Single step genomic evaluation for horses based on a multi-breed reference population

M. Wobbe[1,2], H. Alkhoder[1], Z. Liu[1], S. Vosgerau[3], N. Krattenmacher[3], M. von Depka-Prondzinski[4], E. Kalm[3], R. Reents[1], W. Nolte[5], C. Kühn[5], J. Tetens[6], G. Thaller[3], K.F. Stock[1,2]

[1]IT Solutions for Animal Production (vit), Heinrich-Schroeder-Weg 1, 27283 Verden, Germany, [2]University of Veterinary Medicine Hannover (Foundation), Institute for Animal Breeding and Genetics, Buenteweg 17p, 30559 Hanover, Germany, [3]Kiel University, Institute of Animal Breeding and Husbandry, Olshausenstr. 40, 24098 Kiel, Germany, [4]Werlhof-Institut MVZ, Schillerstr. 23, 30159 Hanover, Germany, [5]Leibniz Institute for Farm Animal Biology (FBN), Institute of Genome Biology, Wilhelm-Stahl-Allee 2, 18196 Dummerstorf, Germany, [6]University of Goettingen, DNTW Functional Breeding, Burkhardtweg 2, 37077 Goettingen, Germany; *mirell.wobbe@vit.de

Abstract

The structure of the equine sector implies the need for collaborative approaches to implement genomic applications in sport horse breeding. Based on a multi-breed reference population of 5,000 horses from five breeds, a single step genomic evaluation system for linear conformation and performance traits was developed. Using two different validation approaches, forward and cross validation, medium (around 0.6) to high (over 0.9) correlations of genomic estimated breeding values (GEBVs) indicated stability of the developed system. Different relatedness between the breeds was reflected in the results, but the benefit of available genotypic data with regard to the variance of the GEBVs was found for each breed. According to the findings of this study, single-step genomic evaluation using a multi-breed reference population is a feasible approach for routine implementation of genomic selection in sport horses.

Introduction

The introduction of genomic selection in animal breeding can be referred to as paradigm shift (Meuwissen et al., 2016). For horse breeding, this method has great potential because it allows shortening the long generation interval and integrating traits with low heritability or which are difficult to record into the breeding programs. Although this has been recognized already years ago (Haberland et al., 2012, Mark et al., 2014), there is until today only one studbook for sport horses with officially published genomic breeding values for a single trait (Knaap, 2017). The major challenges of successfully implementing genomic selection in horse breeding relate to definition of suitable target traits and compiling of a sufficiently large reference population. The structure of the equine sector implies particular interest in genomic systems which allow use of data across studbooks and breeds of riding horses: individual breeding populations are often small, but breeding goals are overlapping and exchange of genetic material is existing worldwide. The aim of this study was to validate the across breed approach for single step genomic evaluation and perform detailed distribution analyses of GEBVs with focus on possible breed differences.

Materials & Methods

The data basis for developing a genomic evaluation system for equine conformation and performance traits was from five breeding organizations of German riding horses: Oldenburg (OL), Oldenburg International (OS), Trakehner (TRAK), Holstein (HOL), and Westphalian

(WESTF). Phenotypes were available from routine linear description of the respective studbooks for which they used the same comprehensive linear scheme. The about 41,600 linear profiles of foals and adult horses, which could be used for this work, had been recorded between 2012 and 2019 in connection with regular evaluation events. To implement a single step genomic evaluation system, pedigree data of all phenotyped horses plus four ancestral generations were considered resulting in a relationship matrix of around 116,500 horses. The references population included 5,000 linearly described horses.

Genotyping was performed with commercially available single nucleotide polymorphism (SNP) arrays of medium density from Illumina (Illumina, 2015), GGPEquine (65,157 SNPs; 788 horses), and GGPEquinePlus (71,947 SNPs; 4,212 horses). After quality control with minor allele frequency of 0.01, call frequency of 0.1, call rate of 0.05 and a threshold of $p < 0.001$ for Hardy Weinberg equilibrium test, N=4,964 horses and N=61,211 SNPs remained for analyses. The distribution of the different breeds in the dataset was: N=1,565 horses from OL, N=1,367 horses from HOL, N=829 horses from WESTF, N=778 horses from TRAK, and N=425 horses from OS. The genomic relationship between the five breeds included in the reference population was analyzed by principal component analysis (PCA plot, python version 3.8.1).

The 23 conformation and 18 performance-related linear traits were defined within age group and analyzed in pairs of analogous traits using the following linear animal models for foals (1) and adult horses (2).

$$Y_{ijklop} = \mu + SB_i + EVENT\text{-}TEAM_j + AGE\_M_k + SEX_l + animal_o + e_{ijklop} \quad (1)$$

$$y_{ijmnop} = \mu + SB_i + EVENT\text{-}TEAM_j + AGE\_Y_m + PTYP_n + animal_o + pe_o + e_{ijmnop} \quad (2)$$

where y is the phenotype; μ is the overall mean; SB is the fixed effect of the i-th studbook (i=1-5); EVENT-TEAM is the fixed effect of the combination of the event and team performing the linear description; AGE_M and AGE_Y are the fixed effects of the age of the foals in months (k=1-5; <1 month, 1-2 months, 2-3 moths., 3-4 months, >4 months) and of the adult horses in years (m=1-4; <3 years, 3-4 years, 4-5 years, >5 years); SEX is the fixed effect of sex for the foals (l=1-2; male female; distinct events for mares and stallions in the group of adult horses); PTYP is fixed effect of the presentation type for the adult horses (n=1-3; in hand, free, under saddle); animal is the random additive genetic effect; pe is the random permanent environmental effect; and e is the random residual effect. We used the software package MiX99 release XI/2019 version 19.1129 (Vuori et al., 2006) for solving the equation model. GEBVs were standardized to a mean of 100 and a genetic standard deviation of 20.

The system was validated in two ways: Ten-fold cross-validation was performed by repeatedly choosing 15% of the genotyped and non-genotyped animals randomly per studbook, sex and trait as validation animals (VA); forward validation considered all horses linearly described in 2018 or later as VA. In each case, phenotypic data of the VA were excluded for the validation runs, and Pearson correlations between the GEBVs from the full run considering all available data and the validation run were determined (python version 3.8.1).

Distribution of the GEBVs were analyzed in more detail in the phenotyped horses born in 2007-2017 (approx. N=28,000). Descriptive statistics (mean, standard deviation, minimum, maximum) were generated for all horses and stratifying by studbook and availability of genotypic data.

Results

The principal component analysis of the genomic relationships among the five breeds in the reference population revealed three clusters. The first cluster comprised most of the horses from WESTF and OL, the second most of the horses from HOL and OS, and the third cluster included the horses from TRAK (figure 1).

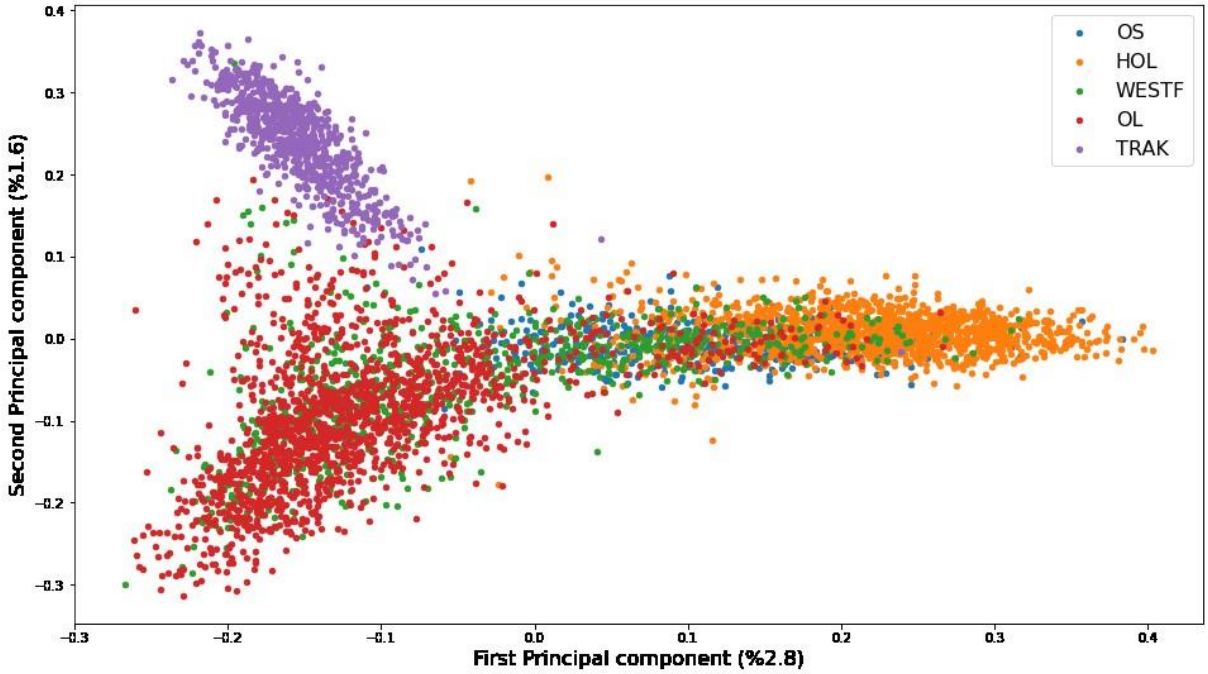

Figure 1. Principal component analysis of the N=4,964 horses in the reference population after quality check, showing the first two principal components.

The mean correlation coefficients between the GEBVs from the full run and the ten cross-validation runs were medium to high, with ranges of 0.60-0.95 for all VA and 0.86-0.99 for genotyped VA. Analyses of correlation coefficients for the genotyped VA by studbook revealed mostly minor differences, but with consistently higher figures for OL (0.88-0.98) and WESTF (0.86-0.99), followed by HOL (0.80-0.97) and OS (0.72-0.96), than for TRAK (0.63-0.94). The forward validation resulted in slightly lower GEBV correlations ranging between 0.56 and 0.89 for all VA and between 0.83 and 0.97 for the genotyped VA.

Analyses of the GEBV distributions showed clearly higher standard deviations in the genotyped horses than in the non-genotyped horses (on average +2.18 for the linear conformation traits and +2.30 for the linear performance traits). Similar results were obtained for each studbook with average differences between standard deviations of GEBVs in genotyped and non-genotyped horses ranging between +1.19 (TRAK) and +2.17 (OS) for conformation and between +1.30 (TRAK) and + 3.02 (OS) for performance traits. The standard deviations in all horses within studbook ranged between 8.68 (HOL) and 13.38 (TRAK) for conformation and between 7.84 (OS) and 11.47 (WESTF) for performance traits (table 1).

Table 1. Means and standard deviations (std.) of GEBV across all performance traits (N=18) and all conformation traits (N=23) for the subset of phenotyped horses born between 2007-2017 by studbook and availability of genotypes.

	all	OL	HOL	WESTF	TRAK	OS
Conformation						
N	27,928	11,209	7,516	2,671	1,487	3,296
mean	100.19	100.73	99.95	100.06	100.15	99.26
std.	9.95	9.82	8.68	8.78	13.38	8.74
Conformation + genotype						
N	4,796	1,556	1,231	824	763	422
mean	100.28	101.02	100.10	100.10	99.80	99.31
std.	12.13	11.71	10.66	10.42	14.57	10.91

Performance						
N	28,560	11,343	7,599	2,874	1,497	3,356
mean	100.46	100.21	102.19	100.53	94.09	100.69
std.	11.27	9.93	8.07	11.47	11.01	7.84
Performance + genotype						
N	4,846	1,556	1,281	824	763	422
mean	100.12	100.67	102.22	101.22	93.67	101.18
std.	13.57	12.04	9.97	13.25	12.31	10.86

Discussion

To meet the challenge of getting a sufficiently large reference population, German horse breeding organizations collaborated to assemble a 5,000 horse multi-breed training set for linear conformation and performance traits. The principal component analysis of the genomic relationships among the five included breeds reflected their relatedness. Different breeding focusses of studbooks on either of the two main disciplines of riding sport can explain the formation of two of the clusters, with one cluster including horses from studbooks with dressage as important breeding goal (WESTF, OL) and another cluster including horses from studbooks focusing on show-jumping (HOL, OS). The remaining cluster included TRAK horses, which is consistent with the breeding policy of a closed studbook where direct genetic links mainly exist through commonly used Thoroughbred or Arabian bloodlines and TRAK horses used in other studbooks. The expected closer relation of TRAK to the dressage cluster than to the cluster of studbooks with special focus on show-jumping was in accordance with other studies (Nolte et al., 2019) also in our sample of genotyped horses.

The reasonable variation of relatedness between the breeds included in the joint reference population implied the need of additional checks of the single step genomic evaluation system. Detailed analyses of the validation results by studbook showed lower GEBV correlations for TRAK than for the other contributing breeds. However, correlations were still in a range, indicating also for this studbook a satisfactory performance of the system for most of the linear target traits. In other species a multi-breed reference population has already proven to be a possible way for smaller populations to start with genomic selection if the respective breeds are related to each other (Lund et al., 2014). This precondition is fulfilled for German warmblood horses. The analyses of the GEBV distributions showed in all breeds higher standard deviations for genotyped than for non-genotyped individuals indicating that genotyping benefits individual characterization of and differentiation between horses in a multi-breed single step genomic evaluation system for linear conformation and performance traits. With the transition to routine SNP genotyping with a medium density array as basis of parentage control, increasing proportions of genotyped horses will support future breeding progress in sport horse breeding.

References

Haberland A.M., König v. Borstel U., Simianer H., König S. (2012) Animal 6: 1369-1376
Illumina GGPEquine (2015) available at: https://www.illumina.com/
Knaap, J. (2017) 8. Pferde-Workshop, Bad Bevensen, Germany
Lund M. S., Su G., Janss L., Guldbrandtsen B., Brøndum R.F. (2014) Livest. Sci. 166:101-110
Mark T., Jönsson L., Holm M., Christianssen K. (2014) Proc of the 10th WCGALP, Vancouver, Canada.
Meuwissen T., Hayes B. and Goddard M. (2016) Anim. Front. 6(1):6-14.
Nolte W., Thaller G., Kühn C. (2019) PLoS ONE 14(4):e0215913.
Vuori K., Strandén I., Lidauer M. and Mäntysaari E. A. (2006) Proc. of the 8th WCGALP, Belo Horizonte, Brazil.