle from Korea and Japan respectively, near the Buryat cattle node.  
292 Other two well-resolved clusters involving Russian breeds and world breeds have grouped  
293 together the Kazakh Whiteheaded breed from Russia with Hereford samples from Russia and  
294 Wales; and the Ala-Tau and Kostroma with two breeds of European origin (Braunveih and  
295 Brown Swiss). Ukrainian Whiteheaded, Gorbatov Red and Istoben formed a separate cluster  
296 on the branch of the European breeds. Yurino formed a cluster with the Pinzgauer cattle from  
297 Austria. Kholmogoy, Black Pied, Tagil, Red Steppe formed a large cluster with the Holstein-  
298 Friesian, French Red Pied and Lithuanian Light Grey breeds.

299 The pairwise analysis of shared haplotypes between the Russian cattle breeds and  
300 taurine breeds of European and Asian origins has identified 39 breeds with a significant level  
301 of haplotype sharing with at least one breed from Russia (LOD>2.5; Figure 4, Suppl. Figure  
302 4). The top 10 world breeds that shared haplotypes at all three levels of the haplotype analysis  
303 were: Brown Swiss, Hereford, Holstein, Braunvieh, Senepol, French Red Pied, Beef Shorthorn,  
304 Maine Anjou, Norwegian Red, and Jersey (Figure 4, Suppl. Figure 4). This analysis has  
305 provided evidence for additional breed relations to the previously described results, and  
306 allowed us to distinguish between older and more recent relationships. Sharing of the short  
307 haplotypes (0-3 Mbp; presumably indicative of older relationships between populations) has  
308 formed two clear large clusters and two smaller clusters of breeds (Figure 4a). The largest  
309 cluster revealed the ancestral relationships between the Northern French, British, and Finnish  
310 breeds with the Yaroslavl, Bestuzhev, Black Pied, Tagil, and Kazakh Whiteheaded breeds from

311 our set. The second large cluster suggested further ancestral relationships between the South-  
312 European breeds from South-East France, Italy, Switzerland with Kostroma and Ala-Tau  
313 breeds from our dataset. Ukrainian Grey cattle samples both from Russia and Serbia shared  
314 short haplotypes with Podolian cattle (Serbia) and Romagnola (Italy) breeds. The Yakut and  
315 Buryat breeds formed a separate small cluster with Japanese Wagyu cattle whereas the Kalmyk  
316 cattle had significant haplotype sharing only with the Beef Shorthorn from England.

317         The longest shared haplotypes (>7 Mbp; likely indicative of recent introgression and  
318 admixture events, Figure 4b) revealed the recent admixture between the Ukrainian Grey  
319 sampled in Russia (Boussaha *et al*, 2015) and the Yakut cattle. The Ukrainian Grey breed  
320 sampled in Serbia (Iso-Touru *et al*, 2016) did not demonstrate this pattern. Both the Kostroma  
321 and Ala-Tau breeds had extensive haplotype sharing with the Brown Swiss and Braunvieh.  
322 Multiple Russian breeds (i.e. Bestuzhev, Black-Pied, Tagil, Yaroslavl, Kholmogory) shared  
323 haplotypes with Holstein-Friesian, Senepol, French Red-Pied Lowland and Normande breeds.

324         To investigate more closely the genetic ancestry of the sampled Turano-Mongolian  
325 breeds (Yakut, Buryat, Kalmyk) we plotted the extent of pairwise haplotype sharing for each  
326 breed from highest to lowest value (Figure 5 and Suppl. Figure 5) for shortest haplotype  
327 segments (0-3 Mbp) including both taurine and indicine world breeds. The closest breeds (>1.5  
328 Mbp average total haplotype length shared per animal) to Yakut cattle were Hanwoo, Buryat,  
329 Wagyu, Qinchan, Mongolian cattle and Morucha demonstrating a pronounced signal on the  
330 plot. The Buryat breed was mostly related to Wagyu, Hanwoo, Yakut, Qinchuan, Ala-Tau and  
331 Mongolian cattle breeds. Thus, Yakut and Buryat breeds showed a close relationship with  
332 taurine Asian breeds (and with each other) confirming their shared ancestry. Our samples of  
333 the Kalmyk cattle demonstrated mostly low values of haplotype sharing with the strongest  
334 relationship to Beef Shorthorn and a much weaker sharing with Wagyu and Welsh Black  
335 breeds. Interestingly, another sampling of the Kalmyk breed (Iso-Touru *et al*, 2016) showed

336 some higher signal values, although it confirmed a relationship between the Kalmyk cattle and  
337 both taurine Asian (Hanwoo) and European (Simmental, Beef Shorthorn, Groningen  
338 Whitehead) breeds (Suppl. Figure 5). In our analysis breeds of known indicine origin did not  
339 demonstrate high degree of haplotype sharing with Turano-Mongolian breeds with signal  
340 values always much lower than the values observed for the top taurine breeds (Figure 5 and  
341 Suppl. Figure 5).

342  
343 **The relationships between the cattle breeds from Russia and the closely related world**  
344 **breeds.**

345 To reveal the fine-structure relationship between the Russian cattle breeds and the set of closely  
346 related world breeds we performed a separate PCA and a fine-scale admixture analysis of the  
347 breeds from Russia and eight world breeds that formed well supported clusters with the Russian  
348 breeds on the collapsed NJ tree (see Figure 3). The first two components of PCA revealed four  
349 major clusters of the breeds (Suppl. Figure 6a). The largest cluster contained the Holstein-  
350 Friesian cattle with the Black Pied and other European and Russian breeds that likely had been  
351 influenced by European dairy cattle genetics during their formation; the second cluster  
352 combined the Buryat, Kalmyk, Ukrainian Gray and Asian taurine breeds (Hanwoo and  
353 Wagyu). Another cluster combined the Kazakh Whiteheaded, Hereford cattle and the last one  
354 - the Yakut cattle. The third principal component of PCA separated the Ukrainian Gray cattle  
355 from the cluster of the taurine breeds of Asian origin and revealed a separate cluster formed by  
356 the Kostroma, Brown Swiss, Ala-Tau and Braunvieh breeds (Suppl. Figure 6b). The most  
357 likely number of populations according to the maximum likelihood estimation was equal to 11  
358 (Suppl. Figure 7). The fineSTRUCTURE results (Figure 6) suggest that the most distant breeds  
359 within this set were the Yakut and Kholmogory, separated from the other breeds at  $K=2$ ,  
360 followed by the Hereford and Kazakh Whitehead group at ( $K=3$ ). The next cluster was formed  
361 by Kostroma and Brown Swiss ( $K=4$ ) followed by the Ukrainian Grey cattle ( $K=5$ ). At  $K=6$  it

363 becomes apparent that the Asian taurine breeds (Hanwoo and Wagyu) cluster with the Kalmyk  
364 and Buryat cattle; the genetic material represented by the Hanwoo had a larger contribution on  
365 the Kalmyk and Buryat breeds than on Wagyu. At  $K=7$  a central cluster of mostly composite  
366 breeds with the influence of Holstein-Friesian/Black Pied genetic material becomes apparent  
367 with the Yaroslavl separating from this cluster at  $K=10$ . We cannot exclude the possibility that  
368 the number of optimal genetic clusters in our analysis has been influenced by the unequal breed  
369 sample size and, in particular, by a small number of individuals collected for the Yurino, Red  
370 Steppe, Red Pied, Gorbatov Red and Istoben breeds. A larger number of samples would be  
371 needed to confirm the genetic composition of these breeds.

372         The Treemix results for the Russian cattle and most related other breeds (Suppl. Figure  
373 8) demonstrated the highest  $\Delta$  likelihood increase for two migration events: the first one from  
374 the Yakut to the Ukrainian Gray breed collected in Russia and the second one from the Holstein  
375 to Tagil breed. Both results were in agreement with the observations made based on the  
376 haplotype sharing. For the Russian breeds combined with known *B. indicus* breed set (Suppl.  
377 Figure 9), the highest gain in likelihood was received for nine migration events without any of  
378 them suggesting migration links between the Russian and indicine cattle populations.

379         The analysis of the  $F_{ST}$  distances between the Russian breeds and those breeds closely  
380 related to them (Suppl. Table 2) revealed a low level of genetic differentiation with the mean  
381 value equal to 0.096 and a range from 0.003 to 0.235. The strongest differentiation involving  
382 a Russian breed was observed between the Podolian and Yakut cattle breeds while the lowest  
383 values were observed between the Red Pied and Finnish Ayrshire ( $F_{ST} = 0.003$ ). The Red Pied  
384 breed had a very low number of samples in our dataset ( $N=2$ , Table 1) and these results should  
385 be taken with caution. Apart from this, the lowest  $F_{ST}$  values were found between the samples  
386 of the Yaroslavl breed collected by us and by (Iso-Touru *et al*, 2016). Surprisingly, the  
387 differentiation between the Black Pied and Holsten breeds ( $F_{ST} = 0.020$ ) was lower than the

388  $F_{ST}$  observed between the Hereford samples from Russia and Wales ( $F_{ST} = 0.029$ ). The Yakut  
389 cattle has consistently demonstrated higher  $F_{ST}$  values with other breeds, with the lowest  
390 differentiation observed with the Buryat cattle followed by the Kalmyk and Hanwoo breeds.  
391 Interestingly, the Wagyu breed, which had a high fraction of haplotypes shared with the Yakut  
392 cattle and was found next to it on the phylogenetic tree, had one of the highest degrees of  
393 differentiation with it ( $F_{ST} = 0.20$ ), following the Hereford and Podolian cattle.

## DISCUSSION

394

395

396 The advent of cost-efficient genotyping SNP arrays has made it possible to reveal the genetic  
397 profiles of various breeds of domesticated species, develop informed strategies of their  
398 improvement on one hand, and learn about the genetic processes accompanying domestication  
399 and breed formation on the other. While most efforts are dedicated to studying popular  
400 commercial breeds, e.g. Texel in sheep (Mucha *et al*, 2015) and Holstein-Friesian in cattle (van  
401 Binsbergen *et al*, 2015), there is a growing interest in the genetics of smaller local breeds  
402 because of the unique adaptations found in their genomes and their potential to contribute to  
403 solving problems in agriculture related to environmental change (e.g. global warming) and  
404 local pathogen resistance (Beynon *et al*, 2015). To this end we performed genotyping of 18  
405 cattle breeds bred in Russia selected on the basis of a likely historical contribution of local  
406 cattle populations onto their contemporary genomes and compared them to commercial and  
407 native breeds previously collected from around the world (Decker *et al*, 2014; Iso-Touru *et al*,  
408 2016). Along with the highly popular abundant Russian breeds (e.g. Black Pied or  
409 Kholmogory) we included highly specialised breeds that demonstrate extensive adaptations to  
410 specific environments (e.g. Yakut) and/or were almost extinct (e.g. Buryat). Therefore, our  
411 current dataset represents the largest and most complete set of the cattle breeds from Russia  
412 available for population genetic studies so far.

413

414

415

416

417

418

419

In agreement with the geographical position of Russia and its historical and trade links,  
the majority of the Russian cattle breeds demonstrated extensive common ancestry with the  
taurine cattle breeds from Europe. As expected to result from the ‘uncontrolled’ and/or  
complex breeding strategies started as early as in the 18<sup>th</sup> century (Dmitriev and Ernst, 1989),  
for most of the Russian breeds we could not clearly identify their sister foreign breeds on the  
phylogenetic tree, except for the European and the Russian cattle being found on the same wide  
polygenetic node. However, there were several examples when our data has confirmed the

420 known historical relationships among the Russian cattle and some foreign breeds  
421 demonstrating robustness of our results. The most profound of these links is between the  
422 Kazakh Whiteheaded and Hereford breeds from both Russia and Europe, well supported by the  
423 known recent breeding history of the Kazakh Whiteheaded. The breed was formed between  
424 1930 and 1950, by crossing of the Turano-Mongolian Kazakh and Kalmyk cattle with Hereford  
425 in the Kazakh Republic of the USSR (Dmitriev and Ernst, 1989). Another example of known  
426 relations and historical breed formation (Dmitriev and Ernst, 1989) was confirmed by  
427 clustering of Kostroma, Brown Swiss, Braunveih and Ala Tau consistently supported by the  
428 structure, phylogenetic, haplotype analysis and population differentiation levels ( $F_{ST}$  ranges  
429 0.032-0.069). A separate node on our phylogenetic tree formed by the Kholmogory, Holstein-  
430 Friesian, Black Pied and several related European breeds further supported by haplotype  
431 sharing likely reflects the historical relations that trace back to the 17th century when the  
432 Kholmogory breed was formed and later interbred with “Dutch cattle” (Dmitriev and Ernst,  
433 1989). However, the structure analysis indicates that the genetic component of the  
434 contemporary Holstein-Friesian breed in Kholmogory is relatively small and that Kholmogory  
435 should be considered genetically distinct, supporting previous observations (Zinovieva *et al*,  
436 2016). On the other hand, our samples of the Russian Black Pied breed demonstrate a very low  
437 differentiation from the Holstein-Friesian ( $F_{ST} = 0.02$ , e.g. lower than between the two sets of  
438 Hereford samples in our analysis) suggesting that the use of imported Holstein-Friesian  
439 sires/semen in Russia could have significantly affected the Black Pied’s genetics. Haplotype  
440 sharing analysis that was based on short haplotype blocks (presumably reflecting ancestral  
441 relationships) has further confirmed a complex history of the Russian cattle breeds of European  
442 origin but allowed to assign them to three major clusters based on predominantly shared  
443 haplotypes. While the largest cluster mostly contained Russian breeds with historical influence  
444 from highly commercial European breeds (e.g. Holstein-Friesian and Angus) and other breeds

445 that could also have been influenced by these multinational breeds, the second one was built  
446 around the related Kostroma, Brown Swiss, Braunveih and Ala Tau breeds with the addition  
447 of several other breeds from France, Italy and Germany. The Ukrainian Gray cattle formed the  
448 last separate cluster shared only with the Podolian and Romadnola breeds confirming the  
449 Ukrainian Gray's position within the primitive Podolian group of cattle breeds (Kushnir and  
450 Glazko, 2009).

451 In addition to extensive links to cattle of European ancestry, PCA suggested that there  
452 are breeds in Russia that have shared ancestry with cattle from Asia. In agreement with this,  
453 the Yakut, Buryat, and Kalmyk cattle clustered with the Turano-Mongolian and other Asian  
454 taurine breeds on the phylogenetic tree and structure plots. While on the structure global dataset  
455 the Yakut breed formed the first breed-specific cluster after the observed divergence of *B.*  
456 *indicus* and African taurines, on the phylogenetic tree, it was found on the same node with  
457 Buryat cattle and other taurine Asian breeds. The exact reason for the Yakut cattle being so  
458 divergent based on structure results is not currently clear but could be related to a combination  
459 of its low historical  $N_e$  combined with long isolation from other breeds. A closer relation of the  
460 Yakut cattle with other divergent Asian Turano-Mongolian breeds may imply their early  
461 separation from the rest of the taurine gene pool or even independent domestication in Asia  
462 (Mannen *et al.*, 2004). Haplotype sharing results further confirm these relationships within the  
463 Turano-Mongolian breed set placing the Yakut cattle on the same cluster with Buryat and  
464 Wagyu and indicating links with Hanwoo and the Mongolian cattle. The Buryat cattle was  
465 considered to be extinct until quite recently when a herd had been discovered in Mongolia and  
466 imported back to Russia to start recovering the breed. Our results indeed demonstrated that the  
467 Buryat shares more haplotypes with Yakut and Wagyu cattle than with the breeds from  
468 Mongolia suggesting its separate origin from Mongolian cattle. Interestingly, the placement of  
469 the third Turano-Mongolian breed on our list, the Kalmyk remains unclear. While it formed a

470 separate cluster within the European cattle on the phylogenetic tree, structure results suggested  
471 a common ancestry with Buryat and Hanwoo breeds. Haplotype sharing showed a strong recent  
472 admixture with the Beef Shorthorn. The latter can be explained by the known use of Shorthorn  
473 to ‘improve’ the Kalmyk cattle in the USSR (Dmitriev *et al*, 1989). This likely had an impact  
474 on the genetics of this breed and affected its position of the phylogenetic tree masking the  
475 expected ancestral relationships that were picked up only by the structure analysis. Another  
476 example of a likely effect of a recent admixture on the genetics of a breed was observed during  
477 comparison of the Ukrainian Gray cattle samples originating from Serbia (Iso-Touru *et al*,  
478 2016) and from Russia (Boussaha *et al*, 2015). While these sample sets cluster together  
479 suggesting that they indeed belong to the same breed, the sample set from Russia demonstrated  
480 a clear evidence of a recent admixture with the Yakut cattle based on the haplotype sharing and  
481 confirmed by TreeMix analysis. The samples from Serbia had no traces of this event.

482         Interestingly, we did not identify any significant evidence of admixture between any of  
483 the Russian breeds and the indicine cattle neither in haplotype sharing nor the Treemix  
484 analyses. However, the structure global plot (K=3) suggested some level of indicine ancestry  
485 in the Turano-Mongolian breeds. This observation may imply a very ancient and probably  
486 weak admixture event not detected by other methods. It is also possible that the  
487 BovineSNP50K array SNP loci (and, as a result, the set of SNPs used in the present work) bias  
488 to taurine and ancient SNPs shared by taurine and indicine populations (McKay *et al*, 2008)  
489 has affected our results to some extent and masked admixture with *B. indicus*. Both scenarios  
490 suggest that more detailed studies involving the whole-genome resequencing of Russian cattle  
491 genomes and their comparison to both the taurine and indicine genome references would be  
492 needed to resolve this issue and shed additional light on the reasons for observed divergence  
493 of the Yakut cattle.

494           When analysed individually or in the context of only the most related world breeds, the  
495 Russian cattle breeds demonstrated a modest level of genetic diversity and comparable  
496 estimates of effective population sizes with other Eurasian breeds (Iso-Touru *et al*, 2016). The  
497 Kholmogory and Yakut breeds are further confirmed as the most genetically distinct within the  
498 set of the breeds from Russia and related Eurasian breeds on the structure results, supported by  
499 high  $F_{ST}$ . Strong influence of Holstein-Friesian genetics became apparent in a separate cluster  
500 of breeds. Yaroslavl breed separated from this cluster at  $K=10$  being the last Russian cattle  
501 breed that demonstrated unique genetics while other breeds including the Black Pied, Tagil,  
502 Bestuzhev, Istoben, Yurino, and Ukrainian Whiteheaded demonstrated different levels of  
503 Holstein-Friesian contribution to their genetics suggesting that these breeds might have left  
504 with a relatively small fraction of alleles from native populations. This was supported by a  
505 relatively low level of population differentiation within this group ( $F_{ST}$  range 0.020-0.094).  
506 However, we cannot exclude that both the SNP loci bias to a small number of taurine breeds  
507 and small sampling sizes for some breeds in our list could have influenced these results.

508           The presence of long runs of homozygosity in the Yakut, Kostroma and Ukrainian Gray  
509 breeds might indicate either a high level of adaptation and specialisation or effects of  
510 inbreeding and low effective population size. Regardless of the reason, this information should  
511 be considered during the development of breeding programs for these populations. The genetic  
512 uniqueness of the highly adapted to harsh climatic conditions Yakut breed should stimulate  
513 and guide its recovery program.

514           Herein we provide the first detailed view on the population genetics of a comprehensive  
515 list of the cattle breeds bred in Russia that potentially have arisen from local cattle populations  
516 and/or could be adapted to harsh environments and climate. Our results demonstrate that some  
517 of the breeds studied have distinct genetic profiles (e.g. Kholmogory, Yakut, Yaroslavl)  
518 making them priority targets for deeper studies to reveal signatures of selection and adaptations

519 related to local environments and for conservation purposes. We also observed that a large  
520 group of breeds had both old and recent influence from commercial European breeds (e.g.  
521 Kostroma, Kazakh Whiteheaded, Istoben) meaning that their genomes could potentially  
522 contain only a small fraction of ancestral alleles, but these could be important for surviving  
523 local conditions and can be used for admixture mapping programs aiming at economically  
524 important traits (Kassahun *et al*, 2015). The links between the Russian breeds and breeds from  
525 other countries presented in this study form a basis for future work on contrasting their genomes  
526 to reveal causative alleles or haplotypes using a right set of related and outgroup populations  
527 for the comparison to avoid ‘signal dilution’ or false positive signals. The uniqueness of the  
528 Yakut breed shown in this study makes it a priority for further detailed studies on one hand,  
529 but makes it difficult to identify the right breeds to contrast it to on the other; implying that  
530 additional, more detailed studies of Russian native cattle breeds of Asian origin in the context  
531 of a larger set of Asian taurine and indicine breeds might be required to fully benefit from their  
532 unique genetics.

533

#### 534 **ACKNOWLEDGEMENTS**

535 The work was supported by the Russian Science Foundation grant (RSF, 16-14-0090).

536

#### 537 **CONFLICT OF INTEREST**

538 The authors declare no conflict of interest.

539

540  
541  
542

## **DATA ARCHIVING**

The data will be submitted to Dryad before the paper publication.

543

## REFERENCES

- 544 Barbato M, Orozco-terWengel P, Tapio M, Bruford MW (2015). SNeP: a tool to estimate  
545 trends in recent effective population size trajectories using genome-wide SNP data. *Front*  
546 *Genet* **6**: 109.
- 547
- 548 Behr AA, Liu KZ, Liu-Fang G, Nakka P, Ramachandran S (2016). pong: fast analysis and  
549 visualization of latent clusters in population genetic data. *Bioinformatics* **32**(18): 2817-2823.  
550
- 551 Beynon SE, Slavov GT, Farre M, Sunduimijid B, Waddams K, Davies B *et al* (2015).  
552 Population structure and history of the Welsh sheep breeds determined by whole genome  
553 genotyping. *BMC Genet* **16**: 65.
- 554
- 555 Boichard D, Brochard M (2012). New phenotypes for new breeding goals in dairy cattle.  
556 *Animal* **6**(4): 544-550.
- 557
- 558 Boussaha M, Letaief R, Yudin N, Voevoda M, Rocha D (2015). The extent of linkage  
559 disequilibrium in the ancient Ukrainian grey cattle breed *Global Journal of Animal Breeding*  
560 *and Genetics* **3**(4): 159-163.
- 561
- 562 Bovine HapMap Consortium (2009). Genome-wide survey of SNP variation uncovers the  
563 genetic structure of cattle breeds. *Science* **324**(5926): 528-532.
- 564
- 565 Browning BL, Browning SR (2013). Improving the accuracy and efficiency of identity-by-  
566 descent detection in population data. *Genetics* **194**(2): 459-471.
- 567
- 568 Buchanan DS, Lenstra JA (2015). Breeds of Cattle. *Genetics of Cattle, 2nd Edition*: 33-66.  
569
- 570 DAD-IS. (2017). *Vol. 2017*. Food and agriculture organization of the United Nations.  
571
- 572 Decker JE, McKay SD, Rolf MM, Kim J, Molina Alcala A, Sonstegard TS *et al* (2014).  
573 Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *PLoS*  
574 *genetics* **10**(3): e1004254.
- 575
- 576 Delaneau O, Zagury JF, Marchini J (2013). Improved whole-chromosome phasing for disease  
577 and population genetic studies. *Nat Methods* **10**(1): 5-6.
- 578
- 579 Dmitriev NG, Ernst LK (1989). Animal Genetics Resources of the USSR *Food and Agriculture*  
580 *Organization of the United Nations*: Rome, Italy.
- 581
- 582 Dmitriev NG, Ernst LK, Food and Agriculture Organization of the United Nations. (1989).  
583 *Animal genetic resources of the USSR*. Food and Agriculture Organization of the United  
584 Nations: Rome.
- 585
- 586 Dunin IM, Dankvert AG (2013). *Breeds and types of farm animals in the Russian Federation*:  
587 Moscow, Russian Federation
- 588
- 589 Edwards CJ, Ginja C, Kantanen J, Perez-Pardal L, Tresset A, Stock F *et al* (2011). Dual origins  
590 of dairy cattle farming--evidence from a comprehensive survey of European Y-chromosomal  
591 variation. *PLoS One* **6**(1): e15922.
- 592

593 Gao Y, Wu H, Wang Y, Liu X, Chen L, Li Q *et al* (2017). Single Cas9 nickase induced  
594 generation of NRAMP1 knockin cattle with reduced off-target effects. *Genome Biol* **18**(1): 13.  
595

596 Gaouar SB, Da Silva A, Ciani E, Kdidi S, Aouissat M, Dhimi L *et al* (2015). Admixture and  
597 local breed marginalization threaten Algerian sheep diversity. *PLoS One* **10**(4): e0122667.  
598

599 Iso-Touru T, Tapio M, Vilkki J, Kiseleva T, Ammosov I, Ivanova Z *et al* (2016). Genetic  
600 diversity and genomic signatures of selection among cattle breeds from Siberia, eastern and  
601 northern Europe. *Anim Genet* **47**(6): 647-657.  
602

603 Kassahun Y, Mattiangeli V, Ameni G, Hailu E, Aseffa A, Young DB *et al* (2015). Admixture  
604 mapping of tuberculosis and pigmentation-related traits in an African-European hybrid cattle  
605 population. *Front Genet* **6**: 210.  
606

607 Kushnir AV, Glazko VI (2009). Gray Ukrainian cattle and their closely related forms.  
608 *Contemporary Problems of Ecology* **2**(3): 288–295.  
609

610 Larkin DM, Yudin N (2016). The genomes and history of domestic animals. *Molecular*  
611 *Genetics, Microbiology and Virology* **31**(4): 197-202.  
612

613 Li JF (2015). A fast neighbor joining method. *Genet Mol Res* **14**(3): 8733-8743.  
614

615 Li MH, Kantanen J (2010). Genetic structure of Eurasian cattle (*Bos taurus*) based on  
616 microsatellites: clarification for their breed classification. *Anim Genet* **41**(2): 150-158.  
617

618 Loftus RT, MacHugh DE, Bradley DG, Sharp PM, Cunningham P (1994). Evidence for two  
619 independent domestications of cattle. *Proc Natl Acad Sci U S A* **91**(7): 2757-2761.  
620

621 Ma L, O'Connell JR, VanRaden PM, Shen B, Padhi A, Sun C *et al* (2015). Cattle sex-specific  
622 recombination and genetic control from a large pedigree analysis. *PLoS genetics* **11**(11):  
623 e1005387.  
624

625 Mannen H, Kohno M, Nagata Y, Tsuji S, Bradley DG, Yeo JS *et al* (2004). Independent  
626 mitochondrial origin and historical genetic differentiation in North Eastern Asian cattle. *Mol*  
627 *Phylogenet Evol.* **32**(2): 539-544.  
628

629 Mason IL (1969). A World Dictionary of Livestock Breed Types and Varieties *Commonwealth*  
630 *Agricultural Bureaux*. Farnham Royal: UK.  
631

632 Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP *et al* (2009).  
633 Development and characterization of a high density SNP genotyping assay for cattle. *PLoS*  
634 *One* **4**(4): e5350.  
635

636 McKay SD, Schnabel RD, Murdoch BM, Matukumalli LK, Aerts J, Coppieters W *et al* (2008)  
637 An assessment of population structure in eight breeds of cattle using a whole genome SNP  
638 panel. *BMC Genet.* **20**(9): 37.  
639

640 Miller SA, Dykes DD, Polesky HF (1988). A simple salting out procedure for extracting DNA  
641 from human nucleated cells. *Nucleic Acids Res* **16**(3): 1215.  
642

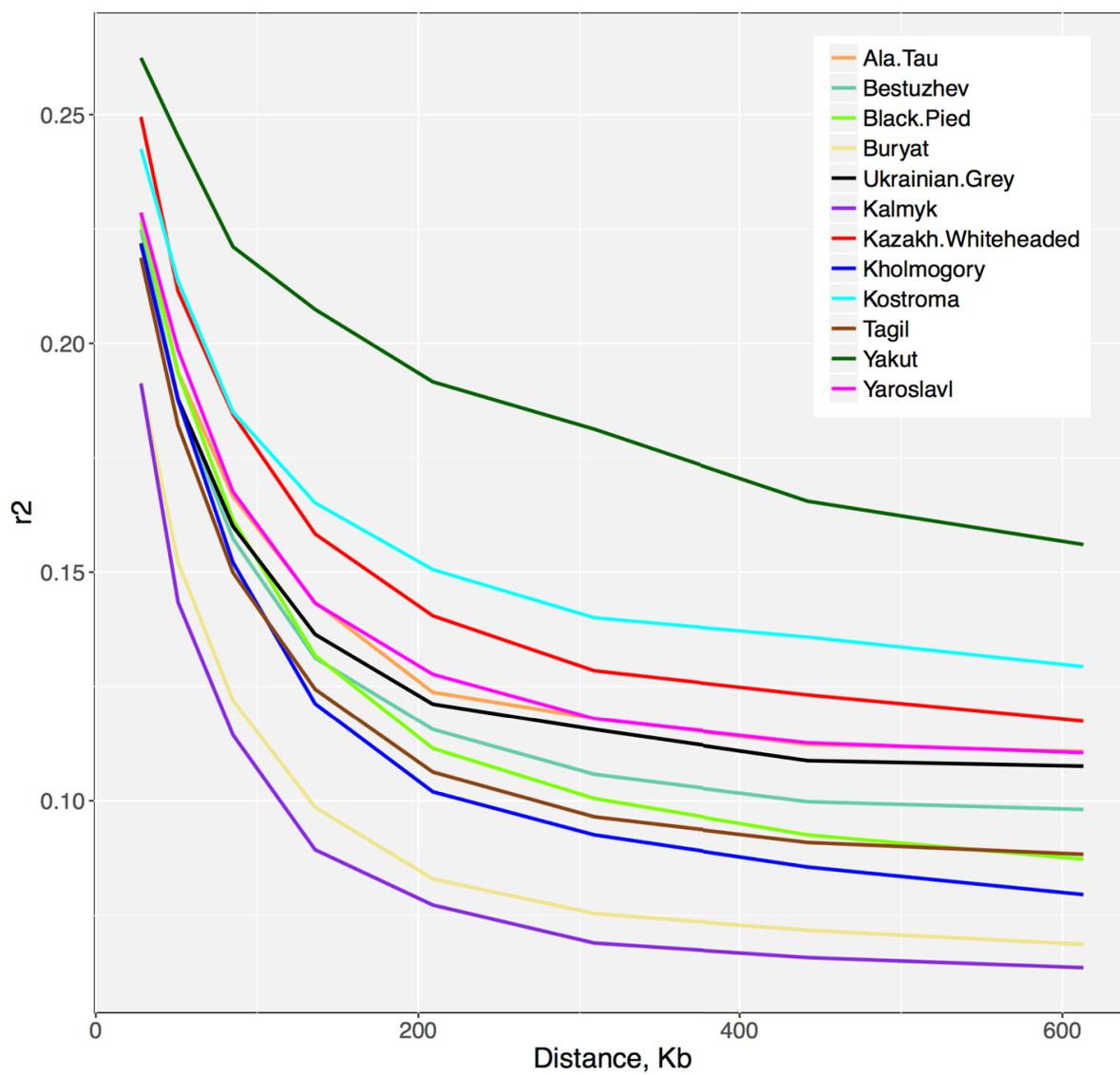
643 Mucha S, Bunger L, Conington J (2015). Genome-wide association study of footrot in Texel  
644 sheep. *Genet Sel Evol* **47**: 35.  
645  
646 Nicolazzi EL, Caprera A, Nazzicari N, Cozzi P, Strozzi F, Lawley C *et al* (2015). SNPchiMp  
647 v.3: integrating and standardizing single nucleotide polymorphism data for livestock species.  
648 *BMC Genomics* **16**: 283.  
649  
650 Patterson N, Price AL, Reich D (2006). Population structure and eigenanalysis. *PLoS genetics*  
651 **2**(12): e190.  
652  
653 Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from  
654 genome-wide allele frequency data. *PLoS genetics*, 8(11): e1002967.  
655  
656 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D *et al* (2007). PLINK: a  
657 tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*  
658 **81**(3): 559-575.  
659  
660 Purfield DC, Berry DP, McParland S, Bradley DG (2012). Runs of homozygosity and  
661 population history in cattle. *BMC Genet* **13**: 70.  
662  
663 Raj A, Stephens M, Pritchard JK (2014). fastSTRUCTURE: variational inference of population  
664 structure in large SNP data sets. *Genetics* **197**(2): 573-589.  
665  
666 Ralph P, Coop G (2013). The geography of recent genetic ancestry across Europe. *PLoS Biol*  
667 **11**(5): e1001555.  
668  
669 Rambaut A, Drummond AJ. (2012). FigTree. Version 1.4.0.  
670 <http://tree.bio.ed.ac.uk/software/figtree>  
671  
672 R Development Core Team (2008). R: A language and environment for statistical computing.  
673 R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL  
674 <http://www.R-project.org>.  
675  
676 Saitou N, Nei M (1987). The neighbor-joining method: a new method for reconstructing  
677 phylogenetic trees. *Mol Biol Evol* **4**(4): 406-425.  
678  
679 Sambrook J, Russell DW, Sambrook J (2006). *The condensed protocols from Molecular*  
680 *cloning : a laboratory manual*. Cold Spring Harbor Laboratory Press: Cold Spring Harbor,  
681 N.Y.  
682  
683 Sved JA, Feldman MW (1973). Correlation and probability methods for one and two loci.  
684 *Theor Popul Biol* **4**(1): 129-132.  
685  
686 Upadhyay MR, Chen W, Lenstra JA, Goderie CR, MacHugh DE, Park SD *et al* (2017). Genetic  
687 origin, admixture and population history of aurochs (*Bos primigenius*) and primitive European  
688 cattle. *Heredity (Edinb)* **118**(2): 169-176.  
689  
690 van Binsbergen R, Calus MP, Bink MC, van Eeuwijk FA, Schrooten C, Veerkamp RF (2015).  
691 Genomic prediction using imputed whole-genome sequence data in Holstein-Friesian cattle.  
692 *Genet Sel Evol* **47**: 71.

693  
694 Yudin NS, Vasil'eva LA, Belyavskaya VA, Aitnazarov RB, Smirnov PN, Heaton M *et al*  
695 (2015). Design of a DNA panel for genomic studies in Russian cattle breeds. *Russian Journal*  
696 *of Genetics: Applied Research* **5**(1): 21-25.  
697  
698 Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS (2012). A high-performance  
699 computing toolset for relatedness and principal component analysis of SNP data.  
700 *Bioinformatics* **28**(24): 3326-3328.  
701  
702 Zinovieva NA, Dotsev AV, Sermyagin AA, Wimmers K, Reyer H, Sölkner J *et al* (2016).  
703 Study of genetic diversity and population structure of five Russian cattle breeds using whole-  
704 genome SNP analysis. *Sel'skokhozyaistvennaya Biologiya* **51**(6): 788-800.  
705

706

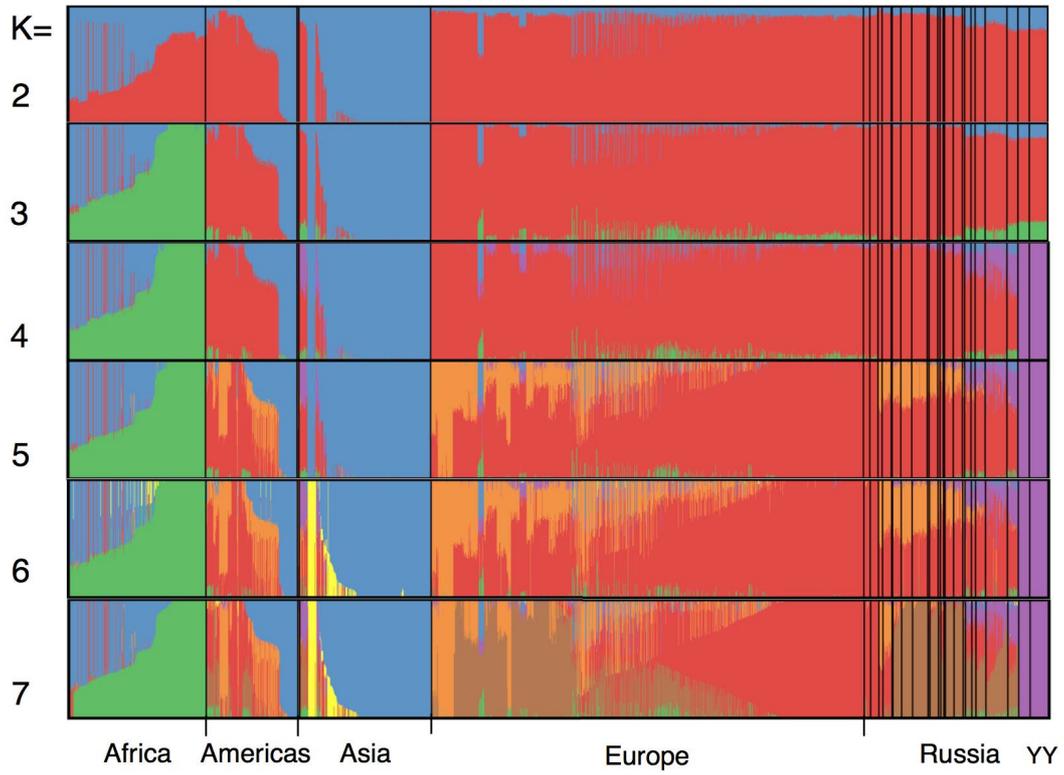
# FIGURES

707



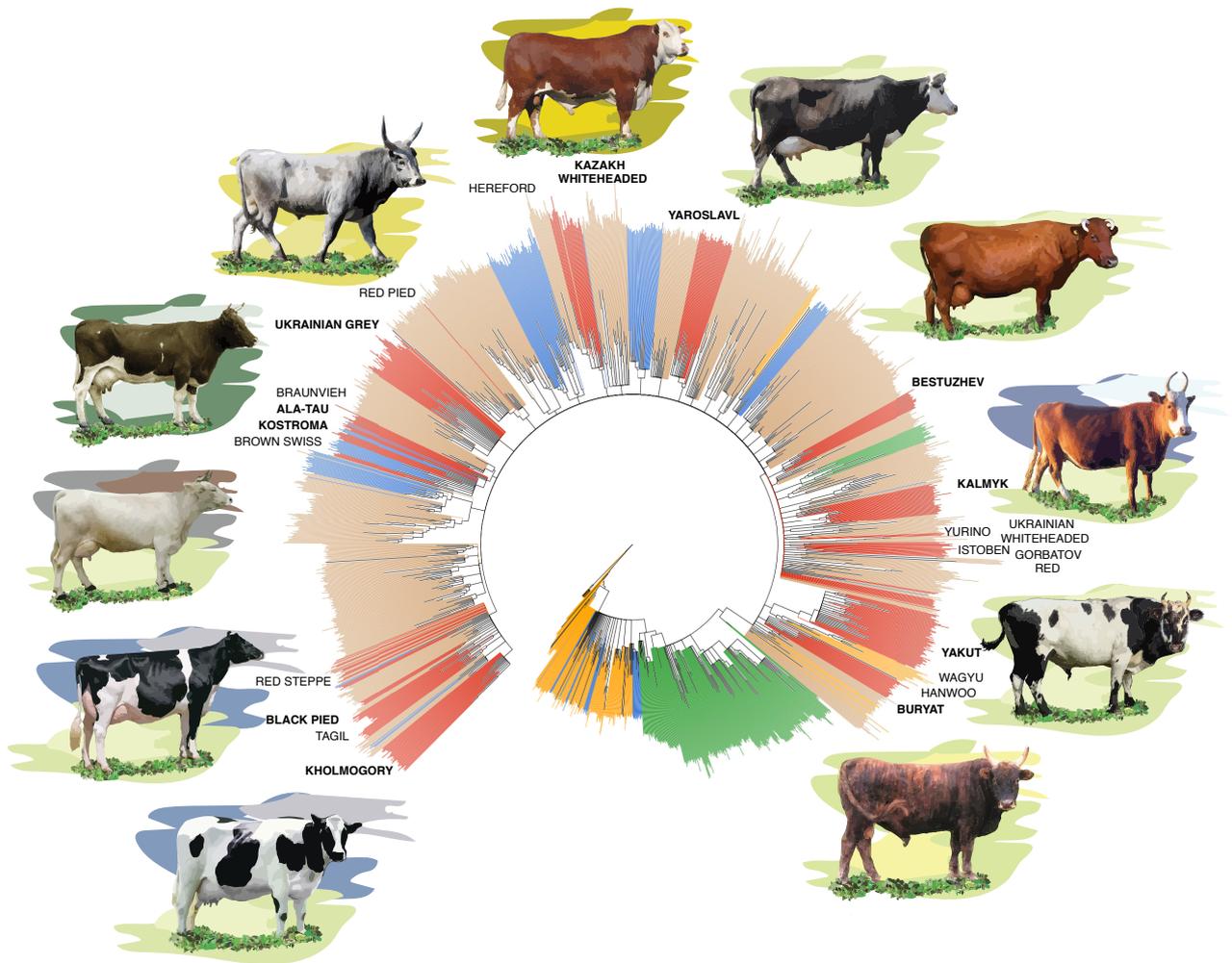
708

709 **Figure 1.**



710

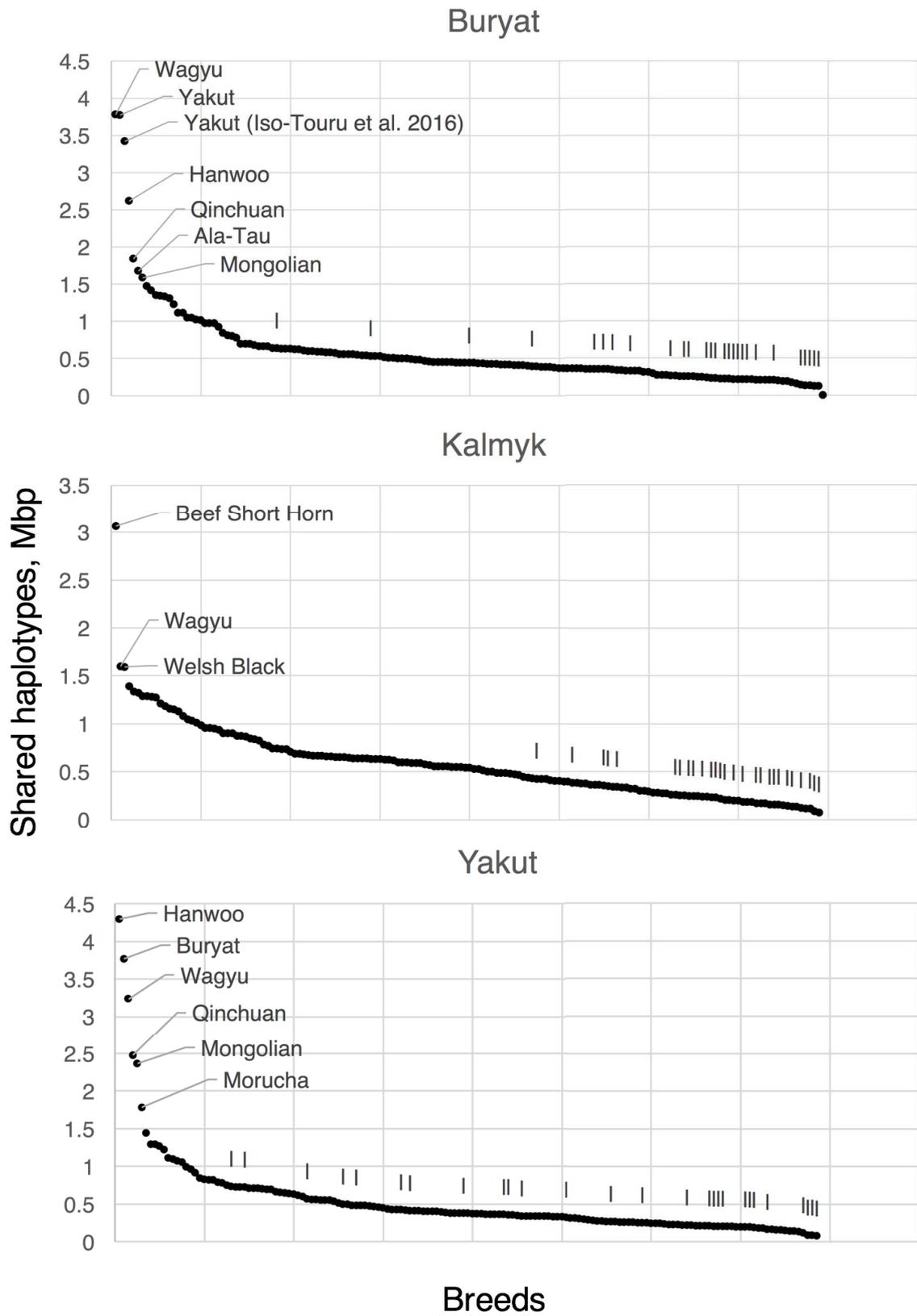
711 **Figure 2.**



712

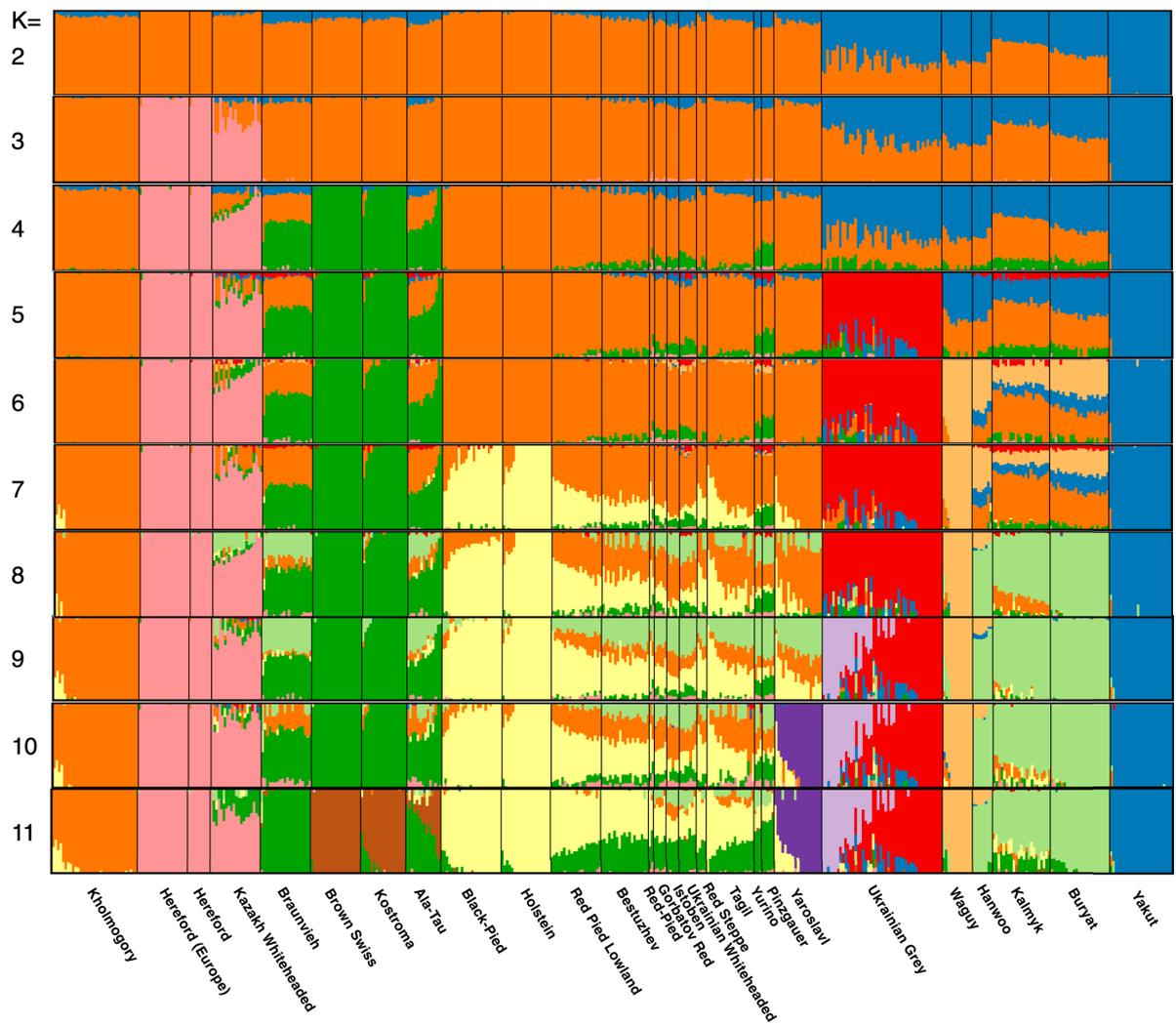
713 **Figure 3.**





748

749 **Figure 5.**



750  
751 **Figure 6.**

## TITLES AND LEGENDS TO FIGURES

752

753

754 **Figure 1.** LD decay plot of the mean  $r^2$  values for Russian breeds with >10 sampled  
755 individuals.

756 **Figure 2.** fastSTRUCTURE results for global cattle diversity and Russian breeds (YY – Yakut  
757 cattle).

758 **Figure 3.** Individual-based neighbor-joining tree of global cattle diversity and Russian breeds.  
759 The nodes with less than 70% bootstrap support were collapsed. Yellow – Asian cattle  
760 (predominately *B. indicus*), green – African cattle (predominantly, taurines) blue – American  
761 cattle, brown – European cattle, red – Russian cattle. The names of Russian and their sister  
762 breeds from other regions are shown. In bold are names of the Russian cattle breeds shown on  
763 the images.

764 **Figure 4.** Haplotype-sharing between the Russian and other taurine breeds for short (A, <3  
765 Mbp) and long (B, >7 Mbp) segments.

766 **Figure 5.** Haplotype sharing between the Turano-Mongolian and all other studied breeds for  
767 short segments (<3 Mbp). Vertical lines indicate positons of *B. indicus* breeds. Sharing with  
768 the Ukrainian Grey was removed from the Yakut breed pairwise comparison due to pronounced  
769 signature of a very recent introgression from the Yakut breed, not identified for the Ukrainian  
770 Gray samples collected in Serbia. Breed names are shown for the largest number of shared  
771 haplotypes (>1.5 Mbp).

772 **Figure 6.** fastSTRUCTURE results for Russian breeds and a set of closely related world  
773 breeds: Hereford, Braunvieh, Brown Swiss, Holstein, Red Pied Lowland, Pinzgauer, Waguy,  
774 Hanwoo.

775

## TABLES

776

777 **Table 1.** Single nucleotide polymorphism, diversity, inbreeding and effective population sizes

778 within the Russian cattle breeds

<b>Breed</b>	<b>No.<sup>a</sup></b>	<b><math>H_e</math></b>	<b><math>F</math></b>	<b><math>P_n</math></b>	<b><math>MAF</math></b>	<b>No. sampling locations</b>
Ala Tau	14	0.338	-0.014	0.937	0.258	2
Bestuzhev	19	0.346	-0.048	0.955	0.265	1
Black Pied	24	0.352	-0.026	0.977	0.269	1
Buryat	24	0.340	-0.026	0.952	0.260	1
Ukrainian Grey <sup>b</sup>	48	0.322	-0.034	0.935	0.244	1
Hereford	9	0.313	-0.070	0.867	0.238	1
Istoben	5	0.304	-0.165	0.812	0.232	1
Kalmyk	23	0.351	-0.027	0.972	0.269	2
Kazakh Whiteheaded	20	0.339	-0.038	0.948	0.259	2
Kholmogory	34	0.343	-0.011	0.965	0.262	2
Kostroma	18	0.315	-0.040	0.900	0.238	2
Gorbatov Red	5	0.322	-0.093	0.855	0.246	2
Red Pied	2	0.271	-0.331	0.650	0.218	1
Red Steppe	4	0.323	-0.148	0.842	0.249	2
Tagil	19	0.350	-0.060	0.967	0.268	1
Ukrainian Whiteheaded	7	0.329	-0.066	0.889	0.252	1
Yurino	3	0.292	-0.212	0.744	0.227	1
Yakut	25	0.273	-0.039	0.818	0.205	1
Yaroslavl	19	0.336	-0.047	0.949	0.256	2
<b>Total/Average</b>	322	0.324	-0.079	0.891	0.248	1.42

779 <sup>a</sup>No. - number of samples,  $H_e$  - expected heterozygosity,  $F$  - mean inbreeding coefficient,  $P_n$  -780 proportion of polymorphic loci,  $MAF$  - minor allele frequency.781 <sup>b</sup>genotypes obtained from (Boussaha *et al*, 2015)