SNP genotyping and population analysis of five indigenous Kazakh sheep breeds

Alexandr Pozharskiy^a, Aigerim Khamzina^a, Dilyara Gritsenko^a, Zhangylsyn Khamzina^a,
Shinara Kassymbekova^a, Nurlybek Karimov^a, Talgat Karymsakov^a, Nurlan Tlevlesov^a

^aKazakh Research Institute of Livestock and Fodder Production, 51, Zhandosov str., 050035, Almaty,
Republic of Kazakhstan

Corresponding author: Alexandr Pozharskiy aspozharsky@gmail.com

3

13 **Abstract**

14 Sheep husbandry is an important branch of agriculture in Kazakhstan. Modern agrarian and breeding
 15 science demands inclusion of molecular genetic and genomic data to supplement traditional methods.
 16 Here we used medium-scale SNP genotyping for the first time to determine the population structure of
 17 five local sheep breeds in Kazakhstan and their relation to global sheep diversity.
 18 Principal component analysis and model-based structure analysis of general population markers
 19 revealed two breed groups. The first group included Akzhayik and Kazak Fine-wool sheep and the
 20 second group had Edilbay, Saryarka and Kazakh Semi-coarse wool sheep. High heterogeneity of
 21 different populations of Akzhayik and Kazakh Semi-coarse wool sheep was observed. A neighbor-
 22 joining tree comparing Kazakh sheep data with the dataset generated by the Sheep HapMap project
 23 supported a close relationship between Kazakh sheep varieties and ancient domestic sheep ancestors.

24

25 **Keywords**

26 Kazakh sheep, Edilbay, SNP

27

28

5

29 **Introduction**

30 Domestic sheep (*Ovis aries* Linnaeus) are one of the most culturally and economically important farm
 31 animals. Sheep were domesticated between 11,000 and 9,000 B.C. in Central and Southwestern Asia
 32 (Tapio et al., 2010; Zeder, 2008). In Kazakhstan, sheep together with horses are important domestic
 33 animals and have a long history of traditional husbandry. According to archaeological data, domestic
 34 sheep were present in regions of Kazakhstan since the early bronze age (Frachetti and Benecke, 2009).
 35 Before the Soviet period, most Kazakh sheep varieties were the result of traditional folk breeding and
 36 selection. In the KazSSR, systematic work for the creation of new sheep breeds creation was carried
 37 out on state farms (Degen, 2013). Sheep husbandry is currently a key branch of the Kazakhstan
 38 economy and provides meat and wool for both domestic use and for export. Modern agriculture
 39 practices involve application of molecular genetic methods as a valuable tool to predict and select
 40 desirable traits and properties. The development of high throughput single-nucleotide polymorphism
 41 (SNP) genotyping technologies has contributed to the growing impact of genome-wide association
 42 studies (GWAS) in animal breeding, genomic prediction and animal selection (Gholizadeh et al.,
 43 2015).
 44 SNPs are among the most common markers in the study of genetic processes. The advantages of SNPs
 45 are the high number of markers that can be processed simultaneously, the ability to cover both coding
 46 and non-coding regions, flexibility in genotyping of non-model organisms and low error rates across a
 47 range of DNA sample quality (Haynes and Latch, 2012). SNP genotyping is a powerful tool for
 48 marker-assisted selection in animal breeding (Seidel, 2010). With increased demand for meat and wool
 49 production by the sheep industry, breeders and geneticists are considering the genetic basis for
 50 interactions between genotype and environmental factors in order to maintain and improve the
 51 distinctive features and biological characteristics of various breeds (Zhang et al., 2013). One approach

7
52 to improving breed features is identification of new gene pools that have potentially beneficial
53 genotypes that can be introduced into the breeding process. From this perspective, indigenous sheep
54 breeds in Kazakhstan are of particular interest, because, although they are well-established, few
55 genomic studies have focused on these breeds.

56 In the present study we considered five sheep breeds that are widely used in Kazakhstan: Edilbay
57 (Edilbayevskaya), Kazakh Fine-wool (Kazakh Fine-fleece), Akzhayik (Akzhayikskaya), Kazakh Semi-
58 coarse wool sheep and Saryarka (Saryarkinskaya). The Edilbay breed is traditionally considered as a
59 descendant of old sheep varieties in Kazakhstan. This breed was created at the end of the 19th century
60 in the Ural region and is thought to have originated from breeding of Astrakhan sheep with local
61 varieties that were naturally selected or the products of folk selection methods, although no
62 documentary data concerning how this breed was established are available. Edilbay sheep are well
63 adapted to the conditions of dry steppes, deserts, and graze nearly year-round (Ermekov and Golodnov,
64 1976). This breed is of special interest because it played a central role in creation of most local sheep
65 breeds in Kazakhstan. Moreover, mitochondrial DNA analysis supports the genetic proximity of
66 Edilbay to the hypothetical ancestor of modern sheep varieties (Hiendleder et al., 1998), although it has
67 not been compared further with other sheep breeds that are common worldwide.

68 On September 28, 1945, the board of the USSR approved the Kazakh fine-wool breed of sheep,
69 generated by V.A. Balmont, G.V. Bakanova and A.P. Pshenichny, as a new independent breed. Work
70 on the breeding of this variety of fine-fleece sheep was begun by V.A. Balmont in regions of
71 southeastern Kazakhstan where Kazakh fat tail queens were crossed with Prekos fine-wool sheep in the
72 autumn of 1931. Animals resulting from this cross were characterized by robust growth, strong
73 constitution and good adaptation to year-round pasture grazing (Balmont, 1957).

74 The Akzhayk meat-wool breed was bred in East Kazakhstan in 1996 using complex reproductive

9
75 crossbreeding of fine-fleece and semi-fine-coarse-haired sheep with rams of the Lincoln and Romney
76 march type and subsequent breeding of animals of the desired type “inside”. For many years, scientific
77 guidance on selection and breeding of the Akzhayk breed was provided by V.V. Terentyev (Terentyev,
78 1987; Traisov et al., 2013).

79 The Saryarka fat tail breed by M.A. Ermekov et al. was approved in 1999. This breed of sheep was
80 created over the period between 1970 and 1998 mainly through in-breeding selection and selection
81 according to breeding traits of the improved Kazakh fat tail sheep with Edilbaev sheep as well as the
82 partial blood line of the Kargaly semi-coarse sheep with subsequent inbreeding of animals of the
83 desired type. These sheep are characterized by a strong constitution, a well-developed skeleton, regular
84 physique, and strong limbs with a dense ungulate horn, which is important for year round pasture
85 grazing. These sheep are found in Central Kazakhstan (Akhatov, 2006; Kanapin et al., 2001; Kanapin
86 and Akhatov, 2000, 1991).

87 Creation of the Kazakh fat tail semi-coarse breed was begun by V.A. Balmont using complex
88 reproductive breeding of local Kazakh coarse-haired sheep with Edilbay, Saraja, Degeress and Tadzhik
89 breeds (Balmont, 1968).

90 Here we present the results of a pilot research project for the introduction of massive SNP genotyping
91 of sheep raised in Kazakhstan. We applied Illumina Beadchip technology for the first time to perform
92 SNP genotyping of local Kazakh sheep breeds. We investigated the genetic diversity and population
93 structure of five local Kazakh sheep breeds (Edilbay, Akzhayik, Saryarka, Kazakh fine-wool, Kazakh
94 semi-coarse wool) using the Illumina OvineSNP50 panel. Furthermore, we compared these data with
95 the Sheep HapMap project dataset compiled by the International Sheep Genome Consortium to
96 compare variability in populations of local breeds in Kazakhstan with global domestic sheep diversity.
97 This work will lay a foundation for the introduction of genomic methods to sheep husbandry in

11
98 Kazakhstan and can be a basis that can be extended to other branches of animal and plant agriculture in
99 Kazakhstan.

100 **Materials and methods**

101 *Collection of samples and DNA extraction*

102 Samples from Edilbay sheep were collected on the “Birlik” (n=794) and “Azhar” (n=470) farms in
103 West Kazakhstan. Samples from Akzhayik sheep were collected on the “Atameken” (n=1110),
104 “Kuanysht” farm (n=500), and “Saltanat” farms (n=410), also in West Kazakhstan. Samples from
105 Kazakh fine wool breed were collected in the Almaty region (624 samples - "PZ" R-Kurty"). For
106 Kazakh fat-tailed semi-coarse-wool sheep, samples were taken on the Altyn Asel farm (n=270) in the
107 Aktobe region, the Otkanzhar farm (n=531) in the Karaganda region and the Khasiev (n=452) and
108 KaraAdyr farms (n=502) in East Kazakhstan. Samples for Saryarka sheep were collected on the Sarysu
109 farm (n=498) in the Karaganda region.

110 Plucks from sheep ears were collected for DNA extraction in Eppendorf tubes, fixed with ethanol and
111 stored at -20 °C. Before processing, the samples were washed twice with phosphate buffered saline to
112 remove fixatives. DNA was extracted using MasterPure Complete DNA & RNA Purification Kits
113 (Lucigen Simplifying Genomics, USA) according to the manufacturer’s protocol and diluted to 50
114 ng/μL.

115 *SNP genotyping and quality control*

116 SNP genotyping was conducted using an Illumina iScan System with ovine SNP50 BeadChips
117 according to the manufacturer’s protocol. HiScan raw intensity data (IDAT) were loaded into the
118 Genotyping module of GenomeStudio v.2.0.4 software (Illumina, 2016) for genotype calling and
119 quality control. Genotypes were inferred using manifest and cluster definition files accessed from the
120 manufacturer’s website (accession date 08.07.2019). After initial SNP clustering and statistical

13
 121 evaluation, samples having a median GC score and call rate of < 0.9 were excluded, and the SNP
 122 statistics was recalculated. All data that remained after filtering (i.e., 50% GC ≥ 0.9 and call rate ≥ 0.9
 123 for samples and SNP) were exported in a PLINK input format (ped+map) using the PLINK Input
 124 Report Plug-in v2.1.4 for GenomeStudio. SNP having minor allele frequency (MAF) < 0.02 and
 125 deviations from Hardy-Weinberg equilibrium at a critical p -value level of 0.0001 were filtered out by
 126 PLINK 1.9 (Purcell and Chang, 2015).

127 *Data analysis*

128 General population statistics including expected and observed heterozygosity, Wright's fixation index
 129 (F_{st}) and linkage disequilibrium test were calculated by with PLINK1.9 and then processed with R (R
 130 Core Team, 2019). Principal component analysis (PCA) was performed using the adegenet package in
 131 R (Jombart, 2008; Jombart and Ahmed, 2011). Model-based analysis of population structure was
 132 conducted using two independent algorithms implemented in ADMIXTURE (Alexander et al., 2009;
 133 Alexander and Lange, 2011) and FastStructure (Raj et al., 2014) software. Both programs were run for
 134 K 's from 2 to 20 with 10 repetitions for cross validation. FastStructure was run in simple prior mode.
 135 Results were visualized with CLUMPAK (Kopelman et al., 2015). To determine the relationship of
 136 Kazakh breeds with other sheep breeds across the world, diversity data were compared with the Sheep
 137 HapMap dataset (SHM) compiled by the International Sheep Genome Consortium (Kijas et al., 2012,
 138 2009). The SHM dataset was retrieved upon request from <https://www.sheephapmap.org/> in accordance
 139 with the ISGC Terms of Access. A subset of Sheep HapMap data that included only those breeds for
 140 which there were at least 20 specimens and available SNPs in the processed Kazakh sheep dataset was
 141 selected, and then SNP physical mapping was updated according to the Illumina OvineSNP50 manifest
 142 file used for genotyping. A total of 50 randomly selected specimens from each Kazakh sheep
 143 population were merged with HapMap data using PLINK. The Ape (Analysis of Phylogenetics and

15
 144 evolution) package for R (Paradis et al., 2004) was used to calculate Neighbor-joining tree of distances,
 145 which was visualized further by FigTree (Rambaut, 2018). R and bash scripts describing conducted
 146 analyses are available in the GitHub repository [https://github.com/ASPOzharsky/kz_ovine_SNP].

147 **Results**

148 Out 6,756 samples and 54,241 SNPs, 4,660 and 30,597, respectively, satisfied the selection criteria. An
 149 additional 4,063 SNPs that had low MAF and/or deviated from HWE ($p < 0.0001$), as well as markers
 150 having an undetermined genomic position and non-autosomal markers were excluded such that 26,534
 151 SNPs were considered for further analysis. Percent of individuals passed quality criteria differs
 152 depending on population (see Table S1 in Supplementary materials). The best sample quality was from
 153 Birlik farm (Edilbay), Kara Adyr (Kazakh Semi-coarse) and Kuanysh farms (Akzhayik). The most
 154 individuals were discarded from the Otkanzhar and Altyn Asel farms (Kazakh Semi-Coarse). The low
 155 quality of these two populations indicates serious violation of storage and transport of animal material
 156 conditions by corresponding farms. However, 59 and 78 samples, respectively, were selected as
 157 suitable for further analysis.

158 First, general genetic diversity statistics including H_e , H_o , and F_{st} were determined for the different
 159 breeds (Table 1). All populations had similar H_o values (average 0.385), which did not substantially
 160 differ from H_e . The lowest within-breed F_{st} was between the two Edilbay populations (0.0052).
 161 Pairwise F_{st} across populations of Kazakh Semi-coarse wool sheep varied between 0.0095 (Altyn, Asel
 162 and Otkanzhar) and 0.0193 (Kara, Adyr and Otkanzhar). F_{st} between breeds did vary and allowed the
 163 population to be separated into two groups that had lower (Edilbay, Saryarka, Kazakh Semi-coarse
 164 wool) and higher (Akzhayik, Kazakh Fine-wool) inter-breed values (Fig. 1a). Separate F_{st} values by
 165 chromosome showed a similar distribution, except for chromosome 16 (Fig. 1b-d, Table S2 in
 166 Supplementary materials).

17
167 Linkage disequilibrium decay was evaluated separately for each breed and each population and
168 expressed as an average R^2 coefficient for 20 kb intervals (Fig. 2). Akzhayik, Edilbay and Kazakh
169 Semi-coarse wool sheep had similar LD decay rates that were close to that for the entire sample. For
170 Edilbay, Akzhayik and Semi-coarse wool breeds, LD decay can be compared between populations.
171 Two Edilbay populations had very similar decay rates. Among Akzhayik sheep, the Saltanat population
172 showed different LD decay from Kuanysh and Atameken. All populations of the semi-coarse wool
173 breed had different LD decay rates.

174 Principal component analysis was next conducted and the first 100 components were retained. All
175 components explained a relatively small part of the total variance (Table S3; Fig. 3B; Fig. S1). The first
176 three PCs explained 15%, 4% and 3.6% of variance, respectively (Fig. 3c-e). The first PC revealed two
177 major clusters that corresponded to the abovementioned groups formed according to F_{st} . Several
178 components had smaller differences between breeds and populations within these two groups (Fig. 3a).

179 ADMIXTURE and FastStructure were run to determine the genetic structure within and between
180 populations and breeds. The cross validation error score did not converge to its minimal value in all
181 runs of K values up to 20 for both algorithms. Thus, the following Ks were selected for examination: 2
182 (minimum), 5 (number of Kazakh breeds), 11 (number of Kazakh breed populations) and 20
183 (maximum) (Fig. 4; see also Figures S2 and S3 in Supplementary materials for all original plots). When
184 $K=2$, both algorithms produced essentially identical results. As with diversity features, these
185 calculations allowed all breeds to be differentiated into two clusters: one having Edilbay, Saryarka and
186 Kazakh Semi-coarse wool sheep, and the second having Akzhayik and Kazakh Fine-wool sheep.

187 Additional K results differed between the two programs. At $K=5$, FastStructure results showed more
188 mixed populations than ADMIXTURE and found effectively no within-population genetic structure
189 with further models. The general population structure revealed by ADMIXTURE was more stable.

19
 190 Populations of sheep that were the same breed but came from different farms showed high
 191 heterogeneity both within and between populations. Akzhayik and Kazakh Semi-coarse breeds had a
 192 highly mixed gene pool in every sampled population. Meanwhile, the Edilbay population from Azhar
 193 was more mixed compared to the Birlik population.

194 To compare diversity of Kazakh sheep breeds to breeds worldwide, we built a genetic distance tree
 195 using the Neighbor-joining method. All Kazakh breeds are shown as a distinct cluster within the entire
 196 tree (Fig. 5). Edilbay is a heterogenous basic group, whereas other breeds appear subsequently as
 197 separate branches. The breed clusters diverged from Edilbay in the following order: Saryarka and
 198 Kazakh Semi-coarse sheep, Akzhayik and Kazakh Fine-wool sheep, followed by additional Asian and
 199 South-West Asian breeds.

200 **Discussion**

201 Comparison of genetic structures between the studied breeds was consistent with the known history of
 202 the respective breeding processes. Edilbay and its descendants, Saryarka and Kazakh Semi-coarse wool
 203 sheep, were shown to be distinct from Akzhayik and Kazakh Fine-wool sheep. On the other hand,
 204 variations in breeds and populations within these groups were less clear. PCA results demonstrated that
 205 the most discriminating first component could explain only 15% of the total variation, whereas other
 206 components described differences between breeds and population even less well. All remaining
 207 components tended asymptotically to the level of 0.5% of explained variance. However, these values
 208 are relatively high with respect to world-wide diversity data (Kijas et al., 2009), which are in agreement
 209 with the weak population structure reported for mitochondrial and genomic SNP data (Kijas et al.,
 210 2012; Meadows et al., 2005). As the ADMIXTURE and FastStructure analysis results showed, all
 211 populations in this study tended to separate from each other. However, neither method were not able to
 212 reached an optimal configuration with the lowest cross-validation error in runs having K=20. Although

21
213 FastStructure was shown more efficient in resolving population structure (Raj et al., 2014), with this
214 method noise increases rapidly with increasing K. Thus, FastStructure appears to be too sensitive to
215 variation, with individual variability overwhelming population integrity. Meanwhile, the
216 ADMIXTURE results were useful for further observations and conclusions, although in the presence of
217 population variations the genetic structure was less clear with increasing individual values for K. The
218 Kazakh Fine-Wool sheep population and Birlik population of Edilbay sheep retained high homogeneity
219 even at maximal K. Akzhayik and Semi-coarse breeds had high heterogeneity between populations
220 from different farms, which was also supported by varying LD decay rates. One noticeable observation
221 for the Semi-coarse breed populations is that the subset of SNPs on chromosome 16 had a differing
222 pattern for between-population F_{st} values (Fig. 1). Populations from Khasiev and Kara Adyr farms
223 differed more not only from other breeds, but also from the Otkanzhar and Altyn Asel populations of
224 the same breed. This finding might imply that all or a significant part of chromosome 16 was
225 introduced from the genetic pool during the breeding process, and lies outside of relationships with
226 other studied breeds. The low number of SNPs included in this study did not allow us to determine the
227 exact origin of chromosome 16, so additional investigations are needed to clarify this finding. Edilbay
228 populations from Birlik and Azhar farms had the lowest F_{st} value across all population pairs and
229 practically identical LD decay rates. The Birlik population was the most homogeneous group, which
230 was supported by PCA. The Azhar population is more mixed, but still expressed high similarity to
231 Birlik. We had only one population for both Saryarka and Kazakh Fine-wool breeds and thus could
232 draw no conclusions about within-breed populational variability. However, when high K_s (near 20)
233 were tested, the Saryarka could be divided into two sub-clusters. This separation was also visible in
234 PCA results, starting from PC 13, and with particular impact of PC 17 (Fig. 3a, S3). Each component
235 in this range was responsible for no more than 1% of all variation, so ADMIXTURE demonstrates high

23
236 sensitivity even to small variability components.

237 The neighbor-joining tree build with the addition of SHM project data allowed data for Kazakh sheep
238 varieties to be placed in the context of global variations in sheep breeds. Considering the hypothesis
239 that domesticated sheep have an Indo-Iranian origin, distribution of all sheep breeds should follow the
240 historical sequence from origins in Asia with expansion into the Mediterranean region, Europe and then
241 Africa. In such a context, the basic position of Edilbay sheep implies that this traditional lineage
242 conserves the genetic pool of hypothetical ancestral domestic sheep varieties. This possibility is
243 supported by previous studies on mitochondrial genetics (Hiendleder et al., 1998). However, in general
244 there is a lack of detailed comparative genetic studies on Kazakh sheep breeds and Edilbay particularly.
245 On the one hand, studies including Kazakh sheep varieties usually limited by a local context
246 (Blackburn et al., 2011; Dossybayev et al., 2018; Mukhametzharova et al., 2018; Ozerov et al., 2008a,
247 2008b). On the other hand, in global between breed diversity studies Kazakh breeds are not being
248 included (Kijas et al., 2012; Meadows et al., 2007, 2005) or have low representation (Hiendleder et al.,
249 2002; Meadows et al., 2011) in global between-breed diversity studies. As such, further in-depth
250 investigation of the importance of Kazakh sheep varieties, particularly for the Edilbay breed, in global
251 sheep diversity is important for understanding the history of sheep domestication. Additional
252 quantitative work involving a larger set of genome-wide markers and qualitative studies such as those
253 that determine haplotypes using reliable mitochondrial and Y-chromosomal markers is also needed.
254 Results from such studies will have practical importance for exploration of Kazakh sheep breeds as a
255 previously understudied source of genetic material for selection, and also for understanding the genetic
256 basis for standardization of Kazakh sheep varieties.

257

258

25 259 **Conclusion**

260 Despite their unique proximity to the geographical and historical origin of domesticated sheep, Kazakh
261 sheep breeds remain yet poorly studied in terms of genetic composition and its relationship to world-
262 wide sheep diversity. Here we used medium-scale SNP genotyping of several sheep breeds in
263 Kazakhstan to lay a basis for further molecular genetic studies. Investigation of the previously
264 uncharacterized genetic pool of Kazakh sheep breeds can have implications not only for breeding
265 practice, but for our fundamental understanding of sheep genetics and domestication history. Our data
266 support that Kazakh breeds, particularly Edilbay, are direct descendants of historical domestic sheep
267 ancestors.

268 The observed heterogeneity in sheep populations from different farms that were sampled in this study
269 raises the issue of standardization among Kazakh sheep breeds. Given that genetic consistency is
270 crucial for successful breeding and selection of sheep and other livestock, effort is needed to compile a
271 detailed genetic description and “passport” for sheep populations in Kazakhstan.

272 273 **Ethical statement**

274 Animal material was collected and provided by qualified personnel of the respective farms as part of
275 routine veterinary care in accordance with local regulations. Collection of samples caused no harm to
276 animal health.

277 **Conflict of interests**

278 The authors have no competing interests to declare.

279 **Acknowledgements**

280 The study was funded by the Ministry of Agriculture of Kazakhstan as a part of a targeted funding
281 program № 267 “Development of efficient breeding methods in branches of animal husbandry”

27
282 The ovine SNP50 HapMap dataset used for the analysis described was provided by the International
283 Sheep Genomics Consortium and obtained from www.sheephapmap.org in agreement with the ISGC
284 Terms of Access.

285 **References**

- 286 Akhatov, A., 2006. Selection of fat tailed coarse wool sheep breed of Saryarka (Sarysu type). Anim.
287 husbandry, forage Prod. Vet. 1, 34–35. [in Russian; original title: Selekcija saryarkinskoj
288 kurdyuchnoj grubosherstnoj porody ovec (sarysujskij tip)]
- 289 Alexander, D.H., Lange, K., 2011. Enhancements to the ADMIXTURE algorithm for individual
290 ancestry estimation. BMC Bioinformatics 12, 246. <https://doi.org/10.1186/1471-2105-12-246>
- 291 Alexander, D.H., Novembre, J., Lange, K., 2009. Fast model-based estimation of ancestry in unrelated
292 individuals. Genome Res. 19, 1655–1664. <https://doi.org/10.1101/gr.094052.109>
- 293 Balmont, V.A., 1968. About application of heterosis in sheep husbandry. Kolos. [in Russian; original
294 title: Ob ispol'zovanii geterozisa v ovcevodstve]
- 295 Balmont, V.A., 1957. Improvement of Kazakh fine-wool sheep breed on experimental farm named
296 after Mynabyev. Proc. Inst. Anim. Husb. Kazakh branch VASKhNIL 4. [in Russian; original title:
297 Sovershenstvovanie kazahskoj tonkorunnoj porody ovec na eksperimental'noj baze im.
298 Mynbaeva]
- 299 Blackburn, H.D., Toishibekov, Y., Toishibekov, M., Welsh, C.S., Spiller, S.F., Brown, M., Paiva, S.R.,
300 2011. Genetic diversity of Ovis aries populations near domestication centers and in the New
301 World. Genetica 139, 1169–1178. <https://doi.org/10.1007/s10709-011-9619-4>
- 302 Degen, A.A., 2013. Karakul sheep production in Kazakhstan: an efficient collective enterprise under
303 the state farm (sovkhoz) system and its collapse with the break-up of the Soviet Union. World
304 Rev. Entrep. Manag. Sustain. Dev. 9, 1. <https://doi.org/10.1504/WREMSD.2013.050932>
- 305 Dossybayev, K., Mussayeva, A., Bekmanov, B., Kulataev, B., 2018. Analysis of Genetic Diversity in
306 three Kazakh Sheep using 12 Microsatellites. Int. J. Eng. Technol. 7, 122.
307 <https://doi.org/10.14419/ijet.v7i4.38.24336>
- 308 Ermekov, M.A., Golodnov, A. V, 1976. Fat tailed sheep of Kazakhstan. Kaynar, Almaty. [in Russian;

29
309 original title: Kurdyuchnye ovcy Kazahstana]

310 Frachetti, M., Benecke, N., 2009. From sheep to (some) horses: 4500 years of herd structure at the
311 pastoralist settlement of Begash (south-eastern Kazakhstan). *Antiquity* 83, 1023–1037.
312 <https://doi.org/10.1017/S0003598X00099324>

313 Gholizadeh, M., Rahimi-Mianji, G., Nejati-Javaremi, A., 2015. Genomewide association study of body
314 weight traits in Baluchi sheep. *J. Genet.* 94, 143–146. <https://doi.org/10.1007/s12041-015-0469-1>

315 Haynes, G.D., Latch, E.K., 2012. Identification of Novel Single Nucleotide Polymorphisms (SNPs) in
316 Deer (*Odocoileus* spp.) Using the BovineSNP50 BeadChip. *PLoS One* 7, e36536.
317 <https://doi.org/10.1371/journal.pone.0036536>

318 Hiendleder, S., Kaupe, B., Wassmuth, R., Janke, A., 2002. Molecular analysis of wild and domestic
319 sheep questions current nomenclature and provides evidence for domestication from two different
320 subspecies. *Proc. R. Soc. London. Ser. B Biol. Sci.* 269, 893–904.
321 <https://doi.org/10.1098/rspb.2002.1975>

322 Hiendleder, S., Mainz, K., Plante, Y., Lewalski, H., 1998. Analysis of mitochondrial DNA indicates
323 that domestic sheep are derived from two different ancestral maternal sources: no evidence for
324 contributions from urial and argali sheep. *J. Hered.* 89, 113–120.
325 <https://doi.org/10.1093/jhered/89.2.113>

326 [software] Illumina, 2016. GenomeStudio Genotyping Module. [https://emea.illumina.com/techniques/](https://emea.illumina.com/techniques/microarrays/array-data-analysis-experimental-design/genomestudio.html)
327 [microarrays/array-data-analysis-experimental-design/genomestudio.html](https://emea.illumina.com/techniques/microarrays/array-data-analysis-experimental-design/genomestudio.html)

328 Jombart, T., 2008. adegenet: a R package for the multivariate analysis of genetic markers.
329 *Bioinformatics* 24, 1403–1405. <https://doi.org/10.1093/bioinformatics/btn129>

330 Jombart, T., Ahmed, I., 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data.
331 *Bioinformatics* 27, 3070–3071. <https://doi.org/10.1093/bioinformatics/btr521>

332 Kanapin, K., Akhatov, A., 2000. Fat tailed coarse wool sheep of Kazakhstan. Evero, Almaty. [in
333 Russian; original title: Kurdyuchnye grubosherstnye ovcy Kazahstana]

334 Kanapin, K., Akhatov, A., 1991. Methods of development of new type of Kazah fat tailed sheep with
335 lightened wool. *J. Agric. Sci. Kazakhstan* 4, 50–53. [in Russian; original title: Methods of

31
336 development of new type of Kazah fat tailed sheep with lightened wool]

337 Kanapin, K., Esentaev, E., Akhatov, A., 2001. Saryarkinskaya kurdyuchnaya grubosherstnaya poroda
338 ovec. Bastau, Almaty, pp. 133–136. [in Russian; original title: Kurdyuchnye ovcy Kazahstana]

339 Kijas, J.W., Lenstra, J.A., Hayes, B., Boitard, S., Neto, L.R.P., Cristobal, M.S., Servin, B., McCulloch,
340 R., Whan, V., Gietzen, K., Paiva, S., Barendse, W., Ciani, E., Raadsma, H., McEwan, J.,
341 Dalrymple, B., Consortium, other members of the I.S.G., 2012. Genome-Wide Analysis of the
342 World's Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection.
343 PLOS Biol. 10, e1001258. <https://doi.org/10.1371/journal.pbio.1001258>

344 Kijas, J.W., Townley, D., Dalrymple, B.P., Heaton, M.P., Maddox, J.F., McGrath, A., Wilson, P.,
345 Ingersoll, R.G., McCulloch, R., McWilliam, S., Tang, D., McEwan, J., Cockett, N., Oddy, V.H.,
346 Nicholas, F.W., Raadsma, H., Consortium, for the I.S.G., 2009. A Genome Wide Survey of SNP
347 Variation Reveals the Genetic Structure of Sheep Breeds. PLoS One 4, e4668.
348 <https://doi.org/10.1371/journal.pone.0004668>

349 Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A., Mayrose, I., 2015. Clumpak: a program
350 for identifying clustering modes and packaging population structure inferences across K. Mol.
351 Ecol. Resour. 15, 1179–1191. <https://doi.org/10.1111/1755-0998.12387>

352 Meadows, J.R.S., Cemal, I., Karaca, O., Gootwine, E., Kijas, J.W., 2007. Five Ovine Mitochondrial
353 Lineages Identified From Sheep Breeds of the Near East. Genetics 175, 1371–1379.
354 <https://doi.org/10.1534/genetics.106.068353>

355 Meadows, J.R.S., Hiendleder, S., Kijas, J.W., 2011. Haplogroup relationships between domestic and
356 wild sheep resolved using a mitogenome panel. Heredity (Edinb). 106, 700–706.
357 <https://doi.org/10.1038/hdy.2010.122>

358 Meadows, J.R.S., Li, K., Kantanen, J., Tapio, M., Sipos, W., Pardeshi, V., Gupta, V., Calvo, J.H.,
359 Whan, V., Norris, B., Kijas, J.W., 2005. Mitochondrial Sequence Reveals High Levels of Gene
360 Flow Between Breeds of Domestic Sheep from Asia and Europe. J. Hered. 96, 494–501.
361 <https://doi.org/10.1093/jhered/esi100>

362 Mukhametzharova, I., Islamov, Y., Shauyenov, S., Ibrayev, D., Atavliyeva, S., Tarlykov, P., 2018.
363 Genetic Characterization of Kazakh Native Sheep Breeds Using Mitochondrial DNA. Online J.

33
364 Biol. Sci. 18, 341–348. <https://doi.org/10.3844/ojbsci.2018.341.348>

365 Ozerov, M.Y., Marzanov, N.S., Tapio, M., Burabaev, A.A., Marzanova, L.K., Amerkhanov, K.A.,
366 Kantanen, Y., 2008a. Genetic features of Kazakh sheep breeds according to microsatellites. Russ.
367 Agric. Sci. 34, 45–48. <https://doi.org/10.3103/S1068367408010175>

368 Ozerov, M.Y., Marzanov, N.S., Tapio, M., Feizullaev, F.R., Burabaev, A.A., Amerkhanov, K.A.,
369 Petrov, S.N., Marzanova, L.K., Gostishchev, S.A., Kantanen, J., 2008b. Microsatellite
370 characterization of closely related fine-wool sheep breeds. Russ. Agric. Sci. 34, 343–347.
371 <https://doi.org/10.3103/S1068367408050182>

372 Paradis, E., Claude, J., Strimmer, K., 2004. APE: Analyses of Phylogenetics and Evolution in R
373 language. Bioinformatics 20, 289–290. <https://doi.org/10.1093/bioinformatics/btg412>

374 [software] Purcell, S., Chang, C., 2015. PLINK. <https://www.cog-genomics.org/plink/1.9/>

375 [software] R Core Team, 2019. R: A Language and Environment for Statistical Computing.
376 <https://www.r-project.org/>

377 Raj, A., Stephens, M., Pritchard, J.K., 2014. fastSTRUCTURE: Variational Inference of Population
378 Structure in Large SNP Data Sets. Genetics 197, 573–589.
379 <https://doi.org/10.1534/genetics.114.164350>

380 [software] Rambaut, A., 2018. FigTree. <http://github.com/rambaut/figtree/>

381 Seidel, G.E., 2010. Brief introduction to whole-genome selection in cattle using single nucleotide
382 polymorphisms. Reprod. Fertil. Dev. 22, 138–144. <https://doi.org/10.1071/RD09220>

383 Tapio, M., Ozerov, M., Tapio, I., Toro, M.A., Marzanov, N., Činkulov, M., Goncharenko, G.,
384 Kiselyova, T., Murawski, M., Kantanen, J., 2010. Microsatellite-based genetic diversity and
385 population structure of domestic sheep in northern Eurasia. BMC Genet. 11, 76.
386 <https://doi.org/10.1186/1471-2156-11-76>

387 Terentyev, V. V., 1987. Development of new type of meat and wool crossbreed sheep in Western
388 Kazakhstan, in: Development of New Breeds of Domestic Animals in Agriculture.
389 Agropromizdat, Moscow, pp. 199–203. [in Russian; original title: Sozdanie novogo tipa myaso-
390 shyorstnyh krossbrednyh ovec v Zapadnom Kazahstane]

35
391 Traisov, B.B., Bozymov, K.K., Esengaliyev, K.G., 2013. Development of sheep husbandry in Western
392 Kazakhstan. Sheep, goats, wool Ind. 2, 91–94. [in Russian; original title: Razvitie ovcevodstva v
393 Zapadnom Kazahstane. Sheep, goats, wool industry]
394 Zeder, M.A., 2008. Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion,
395 and impact. Proc. Natl. Acad. Sci. 105, 11597–11604. <https://doi.org/10.1073/pnas.0801317105>
396 Zhang, L., Liu, J., Zhao, F., Ren, H., Xu, L., Lu, J., Zhang, S., Zhang, X., Wei, C., Lu, G., Zheng, Y.,
397 Du, L., 2013. Genome-Wide Association Studies for Growth and Meat Production Traits in Sheep.
398 PLoS One 8, e66569. <https://doi.org/10.1371/journal.pone.0066569>
399
400
401

402 Table 1. Expected and observed heterozygosity and Write's Fst for sheep population from different farms in Kazakhstan

Breed	Population	He	Ho	Pairwise Fst										Sarysu
				Atameke n	Kuanys h	Saltan at	Birlik	Azhar	Rkurty	Altyn Asel	Kara Adyr	Khasie v	Otkanzh ar	
Akzhayik	Atameken	0.3835	0.3888		0.0103	0.0108	0.0363	0.0349	0.0184	0.0335	0.0373	0.0349	0.0346	0.0381
	Kuanys	0.3831	0.3892	0.0103		0.0134	0.0365	0.0348	0.0187	0.0329	0.0375	0.0346	0.0340	0.0379
	Saltanat	0.3865	0.3915	0.0108	0.0134		0.0301	0.0278	0.0111	0.0252	0.0315	0.0280	0.0261	0.0309
Edilbay	Birlik	0.3727	0.3725	0.0363	0.0365	0.0301		0.0052	0.0345	0.0109	0.0186	0.0155	0.0161	0.0139
	Azhar	0.3742	0.3808	0.0349	0.0348	0.0278	0.0052		0.0322	0.0112	0.0193	0.0161	0.0159	0.0146
Kazakh	Rkurty	0.3866	0.3869	0.0184	0.0187	0.0111	0.0345	0.0322		0.0298	0.0344	0.0313	0.0300	0.0354
Fine-wool														
Kazakh	AltynAsel	0.3764	0.3849	0.0335	0.0329	0.0252	0.0109	0.0112	0.0298		0.0168	0.0131	0.0095	0.0132
	KaraAdyr	0.3711	0.3733	0.0373	0.0375	0.0315	0.0186	0.0193	0.0344	0.0168		0.0122	0.0193	0.0213
	Khasiev	0.3739	0.3769	0.0349	0.0346	0.0280	0.0155	0.0161	0.0313	0.0131	0.0122		0.0155	0.0180
Semi-coarse	Otkanzhar	0.3765	0.3889	0.0346	0.0340	0.0261	0.0161	0.0159	0.0300	0.0095	0.0193	0.0155		0.0168
Saryarka	Sarysu	0.3729	0.3750	0.0381	0.0379	0.0309	0.0139	0.0146	0.0354	0.0132	0.0213	0.0180	0.0168	

Figure 1. Heatmaps of pairwise F_{st} across Kazakh sheep populations

Figure 2. Linkage disequilibrium decay between breeds and populations of the same breed (Akzhayik, Edilbay, Kazakh Semi-coarse breed). The black dashed line represents LD decay for the entire dataset.

Figure 3. Principal Component Analysis of Kazakh sheep populations. a) Heatmap of first 15 components. Populations are designated as follows: Akzhayik: Atameken (1), Kuanysh (2), Saltanat (3); Edilbay: Azhar (4), Birlik (5); Kazakh Fine-wool sheep – R-Rurty (6); Kazakh Semi-coarse wool sheep: Altyn Asel (7), Kara Adyr (8), Khasiev (9), Otkanzhar (10); Saryarka – Sarysu (11). b) Plot of cumulative variance explained by PCs. 95% level is shown by the red line. c,d,e) Pairwise 2D dot plots of the first three components.

Figure 4. Model based population structure analysis using ADMIXTURE (ADM) and FastStructure (FS) software.

Figure 5. Neighbor-joining tree of genetic distances across Kazakh sheep breeds and breeds in the Sheep HapMap (SHM) dataset. All geographical annotations of SHM breeds are in accordance with ISGC sample annotations.

41

425 **Supplementary materials**

426 Table S1: Results of filtering of samples and SNPs by quality

427 Table S2: Wright's fixation index (F_{st}), expected and observed heterozygosity in Kazakh sheep
428 populations by chromosomes

429 Table S3: Explained variance of 100 principal components

430 Figure S1. Heatmap of first 50 principal components

431 Figure S2. CLUMPAK output of ADMIXTURE plots of Kazakh sheep populations

432 Figure S3. CLUMPAK output of FastStructure plots of Kazakh sheep populations

433

a) Overall (26,534 SNP) **b) Chromosome 1 (3,080 SNP)** **c) Chromosome 16 (827 SNP)** **d) Chromosome 26 (521 SNP)**

LD decay by breeds **LD decay in Akzhayik breed** **LD decay in Edilbay breed** **LD decay in Kazakh Semi-coarse breed**

K = 2

K = 5

K = 11

K = 20

1. KazSemiCoarse - Khasiev
2. KazSemiCoarse - KaraAdyr
3. KazSemiCoarse - AltynAsel
4. KazSemiCoarse - Otkanzhar
5. Edilbay - Birlik
6. Akzhayik - Kuanysh
7. KazFineWool - RKurty
8. Akzhayik - Atameken
9. Edilbay - Azhar
10. Akzhayik - Saltanat
11. Saryarka - Sarysu