American Genetic Association

Journal of Heredity, 2018, 384–392
doi:10.1093/jhered/esx114
Original Article
Advance Access publication 26 December 2017

OXFORD

Original Article

Population Networks Associated with Runs of Homozygosity Reveal New Insights into the Breeding History of the Haflinger Horse

Thomas Druml, Markus Neuditschko, Gertrud Grilz-Seger, Michaela Horna, Anne Ricard, Matjaz Mesarič, Marco Cotman, Hubert Pausch, and Gottfried Brem

From the Institute of Animal Breeding and Genetics, University of Veterinary Sciences Vienna, Veterinärplatz 1, A-1210 Vienna, Austria (Druml, Neuditschko, and Brem); the Agroscope, Swiss National Stud Farm, Les Longs Prés, CH-1580 Avenches, Switzerland (Neuditschko); Pöckau 41, A-9601 Arnoldstein, Austria (Grilz-Seger); the Department of Animal Husbandry, Slovak University of Agriculture in Nitra, Nitra-Chrenová, Slovak Republic (Horna); the Institut National de la Recherche Agronomique, UMR 1313 Génétique Animale et Biologie Intégrative, Jouy-en-Josas, France (Ricard); the Institut Français du Cheval et de l'Equitation, Recherche et Innovation, Exmes, France (Ricard); the Clinic for Reproduction and Large Animals, Veterinary Faculty, University of Lubljana, Cesta v Mestni log 47, Ljubljana, Slovenia (Mesarič); the Institute of Preclinical Sciences, Veterinary Faculty, University of Ljubljana, Cesta v Mestni log 47, Ljubljana, Slovenia (Cotman); and the Animal Genomics, ETH Zürich, CH-8092 Zürich, Switzerland (Pausch).

Address correspondence to Thomas Druml at the address above, or e-mail: thomas.druml@vetmeduni.ac.at.

Received September 20, 2017; First decision November 08, 2017; Accepted December 23, 2017.

Corresponding Editor: Ernest Bailey

Abstract

Within the scope of current genetic diversity analyses, population structure and homozygosity measures are independently analyzed and interpreted. To enhance analytical power, we combined the visualization of recently described high-resolution population networks with runs of homozygosity (ROH). In this study, we demonstrate that this approach enabled us to reveal important aspects of the breeding history of the Haflinger horse. We collected high-density genotype information of 531 horses originating from 7 populations which were involved in the formation of the Haflinger, namely 32 Italian Haflingers, 78 Austrian Haflingers, 190 Noriker, 23 Bosnian Mountain Horses, 20 Gidran, 33 Shagya Arabians, and 155 Purebred Arabians. Model-based cluster analysis identified substructures within Purebred Arabian, Haflinger, and Noriker that reflected distinct genealogy (Purebred Arabian), geographic origin (Haflinger), and coat color patterns (Noriker). Analysis of ROH revealed that the 2 Arabian populations (Purebred and Shagya Arabians), Gidran and the Bosnian Mountain Horse had the highest genome proportion covered by ROH segments (306–397 Mb). The Noriker and the Austrian Haflinger showed the lowest ROH coverage (228, 282 Mb). Our combined visualization approach made it feasible to clearly identify outbred (admixture) and inbred (ROH segments) horses. Genomic inbreeding coefficients (F_{ROH}) ranged from 10.1% (Noriker) to 17.7% (Purebred Arabian). Finally it could be demonstrated, that the Austrian Haflinger sample has a lack of longer ROH segments and a deviating ROH spectrum,

which is associated with past bottleneck events and the recent mating strategy favoring out-crosses within the breed.

Keywords: admixture, breed history, Equine BeadChip 700k, Haflinger, Network analysis, ROH

Background

The establishment and implementation of genomic analyses based upon high-density single nucleotide polymorphism (SNP) BeadChip arrays provides new possibilities to ascertain fine-scale population structures. Previous results of SNP array-based studies demonstrated a better resolution of these techniques for the genome-wide determination of diversity parameters, selection signatures, genetic bottlenecks, and realized levels of inbreeding compared to the commonly applied pedigree-based probability coefficients (Peterson et al. 2013a, 2013b; Curik et al. 2014; Burren et al. 2016; Gómez-Romano et al. 2016; Peripolli et al. 2016). To date, many genome-wide population structure and genetic diversity studies were performed in cattle and pig (Purfield et al. 2012; Ferenčaković et al. 2013; Marras et al. 2015; Zavaraez et al. 2015), and to a lesser extent in sheep (Kijas et al. 2012; Herrero-Medrano et al. 2013; Bosse et al. 2015; Traspov et al. 2016) and goat (Al-Mamun et al. 2015; Burren et al. 2016). The introduction of the Equine SNP50 Genotype BeadChip compromising 54 602 SNPs from Illumina® also led to a series of population genetic studies in the horse including the calculation of genetic diversity parameters (Hasler et al. 2011; Khanshour 2013), population structure analyses (Peterson et al. 2013a, 2013b), selection signature analyses and runs of homozygosity (ROH) (Frischknecht et al. 2014; Metzger et al. 2015). However, compared to the aforementioned livestock species only a modest number of genetic diversity analyses were conducted in horses, leaving the population structure of numerous horse breeds undisclosed.

In this study, we aim to reveal the breeding history of the Haflinger horse by visualizing high-resolution population networks associated with ROH using the genotype information of the novel Affymetrix Axiom™ Equine genotyping array.

The Haflinger breed represents the youngest horse breed of Austria, officially founded in the year 1894 in South Tyrol, which nowadays belongs to Italy. The male genealogy of this breed can be traced back to a single founder stallion, the Shagya Arabian 133 El Bedavy XXII, born in 1868, while the female genealogy of the founder population is insufficiently described. Thus, the genetic background of about 200 so-called "Original Haflinger mares" is not known (Druml et al. 2016). It is supposed that due to the geographical overlap between the breeding regions of Haflinger and Noriker horses a significant percentage of these founder mares may have belonged to the central-alpine autochthonous Noriker breed (for further information, see Supplementary File 1). To reveal the different levels of uncertainty in the Haflinger breeding history we specifically sampled contemporary Austrian and Italian Haflinger, Noriker, Gidran (also belonged to the founder population of the Haflinger breed) and Shagya Arabian horses to assemble a comprehensive and representative sample for genetic diversity analyses. Furthermore, we also included a sample of autochthonous Bosnian Mountain Horses, whose ancestors were used during the First and Second World War for restocking the Haflinger breeding population, and a sample of Purebred Arabians to account for more recent introgression/admixture of Austrian Haflingers with Arabian horses.

The population structure including inbreeding and the founder gene pool of Haflinger and Noriker are well described by means of pedigree analyses (Druml et al. 2009, 2016), while the population structure of Gidran, Bosnian Mountain Horse, and Shagya Arabian are unknown. These 3 breeds are characterized by an endangered status, while the Bosnian Mountain Horse with 100 individual worldwide is threatened by extinction.

To conserve the genetic resources of native horse breeds, we additionally explored the application of genome-based diversity measures, especially focusing on the levels of inbreeding, which can be derived via the computation of ROH (Bosse et al. 2015; Peripolli et al. 2016). Although Haflinger horses are internationally bred, a previous study demonstrated, that both the genetic diversity and the variability of the gene pool in the core breeding population of Austria are decreasing, whereas the mean inbreeding coefficient increased from 6.3% in 1980s to 11.9% for horses born after 2000 (Druml et al. 2016). To account for the geographical differences within the gene pool and to ascertain realized levels of inbreeding, we sampled the aforementioned South Tyrolean (Italian subpopulation without Austrian genetic background) and Austrian Haflinger horses. Due to the historical fact, that "Original Haflinger mares," that is, the female founder animals, were supposed to represent unrelated individuals and were registered as such, hence pedigree-based inbreeding coefficients might be underestimated. Within the scope of current genetic diversity analyses, population structure and ROH results are independently analyzed and interpreted. To enhance the analytical power and visualization of genetic diversity studies, we extended the recently introduced high-resolution network visualization approach (Neuditschko et al. 2012; Steining et al. 2015; Neuditschko et al. 2017) by including ROH results in the final population network visualization.

Material and Methods

Sampling and Pedigree Analysis

All animals included in this study were selected to represent the current population structure and demography of the respective breeds, except for the Arabian breeds. In total, we sampled 531 contemporary horses originating from 6 breeds, which have been involved in the formation of the Haflinger population. For the Austrian Haflinger (n = 78) and Noriker (n = 190) populations, we systematically selected informative animals that account for genealogical and family structures. The Haflinger breed is genealogically structured by 7 sire lines, while the Noriker breed can be divided into 5 sire lines and different coat color families due to assortative breeding strategy. In both sample collections, all sire lines are represented, for the Noriker we additionally sampled families representing the single coat color branches (leopard, roan, tobiano, chestnut, black, and bay) within the breed. To effectively distinguish the Italian/South Tyrolean Haflinger population from the Austrian population, we systematically sampled as distantly related individuals as possible (n = 32) without ancestors from the Austrian Haflinger in their pedigrees. The Shagya Arabian sample (n = 33) represents the current breeding population of the Slovak National stud farm of Topol'čianky, and they were selected for this study as their foundation stock was derived from the former k.k. state stud farm Radautz, which was involved in the formation of the Haflinger breed in the

19th century, while the 155 Purebred Arabians have been selected from the French Arabian breeding population. The 23 Bosnian Mountain Horses represent a cross-selection of registered and documented horses of this highly endangered autochthonous breed. The majority (17 horses) belonged to the Bosnian stud farm of Borike, while additional 6 horses were provided from a private stud farm in Slovenia. The 20 Gidran Horses represent the Hungarian nucleus stock which was bred in the Mezöhegyes stud farm.

For the Austrian Haflinger and Noriker, pedigree data were available comprising a total of 51.613 Noriker and 57.021 Haflinger horses. We extracted the pedigree information for the genotyped horses (190 Noriker and 78 Haflinger) ranging up to 23 and 21 generations, respectively. The pedigree information of the horses was used to calculate inbreeding coefficients and complete generation equivalent (Boichard et al. 1997) using the software package ENDOG v.4.6 (Gutiérrez and Goyache 2005). Pedigree data of all other breeds were not available.

SNP Genotyping

The SNP genotypes for the 531 horses were determined using the Affymetrix Axiom™ Equine genotyping array (Schaefer et al. 2017). The returned number of SNPs was 607 796 for each horse. The chromosomal position of the SNPs was determined based on EquCab2.0 reference genome. We did not consider SNPs positioned on the sex chromosomes (X: 28 017 SNPs and Y: 1 SNP) and SNPs without known chromosomal position (30 864 SNPs). SNPs with more than 10% missing genotypes and minor allele frequency (MAF) less than 0.01 were excluded. This resulted in a total of 537 504 SNPs that passed quality control and were used to ascertain the population structure of the breeds. For the ROH and genetic diversity analysis, we did not apply any MAF filter criterion to account for all homozygous SNPs within the breeds, which resulted in a total of 589 172 SNPs.

Population Structure

Principal Component Analysis

To assess the population structure of the sampled horses and determine the optimal number of clusters we performed a principal component analysis (PCA). Here, we applied PCA on a genetic relationship matrix (G) with pairwise identities by state (IBS) between all horses as provided by PLINK v1.7 (Purcell et al. 2007). To determine the number of optimal clusters, we generated k-means clustering results based upon the number of significant components (PCs) increasing the number of clusters (K) from 2 to 15. We used the Horn's parallel analysis (Glorfeld 1995) as implemented in the R package paran (http://www.r-project.org), while the Calinski criterion (Calinski and Harabasz 1974) was applied to determine the optimal number of clusters.

Admixture

To determine the individual level of admixture of the horses, we performed model-based cluster analyses using the program Admixture 1.23 (Alexander et al. 2009). We ran Admixture in unsupervised mode for 100 iterations increasing K from 2 to 15. Convergence between independent runs at the same K was monitored by comparing the resulting log-likelihood scores (LLs) following 100 iterations, and was inferred from stabilized LLs with less than 1 LL unit of variation between runs. Cross validation (CV) error estimation for each K was performed to determine the optimal number of clusters. Admixture results were visualized with the program Distruct 1.1 (Rosenberg 2004) and also integrated in the high-resolution population networks (see below).

ROH and Genetic Diversity Analysis

ROH segments were determined with an overlapping window approach that was implemented in PLINK v1.7 (Purcell et al. 2007) based on the following settings: minimum SNP density was set to one SNP per 50 kb, with a maximum gap length of 100 kb. The final segments were called ROH, if the minimum length of the homozygous segment was greater than 500 kb and comprised more than 80 homozygous SNPs, while 1 heterozygote and 2 missing genotypes were permitted within each segment.

The total number of ROH, length of ROH (in Mb) and the sum of all ROH segments (in Mb) of each horse, were summarized according to the respective breed/subpopulation and ROH length category to compare ROH segment lengths between breeds. Therefore, the ROH segments (L_{ROH}) were divided into the following 7 length classes: 0.5–1, 1–2, >2–4, >4–6, >6–8, >8–10, and >10 Mb. The genomic inbreeding coefficients (F_{ROH}) were calculated following the method described in McQuillan et al. (2008):

$$F_{ROH} = \sum \frac{L_{ROH}}{L_{AUTO}}$$

where the length of the autosomal genome (L_{AUTO}) was set to 2243 GB. Finally, the F_{ROH} values were compared with the individual F_{PED} coefficients by conducting a correlation analysis using proc corr command implemented in the software package SAS v. 9.1 (SAS Institute 2009). Indices of genetic diversity including observed (H_O) and expected (H_E) heterozygosity as well as the inbreeding coefficient (f) were also determined with PLINK v1.7 using the command --het.

High-Resolution Population Networks

To visualize high-resolution population networks, we performed a high-resolution network visualization based upon the aforementioned IBS relationship matrix (G) of the horses. The different components involved in the so-called NetView approach are described in detail by Neuditschko et al. (2012) and Steining et al. (2015). Briefly, we computed genetic distances by subtracting pairwise relationships from 1 and applied the algorithm in its default setting (number of k nearest neighbors k-NN = 10). To illustrate the genetic relatedness between neighboring horses, we associated the thickness of edges (connecting lines) with the proportion of the genetic distance, while thicker edges corresponding to lower genetic distances. To identify highly inbred and outcrossed horses within the respective population networks, we scaled the node size of each horse based on the individual total length size of ROH segments. The node color of each horse represents the individual level of admixture at the selected number of K clusters.

Results

Population Structure

Horn's parallel analysis of the G-matrix resulted in 8 significant principal components (PCs) explaining a total of 82% of the genetic variation. Visualization of the first 3 PCs showed that the horses were clearly separated according to their genetic origin (Figure 1). The first PC explaining 61% of the genetic variation differentiates

between the 2 poles oriental and occidental, that is, the draft (Noriker, Haflinger) and Arabian breeds (Purebred Arabian, Shagya Arabian), whereas within each pole slightly overlaps were detected (Purebred Arabian and Shagya Arabian; Haflinger and Noriker).

The second PC (explaining 12% of the genetic variation) distinguished the samples according to their breed memberships, while PC3 (explaining 4% of the genetic variation) further substructured Noriker and Haflinger horses and separated Bosnian Mountain and Gidran horses from the 2 Arabian breeds. Both applied methods to determine an optimal number of distinct clusters (Calinski criterion and cross validation error) do not provide a consistent solution. Increasing *K* from 2 to 6 the Calinski criterion based upon the *k*-means clustering result on the 8 significant PCs, suggested an optimal number of *K* = 4, assigning the horses into 2 occidental groups (Haflinger and Noriker) and 1 oriental group (Purebred and Shagya Arabians), whereas the fourth genetic cluster (Bosnian cluster) represented a transition group. Further increasing *K* from 6 to 15 a new optimum at *K* = 10 (Supplementary File 2) was found. The cross-validation error as implemented in the program Admixture identified an optimal number of *K* = 12, however, increasing the number of *K* in the analysis, additional optimal cluster solutions at *K* = 15 were suggested (Supplementary File 2). Despite, the inconsistencies in the determination of optimal number of clusters, we considered *K* = 10 as an optimal solution for both methods (PCA and Admixture), whereas the cluster result at *K* = 7 represents the sampled breed/subpopulations.

The first level of the model-based clustering using Admixture (*K* = 2) confirmed the findings of PC1, differentiating gradually between the oriental group (blue cluster) and the occidental group (purple cluster) (Figure 2), where South Eastern Europe horse populations (here represented by the Bosnian Mountain Horse) illustrated the East-West transition with regard to geographical autochthonous gene pool and breeding management in the former Habsburg k.u.k. monarchy.

At the second level (*K* = 3), Haflinger samples formed a distinct cluster, while Haflinger horses originating from Italy/South Tyrol showed higher levels of admixture with the Noriker/occidental cluster. The third level (*K* = 4) identified a South-Eastern European cluster including all Bosnian samples simultaneously showing a high level of admixture with the Gidran and Shagya Arabian samples. This admixture can be explained by the fact, that numerous autochthonous founder mares from South-Eastern and Eastern Europe were involved at an early stage into the formation of the Gidran and Shagya Arabian breed. At *K* = 5 and *K* = 6 the Shagya Arabian and the Gidran horses built a distinct cluster. At the additional 2 levels of clustering (*K* = 7 and *K* = 9) the Haflinger and Noriker horses were further sub-structured according to geographical origin (Italy and Austria) and coat color patterns, hereby identifying 3 distinct subpopulations within the Noriker sample. Finally at the optimal number of clusters *K* = 10 Purebred Arabians were substructured in 2 distinct clusters.

ROH and Genetic Diversity Analysis

The analysis of ROH revealed population-specific measures of genetic diversity. The overall mean genome length covered by ROH comprised 305.1 Mb (±119.2 Mb; max. 922.1 Mb; min. 4.3 Mb), and the overall mean number of ROHs (n_{ROH}) was 212.4 per horse (±64.6; max. 339 ROHs; min. 7 ROHs). The highest mean genome length covered by ROH (396.5 Mb) and the highest average number of ROH (278.5) was found in the Purebred Arabian, followed by the Shagya Arabian population (mean L_{ROH} 355.1 Mb, mean n_{ROH} 259.0). The Noriker breed showed the lowest values with a mean L_{ROH} of 227.5 Mb and overall 165.0 ROH segments, followed by the Austrian Haflinger (mean L_{ROH} 282.1 Mb, mean n_{ROH} 208.5) (Table 1).

The main part (54.7–61.7%) of ROH segments per breed in our sample had a mean length between 500 kb and 1 MB. The highest proportion of the longest segments (>10 MB) was found in the Bosnian Mountain Horse population (5.4%), the Italian Haflinger (1.3%) and the Gidran (1.2%). Small segments (1–2 Mb) did occur at highest frequencies in the Austrian Haflinger sample (27.4%) and the Arabian samples (26.3% for both), whereas the lowest proportion (17.5%) of ROHs from 1 to 2 Mb (Table 2) was found in the Bosnian Mountain Horse. Overall the Bosnian Mountain Horse showed the highest proportion of ROHs longer than 4 Mb (12.9%), whereas the Austrian Haflinger was characterized by the lowest proportion (4.3%) of ROHs in this length classes. On the contrary, the Italian Haflinger sample had the second highest proportion of ROH longer than 4 Mb comprising 8.8%.

The estimation of inbreeding using the proportion of the genome covered by ROH (>0.5 Mb) resulted in following picture:

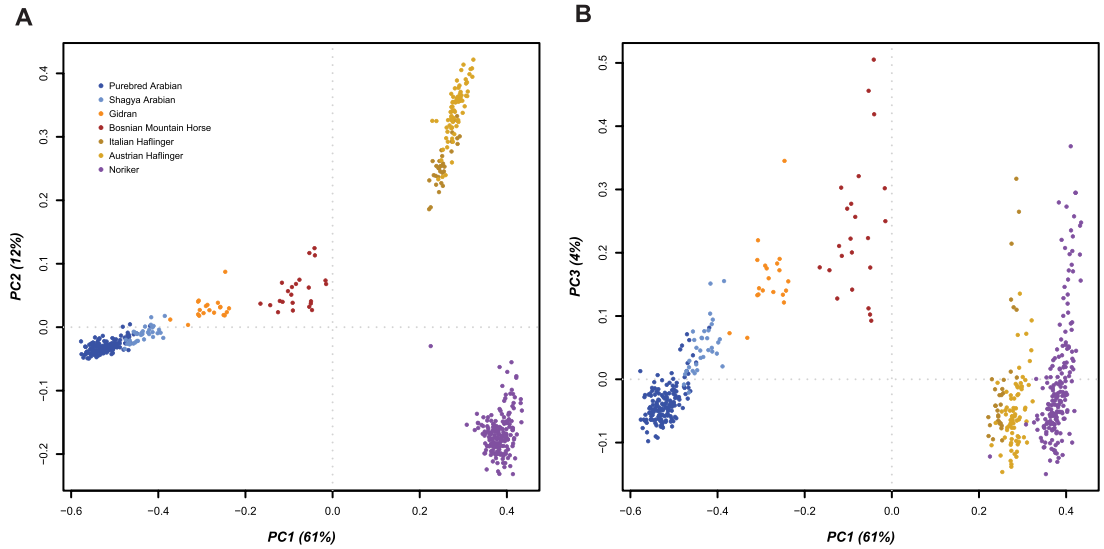

Figure 1. Visualization of the dataset on the first 3 PCs together explaining 77% of the variation (A = scatter plot of PC1 and PC2; B = scatterplot of PC1 and PC3). See online version for full colors.

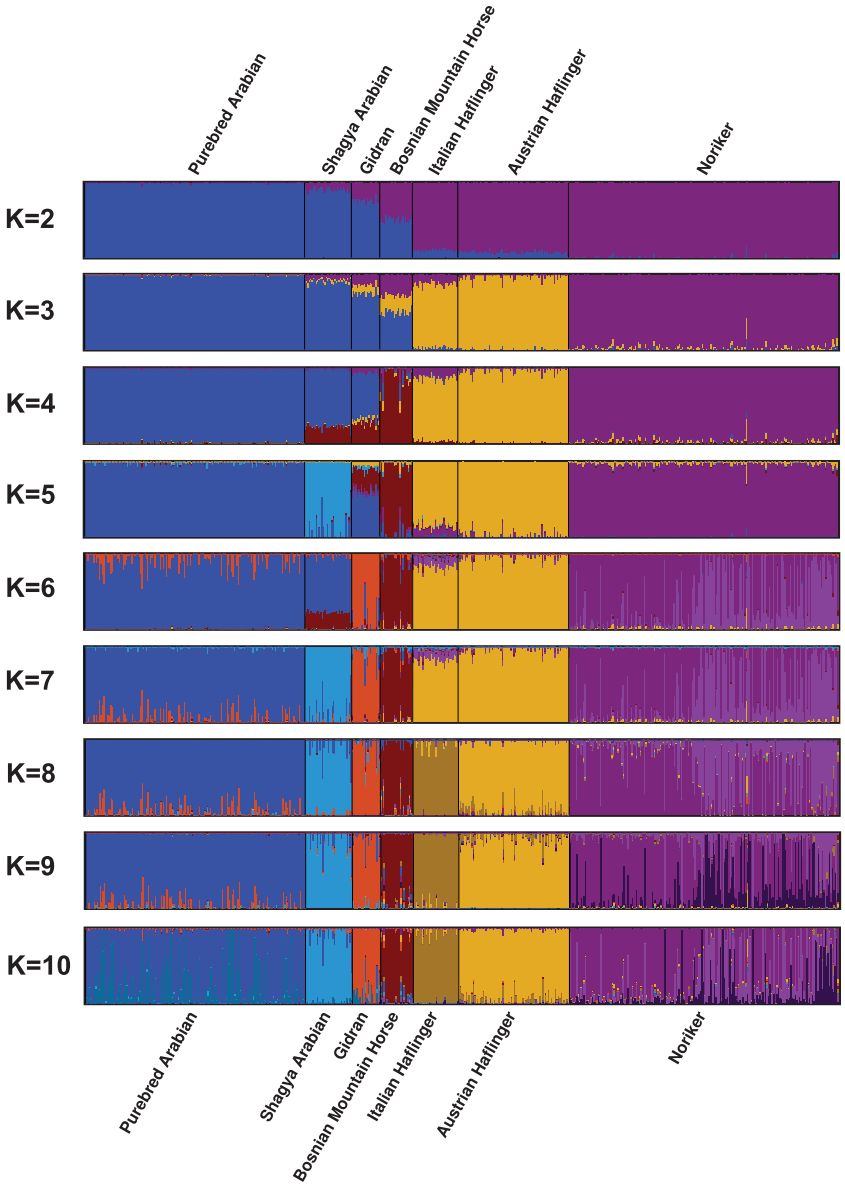

Figure 2. Graphical representation of individual cluster membership coefficients for Admixture runs increasing K = 2–10 (the optimal number of clusters). See online version for full colors.

Table 1. Mean genome length (in kb) covered by ROH (standard deviation (SE), minimum and maximum values), mean number of ROH (standard deviation, minimum and maximum) for the samples Haflinger (Austria, Italy), Noriker, Bosnian Mountain Horse, Gidran, Shagya Arabian, and Purebred Arabian

Population	*N*	Mean L_{ROH} Kb	SE	Min	Max	Mean n_{ROH}	SE	Min	Max
Haflinger Austria	78	282 095	90 951	71 451	524 020	208.5	39.3	97	326
Haflinger Italy	32	316 654	109 873	54 411	486 520	188.3	49.0	59	325
Noriker	190	227 491	104 078	4 257	459 721	165.0	55.7	7	339
Bosnian Mountain Horse	23	305 291	199 795	65 967	922 076	133.1	22.0	94	178
Gidran	20	321 972	116 251	143 198	756 258	217.1	26.8	158	279
Shagya Arabian	33	355 131	72 517	195 821	498 701	259.0	20.1	209	300
Purebred Arabian	155	396 501	62 456	236 525	633 472	278.5	18.4	222	323
All	531	305 081	119 239	4 257	922 076	212.4	64.6	7	339

The Arabian populations were characterized by the highest F_{ROH} values ranging from 17.7% (Purebred Arabian) to 15.8% (Shagya Arabian). The F_{ROH} values were lower in the Gidran (14.3%), the Bosnian Mountain Horse (13.6%) and the Italian Haflinger (14.1%) breeds. The Noriker and the Austrian Haflinger sample had the lowest F_{ROH} values of 10.1% and 12.6%, respectively (Table 3). For the

2 Austrian Haflinger and Noriker breeds, pedigree records were available (complete generation equivalent of 9 generations in both cases). The correlation between the pedigree based inbreeding coefficients F_{PED} and F_{ROH} was 0.38 and 0.19, respectively.

To quantify more recent inbreeding, we calculated F_{ROH} considering ROHs longer than 5 Mb. The highest recent inbreeding was observed in the Bosnian Mountain Horse (mean $F_{ROH>5Mb}$ of 10.1%), followed by the Italian Haflinger (mean $F_{ROH>5Mb}$ of 5.1%), the Purebred Arabian (mean $F_{ROH>5Mb}$ of 3.8%) and the Gidran (mean $F_{ROH>5Mb}$ of 3.8%). Lowest recent inbreeding was obtained for the Austrian Haflinger (mean $F_{ROH>5Mb}$ of 2.1%) and the Noriker sample (mean $F_{ROH>5Mb}$ of 2.9%). Moreover, between 91% and 95% of all horses of the Purebred and Shagya Arabians and Gidran horses had ROHs longer than 5 Mb, whereas the proportion of animals carrying long ROHs was less in all other breeds (65% of Noriker horses, 74% of Austrian Haflinger horses).

Observed heterozygosity ranged from 0.256 to 0.326 (Table 3). The highest heterozygosity was found for Gidran (0.326), Shagya Arabian (0.312), and Bosnian Mountain Horse (0.305) samples, lowest values were observed for Noriker (0.256) and Austrian Haflingers (0.273).

High-Resolution Network Visualization

Finally, we integrated the individual levels of admixture at the respective number of *K* clusters (*K* = 7 and *K* = 10) and individual total ROH segment length into a high-resolution population network. Compared to the previous applied cluster approaches (PCA and Admixture), NetView clearly separated the Shagya Arabian samples into 2 distinct population groups, while the smaller population cluster, comprises particularly highly related individuals (halfsibs) (Figure 3). The high-resolution network structure additionally illustrated, that the remaining Shagya Arabians show high levels of admixture with Purebred Arabians, while especially the 2 cross-link horses, acting as a hub between the Shagya Arabian and Purebred Arabian population cluster, showing admixture levels greater than 50% with Purebred Arabians. This result reflects the studbook entries of these 2 horses, which were Shagya Arabian × Purebred Arabian crossbreds according to pedigree information. The nearest Shagya Arabian horse next to 1 of the 2 crossbreds showed a 25% admixture level with Purebred Arabians, hereby representing a father-offspring pair, as these 2 horses are connected via a thick edge, which represents a close genetic distance between horses. Such crosslink horses acting as a hub between the breeds were also identified within the Noriker, and for Bosnian Mountain Horse and Gidran population cluster. According to pedigree information the crosslink animal between Noriker and Haflinger is a crossbred horse between a Noriker stallion and a Haflinger mare of unknown origin, as depicted in the high-resolution network visualization.

The network of the Haflinger population showed a clear genealogical separation between horses originating from Austria (down left) and Italy (top right) and that particularly horses from Italian population were closely related with Arabian and Noriker samples according to the individual levels of admixture at *K* = 7 (Figure 3A). The network of the Noriker horses also indicated putative substructures within this breed. Taking into account the individual levels of admixture at *K* = 10 (Figure 3B), the Noriker substructures were in concordance with the different coat color patterns of this breed. The light purple cluster is represented by Leopard spotted (LP) horses, whereas next to these families the dark purple cluster is mainly build by outcross Leopards and LP-related black horses. The major purple cluster especially involves families with chestnut and bay coat color. In the case of relationship between Gidran and the substructure of the French Arabian sample, we could observe crosslink animals and an admixture on both sides. Compared to the Purebred Arabian substructure it can be noticed, that the 2 Haflinger subpopulations became more apparent by 2 distinct genetic clusters.

We have associated the node size of individuals with the total length of ROH to characterize the horses as a result of recent so called

Table 2. Distribution of ROH segments of different length classes within the samples Haflinger (Austria and Italy), Noriker, Bosnian Mountain Horse, Gidran, Shagya Arabian, and Purebred Arabian

Breed/ROH	0.5–1 MB	1–2 MB	2–4 MB	4–6 MB	6–8 MB	8–10 MB	>10 MB
Haflinger Austria	56.2	**27.4**	12.1	2.9	**0.9**	**0.3**	**0.2**
Haflinger Italy	**54.7**	23.9	**12.5**	**4.3**	**2.2**	1.0	1.3
Noriker	59.5	24.6	11.0	2.8	1.0	0.4	0.6
Bosnian Mountain Horse	**61.7**	**17.5**	**7.9**	3.9	1.9	**1.7**	**5.4**
Gidran	58.9	24.6	10.1	3.3	1.5	0.5	1.2
Shagya Arabian	58.8	26.3	10.0	2.6	1.1	0.5	0.7
Purebred Arabian	58.7	26.3	10.1	2.5	1.1	0.5	0.9

Highest and lowest values in the respective RHO segment class are expressed in bold.

Table 3. Mean F_{ROH} values (standard deviation, minimum and maximum) for the samples Haflinger (Austria, Italy), Noriker, Bosnian Mountain Horse, Gidran, Shagya Arabian, Purebred Arabian, and mean F_{PED} values (standard deviation, minimum, and maximum), observed heterozygosity (*HO*), expected heterozygosity (*HE*) and inbreeding coefficient (*f*)

Population	*N*	Mean F_{ROH}	SE	Min	Max	Mean F_{PED}	SE	Min	Max	H_O	H_E	*f*
Haflinger Austria	78	12.6	4.1	3.2	23.4	11.8	3.3	2.4	19.5	0.273	0.272	−0.004
Haflinger Italy	32	14.1	4.9	2.4	21.7					0.294	0.285	−0.033
Noriker	190	10.1	4.6	0.2	20.5	5.0	1.7	1.2	11.9	0.256	0.258	0.001
Bosnian Mountain Horse	23	13.6	8.9	2.9	41.1					0.305	0.293	−0.040
Gidran	20	14.4	5.2	6.4	33.7					0.326	0.311	−0.050
Shagya Arabian	33	15.8	3.2	8.7	22.2					0.312	0.297	−0.049
Purebred Arabian	155	17.7	2.8	10.5	28.2					0.279	0.279	0.003
All	531	13.6	5.3	0.2	41.1							

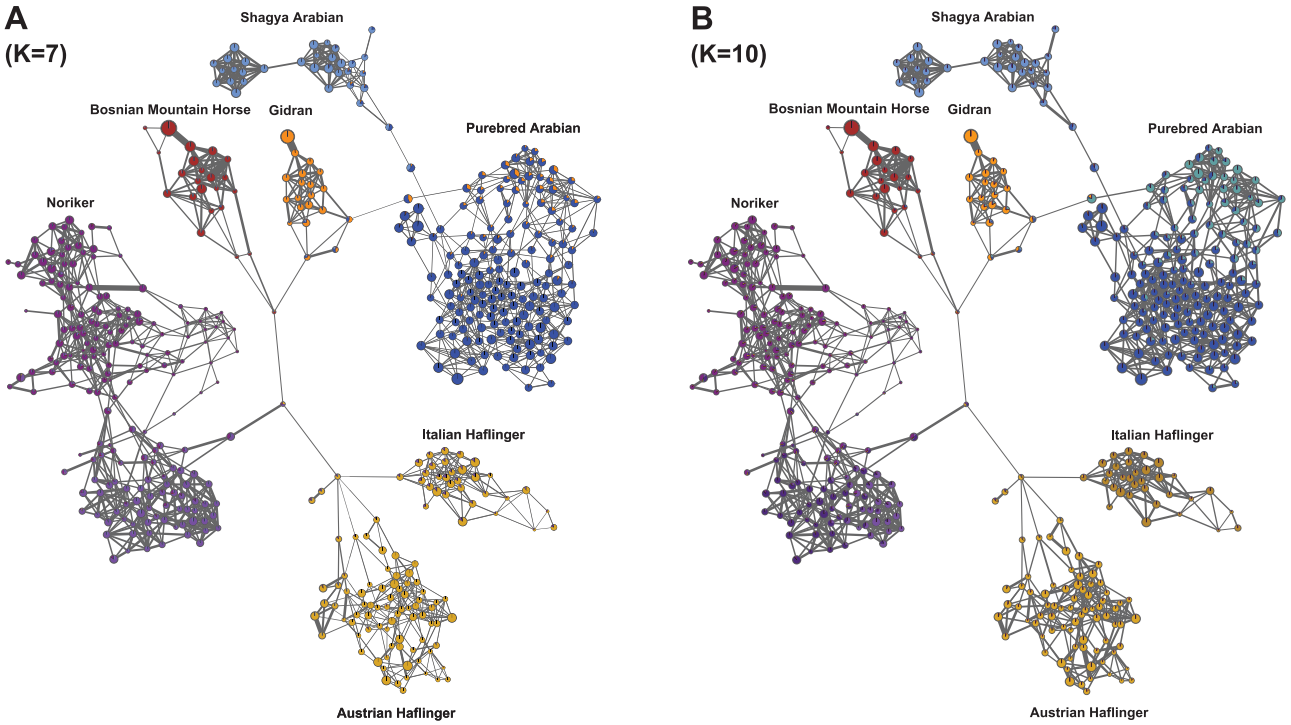

Figure 3. High-resolution population structure of Haflinger, Noriker, Bosnian Mountain Horse, Gidran, Shagya Arabian, and Purebred Arabian horses. Individual cluster membership coefficients from Admixture runs $K = 7$ (network A) and $K = 10$ (network B) are overlaid on the nodes representing an individual. The size of the node is in relation to the total length size of homozygous segments (LROH) and the links between nodes represent the genetic distances between animals. The thicknesses (edge) of the links are proportional to the pairwise relationship coefficients. See online version for full colors.

out-breeding (short homozygous segments—smaller nodes) or of inbreeding/assortative breeding (longer homozygous segments—bigger nodes). This effect of breeding practices is evident exemplarily in the aforementioned cross-link Noriker × Haflinger horse, but also in other examples like the cross-link animals between the 2 Haflinger populations, the Bosnian Mountain Horses, the Gidran or the cross-link animals between the purple and dark purple Noriker cluster. Within the Noriker sample a coat color gradient from black Leopards (light purple node color), Leopards and spotted blankets (dark purple node color) over bay and black horses, to higher selected and inbred chestnut horses (purple node color), correspond with the individual ROH segment length, individual genetic relationship, and admixture levels (Figure 3B). For the Bosnian Mountain Horse and the Gidran sample, 2 highly inbred animals (F_{ROH} of 41.1% Bosnian Montain Horse and of 33.7% Gidran) and its genealogical relationship could be identified.

Discussion

The high-resolution population network analyses using high-density SNP data revealed new aspects of the Haflinger breeding history. Whereas the proportion of Arabian genes that were introgressed into the Haflinger breed have been estimated to range from 4.4% to 6.9% of the gene pool using pedigree data (Gandini et al. 1997; Druml et al. 2016), the contribution of the Noriker horse to the Haflinger breed has not been adequately quantified. Based upon PCA and model-based clustering, we could show a genetic admixture between Haflinger and Noriker ($K = 21$). According to the SNP information both Haflinger samples were separated into 2 distinct groups, which were characterized by a higher within-group relationship and a clear genealogical separation, leading to the conclusion that both samples represent 2 isolated populations equivalent to distinct breeds. This result also reflects the breeding history, as the Italian sample includes only animals which belong to the Italian stud book that was kept nearly closed for introgression of Austrian Haflingers from 1919 to 1994. Within the Austrian Haflinger, we additionaly detected 2 horses with a stable and distinct Purebred Arabian contribution (Figure 3). The high genetic variability within the Noriker breed is represented by a 2-fold structure and the existence of 3 distinct genetic clusters ($K = 2$, $K = 6$, $K = 9$). The clusters assignments at $K = 9$ represents the family structures according to different coat color breeding on a long term. Similar results for this differentiation have already been shown by pedigree analysis (Druml et al. 2009). According to model based-clustering, the highest genetic distance was identified between leopard spotted and chestnut families. A recent study (Druml et al. 2017) documented coat color specific segregation of the GYS1 mutation in the Noriker breed by means of pedigree analysis, which is also associated with the genetic distance between horses from leopard spotted and chestnut families.

The high-resolution network visualization combined with L_{ROH} measures clearly identified cross-link animals between and within populations. Besides these out-bred animals, characterized by a low genome-wide L_{ROH} and a higher level of admixture we observed the highest proportion of shortest ROHs (<2 Mb) in the Austrian Haflinger population, simultaneously showing the lowest proportion of ROH >6 Mb. Taking into account, that ROHs of a length from 5 to 16.6 Mb characterize inbreeding events that have taken place 10–13 generations to 3–3.9 generations ago (Arias et al. 2009; Thompson 2013), the relative lack of these ROH classes within the Austrian Haflinger sample indicates out-crossing events simultaneously minimizing inbreeding by matings of genetically distant horses throughout the recent breeding history (past 5–7 generations). The observed high proportion of short ROHs (<4 Mb) illustrated the effect of a reduced

population size and bottleneck events (Kirin et al. 2010), which could be noticed in all sampled horses included in this study, however, at a lower degree compared to the Austrian Haflinger. Furthermore, Italian Haflinger and Austrian Haflinger samples had a different ROH length class composition, whereas Italian Haflingers showed higher proportions of ROH >2 Mb to >10 Mb which resulted in a higher F_{ROH} of 14.1%. Whereas the highest proportion of long ROHs (>10 Mb) and genome coverage by ROH were found in the Bosnian Mountain Horse and the Arabian samples, the Noriker sample had the shortest genome-wide L_{ROH}. These results were in concordance with pedigree-based inbreeding estimates from Druml et al. (2009) and Cervantes et al. (2008). The composition of ROH length classes were similar in the Purebred Arabian, Shagya Arabian, Gidran, and Noriker breeds and reflected the applied breeding scheme (closed stud book) and population history (crisis of horse breeding in 1970s). Within the whole dataset, the Austrian Haflinger as represented by ROH length classes marked a specific system of breeding management characterized by minimizing coancestry in matings up to 5–7 generations based on geographical distance or on divergence in generation. The samples of Bosnian Mountain Horse or Gidran clearly showed the effects of extremely limited census, for instance, the purebred Bosnian samples represent a typical collection of the former state stud farm Borike, whose breeding herd comprised only about 30 horses on a long term.

Observed and expected heterozygosity in the populations under study were slightly higher in Gidran, Bosnian Mountain Horse, Italian Haflinger, and Shagya Arabian, samples which were characterized by smaller sample size and a proportional higher number of admixed individuals (Figure 3). The Noriker, Austrian Haflinger, and the Purebred Arabian, samples with lower admixture and closed stud book, showed lower levels of observed heterozygosity. Especially the Noriker sample, characterized by the lowest genome-wide ROH coverage and the lowest level of heterozygosity, underlines the effect of population structure (see cluster assignment at $K = 9$) on measures of inbreeding. Thus, the proportion of heterozygous loci/animals without the consideration of additional diversity parameters can bias the estimation of genetic diversity. On average heterozygosity, levels were comparable to values as represented by Petersen et al. (2013a), although in this study significantly lower numbers of SNPs (6.028–26.171) were used.

The results of the ROH analyses revealed new insights into population structure and history of the horse breeds studied but a comparison between our samples and previously published studies is difficult, because of the limited number of works concerning equines and different sampling strategies and methods. For instance, Metzger et al. (2015) used a 50 SNP sliding window to identify ROH coverage ranging from 476 Mb to 798 Mb comprising between 2804 and 4175 ROHs per horse in a sample of 10 horses. HD SNP panels in cattle revealed a mean genome length covered by ROH from 80 to 198 Mb (Purfield et al. 2012; Zavarez et al. 2015), whereas the analysis of 50k SNP panels in cattle resulted in higher ROH coverage as reported by Ferenčaković et al. (2013), Kim et al. (2013), and Marras et al. (2015). Regarding the 50k SNP chip Purefield et al. (2012) suggest that this panel is appropriate for the identification of ROHs longer than 5 Mb, whereas HD SNP panels have a higher sensitivity to detect short ROHs, but might not reveal consistent patterns (Curik et al. 2014).

The extent of autozygosity detected via ROH in the Austrian horse samples resulted in a calibration comparable to previous results from pedigree analysis in Haflinger and Noriker horses (Gandini et al. 1992; Druml et al. 2009, 2016). The correlation between F_{ROH} and F_{PED} was 0.19 and 0.38 in Noriker and Austrian Haflinger, respectively, and thus significantly lower than reported (from 0.53 to 0.71) in the review of Peripolli et al. (2016). The proportion of ROHs in length classes shorter than 2 Mb was 84% for Austrian Haflingers and Noriker horses. It should be underlined that these distributions were derived from analyzing adjustments which previously have been applied on HD panels in cattle and humans (Lencz et al. 2007; Kirin et al. 2010; Purefield et al. 2012). The low correlation measured between F_{ROH} and F_{PED} may be due to the high proportion of ROHs shorter than 2 Mb in both breeds, a fact which is pointed out by Kirin et al. (2010), who mention that sum of ROHs shorter than 5 Mb do not correlate well with pedigree-based inbreeding. In the Noriker breed F_{PED} is about 40% smaller than F_{ROH}, a deviation which can be expected according to the findings in literature (Kirin et al. 2010; Purfield et al. 2012; Ferenčaković et al. 2013). Short ROH segments (<2 Mb) indicate inbreeding which took place prior documented breed history, and thus cannot be described by a pedigree based inbreeding coefficient. Practically, the lack of longer ROH segments and the deviating ROH spectrum in the Austrian Haflinger sample may be explained by several bottleneck situations and by the recent mating strategy favoring out-crosses within the breed.

Conclusions

In this work, we could demonstrate that the NetView approach in combination with Admixture and ROH information is a valuable tool to reveal new insights into the breeding history of the Haflinger horse. However, it should be noticed that, the analysis of mtDNA profiles would provide additional insights into the female founder gene pool. Previous structure analyses of equine populations on genome-wide SNP panels concentrated on the estimation of *F*-statistics within/between populations and on parsimonious phylogenies to cluster individuals or groups according to genetic distance metrics. Such derived phylogenetic relationships further assist in the verification process of novel phenotype associated mutations or alleles, taking into account that the global phylogenetic framework assumes neutral segregation of loci. In this study, we could demonstrate that sample size and sampling methods are not trivial within the analyses of genetic structure. The combination of genealogical aspects (Netview) with gene pool characterization (Admixture) and diversity measures (ROH) was capable to explain aspects of complex development of a composite breed within a specific economic and political environment. Furthermore, we demonstrate that population networks associated with ROH supports the conservation of autochthonous and small livestock populations by optimizing breeding management decisions.

Supplementary Material

Supplementary data are available at *Journal of Heredity* online.

Funding

This work was supported by the Austrian Research Promotion Agency (FFG) and Xenogenetik (contract number 843464).

Acknowledgments

The authors want to thank the ARGE Noriker and the Südtiroler Haflingerpferdezuchtverband for assistance in data collection and Ms Veronika Lugmayr and Mr Lutz Plobner for sample management.

Data Availability

The primary data of this study belongs to several breeding associations and research groups and are available for academic use upon request.

References

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 19:1655–1664.

Al-Mamun HA, Clark SA, Kwan P, Gondro C. 2015. Genome-wide linkage disequilibrium and genetic diversity in five populations of Australian domestic sheep. *Genet Sel Evol*. 47:90.

Arias JA, Keehan M, Fisher P, Coppieters W, Spelman R. 2009. A high density linkage map of the bovine genome. *BMC Genet*. 10:18.

Boichard D, Maignel L, Verrier E. 1997. The value of using probabilities of gene origin to measure genetic variability in a population. *Gen Sel Evol*. 29:5–23.

Bosse M, Megens H-J, Madsen O, Crooijmans RPMA, Ryder OA, Austerlitz F, Gorenen MAM, de Cara MAR. 2015. Using genome-wide measures of coancestry to maintain diversity and fitness in endangered and domestic pig populations. *Gen Res*. 25:1–12.

Burren A, Neuditschko M, Signer-Hasler H, Frischknecht M, Reber I, Menzi F, Drögemüller C, Flury C. 2016. Genetic diversity analyses reveal first insights into breed-specific selection signatures within Swiss goat breeds. *Anim Genet*. 47:727–739.

Calinski T, Harabasz J. 1974. A dendrite method for cluster analysis. *Commun Stat*. 3: 1–27.

Cervantes I, Molina A, Goyache F, Guttierez JP, Valerea M. 2008. Population history and genetic variability in the Spanish Arab Horse assessed via pedigree analysis. *Liv Sci*. 113: 24–36.

Curik I, Ferencakovic M, Sölkner J. 2014. Inbreeding and runs of homozygosity: a possible solution to an old problem. *Liv Sci*. 166: 26–34.

Druml T, Baumung R, Sölkner J. 2009. Pedigree analysis in the Austrian Noriker draught horse: genetic diversity and the impact of breeding for coat colour on population structure. *J Anim Breed Genet*. 126:348–356.

Druml T, Grilz-Seger G, Neuditschko M, Brem G. 2017. Association between population structure and allele frequencies of the glycogen synthase 1 mutation in the Austrian Noriker draft horse. *Anim Genet*. 48:108–112.

Druml T, Sauer K, Elsbacher J, Grilz-Seger G, Brem G. 2016. Analyse des Genpools, der genetischen Diversität und der Inzuchtverhältnisse der österreichischen Haflingerpopulation. *Züchtungskunde*. 88: 379–394.

Ferenčaković M, Hamzić E, Gredler B, Solberg TR, Klemetsdal G, Curik I, Sölkner J. 2013. Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations. *J Anim Breed Genet*. 130:286–293.

Frischknecht M, Neuditschko M, Jagannathan V, Drögemüller C, Tetens J, Thaller G, Leeb T, Rieder S. 2014. Imputation of sequence level genotypes in the Franches-Montagnes horse breed. *Genet Sel Evol*. 46:63.

Gandini GC, Bagnato A, Miglior F, Pagnacco G. 1992. Inbreeding in the Italian Haflinger horse. *J Anim Breed Genet*. 109: 433–443.

Gandini GC, Samore A, Pagnacco G. 1997. Genetic contribution of the Arabian to the Italian Haflinger horse. *J. Anim. Breed. Genet*. 114: 457–464.

Glorfeld LW. 1995. An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educ Psychol Measure*. 55: 377–393.

Gómez-Romano O, Villanueva JB, Fernández JA, Woolliams JA, Pong-Wong R. 2016. The use of genomic coancestry matrices in the optimisation of contributions to maintain genetic diversity at specific regions of the genome. *Gen Selec Evol*. 48: 2.

Gutiérrez JP, Goyache F. 2005. A note on ENDOG: a computer program for analysing pedigree information. *J Anim Breed Genet*. 122: 357–360.

Hasler H, Flury C, Menet S, Haase B, Leeb T, Simianer H, Poncet PA, Rieder S. 2011. Genetic diversity in an indigenous horse breed: implications for mating strategies and the control of future inbreeding. *J Anim Breed Genet*. 128:394–406.

Herrero-Medrano JM, Megens HJ, Groenen MAM, Ramis G, Bosse M, Perez-Enciso M, Crooijmans RBMA. 2013. Conservation genomic analysis of domestic and wild pig populations from the Iberian peninsula. *BMC Genet*. 14: 106. Available from: URL http://www.biomedcentral.com/1471–2156/14/10.

Khanshour AM. 2013. *Genetic diversity and population structure of the Arabian horse populations from Syria and other countries* [doctoral dissertation]. [College Station (TX)]: Texas A&M University.

Kijas JW, Lenstra JA, Hayes B, Boitard S, Porto Neto LR, San Cristobal M, Servin B, McCulloch R, Whan V, Gietzen K, *et al.*; International Sheep Genomics Consortium Members. 2012. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biol*. 10:e1001258.

Kim ES, Cole JB, Huson H, Wiggans GR, Van Tassell CP, Crooker BA, Liu G, Da Y, Sonstegard TS. 2013. Effect of artificial selection on runs of homozygosity in U.S. Holstein cattle. *PLoS One*. 8:e80813.

Kirin M, McQuillan R, Franklin CS, Campbell H, McKeigue PM, Wilson JF. 2010. Genomic runs of homozygosity record population history and consanguinity. *PLoS One*. 5:e13996.

Lencz T, Lambert C, DeRosse P, Burdick KE, Morgan TV, Kane JM, Kucherlapati R, Malhotra AK. 2007. Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proc Natl Acad Sci U S A*. 104:19942–19947.

Marras G, Gaspa G, Sorbolini S, Dimauro C, Ajmone-Marsan P, Valentini A, Williams JL, Macciotta NP. 2015. Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. *Anim Genet*. 46:110–121.

McQuillan R, Leutenegger AL, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, Smolej-Narancic N, Janicijevic B, Polasek O, Tenesa A, *et al.* 2008. Runs of homozygosity in European populations. *Am J Hum Genet*. 83:359–372.

Metzger J, Karwath M, Tonda R, Beltran S, Águeda L, Gut M, Gut IG, Distl O. 2015. Runs of homozygosity reveal signatures of positive selection for reproduction traits in breed and non-breed horses. *BMC Genomics*. 16:764.

Neuditschko M, Khatkar MS, Raadsma HW. 2012. NetView: a high-definition network-visualization approach to detect fine-scale population structures from genome-wide patterns of variation. *PLoS One*. 7:e48375.

Neuditschko M, Raadsma HW, Khatkar MS, Jonas E, Steinig EJ, Flury C, Signer-Hasler H, Frischknecht M, von Niederhäusern R, Leeb T, *et al.* 2017. Identification of key contributors in complex population structures. *PLoS One*. 12:e0177638.

Peripolli E, Munari DP, Silva MVGB, Lima ALF, Irgang R, Baldi F. 2016. Runs of homozygosity: current knowledge and applications in livestock. *Anim Genet*. 48:255–271.

Petersen JL, Mickelson JR, Cothran EG, Andersson LS, Axelsson J, Bailey E, Bannasch D, Binns MM, Borges AS, Brama P, *et al.* 2013a. Genetic diversity in the modern horse illustrated from genome-wide SNP data. *PLoS One*. 8:e54997.

Petersen JL, Mickelson JR, Cleary KD, McCue M. 2013b. The American Quarter Horse: population structure and relationship to thouroughbred. *J Hered*. 105:148–162.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, *et al.* 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 81:559–575.

Purfield DC, Berry DP, McParland S, Bradley DG. 2012. Runs of homozygosity and population history in cattle. *BMC Genet*. 13:70. Available from: URL http://www.biomedcentral.com/1471–2156/13/70.

Rosenberg NA. 2004. Distruct: a program for the graphical display of population structure. *Mol Ecol Notes*. 4:137–138.

SAS Institute. 2009. *SAS version 9.1*. Cary (NC): SAS Institute, Inc.

Schaefer RJ, Schubert M, Bailey E, Bannasch DL, Barrey E, Bar-Gal GK, Brem G, Brooks SA, Distl O, Fries R, *et al.* 2017. Developing a 670k genotyping array to tag ~2M SNPs across 24 horse breeds. *BMC Genomics*. 18:565.

Steinig EJ, Neuditschko M, Khatkar MS, Raadsma HW, Zenger KR. 2015. Netview p: a network visualization tool to unravel complex population structure using genome-wide SNPs. *Mol Ecol Res*. 16:216–227.

Thompson EA. 2013. Identity by descent: variation in meiosis, across genomes, and in populations. *Genetics*. 194:301–326.

Traspov A, Deng W, Kostyunina O, Ji J, Shatokhin K, Lugovoy S, Zinovieva N, Yang B, Huang L. 2016. Population structure and genome characterization of local pig breeds in Russia, Belorussia, Kazakhstan and Ukraine. *Genet Sel Evol*. 48:16.

Zavaraez LB, Utsunomiya YT, Carmo AS *et al.* 2015. Assessment of autozygosity in Nellore cows (*Bos indicus*) through high density SNP genotypes. *Front Genet*. 6:5.