

Performance trait analysis and genetic diversity of the SA Boerperd

N. Breytenbach, J.P. Grobler & H. Bindeman[#]

Department of Genetics, University of the Free State (UFS), PO Box 339, Bloemfontein, South Africa

(Received 27 November 2018; Accepted 9 July 2019; First published online 10 March 2020)

Copyright resides with the authors in terms of the Creative Commons Attribution 4.0 South African Licence.

See: <http://creativecommons.org/licenses/by/4.0/za>

Condition of use: The user may copy, distribute, transmit and adapt the work, but must recognise the authors and the South African Journal of Animal Science.

Abstract

This study determined the occurrence and frequency of mutations that influence performance traits in the SA Boerperd horse and evaluated genetic diversity within the breed. Two gait-associated and two height-associated mutations were investigated. Seventeen microsatellite markers were genotyped for 363 horses and used to assess genetic diversity. The C-allele of the height-associated single nucleotide polymorphism (SNP) BIEC2_808543 was present in 5% of the population, and 57% of the horses possessed the A-allele of BIEC2_1105377, which has been also associated with height. However, these SNPs did not significantly affect measured phenotypic height within the breed. The minor alleles for SNP *DMRT3_Ser301STOP* and BIEC2_620109, which have been associated with gaitedness, occurred with frequencies of 0.105 and 0.091, respectively, within the breed. The microsatellite data revealed observed heterozygosity ($H_o = 0.679$) to be similar to that found 15 years ago, while the level of inbreeding had decreased from 8.4% to 3.2%. However, allelic richness had declined from 4.212 to 3.804. Analysis of the population structure revealed that two distinct founder populations have contributed to the present-day breed. Compared with nine European breeds, the SA Boerperd had above average levels of heterozygosity and a high number of private alleles (17.6%). A high degree of variation remained in the SA Boerperd, despite selective breeding, and levels of inbreeding were still manageable. Results obtained in this study can be used by SA Boerperd breeders to develop the breed, while simultaneously conserving its genetic potential.

Keywords: gaitedness, height, horses, inbreeding, selective breeding

[#]Corresponding author: vdwesthm@ufs.ac.za

Introduction

The SA Boerperd originated in South Africa and gained breed status officially in 1980 (Nel, 2014). This hardy breed was originally developed as working animals for farmers, but today it plays an important economic role by contributing to agriculture (Nel, 2014), tourism (Equus Journeys, Shropshire, United Kingdom; Pakamisa, 2016), and the horse show industry (Louw, 2008; Du Toit, 2010).

The SA Boerperd show industry is extremely competitive and each horse must possess certain performance-based characteristics to be able to partake in specific disciplines. The five-gaited class is one such discipline, and requires horses to showcase five varying movement styles or gaits. In general, most equine breeds can display only three gaits, namely walk, trot and canter (Robilliard *et al.*, 2007). The display of additional gaits, also referred to as gaitedness, can be acquired through training (Bekker, 2012), but is largely possible because of the presence of specific polymorphisms in the genome of the horse (Andersson *et al.*, 2012). Certain polymorphisms, more specifically single nucleotide polymorphisms (SNPs) give horses the potential to display gaits that are uncommon to most equine breeds (Andersson *et al.*, 2012). Two uncommon gaits, namely the rack and slow-gait, are required of SA Boerperd horses that partake in five-gaited classes (Louw, 2008).

In addition to SNPs for gaitedness, other trait-related polymorphisms that affect competitive performance are those influencing size, since size can affect how well a horse performs in an event. The height of a horse has been linked to its execution of certain gaits (Signer-Hasler *et al.*, 2012), which can affect how it performs in five-gaited classes. It can also influence performance in disciplines such as show jumping and events courses, in which the SA Boerperd also takes part (SA Boerperd Breeders Society, 2016). This performance trait is influenced by multiple major and minor contributing SNPs (Makvandi-Nejad *et al.*, 2012; Signer-Hasler *et al.*, 2012).

Few studies have focused on the genetic composition of the SA Boerperd, with one study investigating performance-based traits on a genetic basis (Promerová *et al.*, 2014). This lack of investigation into the breed has gone unnoticed by most breeders, because the traditional phenotype-based method of selecting horses has served them well over the years (Du Toit, 2010; Swart, 2010). Yet, with jumping and equitation being two of the most popular equestrian disciplines in South Africa (Equestrian, 2017), the demand for equine top performers can only increase. Screening for performance-influencing SNPs in individuals would be a cost- and time-effective method to meet this demand. Individuals that have the relevant SNPs can be used for breeding to ensure the continuance of favourable variants in the offspring.

Owing to the SA Boerperd's uniqueness, it is also of value to monitor the breed's genetic diversity levels to ensure its overall fitness and continued existence. The stud book registry of the SA Boerperd was officially closed in 1998 (Swart, 2010; Nel, 2014) and all individuals are the progeny of eleven founding bloodlines (Louw, 2008; Du Toit, 2010). Thus, no new genetic material has been added to the breed in approximately twenty years. This is not necessarily disadvantageous, but if the breeding of the population is not well managed, it could increase the probability of inbreeding and the loss of fitness (Lacy, 1987).

Microsatellite markers, or short tandem repeat markers (STRs), are effective tools to evaluate genetic diversity. These markers are less likely to misalign owing to their repetitive nature (Goldstein & Schlötterer, 1999) and are thus reliable when determining genotypes. To date, only one study has investigated the diversity of the SA Boerperd using microsatellites (Botha, 2002), and it revealed that the breed is highly diverse. This is in line with expectations for a newly established breed with a large founder population (Allendorf *et al.*, 2013). However, after years of diligently breeding certain phenotypic characteristics for admission to the studbook (Louw, 2008), it is possible that the level of diversity of the SA Boerperd may have decreased.

Based on this motivation, the current study screened for the occurrence of two height-associated SNPs and two gait-associated SNPs in the breed. The genetic diversity and differentiation within the breed were assessed using 17 microsatellite markers to determine the presence of subpopulations, compare the present level of genetic diversity with that of previous years, and to determine the breed's sustainability by comparison to other breeds.

Materials and Methods

This study was conducted with ethical approval (UFS-AED2016/0033) from the Interfaculty Animals Ethics Committee of the University of the Free State (UFS). Permission was also granted by the SA Boerperd Breeders Society of South Africa.

The sample population comprised 100 horses from 12 studs from various locations across South Africa (named ST1 - ST12). Care was taken to sample as many unrelated individuals as possible, but a few individuals had a sire in common based on progeny information obtained from the Breeders Society. Either mane or tail hair were pulled from each horse and temporarily stored in a paper envelope until DNA analysis. In addition to hair sampling, the individual's gender, height and physical gait status (three- or five-gaited) were documented.

The four SNPs that were investigated, chromosomal positions, primer sequences used, and annealing temperatures are presented in Table 1. The reference sequences of the SNPs were extracted from the EquCab2.0 database (Broad Institute, 2008), and Primer3Plus 4.0.0 software (Untergasser *et al.*, 2012) was used to design the primer-pairs of BIEC2_1105377 and BIEC2_620109.

Direct polymerase chain reaction (PCR) was performed on hair bulbs using either the KAPA HiFi HotStart ReadyMix PCR Kit or the KAPA2G Robust HotStart ReadyMix PCR Kit (Sigma-Aldrich Chemie GmbH, Munich, Germany). To each tube, 4.6 µl deionized H₂O, 7 µl HiFi/Robust Ready Mix, 0.7 µl of the forward primer [10µM] and 0.7 µl of the reverse primer [10µM] were added to 2 - 4 hair bulbs of each horse. This procedure was repeated for each of the four SNPs, per horse. The PCR protocol described by He *et al.* (2015) was used, but the annealing temperatures were changed for each primer-pair, as indicated in Table 1. The Bio-Spin® PCR Purification Kit (Hangzhou Bioer Technology Co. Ltd., Hangzhou, P. R. China) was used according to manufacturer specifications to remove the forward and reverse primers in the PCR product. The nucleotide sequence of each horse was obtained using a BigDye® Terminator Cycle Sequencing Kit (Thermo Fisher Scientific, Waltham, Massachusetts, USA), with reaction mixtures consisting of 7.3 µl of deionized H₂O, 0.5 µl buffer, 1 µl BigDye®, 0.2 µl forward/reverse primer, and 1 µl of purified PCR product. The pre-sequencing protocol of ZR DNA Sequencing Clean-up Kit™ (Zymo Research Corp., Irvine, California, USA) was used after amplification according to manufacturer specifications. Lastly, the eluted DNA was sequenced using a 3130xl Genetic Analyzer (Thermo Fisher Scientific, Waltham, Massachusetts, USA).

Table 1 SNPs screened in SA Boerperd populations, chromosomal positions, primer sequences, and annealing temperatures

SNP	Nucleotide Position	Primer sequences	Annealing temperature	Reference
Height-associated				
BIEC2_808543	ECA3 (105 547 002)	P _F : 5'-TGG AGT CAG TTG GGT TTA ATG-3' P _R : 5'-GAC CGG ATA GCA TAG AGA GAG-3' P _F : 5'-CAG GAC AAC CTC CCT CAC CA-3' P _R : 5'- TAC GTT GGC TGT CTT GGG TG-3'	56 °C	He <i>et al.</i> , 2015
BIEC2_1105377	ECA9 (74 798 143)		62 °C	Untergasser <i>et al.</i> , 2012
Gait-associated				
DMRT3_Ser301STOP	ECA23 (22 999 655)	P _F : 5'-CGA CAA AGA CAC CGA CCA GA-3' P _R : 5'-CCG ATC CCA CGG ACC ATT-3' P _F : 5'- GCC TCT CAC CCA GAC ACC AT-3' P _R : 5'- ATC TTA TGG CAC ACG GCA CC-3'	62 °C	Han <i>et al.</i> , 2015
BIEC2_620109	ECA23 (22 967 656)		60 °C	Untergasser <i>et al.</i> , 2012

P_F: forward primer; P_R: reverse primer

The DNA sequence of each horse was verified and analysed with GENEIOUS 8.1.7 software (Kearse *et al.*, 2012). The average height of horses associated with a specific genotype was calculated in Microsoft Office Excel. Two height ranges (below and above 160 cm) were used, based on the observation by Metzger *et al.* (2013) who found the CC-genotype (BIEC2_808543) in horses taller than 160 cm. GenAIEx 6.503 (Peakall & Smouse, 2006; 2012) was used to count the number of gaited horses associated with SNP DMRT3_Ser301STOP and to determine the allele and genotype frequencies of the four SNPs. Data of the SNPs were analysed using an r x c Fisher's exact test (Mehta & Patel, 1983) to evaluate whether the proportions of two variables were independently different in the small sample sizes (McDonald, 2014). Some samples could not be amplified for SNP BIEC2_620109, and were thus not included in the test.

The microsatellite data of 363 SA Boerperd horses representing 12 studs, were obtained from Unistel Medical Laboratories (UML) to evaluate genetic diversity levels. This data were originally used for parentage analysis, but not population genetic analysis. The UML microsatellite panel consists of 17 loci (AHT4, AHT5, ASB2, ASB17, ASB23, CA425, HMS1, HMS2, HMS3, HMS6, HMS7, HTG4, HTG6, HTG7, HTG10, LEX3, and VHL20). Data obtained by Botha (2002) for 81 SA Boerperd horses were compared to data from this study. Botha (2002) investigated 12 microsatellite markers and of those, nine could be compared with UML data with a high degree of confidence. The three excluded markers had no alleles in common between the populations sampled in 2002 and 2017. As an additional diverse reference group, the data of a 17-loci panel used by Van de Goor *et al.* (2011) for nine of 35 equine populations sampled across Europe, were included. The nine populations were chosen, because they had a historical influence on the origin of the SA Boerperd (Louw, 2008; Du Toit, 2010; Swart, 2010; Nel, 2014). Microsatellite marker data were entered into Microsoft Office Excel and analysed by the GenAIEx 6.503 add-in (Peakall & Smouse, 2006; 2012). The software was used to determine allelic variability and frequency, as well as heterozygosity. Wright's F-statistic values were calculated with ARLEQUIN 3.5.2.2 (Excoffier *et al.*, 2005). The program STRUCTURE 2.3.4 (Pritchard *et al.*, 2000) was used to determine population structure of the SA Boerperd. Five independent runs were repeated for K-value of 1 - 10 and a burn-in period of 20,000 steps followed by 100,000 MCMC (Markov Chain Monte Carlo) iterations were used. Structure Harvester 0.6.94 (Earl & von Holdt, 2012) was used to determine the true value of K.

Results

Just more than half (51%) the horses sampled for performance trait analysis were male. Of the horses sampled 86% showed the three-gaited phenotype. The sample ranged in height from 140 - 167.6 cm with an average height of 153.1 cm in the males and 151.5 cm in the females. The allele and genotype frequencies

for both size- and gait-associated SNPs are shown in Table 2. The wild-type alleles of all four SNPs were predominant. Due to amplification difficulties, SNP BIEC2_620109 data were generated for 88 horses only.

Table 2 Distributions of the genotype and allelic frequencies of the four investigated SNPs in the SA Boerperd

SNP	N	Allele	Allele frequency	Genotype	Genotype frequency
Height-associated					
BIEC2_808543	100	T*	0.975	TT	0.950
		C	0.025	TC	0.050
				CC	0.000
BIEC2_1105377	100	G*	0.670	GG	0.430
		A	0.330	GA	0.480
				AA	0.090
Gait-associated					
DMRT3_Ser301STOP	100	C*	0.895	CC	0.790
		A	0.105	CA	0.210
				AA	0.000
BIEC2_620109	88	C*	0.909	CC	0.818
		T	0.091	CT	0.182
				TT	0.000

N: sample size, *Wild-type allele

The $r \times c$ Fisher's exact test (Mehta & Patel, 1983) data are presented in Table 3. The SNP genotypes of BIEC2_808543 and DMRT3_Ser301STOP were significantly associated with their specific traits at a 95% significance level.

Table 3 SNP data showing the relationship between genotype and trait for three SNPs

Height				Gait					
SNP BIEC2_808543 (T > C)			BIEC2_1105377 (G > A)		DMRT3_Ser301STOP (C > A)				
Genotype	<160 cm	>160 cm	Genotype	<160 cm	>160 cm	Genotype	3-gaited	5-gaited	
TT	83	12	GG	40	3	CC	74	5	
TC	2	3	GA	39	9	CA	12	9	
CC	0	0	AA	6	3	AA	0	0	
$P = 0.0232$			$P = 0.0659$			$P = 0.0002$			

Table 4 shows the genetic diversity of 363 horses from 12 SA Boerperd studs. The mean number of alleles (N_a) over all studs ranged from 3.471 (Stud8) to 6.529 (Stud7). Taking sample size into consideration, the N_e -values showed that Stud2 had the most alleles per locus (3.937), whilst Stud9 had the least (2.973). Of all the studs, Stud12 had the highest observed heterozygosity (H_o) at 0.735, and the rest of the studs had values higher than 0.640. Unbiased expected heterozygosity (H_e) was highest for Stud2, which correlated with its high N_e -value. Stud12 was the most outbred stud with a F_{IS} -value of -0.118. The lower value of Stud8 was disregarded due to its small sample size and obvious bias. The stud that showed the most significant deviations from random mating based on heterozygote deficit was Stud6 (0.074) with an inbreeding value of $F_{IS} = 0.074$.

Table 4 Genetic diversity of 12 SA Boerperd studs, based on 17 microsatellite markers

Stud	N	N _a (SD)	N _e (SD)	H _o (SD)	H _z (SD)	F _{IS}
1	33	5.529 (0.286)	3.116 (0.230)	0.689 (0.040)	0.658 (0.028)	-0.061
2	23	6.176 (0.395)	3.937 (0.372)	0.680 (0.043)	0.727 (0.026)	0.050
3	31	5.353 (0.383)	3.141 (0.252)	0.649 (0.040)	0.654 (0.032)	-0.003
4	35	5.765 (0.338)	3.487 (0.311)	0.643 (0.034)	0.686 (0.028)	0.043
5	14	4.706 (0.329)	3.063 (0.260)	0.683 (0.053)	0.652 (0.039)	-0.082
6	61	6.471 (0.463)	3.876 (0.320)	0.655 (0.036)	0.716 (0.026)	0.074
7	29	6.529 (0.385)	3.643 (0.316)	0.713 (0.039)	0.696 (0.036)	-0.048
8	3	3.471 (0.286)	2.977 (0.302)	0.725 (0.065)	0.722 (0.052)	-0.215
9	12	4.588 (0.243)	2.973 (0.188)	0.681 (0.045)	0.667 (0.028)	-0.063
10	26	5.706 (0.460)	3.511 (0.337)	0.646 (0.040)	0.682 (0.033)	0.023
11	16	4.941 (0.369)	3.317 (0.243)	0.685 (0.034)	0.691 (0.027)	-0.028
12	80	6.353 (0.461)	3.276 (0.276)	0.735 (0.039)	0.662 (0.031)	-0.118*

N: sample size; SD: standard deviation; N_a: number of alleles; N_e: number of effective alleles; H_o : observed heterozygosity; H_z: unbiased heterozygosity; F_{IS}: inbreeding coefficient

*P < 0.05

The overall genetic differentiation of the SA Boerperd studs was calculated as $F_{ST} = 0.061$. A critical $P = 0.001$ was used for pairwise F_{ST} -values after Bonferroni correction. Based on the F_{ST} and p-values, most of the studs were significantly differentiated from one another. All pair-wise combinations between Stud8 and the rest of the studs, with the exception of Stud1, showed F_{ST} -values and p-values that did not support the hypothesis of significant differentiation. This could have been influenced by the small sample size of Stud8.

The Bayesian assignment approach from STRUCTURE showed that the data obtained from twelve SA Boerperd studs had the highest probability of representing two clusters (Figure 1). The presence of two main clusters was based on the ΔK generated by Structure Harvester (not shown). Based on the composition of the bar plot, Cluster1 consisted primarily of the individuals from Stud1 – Stud11, and Cluster2 of horses from Stud12.

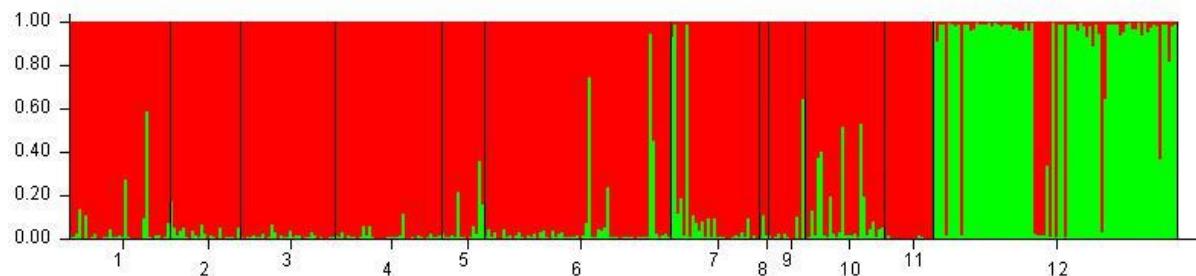


Figure 1 Graphical representation of the population structure of 363 SA Boerperd horses from 12 studs (K = 2), from STRUCTURE. Each vertical line with its colour segments indicate the inferred membership of an individual to a specific founder population

The genetic diversity of the SA Boerperd calculated in this study (conducted in 2017) was compared to diversity as estimated in 2002. Some of the investigated loci proved to be unsuitable for comparison, thus the data from nine loci (AHT4, AHT5, ASB2, HMS3, HMS6, HMS7, HTG4, HTG7, VHL20) are shown in Table 5. Even though the 2002 population had considerably fewer individuals than the 2017 sample, its N_e was slightly higher. The breed's inbreeding coefficient showed a significant change over time. The number of heterozygotes seemed to have increased over the years, but the larger sample size of 2017 may have influenced this value.

Table 5 Genetic diversity of the SA Boerperd population sampled in 2002 and 2017

Year	N	N _a (SD)	N _e (SD)	H _o (SD)	H _z (SD)	F _{IS}
2002	81	7.556 (0.626)	4.212 (0.474)	0.672 (0.037)	0.739 (0.033)	0.084
2017	363	7.444 (0.626)	3.804 (0.461)	0.679 (0.044)	0.701 (0.041)	0.032

N: sample size; SD: standard deviation; N_a: number of alleles; N_e: number of effective alleles; H_o: observed heterozygosity; H_z: unbiased heterozygosity; F_{IS}: inbreeding coefficient

The European equine breeds chosen to compare to the SA Boerperd belonged to the Tennessee Walker, Andalusian, Standardbred, Appaloosa, Arabian, Friesian, Hackney and Thoroughbred. The Icelandic Horse was also included, because it is a gaited breed like the SA Boerperd. As indicated in Table 6, the SA Boerperd had the highest N_a (8.000) and the Thoroughbred the lowest at 5.118. Yet when taking sample size into consideration, the Appaloosa had the highest N_e of 4.499 and the Friesian the lowest (2.163). Both H_o and H_z were also highest for the Appaloosa, and lowest for the Friesian. The SA Boerperd had average heterozygosity values of H_z = 0.722. The number of private alleles within each breed were also calculated. The Standardbred, Arabian and Icelandic shared the highest value at 23.5%. The SA Boerperd had the second highest proportion at 17.6%. The least inbred breed was the Tennessee Walker with a value of F_{IS} = -0.019, whilst the breed with the highest inbreeding value was the Hackney with F_{IS} = 0.060.

Table 6 Genetic diversity of the SA Boerperd compared to nine European horse breed populations, based on 17 microsatellite markers

Breed	N	N _a (SD)	N _e (SD)	H _o (SD)	H _z (SD)	PA (SD)	F _{IS}
SA Boerperd	363	8.000 (0.536)	4.123 (0.378)	0.682 (0.030)	0.722 (0.027)	0.176 (0.095)	0.053
Tennessee Walker	23	5.176 (0.312)	3.204 (0.237)	0.670 (0.033)	0.674 (0.026)	0.000 (0.000)	-0.019
Andalusian	67	6.824 (0.456)	3.405 (0.292)	0.658 (0.037)	0.670 (0.033)	0.059 (0.059)	0.000
Standardbred	997	7.588 (0.500)	3.764 (0.254)	0.690 (0.032)	0.710 (0.025)	0.235 (0.136)	0.024
Appaloosa	99	7.706 (0.460)	4.499 (0.267)	0.717 (0.034)	0.765 (0.019)	0.118 (0.081)	0.058
Arabian	615	7.235 (0.497)	3.412 (0.242)	0.645 (0.030)	0.678 (0.028)	0.235 (0.106)	0.043
Friesian	781	5.235 (0.450)	2.163 (0.262)	0.439 (0.053)	0.447 (0.051)	0.059 (0.059)	0.023
Hackney	141	6.118 (0.453)	3.210 (0.231)	0.618 (0.030)	0.661 (0.027)	0.059 (0.059)	0.060
Thoroughbred	55	5.118 (0.331)	3.542 (0.276)	0.667 (0.031)	0.695 (0.025)	0.000 (0.000)	0.026
Icelandic	134	7.176 (0.516)	3.945 (0.383)	0.659 (0.033)	0.705 (0.032)	0.235 (0.106)	0.049

N: sample size; SD: standard deviation; N_a: number of alleles; N_e: number of effective alleles; H_o: observed heterozygosity; H_z: unbiased heterozygosity; F_{IS}: inbreeding coefficient

The pair-wise population F_{ST}-values and p-values among the ten breeds are presented in Table 7 (with a critical P =value of 0.001, after Bonferroni correction). The pair-wise combinations between all the breeds showed F_{ST}-values and p-values that supported the hypothesis of significant differentiation. The Friesian was consistently shown to be the furthest related from all other breeds. The closest associations, however, were found between the Appaloosa and Hackney, as well as the Appaloosa and SA Boerperd, with an F_{ST}-value of 0.070. In addition to the Appaloosa, the SA Boerperd also showed a close association with the Andalusian and Standardbred.

Table 7 Pairwise genetic distance between 10 breeds of horse as indicated by F_{ST} statistics. P -values are indicated in the brackets

	TEN	AND	STA	APP	ARA	FRI	HAC	THO	ICE
AND	0.105 (0.000)								
STA	0.124 (0.000)	0.112 (0.000)							
APP	0.071 (0.000)	0.095 (0.000)	0.107 (0.000)						
ARA	0.126 (0.000)	0.099 (0.000)	0.097 (0.000)	0.106 (0.000)					
FRI	0.373 (0.000)	0.352 (0.000)	0.314 (0.000)	0.308 (0.000)	0.327 (0.000)				
HAC	0.117 (0.000)	0.138 (0.000)	0.153 (0.000)	0.070 (0.000)	0.151 (0.000)	0.318 (0.000)			
THO	0.118 (0.000)	0.116 (0.000)	0.084 (0.000)	0.112 (0.000)	0.099 (0.000)	0.403 (0.000)	0.166 (0.000)		
ICE	0.124 (0.000)	0.157 (0.000)	0.167 (0.000)	0.065 (0.000)	0.174 (0.000)	0.338 (0.000)	0.099 (0.000)	0.180 (0.000)	
SAB	0.086 (0.000)	0.078 (0.000)	0.083 (0.000)	0.070 (0.000)	0.093 (0.000)	0.308 (0.000)	0.100 (0.000)	0.087 (0.000)	0.126 (0.000)

TEN: Tennessee Walker; AND: Andalusian; STA: Standardbred; APP: Appaloosa; ARA: Arabian; FRI: Friesian; HAC: Hackney; THO: Thoroughbred; ICE: Icelandic Horse; SAB: SA Boerperd

Based on the ΔK and mean likelihood values for various values of K in STRUCTURE, the 10 equine breeds had the highest probability for belonging to two clusters. On closer inspection, this arrangement appeared to be based on the separation of the cold-blooded Friesian from the more hot-blooded breeds (Figure 2). Alternatively, assignments from $K = 4$ and $K = 10$ were also examined (Figure 2). These clustering results were included since the likelihood scores at $K = 4$ showed an elevated ΔK -value, and because the genomic data originated from 10 different breeds.

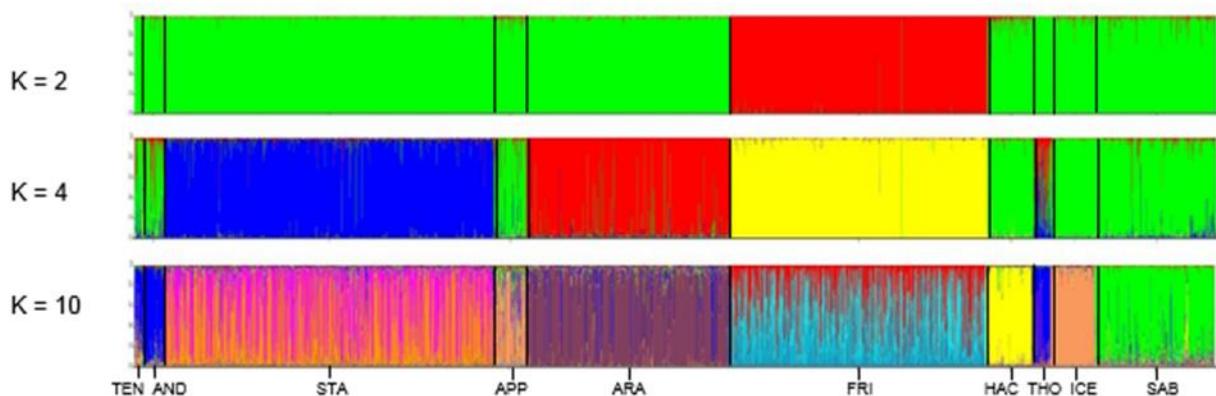


Figure 2 STRUCTURE-based clustering output for three values of K in 3,275 horses belonging to 10 breeds. Each individual is represented by one vertical line, which the colours indicate the proportional assignment to one or more founder populations

TEN: Tennessee Walker; AND: Andalusian; STA: Standardbred; APP: Appaloosa; ARA: Arabian; FRI: Friesian; HAC: Hackney; THO: Thoroughbred; ICE: Icelandic Horse; SAB: SA Boerperd

Discussion

Metzger *et al.* (2013) found that horses across many breeds measuring 160 cm and less at the withers (the ridge between the shoulder blades), displayed the highest frequency for the wild-type T-allele when genotyped for the SNP BIEC2_808543. In comparison, individuals higher than 160 cm tended to have a higher frequency for the mutated C-allele (Metzger *et al.*, 2013). Similar results were obtained for the SA Boerperd in this study. The T-allele (BIEC2_808543) was the predominant allele, which was in line with the

medium height range of 140 – 160 cm observed for the breed. Those individuals that possessed the mutated C-allele (BIEC2_808543) were of average height, the tallest being 167.6 cm.

The frequency of the mutated BIEC2_1105377 allele (A) in the SA Boerperd was much higher than that of the minor BIEC2_808543 allele (C). It also did not seem as if the presence of one or two of the A-alleles (BIEC2_1105377) within an individual was dependant on the presence of the minor C-allele (BIEC2_808543), as was the case with the gait-associated SNPs (*DMRT3_Ser301STOP* and BIEC2_620109). This indicated that the height-associated SNPs were inherited independently of one another, most likely since they occurred on different chromosomes. It was noteworthy that the T-allele (BIEC2_808543) was always accompanied by at least one A-allele (BIEC2_1105377), though this might have been coincidental.

Most horses that possessed both height-associated SNPs (BIEC2_808543 & BIEC2_1105377) were shorter than 160 cm, which supported the findings of Metzger *et al.* (2013). In addition, the average height for each size-associated genotype revealed that horses with or without the mutated A-allele at SNP BIEC2_1105377 (TT/GA & TT/AA) averaged about 152 cm, and horses with the mutated C-allele at SNP BIEC2_808543 were much closer to 160 cm (TC/GA & TC/AA). Since the latter mutation was only observed in five individuals, and three of those were taller than 160 cm, it would be premature to accept that BIEC2_808543 contributes significantly to size without testing additional tall horses. Furthermore, it could not be established whether BIEC2_808543 attributed 1 cm and BIEC2_1105377 0.5 cm to height, as was found by Signer-Hasler *et al.* (2012).

Overall, the individual level of influence that BIEC2_808543 and BIEC2_1105377 had on height at withers, was uncertain. Based on the findings of the current study, these two SNPs alone appeared not to be definitive predictors of height in the SA Boerperd due to the lack of a clear association. It was thus reasonable to assume that combinations of strongly associated and linked SNPs, which formed different haplotypes contributed more to height and was a better indication of the size of a horse. Such haplotypes were identified in both the LCORL and HMGA2 genes (Makvandi-Nejad *et al.*, 2012), though the exact measurements of the individuals were not documented to determine the validity of this prediction.

Most gaited breeds are fixed for the A-allele of SNP *DMRT3_Ser301STOP* (Andersson *et al.*, 2012; Promerová *et al.*, 2014). The SA Boerperd, however, had a very low frequency of for this allele (10.5%). Promerová *et al.* (2014) records a similar value for the Boer Pony, which possibly includes both the SA and Cape Boerperd. Additionally, none of the Boer Ponies possessed the AA-genotype (Promerová *et al.*, 2014), as was the case in the current study. The Fisher's exact test indicated a significant difference of association between the two gait categories, where most horses were three-gaited. It thus seemed that the SA Boerperd was not a true gaited breed due to the low frequency of the A-allele (*DMRT3_Ser301STOP*) and the breed's bias towards being three-gaited. This can partially be explained by the nature of selective breeding. Currently, there is a very small market for five-gaited individuals within the SA Boerperd breeders community, since fewer five-gaited classes are hosted per show than three-gaited ones (SA Boerperd Breeders Society, 2017).

Based on the findings of the study, *DMRT3_Ser301STOP* and BIEC2_620109 seemed to be co-inherited, which is plausible since the two SNPs are significantly linked (Promerová *et al.*, 2014). The frequencies for both SNPs would have been identical, if not for the lack of BIEC2_620109 data. Additionally, the genotypes consisting of both wild-type alleles (C-C) and both mutant alleles (A-T) were present. However, Promerová *et al.* (2014) found that the mutated allele of either SNP occurred along with the wild-type allele of the other, though at very low frequencies. The same study also found that only 0.3% of individuals across various breeds had the SNP *DMRT3_Ser301STOP* (A-allele) and the wild-type C-allele at BIEC2_620109. Also of significance was that SNP *DMRT3_Ser301STOP* was absent in some five-gaited individuals in the current study. Similar findings were made by Promerová *et al.* (2014), who theorised that either another SNP affected gaitedness, or that multiple genes affected the trait.

The levels of diversity in all the studs were broadly similar, despite the varying sample sizes. One hundred and thirty-six alleles were detected across the 17 loci for the SA Boerperd, and roughly three alleles occurred on average per locus for each stud. The number of heterozygotes within each stud did not deviate meaningfully from what was expected. Relatively low F_{IS} -values and a surplus of heterozygotes were found for most studs. Of significance was Stud12, which had the most heterozygotes and lowest inbreeding coefficient. The differentiation estimates revealed that most studs were genetically similar, yet Stud11 and Stud12 were genetically less similar, as were Stud1 and Stud11. This could have been due to the geographical locations of these studs, which were located in different provinces, and that stud owners probably did not obtain new breeding stock from each other.

Further structural analysis of the SA Boerperd population revealed that Stud12 was significantly different from the overall population. This population was situated geographically close to some of the other studs, thus the differentiation was most likely due to selective breeding. It was possible that the owner of this

stud selected individuals of certain phenotypes to help conform the rest of the stock to a specific standard. This practice might have led to the creation of a genetically distinct subpopulation within the breed. Although the creation of such a homogenous group can threaten sustainability (Lacy, 1987), it seemed that the selection process of this particular stud ensured heightened genetic diversity when considering the negative inbreeding value.

The comparison of the genetic data of the SA Boerperd in 2017 with that of 2002 revealed that the number of heterozygotes stayed constant, despite the larger sample size used in 2017. Of concern is the slight decrease in the breed's allelic richness, which was unexpected given the increased sample size. Despite this, the inbreeding coefficient showed a promising improvement (though this could also have been due to the larger number of individuals). Nevertheless, the overall genetic diversity estimates of the SA Boerperd demonstrated its variability and sustainability.

The comparison of the overall genetic diversity of the SA Boerperd to that of other breeds revealed a healthy, sustainable level. Compared to other breeds, the SA Boerperd possessed the highest mean number of alleles per locus, proving its long-term fitness (Allendorf *et al.*, 2013) and suggesting that selective pressures had little effect on the breed over the years. The SA Boerperd also had a significant number of private alleles, which makes it a distinct and unique breed (Allendorf *et al.*, 2013). The 0.053 inbreeding value of the breed is most likely caused by the substructure found in the breed. Significantly positive values of F_{IS} were also found within Standardbreds due to substructure (Cothran *et al.*, 1987). Pairwise differentiation values and structure analysis showed that the SA Boerperd and Friesian were the least genetically similar, which corresponds to the findings of Botha (2002). Despite not being able to determine a 'true' value for K, this study showed that the SA Boerperd was most closely related to the Andalusian and Standardbred. Since the breed was possibly influenced by American breeds not included in this study, it would be beneficial to compare the diversity level of the SA Boerperd with such breeds to achieve a more conclusive estimation of diversity and relationships.

Conclusions

Polymorphisms at the SNPs BIEC2_808543 and BIEC2_1105377 were present within SA Boerperd individuals, but it could not be determined whether they were the main contributors to height. Screening horses for these two mutations with the aim of breeding individuals with a height of more than 160 cm will thus not be meaningful for breeders, unless future research involving a much larger sample population finds evidence to the contrary. This also extends to both gait-associated SNPs (DMRT3_Ser301STOP & BIEC2_620109). Although most five-gaited individuals possessed the SNPs, about a third did not. Since the frequencies of the mutated alleles were very low within the breed, it was difficult to determine whether this was a random phenomenon. Future research is thus needed to investigate the influence that other mutations may have on the expression of gaitedness. It is not recommended that breeders test their horses solely for the mutations investigated in this study. If testing for performance-related characteristics become more of a priority in the future, then whole genome data should be considered. The genetic diversity of the SA Boerperd was comparable to the level determined in previous years, and also to other equine breeds.

Acknowledgements

The authors acknowledge Unistel Medical Laboratories for providing the STR marker data of the registered SA Boerperd horses. DNA data of multiple equine breeds were provided by L.H.P. van de Goor (Dr. Van Haeringen Laboratorium, The Netherlands) and Dr K. Ehlers (University of the Free State). The authors thank the SA Boerperd Breeders Society for its support and breeders for allowing samples to be taken from their horses.

Authors' Contributions

NB designed the research project, collected samples, did laboratory work, data analysis and wrote the manuscript (as part of her MSc programme). HB and JPG assisted with data analysis, editing the manuscript, and final approval of the version to be published.

Conflict of Interest Declaration

None of the authors have a financial or other relationship with people or organisations that could influence inappropriately or bias the contents of this paper.

References

- Allendorf, F.W., Luikart, G. & Aitken, S.N., 2013. Conservation and the genetics of populations. Wiley-Blackwell, Chichester.
- Andersson, L. S., Larhammar, M., Memic, F., Wootz, H., Schwochow, D., Rubin C-J., Patra, K., Arnason, T., Wellbring, L., Hjälm, G., Imsland, F., Petersen, J.L., McCue, M.E., Mickelson, J.R., Cothran, G., Ahituv, N., Roepstorff, L., Mikko, S., Vallstedt, A., Lindgren, G., Andersson. L. & Kullander. K., 2012. Mutations in DMRT3 affect locomotion in horses and spinal circuit in mice. *Nature*. 488, 642-646. doi: 10.1038/nature11399

Bekker, L., 2012. Vyfgang. In: SA Boerperd: Die perd vir die toekoms: 31-34. Edited by S. Louw. SA Boerperdtelersgenootskap, Bloemfontein.

Botha, K., 2002. Genetic relationships of seven horse breeds in South Africa based on DNA markers. Dissertation (M.Sc.): University of the Free State, Bloemfontein.

Broad Institute., 2008. FTP Directory: Distribution horse SNP release. Available from: ftp://ftp.broadinstitute.org/distribution/horse_snp_release/v2/ Accessed 4 June 2016.

Cothran, E.G., MacCluer, J.W., Weitkamp, L.R. & Bailey, E., 1987. Genetic differentiation associated with gait within American Standardbred horses. *Anim. Genet.* 18, 285-296.

Du Toit, K., 2010. Boerperd: Die geskiedenis van die S.A. Boerperd. Pretoria: West Publishers.

Earl, D.A. & Von Holdt, B.M., 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* 4(2), 359-361.

Equestrian., 2017. Five popular disciplines of the South African Equestrian Federation – Dressage, show jumping, endurance ride, eventing and equitation. Available at: <https://www.equestrian.co.za/five-popular-disciplines-of-the-south-african-equestrian-dressage-show-jumping-endurance-ride-eventing-and-equitation/>

Excoffier, L., Laval, G. & Schneider, S., 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evol. Bioinform. Online.* 1, 4750.

Finno, C.J. & Bannasch, D.L., 2014. Applied equine genetics. *Equine Veterinary Journal* 46(5), 538-544.

Goldstein, D. and Schlötterer, C. 1999. Microsatellites: Evolution and applications.: Oxford University Press, New York.

Han, H., Zeng, L., Dang, R., Lan, X., Chen, H. & Lei, C., 2015. The DMRT3 gene in Chinese horse breeds. *IFAG* 46, 340-342.

He, S., Zhang, L., Li, W. and Liu, M. 2015. BIEC2-808543 SNP in the LCORL gene is associated with body conformation in the Yili horse. *Anim. Biotechnol.* 26, 289 – 291.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. & Drummond, A., 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *BMC Bioinformatics.* 28(12), 1647-1649.

Lacy, R.C., 1987. Loss of genetic diversity from managed populations: Interacting effects of drift, mutation, immigration, selection, and population subdivision. *Conserv. Biol.* 1, 143-158.

Louw, S., 2008. SA Boerperd – Indigenous breed to southern Africa. In: SA Boerperd: The ideal sport and pleasure horse. Ed: S. Louw, Kejafa Knowledge Works, Maanhaarrand. pp. 22-29.

Makvandi-Nejad, S., Hoffman, G.E., Allen, J.J., Chu, E., Gu, E., Chandler, A.M., Loredo, A.I., Bellone, R.R., Mezey, J.G., Brooks, S.A. & Sutter, N.B., 2012. Four loci explain 83% of size variation in the horse. *PLoS One* 7(7), e39929. doi: 10.1371/journal.pone.0039929

McDonald, J.H., 2014. Handbook of Biological Statistics, 3rd edition. Maryland: Sparky House Publishing.

Mehta, C.R. & Patel, N.R., 1983. A network algorithm for performing Fisher's exact test in r x c contingency tables. *J. Am. Stat. Assoc.* 78(382), 427-434.

Metzger, J., Schrimpf, R., Philipp, U. & Distl, O., 2013. Expression levels of LCORL are associated with body size in horses. *PLoS One* 8(2), e56497. doi: 10.1371/journal.pone.0056497

Nel, C., 2014. SA Boerperd Breed History. SA Boerperd Breeders Society. Available at: <http://www.saboerperd.com/p2/history/sa-boerperd-breed-history.html> [Accessed 5 July 2016].

Pakamisa., 2016. Horseback Safaris. Available at: <http://www.pakamisa.co.za/horseback-safaris.php>

Peakall, R. & Smouse P.E., 2006. GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Resour.* 6, 288-295.

Peakall, R. & Smouse P.E., 2012. GENALEX 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. *BMC Bioinformatics* 28, 2537-2539.

Pritchard, J.K., Stephens, M. & Donnelly, P., 2000. Inference of population structure using multi-locus genotype data. *Genetics*. 155, 945-959.

Promerová, M., Andersson, L.S., Juras, R., Penedo, M.C.T., Reissmann, M., Tozaki, T., Bellone, R., Dunner, S., Hořín, P., Imsland, F., Imsland, P., Mikko, S., Modrý, D., Roed, K.H., Schwochow, D., Vega-Pla, J.L., Mehrabani-Yeganeh, H., Yousefi-Mashouf, N., Cothran, E.G., Lindgren, G. & Andersson, L., 2014. Worldwide frequency distribution of the 'Gait keeper' mutation in the DMRT3 gene. *Anim. Genet.* 45, 274-282.

Robilliard, J.J., Pfau, T. & Wilson, A.M. 2007. Gait characterization and classification in horses. *Journal of Experimental Biology*, 210: 187-197.

SA Boerperd Breeders Society., 2016. SA Boerperd as show horse and sport horse. Available at: <http://www.saboerperd.com/p26/horses/sa-boerperd-as-show-and-sport-horse.html> [Accessed 25 Jan. 2017].

SA Boerperd Breeders Society., 2017. Show Classes (Comprehensive & Minimum) June 2017, 2nd ed. [pdf] Available at: <http://www.saboerperd.com/images/photos/2017-C4-2-Meesterklaslys-Junie2017a.pdf>

Signer-Hasler, H., Flury, C., Haase, B., Burger, D., Simianer, H., Leeb, T. & Rieder, S. 2012. A genome-wide association study reveals loci influencing height and other conformation traits in horses. *PLoS One* 7(5), e37282. doi: 10.1371/journal.pone.0037282

Swart, S. 2010. Riding High: Horses, Humans and History in South Africa. Johannesburg: Wits University Press.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M. & Rozen, S.G., 2012. Primer3 – New capabilities and interfaces. *Nucleic Acids Res.* 40(15), 1-12.

van de Goor, L.H.P., van Haeringen, W.A. & Lenstra, J.A., 2011. Population studies of 17 equine STR for forensic and phylogenetic analysis. *Anim. Genet.* 42(6), 627-633. doi:10.1111/j.1365-2052.2011.02194.x