ANIMAL GENETICS · ORIGINAL PAPER

Genetic and genomic characterization followed by single-step genomic evaluation of withers height in German Warmblood horses

Sarah Vosgerau[1] · Nina Krattenmacher[1] · Clemens Falker-Gieske[2] · Anita Seidel[1] · Jens Tetens[2,3] · Kathrin F. Stock[4,5] · Wietje Nolte[6,7] · Mirell Wobbe[4,5] · Iulia Blaj[1] · Reinhard Reents[4] · Christa Kühn[6,8] · Mario von Depka Prondzinski[9] · Ernst Kalm[1] · Georg Thaller[1]

Abstract

Reliability of genomic predictions is influenced by the size and genetic composition of the reference population. For German Warmblood horses, compilation of a reference population has been enabled through the cooperation of five German breeding associations. In this study, preliminary data from this joint reference population were used to genetically and genomically characterize withers height and to apply single-step methodology for estimating genomic breeding values for withers height. Using data on 2113 mares and their genomic information considering about 62,000 single nucleotide polymorphisms (SNPs), analysis of the genomic relationship revealed substructures reflecting breed origin and different breeding goals of the contributing breeding associations. A genome-wide association study confirmed a known quantitative trait locus (QTL) for withers height on equine chromosome (ECA) 3 close to *LCORL* and identified a further significant peak on ECA 1. Using a single-step approach with a combined relationship matrix, the estimated heritability for withers height was 0.31 (SE = 0.08) and the corresponding genomic breeding values ranged from − 2.94 to 2.96 cm. A mean reliability of 0.38 was realized for these breeding values. The analyses of withers height showed that compiling a reference population across breeds is a suitable strategy for German Warmblood horses. The single-step method is an appealing approach for practical genomic prediction in horses, because not many genotypes are available yet and animals without genotypes can by this way directly contribute to the estimation system.

Keywords Genomic selection · Horse breeding · Withers height · Single-step · German Warmblood breeds

✉ Sarah Vosgerau
  svosgerau@tierzucht.uni-kiel.de

1   Institute of Animal Breeding and Husbandry, Kiel University, Olshausenstr. 40, 24098 Kiel, Germany

2   Department of Animal Science, University of Göttingen, Burckhardtweg 2, 37077 Göttingen, Germany

3   Center for Integrated Breeding Research (CiBreed), University of Göttingen, Albrecht-Thaer-Weg 3, 37075 Göttingen, Germany

4   IT Solutions for Animal Production (Vit), Heinrich-Schroeder-Weg 1, 27283 Verden, Germany

5   Institute for Animal Breeding and Genetics, University of Veterinary Medicine Hannover, Buenteweg 17, 30559 Hannover, Germany

6   Institute of Genome Biology, Leibniz Institute for Farm Animal Biology (FBN), Wilhelm-Stahl-Allee 2, 18196 Dummerstorf, Germany

7   Saxon State Office for Environment, Agriculture and Geology, Schlossallee 1, 01468 Moritzburg, Germany

8   Faculty of Agricultural and Environmental Sciences, University Rostock, Justus-von-Liebig-Weg 6, 18059 Rostock, Germany

9   Werlhof-Institut MVZ, Schillerstr. 23, 30159 Hannover, Germany

Introduction

Genomic selection was first proposed 20 years ago by Meuwissen et al. (2001). About 10 years later, advances in molecular genetic technologies and bioinformatics have enabled the implementation of routine genomic evaluation and selection. In the main livestock species, genomic selection is widely applied, whereas in horses, it is still a scarcely used breeding tool. However, the potential of genomic selection in sport horse breeding is high due to the long generation interval (Haberland et al. 2012). The availability of sufficiently large reference populations for the most important breeding goal traits is a major limiting factor for developing genomic selection in horses. Among other parameters such as heritability of the target traits and genetic relationship, the size of the reference population is crucial for the reliability of, and thus, the benefit from genomic breeding values (Meuwissen et al. 2001; Goddard 2009; Van den Berg et al. 2019). In sport horse breeding, where many populations are too small to establish a sufficiently large reference population, collaborative approaches and the development of a reference population across genetically linked breeds are highly reasonable. In dairy cattle, international cooperations were also successfully used to increase the number of animals in the reference population and enable genomic selection in smaller populations (Lund et al. 2016). In order to meet this requirement, a joint project of five German Warmblood breeding associations and partners from science and industry was launched in 2017. The aim of this project is to implement genomic selection in horses by establishing a suitable reference population across genetically linked populations of distinct breeds of Warmblood horses.

To verify the approach and develop the genomic evaluation system, it was decided to use a reference trait as a starting point, similar to milk performance traits in the first application of genomic prediction in dairy cattle. Due to its clear definition, objective measurement, and known genetic characteristics, withers height was chosen for this purpose. In the literature, heritability estimates for withers height of 0.72 for Franches-Montagnes horses (Signer-Hasler et al. 2012), 0.53 for German Warmblood horses (Tetens et al. 2013) and 0.49 for Hanoverian Warmblood horses (Stock and Distl 2006) were reported. A genome-wide association study with 1077 Franches-Montagnes horses revealed two trait-influencing genomic regions (quantitative trait locus (QTL)) on an equine chromosome (ECA) 3 and ECA 9 for withers height (Signer-Hasler et al. 2012). In particular, the QTL on ECA 3 localized near the *LCORL* (ligand-dependent nuclear receptor corepressor-like) gene has later been confirmed in other studies and horse populations (Metzger et al. 2013; Tetens et al.

2013). Furthermore, the *LCORL* gene is known to play a significant role in body size in a wider range of species, including for example humans (Lango Allen et al. 2010) and cattle (Pryce et al. 2011).

A single-step approach for estimating genomic breeding values has been proposed to make optimal use of available data (Aguilar et al. 2010; Christensen and Lund 2010). With this method, all available genotype, phenotype, and pedigree information can be employed simultaneously (Legarra et al. 2014). Not only the performance information from the genotyped but also from the non-genotyped animals is directly included. All existing information can thus be used optimally and not only genotyped animals contribute to the reference population. Since systematic SNP genotyping is not yet common in the horse, the situation with far more non-genotyped than genotyped animals is likely to persist for a longer time period even in those populations interested in implementing genomic selection. Therefore, the use of single-step methodology can be seen as particularly relevant and straightforward for this species (Mark et al. 2014; Stock et al. 2016).

The objective of this study was to genetically and genomically characterize withers height and to develop a system for estimating genomic breeding values for withers height in German Warmblood horses using the single-step approach and a reference population comprising horses from five distinct breeds. For this purpose, preliminary data reflecting the overall structure of the reference population were evaluated.

Material and methods

Dataset

For this study, 2157 mares born between 1997 and 2015 from five different German Warmblood horse breeds were both, genotyped, and phenotyped. The distribution of breeds in the dataset was as follows: 1001 mares from Holstein (HOL), 453 from Oldenburg (OL), 100 from Oldenburg International (OS), 427 from Trakehner (TRAK), and 176 from Westfalian (WESTF) studbook. The two Oldenburg breeding associations were considered separately, as they have a different focus with OS predominantly concentrating on the breeding of show jumping horses and OL putting more emphasis on dressage ability. Selection of horses contributing to the reference population was based upon two criteria: (i) a low level of preselection (which was achieved by selecting mares instead of stallions) and (ii) the current genetics of German riding and sport horses should be represented as comprehensive as possible. Therefore, from a larger number of broodmares, those with the lowest pedigree relationships were selected. Genomic data were made available through molecular genetic analyses of genome-wide

single nucleotide polymorphisms (SNPs) using commercially available SNP arrays. The GGP Equine 70 k BeadChip® (Neogen/Illumina) containing 65,157 SNPs was used for genotyping of 193 mares, in the following referred to as the first cohort. The remaining 1964 mares (second cohort) were genotyped using the GGP Equine Plus BeadChip® (Neogen/Illumina) which became available more recently and included 6790 additional SNPs, so in total 71,947 SNPs.

Quality control and filtering were performed separately for each cohort using the software PLINK v1.90b6.4 (Purcell et al. 2007). Only autosomal SNPs were considered in the subsequent analyses. Genotyping results were filtered based on Illumina's GenCall score and applying a threshold of 0.15, which resulted in a mean call rate higher than 0.95 of those SNPs kept for analyses. Markers with a minor allele frequency less than 0.01 and a genotyping rate less than 90% as well as SNPs with a significant deviation from Hardy–Weinberg-Equilibrium ($p(\chi^2) < 0.001$) were discarded. Additionally, duplicate SNPs ($n = 55$) were removed. After quality control, the numbers of remaining SNPs were 58,562 for the first cohort and 62,070 for the second cohort. To generate a homogeneous SNP dataset, all animals were imputed to a uniform level using the program Beagle 5.0 (Browning et al. 2018); thus, the additional SNPs on the GGP Equine Plus BeadChip® are also available for the first cohort.

The final data set used for further analyses included 2113 mares (HOL ($n = 982$), OL ($n = 444$), OS ($n = 98$), TRAK ($n = 416$), and WESTF ($n = 173$)) and 62,070 SNPs. Pedigree information contained three ancestral generations, resulting in a total pedigree size of 14,632 individuals. In particular, the selected mares descended from 608 sires with 1 to 82 offspring (mean 3.5) and from 2006 mares with 1 to 3 offspring (103 mares with 2 offspring, 2 mares with 3 offspring). Table 1 gives an overview on the numbers of sires and dams overall and by breeding association. The most recent measurements of withers height were for 1991 mares from evaluations at studbook registration (SBR) and for 122 mares from supplementary data recording in connection with the mare performance test (MPT). Measurements were taken between 2014 to 2020, when the mares were on average 4.4 ± 2.6 years old. Withers height ranged from 154 to 181 cm with a mean of 167.6 ± 3.5 cm. Figure 1 and Table 1 give more detailed information on the distribution of the investigated trait.

Genomic analyses

To examine the suitability of a joint reference population across German Warmblood breeds, the SNP-derived genomic population structure among the 2113 mares was analyzed. First, a principal component analysis (PCA) of the realized genomic relationship matrix across the different subpopulations was conducted using GCTA 1.91.6 beta1 (Yang et al. 2011). Second, using PLINK (Purcell et al. 2007), the linkage disequilibrium (LD; r^2 according to Hill and Robertson 1968) and allele frequencies were determined for the individual breeding associations. Subsequently, a genome-wide association scan for loci affecting withers height was performed to confirm the major QTL for withers height mapped on ECA 3 (Signer-Hasler et al. 2012; Metzger et al. 2013; Tetens et al. 2013). Therefore, a genome-wide association study (GWAS) was conducted with a leave-one-chromosome-out (LOCO) approach in GCTA (Yang et al. 2011) and the following mixed linear model:

$$y_{ijk} = \mu + \text{age class}_i + \beta \times \text{SNP}_j + g^- + e_{ijk}$$

where y_{ijk} is the phenotype (mares' withers height; $k = 1, \ldots, 2,113$); μ is the overall mean; age class$_i$ represents the fixed effect of the ith age class at measurement ($i = 1, 2, 3, 4, 5$), the first age class contains mares younger than 3 years, the second class includes mares between 3 and 4 years old, the third between 4 and 5 years old, the fourth between 5 and 6 years old, and the fifth group includes all mares older than 6 years old; β is the effect of the candidate SNP; SNP$_j$ represents the SNP genotype indicator variable ($j = 0, 1, 2$; number of the respective target alleles of the SNP); g^- is

Table 1 Distribution of the 2113 mares in the reference population by breeding association (*HOL* Holstein horse breed, *OL* Oldenburger horse breed, *OS* Oldenburger International horse breed, *TRAK* Trakehner horse breed, *WESTF* Westfalian horse breed) and across paternal and maternal progeny groups and descriptive statistics for withers height

Breeding association	n	Sires	Dams	Mean	SD	Min	Max
HOL	982	243	933	167.8	3.6	154	181
OL	444	148	424	168.3	3.1	160	180
OS	98	70	91	167.4	3.1	160	175
TRAK	416	145	389	166.6	3.4	157	178
WESTF	173	97	170	167.7	3.1	160	178
Total	2113	608	2006	167.6	3.5	154	181

Mean, standard deviation (SD), minimum (Min), and maximum (Max) were given in centimeters

Fig. 1 Distribution of withers height for 2113 mares. *HOL* Holstein horse breed ($n = 982$), *OL* Oldenburger horse breed ($n = 444$), *OS* Oldenburger International horse breed ($n = 98$), *TRAK* Trakehner horse breed ($n = 416$), *WESTF* Westfalian horse breed ($n = 173$)
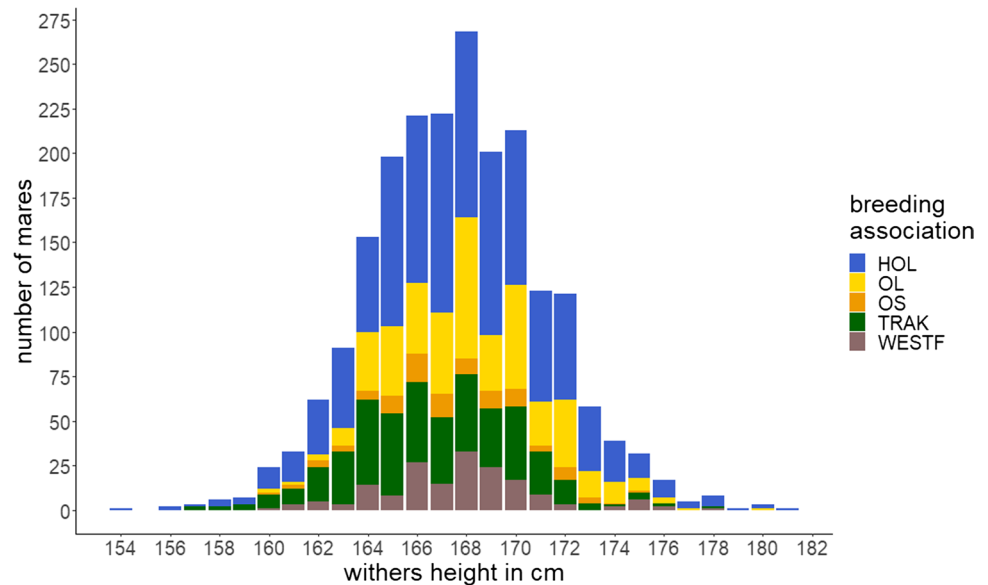
the additive genetic effect based on a genomic relationship matrix calculated using all SNPs except those on the chromosome where the candidate SNP is located; and e_{ijk} is the random residual effect. To partition the variance onto different chromosomes and the SNPs identified as significant, a genomic relationship matrix was built separately for the 31 autosomes and the significant SNPs in the GCTA software (Yang et al. 2011). Variance components were estimated using the restricted maximum likelihood analysis. In addition, a conditional association study was performed for the data set. Therefore, single highly associated SNPs were included as fixed effects in the LOCO analysis.

Genomic prediction

For the estimation of genomic breeding values, a single-step approach was applied. Therefore, pedigree, genotype, and phenotype information were combined for a joint evaluation in an animal model with the BGLR package (Pérez and de los Campos 2014) in the R statistical software (R Core Team 2019). Phenotypes were only available for the genotyped mares ($n = 2113$). First, the numerator relationship matrix (**A**) based on the pedigree information ($n = 14,632$) and the realized genomic relationship matrix (**G**) based on the genotyped animals ($n = 2113$) were built and used to generate a combined relationship matrix (**H**) for all individuals (Legarra et al. 2009). For this approach, the 14,632 horses can be divided into two groups. The first group with non-genotyped horses ($n = 12,519$) and the second group with genotyped horses ($n = 2113$). The inverse of the combined matrix can be computed as (Aguilar et al. 2010; Christensen and Lund 2010):

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}$$

where $\mathbf{A}^{-1}$ is the inverse of the pedigree relationship matrix for all individuals; $\mathbf{G}^{-1}$ is the inverse of the genomic relationship matrix for genotyped individuals; and $\mathbf{A}_{22}^{-1}$ is the inverse of the pedigree relationship matrix for all genotyped individuals. The single-step model can be written as:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e}$$

where **y** is the phenotype vector (mares' withers height); **X** is a design matrix for the effects of age class; **b** is the corresponding vector of fixed effects; **Z** is a design matrix for the random animal effects; **a** is the corresponding vector of additive genetic random effects; and **e** is the random residual effect. To improve the computation speed, the matrix **H** can be transformed such that the random animal effects become independent (Waldmann et al. 2008). To this end, a Cholesky decomposition was used, which led to an equivalent animal model to obtain genomic predictions in a single step:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{F}_a\mathbf{c}_a + \mathbf{e}$$

where $\mathbf{F}_a = \mathbf{ZH}^{1/2}$ and $\mathbf{c}_a = \mathbf{H}^{-1/2}\mathbf{a}$. The reliabilities of the individual genomic breeding values were computed with the following equation:

$$r^2 = 1 - \frac{\text{PEV}}{\sigma_a^2}$$

where PEV is the prediction error variance of the breeding values and σ_a^2 is the additive genetic variance, both estimated in BGLR (Pérez and de los Campos 2014).

Results

The present data set was used to verify whether a reference population across breeds could be applied for genomic selection in German Warmblood horses. For this purpose, the genomic relationship was analyzed. Afterwards, this reference population was used to check if a known signal for withers height could be found on ECA 3. Furthermore, a genomic breeding value estimation was performed using a single-step approach to determine whether acceptable reliabilities for withers height could be achieved.

Principal component analysis

The first two eigenvectors of the genomic relationship matrix were used to visualize the relationship at the genomic level. Figure 2 shows the results of the principal component analysis for the 2113 mares. The first principal component (PC1) explained 2.85% and the second principal component (PC2) explained 1.49% of the variation in the data. The PCA revealed three different clusters for the five breeds HOL, OL, OS, TRAK, and WESTF. Substantial overlap was found for OL and WESTF. HOL overlapped with OS along PC1 and PC2. TRAK displayed the most distinct separation. Analyses of the LD (measured as r^2) and allele frequencies underlined the results described above, as the breeds within the clusters showed similar values. This is exemplarily shown for the top associated SNPs in Supplementary table S1.

GWAS for withers height

The GWAS in 2113 mares revealed a single genome-wide significant signal on ECA 3, with the three SNPs BIEC2_808543 (base pair position (EquCab 2.0) 105,547,002 bp), BIEC2_808500 (105,363,241 bp), and BIEC2_808466 (105,163,077 bp) showing the highest negative decadic logarithms of the empirical p-values ($-\log_{10}(p)$ values; Fig. 3). The SNP BIEC2_808543, which is highest associated with withers height, is located near the *LCORL* gene and had a minor allele frequency of 0.49. On ECA 1, the SNP BIEC2_23633 (56,208,092 bp) reached the significance level of $P \leq 0.05$. Analyses of the SNP effects indicated that a few SNPs are strongly associated with the expression of withers height in horses while many SNPs can only explain a minor part of the genetic variance for the trait. The three SNPs on ECA 3 mentioned above together explained 22.8% of the variance. By including the SNP BIEC2_808543 as a fixed effect in the LOCO analysis, the significant signal on ECA 3 dropped completely, only the peak on ECA 1 was above the significant threshold (Fig. 4).

Genomic breeding values for withers height

Variance component estimation was done for all 14,632 horses in the pedigree using the transformed matrix H, resulting in a value of 3.95 for the additive genetic variance and a value of 8.62 for the residual variance, and, thus, a heritability of 0.31 with a standard error (SE) of 0.08. The genomic breeding values estimated in this study ranged from − 2.94 to 2.96, with the majority of horses having a value close to 0. The genomic breeding values and the underlying phenotype showed a correlation of $r = 0.97$ for the 2113 mares (Fig. 5). The realized reliabilities (r^2) for the genomic breeding values are shown in Table 2. Figure 6 gives an additional overview of the reliabilities for the genotyped and phenotyped mares and the sires of these mares grouped according to the number of phenotyped offspring.

Fig. 2 Results of the principal component analysis for 2113 mares. *HOL* Holstein horse breed ($n=982$), *OL* Oldenburger horse breed ($n=444$), *OS* Oldenburger International horse breed ($n=98$), *TRAK* Trakehner horse breed ($n=416$), *WESTF* Westfalian horse breed ($n=173$)

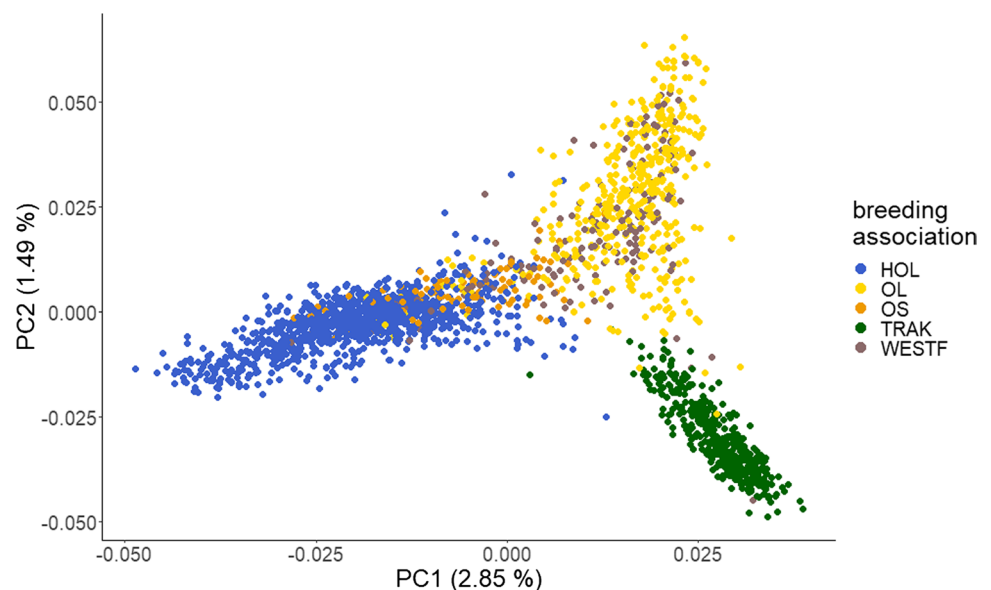**Fig. 3** Results of the genome-wide association study for withers height for 2113 mares. *HOL* Holstein horse breed (*n* = 982), *OL* Oldenburger horse breed (*n* = 444), *OS* Oldenburger International horse breed (*n* = 98), *TRAK* Trakehner horse breed (*n* = 416), *WESTF* Westfalian horse breed (*n* = 173). The blue line represents the significance threshold of *P* = 0.01. The red line represents the significance threshold of *P* = 0.05
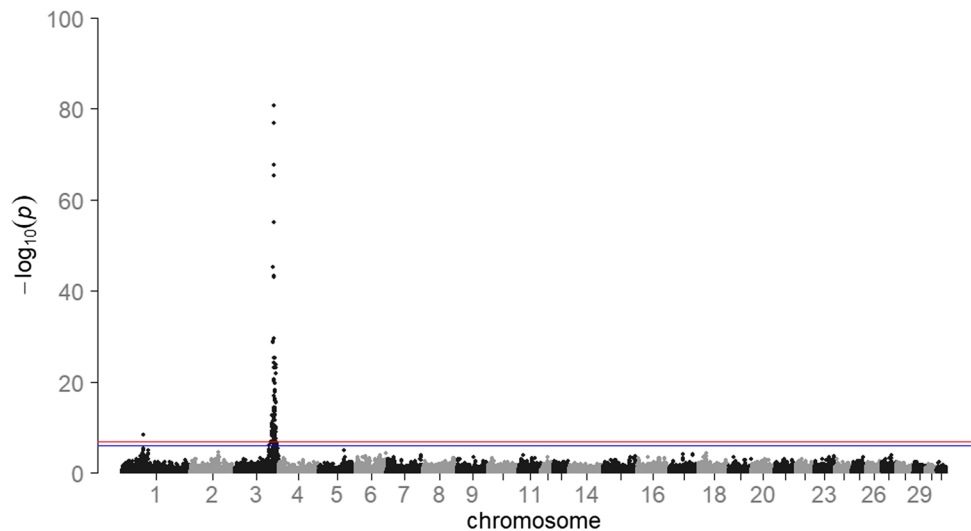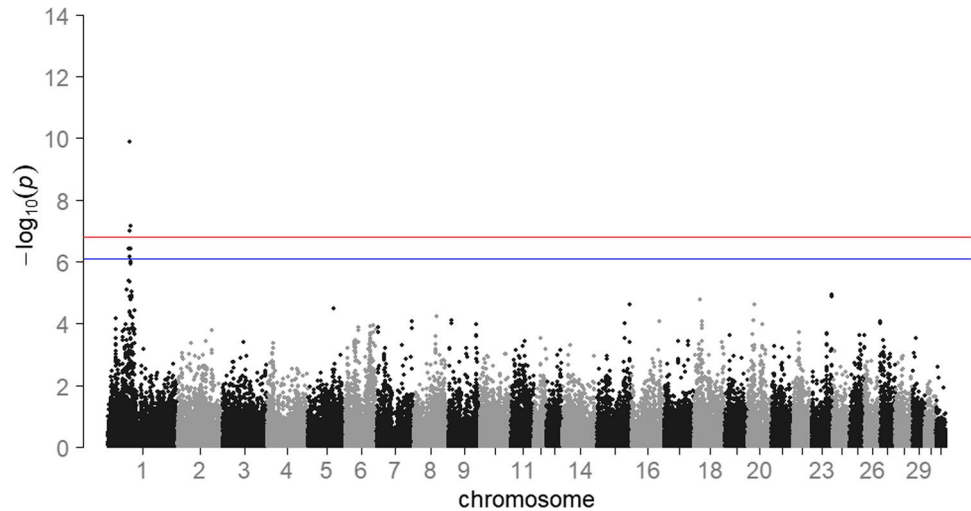

Fig. 4 Results of the conditional genome-wide association study for withers height for 2113 mares. *HOL* Holstein horse breed (*n* = 982), *OL* Oldenburger horse breed (*n* = 444), *OS* Oldenburger International horse breed (*n* = 98), *TRAK* Trakehner horse breed (*n* = 416), *WESTF* Westfalian horse breed (*n* = 173). The blue line represents the significance threshold of *P* = 0.01. The red line represents the significance threshold of *P* = 0.05

Fig. 5 Correlation between genomic breeding value and withers height for 2113 mares. *HOL* Holstein horse breed (*n* = 982), *OL* Oldenburger horse breed (*n* = 444), *OS* Oldenburger International horse breed (*n* = 98), *TRAK* Trakehner horse breed (*n* = 416), *WESTF* Westfalian horse breed (*n* = 173)
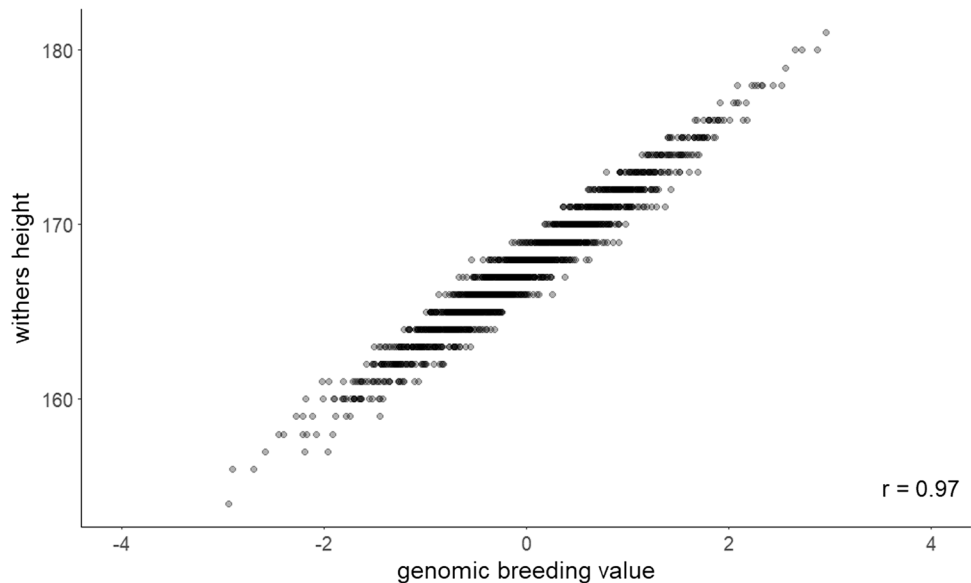
Table 2 Realized reliabilities (r^2) for the genomic breeding values for withers height for 2113 mares

(Sub)sample	n	Mean	Min	Max
All horses in the pedigree	14,632	0.13	0	0.89
Horses without phenotype an genotype	12,519	0.09	0	0.89
Horses with phenotype and genotype	2113	0.38	0.14	0.50

Discussion

In this study, 2113 mares with 62,070 SNPs and data for withers height were available. Data was contributed by five different breeding associations which were not equally represented in the present reference population. Instead, breeding associations with a large total number of broodmares were stronger represented in the reference population than breeding associations with only a small number of broodmares.

Genomic relationship and potentials and limitations for building a joint reference population

The data set consisted of data from five different horse breeding associations, all of them being subpopulations of the German Warmblood horse. In Germany, conventional breeding values for dressage and jumping performance are routinely estimated jointly for all Warmblood horse breeds. Thus, a joint reference population is the obvious solution to guarantee a fast implementation of genomic selection, especially because most of the subpopulations are much too small to build their own reference population. In dairy cattle breeding, several studies have shown that establishing a joint reference population across related breeds can increase the

reliability of genomic breeding values, especially for small breeds (Brøndum et al. 2011; Heringstad et al. 2011; Zhou et al. 2014). However, each breeding association has its own breeding goal and characteristics. On the other hand, there is a growing effort to record new traits (e.g., linear and health traits) using standardized protocols across the different breeding associations.

Figure 2 shows that horses are genomically similar across breeds. Therefore, horses from the five breeding associations and subpopulations used in this study can safely be regarded as one large cohort. Horses from breeding associations with a focus on similar characteristics, e.g., aptitude for a certain equestrian discipline, are most closely related on the genomic level. This is most probably due to the shared use of appropriately qualified sires approved for more than one breeding association. HOL and OS, which both put a strong emphasis on jumping ability in their breeding program, were more related to each other than to OL, WESTF, and TRAK. OL and WESTF, on the other hand, put more weight on the suitability of horses for dressage; this was also reflected in a close relationship at the genomic level. TRAK horses showed comparatively little genomic relationships to the horses registered by the other associations, with a closer relationship to OL and WESTF than to HOL and OS. This result can be explained by the fact that TRAK is the only horse breed bred in accordance with the principles of pure breeding, existing as a partially closed population since 1732. In TRAK, only introgression of English and Arabian Thoroughbred, Shagya, and Anglo-Arabian as breeding animals is allowed (Teegen et al. 2009). These breeds are also used in many other Warmblood breeding associations to refine their horses in a certain direction. In addition to TRAK, the HOL breed also has a closed studbook. On the mare side, it is strictly closed. Currently, no non-Holstein

Fig. 6 Reliabilities for the genotyped and phenotyped mares (n=2113; red points) and the sires of these mares (n=608; blue points) grouped according to the number of phenotyped offspring
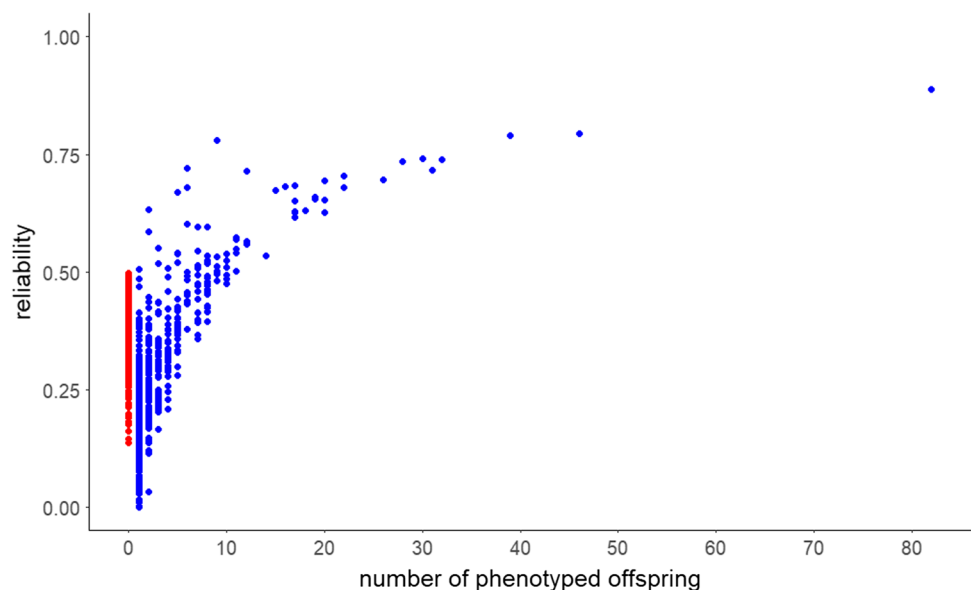
mares can be registered in the studbook. On the stallion side, the breeding is also characterized by the use of English Thoroughbreds and the Arabian horse breed. The use of stallions with foreign blood is minimized. Only a few foreign sires that show exceptional performance in sport and/or in breeding are approved (Roos et al. 2015). The other three breeding associations (OL, OS, and WESTF) are characterized by a more open breeding program and allow the use of foreign sires and dams from other Warmblood breeds, e.g., from one another and from HOL and TRAK.

Withers height as a reference trait

In sport horse breeding, there are many traits that are difficult to capture or are strongly subjective. Since the success of genomic breeding value estimation and selection strongly depends on the accurate recording of the trait (Daetwyler et al. 2010), withers height is particularly suitable as a reference trait as it is a well-studied as well as an easily and objectively measurable trait. Height, like many other quantitative traits in livestock breeding, is genetically determined by a few SNPs with major and many SNPs with small effects (Hayes and Goddard 2001). In this study, one previously reported genomic region including an important QTL for withers height in horses was confirmed. We identified the SNP BIEC2_808543 on ECA 3 as the SNP highest associated with withers height. In a study of 214 Hanoverian horses by Metzger et al. (2013), the same SNP was significantly associated with withers height with a similar minor allele frequency of 0.45 (vs. 0.49 in the present study). The results of the GWAS confirmed the assumption that there is a genetic variant close to the *LCORL* gene that plays a major role in the expression of withers height in horses (Signer-Hasler et al. 2012; Metzger et al. 2013; Tetens et al. 2013). The association signal could be identified consistently across the study populations. To gather further statistical evidence for the causality of the QTL on ECA 3, a conditional association analysis was conducted (Cohen-Zinder et al. 2005). When including the top SNP (BIEC2_808543) as a fixed effect, no peak was observed above the significance threshold on ECA 3. This suggests that there is only one QTL responsible for withers height in this region. Besides the expected signal on ECA 3, a single SNP on ECA 1 was also significantly associated with withers height. This peak also persisted and got even more pronounced in the conditional GWAS meaning that there is additional genetic variation due to this chromosome. Signer-Hasler et al. (2012) reported that a large fraction of the variance for height in horses could be explained by ECA 1, although they did not find any significantly associated SNPs in this area. A further QTL on ECA 9, also found in the latter study, could not be identified with the present dataset.

The analysis of variance revealed a heritability of 0.31 (SE = 0.08) for the 2113 mares when using the combined relationship matrix, whereas the SNP-based heritability was higher (0.50; SE = 0.05) and more similar to the estimates reported in the literature. Signer-Hasler et al. (2012) estimated a SNP-based heritability of 0.72 for 1077 Franches-Montagnes horses using the Equine 50 k BeadChip® (Illumina). Withers height of the horses ranged from 145 to 165 cm. Tetens et al. (2013) also used the Equine 50 k BeadChip® (Illumina) and calculated a heritability of 0.53 for 782 German Warmblood stallions with a withers height of 158 to 176 cm. In a study by Stock and Distl (2006), a heritability of 0.49 (SE = 0.02) was determined for 20,768 Hanoverian horses with a withers height ranging between 143 and 185 cm.

The objective of this study was to quantify the reliability for first genomic breeding values for the reference trait withers height. For this purpose, a single-step procedure was used. In horse breeding, this is the method of choice because the information from non-genotyped horses can also be incorporated directly and also information from genotyped horses (which are not phenotyped). Especially in the initial phase of genomic selection, this is an advantage, because not many genotypes of horses are available yet (Mark et al. 2014; Stock et al. 2016). Unfortunately, the full potential of this method could not be exploited here because only phenotypes of genotyped horses were available. With the present data set, genomic breeding values for the 2113 mares could be estimated with an average reliability of 0.38. Running a classical BLUP without genomic data yielded in average reliabilities of 0.35 for the same animals. Thus, reliabilities do not increase much by including genotype information. Advantages would possibly be more pronounced, if none phenotyped horses that were genotyped were included in the data, which was however not done in this study. Anyway, the moderate reliability of 0.38 for the genomic breeding values was within the expected range and is comparable to the initial phase of genomic breeding value estimation in dairy cattle breeding for equivalent traits (Luan et al. 2009; VanRaden et al. 2009). Considering all animals in the pedigree, the highest reliability estimates for the genomic breeding values were achieved for non-genotyped and non-phenotyped sires with several genotyped and phenotyped daughters. One sire was represented by 82 offspring in the reference population which resulted in the most reliable breeding value (maximum reliability of 0.89; cf. Figure 6). As expected, the daughters of this sire had the highest reliabilities among the genotyped horses with a mean value of 0.45. It should be noted that many non-genotyped but also some genotyped horses had a reliability of the genomic breeding value close to zero. The latter was observed for mares with genotypes and phenotypes from all five breeding associations and is probably due to the fact that

phenotypes were only available for genotyped horses in the reference population, and, furthermore, could be attributed to the selection of mares for the reference population. Here, care was taken to ensure that the animals were as unrelated as possible to each other, thereby, aiming to establish a reference population that represents almost the entire gene pool. However, as a result, relatedness within and between the populations was low with only a small number of common sires. Further enlargement of the current reference population should be done following a selection strategy which puts more focus on close enough relationship structures, e.g., by choosing more mares descending from stallions that were used as sires in more than one subpopulation. In a simulation study by Plieschke et al. (2016), selective genotyping of daughters of certain bulls increased the reliability of predicted breeding values. In dairy cattle breeding, the number of genotyped reference bulls with daughter proofs has been shown to be the most important parameter in determining the reliability of genomic predictions for selection candidates (Goddard and Hayes 2009). A targeted genotyping of stallions from the current breeding population and measurement of their withers height would be useful to further increase the reliability of the calculated genomic breeding values. This is important because low reliability can limit the genetic gain, especially for young animals that are not progeny tested (Thomasen et al. 2014).

In general, our study demonstrates the benefits of using genomic information in horse breeding and shows that using a joint reference population is a suitable strategy for German Warmblood horses. However, the reference population should be further enlarged. Particular attention will be given to the effects of the different subpopulations in further studies aiming to optimize the structure of the reference population. This will, in turn, result in better reliabilities of the genomic breeding values and, thus, increase the acceptance of the horse breeders, which is important for a successful implementation for genomic selection.

Declarations

References

Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ (2010) Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J Dairy Sci 93:743–752. https://doi.org/10.3168/jds.2009-2730

Brøndum RF, Rius-Vilarrasa E, Strandén I, Su G, Guldbrandtsen B, Fikse WF et al (2011) Reliabilities of genomic prediction using combined reference data of the Nordic Red dairy cattle populations. J Dairy Sci 94:4700–4707. https://doi.org/10.3168/jds.2010-3765

Browning BL, Zhou Y, Browning SR (2018) A one-penny imputed genome from next generation reference panels. Am J Hum Genet 103:338–348. https://doi.org/10.1016/j.ajhg.2018.07.015

Christensen OF, Lund MS (2010) Genomic prediction when some animals are not genotyped. Genet Sel Evol 42:2. https://doi.org/10.1186/1297-9686-42-2

Cohen-Zinder M, Seroussi E, Larkin DM, Loor JJ, Everts-van der Wind A, Lee JH et al (2005) Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. Genome Res 15:936–944. https://doi.org/10.1101/gr.3806705

Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA (2010) The impact of genetic architecture on genome-wide evaluation methods. Genetics 185:1021–1031. https://doi.org/10.1534/genetics.110.116855

Goddard M (2009) Genomic selection: prediction of accuracy and maximisation of long term response. Genetica 136:245–257. https://doi.org/10.1007/s10709-008-9308-0

Goddard ME, Hayes BJ (2009) Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat Rev Genet 10:381–391. https://doi.org/10.1038/nrg2575

Haberland AM, König von Borstel U, Simianer H, König S (2012) Integration of genomic information into sport horse breeding programs for optimization of accuracy of selection. Animal 6:1369–1376. https://doi.org/10.1017/S1751731112000626

Hayes BJ, Goddard ME (2001) The distribution of the effects of genes affecting quantitative traits in livestock. Genet Sel Evol 33:209–229. https://doi.org/10.1186/1297-9686-33-3-209

Heringstad B, Su G, Solberg TR, Guldbrandtsen B, Svendsen M, Lund MS (2011) Genomic predictions based on a joint reference population for Scandinavian red breeds. In: Book of Abstracts of the 62nd Annual Meeting of the European Federation of Animal Science. 29 August-2 September 2011; Stavanger. https://www.wageningenacademic.com/pb-assets/wagen/files/e-books/open_access/9789086867318eaap2011-e.pdf/

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. Theor Appl Genet 38:226–231

Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F et al (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature 467:832–838. https://doi.org/10.1038/nature09410

Legarra A, Aguilar I, Misztal I (2009) A relationship matrix including full pedigree and genomic information. J Dairy Sci 92:4656–4663. https://doi.org/10.3168/jds.2009-2061

Legarra A, Christensen OF, Aguilar I, Misztal I (2014) Single step, a general approach for genomic selection. Livest Sci 166:54–65. https://doi.org/10.1016/j.livsci.2014.04.029

Luan T, Woolliams JA, Lien S, Kent M, Svendsen M, Meuwissen THE (2009) The accuracy of genomic selection in Norwegian Red cattle assessed by cross-validation. Genetics 183:1119–1126. https://doi.org/10.1534/genetics.109.107391

Lund MS, van den Berg I, Ma P, Brøndum RF, Su G (2016) Review: how to improve genomic predictions in small dairy cattle populations. Animal 10:1042–1049. https://doi.org/10.1017/S1751731115003031

Mark T, Jönsson L, Holm M, Christiansen K (2014) Towards genomic selection in Danish Warmblood horses: expected impacts and selective genotyping strategy. In: Proceedings of the 10th World Congress on Genetics Applied to Livestock Production: 17–22 August 2014; Vancouver. http://www.wcgalp.org/system/files/proceedings/2014/towards-genomic-selection-danish-warmblood-horses-expected-impacts-and-selective-genotyping-strategy.pdf

Metzger J, Schrimpf R, Philipp U, Distl O (2013) Expression levels of LCORL are associated with body size in horses. PLoS One 8:e56497. https://doi.org/10.1371/journal.pone.0056497

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829

Pérez P, de los Campos G (2014) Genome-wide regression and prediction with the BGLR statistical package. Genetics 198:483–495. https://doi.org/10.1534/genetics.114.164442

Plieschke L, Edel C, Pimentel ECG, Emmerling R, Bennewitz J, Götz KU (2016) Systematic genotyping of groups of cows to improve genomic estimated breeding values of selection candidates. Genet Sel Evol 48:73. https://doi.org/10.1186/s12711-016-0250-9

Pryce JE, Hayes BJ, Bolormaa S, Goddard ME (2011) Polymorphic regions affecting human height also control stature in cattle.

Genetics 187:981–984. https://doi.org/10.1534/genetics.110.123943

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D et al (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575. https://doi.org/10.1086/519795

R Core Team (2019) R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. https://www.r-project.org/

Roos L, Hinrichs D, Nissen T, Krieter J (2015) Investigations into genetic variability in Holstein horse breed using pedigree data. Livest Sci 177:25–32. https://doi.org/10.1016/j.livsci.2015.04.013

Signer-Hasler H, Flury C, Haase B, Burger D, Simianer H, Leeb T et al (2012) A genome-wide association study reveals loci influencing height and other conformation traits in horses. PLoS ONE 7:e37282. https://doi.org/10.1371/journal.pone.0037282

Stock KF, Distl O (2006) Genetic correlations between conformation traits and radiographic findings in the limbs of German Warmblood riding horses. Genet Sel Evol 38:657–671. https://doi.org/10.1051/gse:2006027

Stock KF, Jönsson L, Ricard A, Mark T (2016) Genomic applications in horse breeding. Anim Front 6:45–52. https://doi.org/10.2527/af.2016-0007

Teegen R, Edel C, Thaller G (2009) Population structure of the Trakehner horse breed. Animal 3:6–15. https://doi.org/10.1017/S1751731108003273

Tetens J, Widmann P, Kühn C, Thaller G (2013) A genome-wide association study indicates LCORL/NCAPG as a candidate locus for withers height in German Warmblood horses. Anim Genet 44:467–471. https://doi.org/10.1111/age.12031

Thomasen JR, Egger-Danner C, Willam A, Guldbrandtsen B, Lund MS, Sørensen AC (2014) Genomic selection strategies in a small dairy cattle population evaluated for genetic gain and profit. J Dairy Sci 97:458–470. https://doi.org/10.3168/jds.2013-6599

Van den Berg I, Meuwissen THE, MacLeod IM, Goddard ME (2019) Predicting the effect of reference population on the accuracy of within, across, and multibreed genomic prediction. J Dairy Sci 102:3155–3174. https://doi.org/10.3168/jds.2018-15231

VanRaden PM, van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF et al (2009) Invited review: reliability of genomic predictions for North American Holstein bulls. J Dairy Sci 92:16–24. https://doi.org/10.3168/jds.2008-1514

Waldmann P, Hallander J, Hoti F, Sillanpää MJ (2008) Efficient Markov chain Monte Carlo implementation of Bayesian analysis of additive and dominance genetic variances in noninbred pedigrees. Genetics 179:1101–1112. https://doi.org/10.1534/genetics.107.084160

Yang J, Hong Lee S, Goddard ME, Visscher PM (2011) GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 88:76–82. https://doi.org/10.1016/j.ajhg.2010.11.011

Zhou L, Heringstad B, Su G, Guldbrandtsen B, Meuwissen THE, Svendsen M et al (2014) Genomic predictions based on a joint reference population for the Nordic Red cattle breeds. J Dairy Sci 97:4485–4496. https://doi.org/10.3168/jds.2013-7580