Genetic diversity within and relationships among Dutch horse populations

A. Schurink[1,2], M. Shrestha[1,3], S. Eriksson[3], M. Bosse[1], H. Bovenhuis[1], J.K. Oldenbroek[2], J.J. Windig[1,2], S.J. Hiemstra[2], W. Back[4,5], A.M. Johansson[3], & B.J. Ducro[1]

[1] Wageningen University & Research, Animal Breeding and Genomics, P.O. Box 338, 6700 AH Wageningen, the Netherlands
anouk.schurink@wur.nl (Corresponding Author)
[2] Wageningen University & Research, Centre for Genetic Resources, the Netherlands (CGN), P.O. Box 338, 6700 AH Wageningen, the Netherlands
[3] Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, P.O. Box 7023, 75007 Uppsala, Sweden
[4] Department of Equine Sciences, Faculty of Veterinary Medicine, Utrecht University, Yalelaan 112-114, 3584 CM Utrecht, the Netherlands
[5] Department of Surgery and Anaesthesiology of Domestic Animals, Faculty of Veterinary Medicine, Ghent University, Salisburylaan 133, B-9820 Merelbeke, Belgium

Summary

Genotype data from 9 horse populations sampled in the Netherlands with about 20 horses per population were analysed to quantify genetic relationships and diversity and to increase our understanding of their evolutionary history. Level of inbreeding was low in populations that allowed introgression like the royal Dutch sport horse (KWPN) population. Rather high and more recent inbreeding, as indicated by the contribution of long runs of homozygosity (ROHs) to the genome, was detected in closed populations such as the Friesian horse population. The multidimensional scaling analysis firstly separated the warmblood and coldblood populations, and secondly the Friesian horse, Icelandic horse and Shetland pony populations from the other populations. Genetic differentiation was smallest among the four Dutch warmblood horse populations, while differentiation from other populations was largest for the Friesian horse and Shetland pony populations. ROHs aided in understanding breeding history as it enabled us to differentiate between more recent and past inbreeding. Estimated diversity within and most of the quantified relationships among the populations were as expected based on their history and classifications. Our study shows how admixture, drift and demography have shaped the genome-wide diversity of these populations in the Netherlands. Future research will identify which populations comprise unique genes and contribute significantly to the total genetic diversity to ultimately be able to make well-informed decisions in genetic conservation.

Keywords: genetic diversity, horses, population structure, runs of homozygosity

Introduction

Genetic diversity in the horse is largely present as variation between breeds or types, thereby leaving relatively little to within breed variation (McCue *et al.*, 2012). The evolution from the first domesticated horses into modern breeds and the ongoing formation of breeds is driven by the large and wide variety of historic and current purposes of keeping horses (Librado *et al.*, 2016) and by geographic and environmental conditions. Exchange of genetic material between

breeds decreased with the emergence of studbooks in the late 19th century, as breeding goals and standards became more formalized and narrower. Some breeds became closed and preservation of specific breed characteristics was prioritized (Bowling & Ruvinsky, 2000). Divergence between populations achieved in such relatively short time period is often at the expense of genetic diversity.

The spectrum of modern horse breeds encompasses breeds with a long history of development in isolation like Icelandic horses and Shetland ponies (Hendricks, 2007) as well as relatively recently formed breeds. Research on the genomic make-up of horse populations with varying development and background can teach us about breeding history and give valuable information on how best to conserve or improve populations for the future. The aims of this study were to quantify genetic relationships and diversity within and among 9 horse populations in the Netherlands and to increase our understanding of their evolutionary history using genome-wide SNP data.

Material and methods

Horses, genotypes and quality control

Data contained 184 horses from 9 populations in the Netherlands with about 20 horses per population (Table 1). Belgian draft horses, Gelder horses, Groningen horses, harness horses, KWPN sport horses and Lipizzaner horses were sampled with focus on genetic diversity research. Friesian horses, Shetland ponies and Icelandic horses were originally sampled with focus on genome-wide association studies. Individuals from these populations were selected based on pedigree information.

Horses were genotyped with either the Illumina® EquineSNP50 Genotyping Beadchip (54,602 SNPs) or the Illumina® equine HD array (65,157 SNPs). For our study we used 45,986 SNPs that were common to both arrays. Quality control was performed with PLINK software (v1.9; Purcell, 2007; Purcell *et al.*, 2007). SNPs on the X chromosome or with a call-rate <90% were removed. All horses were retained as genotyping rate was >90%.

Estimated parameters to quantify relationships and diversity

Parameters to quantify genetic relationships and diversity within and among the populations were estimated using PLINK software v1.9 (Purcell, 2007; Purcell *et al.*, 2007). Inbreeding coefficients (f_i) were estimated based on expected and observed homozygosity (Table 1) after SNPs were pruned for linkage disequilibrium. Inbreeding coefficients ($f_{i,ROH}$) were also estimated as the proportion of the autosomal genome that was homozygous based on ROHs (McQuillan *et al.*, 2008). ROHs were identified using a sliding window approach. In short, at each window position it was determined whether the window was homozygous and fulfilled several criteria (e.g. no heterozygous genotypes allowed and a minimum of 50 subsequent homozygous SNPs to call an ROH; Table 2). ROHs aid in understanding breeding history as it enables to differentiate between more recent (long ROHs, limited recombination) and past inbreeding. The genetic relationships were examined using genome-wide identity-by-state distances between the 184 horses and visualized in a multidimensional scaling plot. To obtain quantitative measures of genetic differentiation between the populations, fixation index F_{ST} was estimated.

Results and discussion

The mean inbreeding coefficients across the populations were 8.8% based on expected and observed homozygosity and 9.7% based on ROHs (Table 1). In total 6,170 ROHs were identified with an average length of 6.5Mb (Table 2). The contribution of short, medium and long ROHs to inbreeding differed between populations (Figure 1). The first component of the multidimensional scaling plot separated the warmblood and coldblood populations, the second component separated the Friesian horse, Icelandic horse and Shetland pony populations from the other populations (Figure 2). Distinct clusters were found for most of the populations except the four Dutch warmblood horse populations. Limited genetic differentiation between these warmblood horse populations was also observed based on F_{ST} values (Figure 3), while differentiation from other populations was largest for the Shetland pony and Friesian horse populations. Most of the quantified relationships among populations were as expected based on their history and classifications of the populations (e.g. open or closed).

The Shetland pony and Icelandic horse populations clustered together (Figure 2) like in the analyses by van de Goor *et al.* (2011) and Petersen *et al.* (2013). These more ancient populations possibly shared ancestors back in time (Hendricks, 2007). Estimated inbreeding in the Shetland pony population was rather high (Table 1) and is comparable to estimations by van de Goor *et al.* (2011). Inbreeding within this population seemed of more recent origin as many long ROHs were observed (Figure 1). In contrast, inbreeding levels in the Icelandic horse population were low (Table 1) and many short ROHs contributed to inbreeding (Table 2), possibly representing inbreeding in the past. Similarly, Petersen *et al.* (2013) identified a high level of genetic diversity still present within the Icelandic horse population. They suggested that a large census population size contributed to the diversity despite bottlenecks caused by isolation of the breed for several hundreds of years and natural disasters like fluorine poisoning due to a volcano eruption in 1783 (Hendricks, 2007).

The limited genetic differentiation between the Dutch warmblood populations confirms their origin (Figure 2, 3). Groningen and Gelder horses were used for agricultural purposes in different areas of the Netherlands several decades ago. Gelder horses were introgressed with Hackney horses and American Saddlebred horses to breed harness horses with specific movement and posture, but also with Thoroughbred horses to breed sport horses. Inbreeding levels within the Groningen, Gelder and KWPN sport horse populations were low (Table 1) likely because introgression is allowed. Likewise, short ROHs determined a large part of inbreeding within the genome of these horses (Figure 1) and the on average shortest ROHs were found within these populations (Table 2). In the harness horse population introgression is allowed and performed, but estimated inbreeding levels were higher compared to the other Dutch warmblood populations (Table 1). The proportion of the genome covered by long ROHs was considerable, indicating recent inbreeding (Figure 1). Indeed, the recent inbreeding rate based on pedigrees was substantial (ΔF = 1.36%; Schurink *et al.*, 2012). Although introgression is allowed, it is not always executed optimally as well performing outbred stallions are often mated to many dams (Schurink *et al.*, 2012).

Inbreeding within the Friesian horse population was comparatively high (Table 1). Almost a quarter of all ROHs were identified in this population (Table 2) and inbreeding seemed recent (Figure 1). The Friesian horse population is a unique breed, largely isolated from the other populations, but least distant from the Belgian draft horses (Figure 2, 3), like in van de Goor *et al.* (2011) who investigated 35 horse populations using short tandem repeats. Ducro (2011) stated that inbreeding in the Friesian horse population is primarily caused by

genetic drift from a small effective population size during several generations since the foundation of the breed, but also unequal founder contribution, selection and bottlenecks contributed to inbreeding. Moderate levels of inbreeding were also observed in the Belgian draft horse population, comparable to estimates from van de Goor *et al.* (2011). The considerable contribution of long ROHs indicate a relative recent change in the breeding program of this population resulting in loss of diversity.

Genetic diversity and ROH characteristics within the Lipizzaner horse population were in between the Dutch warmblood horse populations harbouring the greatest diversity within this study and the more inbred coldblood populations (Table 1, 2 and Figure 1). The Lipizzaner horse population was genetically least distant from the Groningen horse population (Figure 2, 3). Despite their origin from Spanish horses, the Lipizzaner horses were genetically more distant from the Iberian populations compared to the Groningen horse population in the study by van de Goor *et al.* (2011).

Our study shows how admixture, drift and demography have shaped the genome-wide diversity of several horse populations in the Netherlands. Estimated diversity within and most of the quantified relationships among the populations were as expected based on their history and classifications, although some were less obvious (e.g. Lipizzaner horses). ROHs aided in understanding breeding history as it enabled us to differentiate between more recent and past inbreeding. Future research will determine which populations comprise unique genes and contribute significantly to the genetic diversity that is present to ultimately be able to make well-informed decisions in genetic conservation.

List of References

Bowling, A.T. & A. Ruvinsky, 2000. The genetics of the horse. CABI Publishing, Wallingford, Oxon, UK.

Ducro, B.J., 2011. Relevance of test information in horse breeding. PhD thesis, 2011. http://edepot.wur.nl/168395, ISBN 978-90-8585-855-3.

Hendricks, B.L., 2007. International Encyclopedia of Horse Breeds. University of Oklahoma Press, Norman, Oklahoma, USA.

Librado, P., A. Fages, C. Gaunitz *et al.*, 2016. The evolutionary origin and genetic makeup of domestic horses. Genetics 204: 423-434.

McCue, M.E., D.L. Bannasch, J.L. Petersen *et al.*, 2012. A high density SNP array for the domestic horse and extant Perissodactyla: utility for association mapping, genetic diversity, and phylogeny studies. PLoS Genet. 8: e1002451.

McQuillan, R., A.L. Leutenegger, R. Abdel-Rahman *et al.*, 2008. Runs of homozygosity in European populations. Am. J. Hum. Genet. 83: 359-372.

Petersen, J.L., J.R. Mickelson, E.G. Cothran *et al.*, 2013. Genetic diversity in the modern horse illustrated from genome-wide SNP data. PLoS One 8: e54997.

Purcell, S., 2007. PLINK v1.9, http://pngu.mgh.harvard.edu/purcell/plink/

Purcell, S., B. Neale, K. Todd-Brown *et al.*, 2007. PLINK: a toolset for whole-genome association and population-based linkage analysis. Am. J. Hum. Genetic. 81: 559-575.

Schurink, A., D.J.G. Arts & B.J. Ducro, 2012. Genetic diversity in the Dutch Harness horse population using pedigree analysis. Livest. Sci. 143: 270-277.

van de Goor, L.H.P., W.A. van Haeringen & J.A. Lenstra, 2011. Population studies of 17 equine STR for forensic and phylogenetic analysis. Anim. Genet. 42: 627-633.

Table 1. Inbreeding coefficient within and across 9 horse populations sampled in the Netherlands. Number of horses and mean, SD, minimum and maximum inbreeding coefficient based on observed and expected number of homozygous SNPs (f_i) and based on the proportion of autosomal genome that is homozygous using runs of homozygosity ($f_{i,ROH}$).

		Inbreeding coefficient (f_i)[1]				Inbreeding coefficient ($f_{i,ROH}$)[2]			
Population	n	Mean	SD	Min	Max	Mean	SD	Min	Max
Belgian draft horse	23	10.3	3.3	4.8	18.8	10.1	3.1	6.3	18.8
Friesian horse	20	25.5	3.7	18.9	33.5	22.3	4.5	9.3	30.2
Gelder horse	20	3.1	4.3	-3.7	13.9	5.9	2.8	2.1	12.3
Groningen horse	20	2.9	3.0	-3.9	9.2	6.2	2.3	2.9	12.3
Harness horse	20	8.4	5.0	1.6	18.4	9.7	4.1	5.2	16.9
Icelandic horse	20	5.9	2.9	1.5	10.2	4.1	2.7	0.6	9.4
KWPN sport horse	18	-1.5	3.2	-6.0	5.5	5.3	2.4	1.6	11.3
Lipizzaner horse	23	6.8	2.5	2.8	12.7	9.0	2.5	4.9	14.1
Shetland pony	20	17.4	6.9	8.7	36.4	14.4	6.6	6.6	33.1
Across populations	184	8.8	8.6	-6.0	36.4	9.7	6.4	0.6	33.1

[1] Inbreeding coefficient of each horse (f_i) was estimated as $f_i = \frac{(O_i - E_i)}{(L_i - E_i)}$, where $E_i = \sum_{j=1}^{L_i} 1 - 2p_j(1 - p_j)$, O_i is the observed number of homozygous SNPs for individual i, and E_i is the expected number of homozygous SNPs based on genotype data from all horses. L_i is the number of genotyped SNPs and p_j is the major allele frequency of a SNP at locus j (PLINK version 1.9 (Purcell, 2007; Purcell *et al.*, 2007)).

[2] Inbreeding coefficient of each horse based on the proportion of the autosomal genome that is homozygous ($f_{i,ROH}$) was estimated as $f_{i,ROH} = \sum L_{i,ROH} / L_{AUTOSOME}$, where $\sum L_{i,ROH}$ is the sum of the lengths of the ROHs for individual i and $L_{AUTOSOME}$ is the length of the autosomal equine genome covered by the SNP array (McQuillan *et al.*, 2008), in this case 2,242,939,370 bp.

Table 2. Total, mean, standard deviation, minimum and maximum number and length of ROHs (in Mb) within and across 9 horse populations sampled in the Netherlands.

	Number of ROH[1]					Length of ROH, Mb			
Population	Total	Mean	SD	Min	Max	Mean	SD	Min	Max
Belgian draft horse	836	36.3	6.8	25	53	6.2	5.5	1.6	39.9
Friesian horse	1,485	74.3	9.5	51	89	6.7	5.4	1.7	47.5
Gelder horse	443	22.2	7.0	13	39	6.0	5.1	1.9	55.4
Groningen horse	509	25.5	7.7	14	46	5.5	4.0	1.7	33.1
Harness horse	560	28.0	8.9	13	44	7.8	6.8	1.7	52.8
Icelandic horse	302	15.1	6.7	5	28	6.1	6.0	1.7	56.4
KWPN sport horse	400	22.2	6.7	10	34	5.4	4.2	1.9	32.8
Lipizzaner horse	729	31.7	5.6	22	43	6.4	4.9	1.9	48.2
Shetland pony	906	45.3	8.9	29	63	7.2	7.7	1.9	91.6
Across populations	6,170	33.5	18.1	5	89	6.5	5.8	1.6	91.6

[1] ROHs were called using a sliding window approach (PLINK version 1.9 (Purcell, 2007; Purcell *et al.*, 2007)). A window of 50 SNPs was slid across the genome, one SNP at a time. At each window position it was determined whether the window was homozygous. At maximum two missing genotype calls were allowed, but no heterozygous genotypes. Per SNP, the percentage of windows being homozygous was calculated and needed to be at least 5% to be defined present in a homozygous segment. Segments needed to contain at least 50 SNPs, had to be at least 100kb in length and had to contain at least 1 SNP per 5,000kb to be called an ROH. Distance between two adjacent SNPs was allowed to be at maximum 5,000kb.

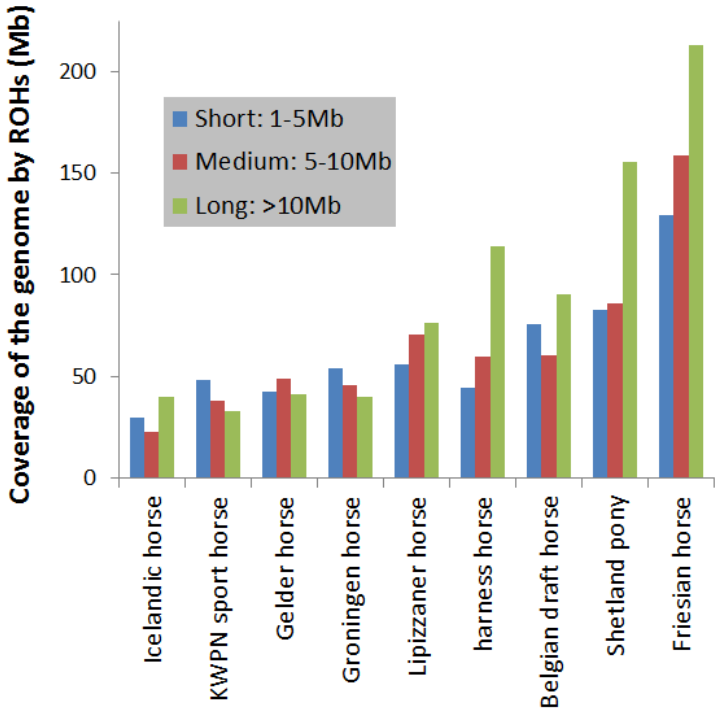

Figure 1. Mean total size of the genome (in Mb) covered by ROHs of different size per individual within 9 horse populations sampled in the Netherlands. For example, on average 213Mb of the genome of a Friesian horse was covered by long ROHs (each >10Mb in size).

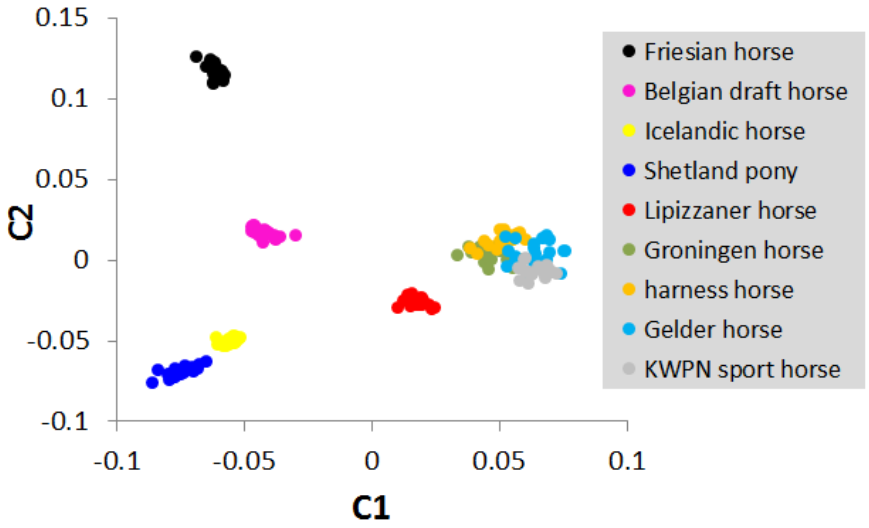

Figure 2. Multidimensional scaling plot of 184 horses from 9 populations sampled in the Netherlands. The first component (C1) explained 8.8% of the variance present, the second (C2) 7.2%.

	Belgian draft horse	Lipizzaner horse	Friesian horse	Groningen horse	Harness horse	Gelder horse	KWPN sport horse	Icelandic horse	Shetland pony
Belgian draft horse		0.122	0.163	0.098	0.116	0.104	0.097	0.110	0.136
Lipizzaner horse	8,091		0.183	0.097	0.121	0.106	0.084	0.118	0.143
Friesian horse	7,751	7,707		0.155	0.171	0.164	0.155	0.176	0.202
Groningen horse	8,151	8,062	7,827		0.070	0.055	0.045	0.100	0.124
Harness horse	8,063	7,997	7,646	7,987		0.065	0.076	0.119	0.145
Gelder horse	8,083	8,012	7,723	7,990	7,905		0.049	0.108	0.132
KWPN sport horse	8,316	8,229	8,090	8,200	8,183	8,173		0.094	0.117
Icelandic horse	8,240	8,228	7,954	8,279	8,195	8,236	8,421		0.101
Shetland pony	8,105	8,051	7,727	8,118	8,020	8,073	8,297	8,101	

F_{ST} value

0.05

0.10

0.15

0.20

Figure 3. Genetic differentiation between 9 populations sampled in the Netherlands. Mean F_{ST} values between the populations (above diagonal) and number of SNPs used to estimate these values (below diagonal) based on a dataset pruned for LD.