



Original Article

# Analysis of Genetic Diversity in the American Standardbred Horse Utilizing Short Tandem Repeats and Single Nucleotide Polymorphisms

Elizabeth Esdaile<sup>\*,\*</sup>, Felipe Avila<sup>\*,\*</sup> and Rebecca R. Bellone<sup>\*</sup>

From the Veterinary Genetics Laboratory, School of Veterinary Medicine, University of California–Davis, Davis, CA, USA (Avila, Esdaile, and Bellone) and Department of Population Health and Reproduction, School of Veterinary Medicine, University of California—Davis, Davis, CA, USA (Bellone)

Address correspondence to R. Bellone at the address above, or e-mail: [rbellone@ucdavis.edu](mailto:rbellone@ucdavis.edu).

\*These authors contributed equally to the work.

Received July 29, 2021; First decision September 1, 2021; Accepted November 10, 2021.

Corresponding Editor: Ernest Bailey

## Abstract

American Standardbreds were developed as a harness racing horse breed. The United States Trotting Association closed the studbook in 1973 and implemented a book size cap in 2009. This study aimed to investigate genetic diversity in the American Standardbred after the studbook cap was introduced using short tandem repeats (STRs) and single-nucleotide polymorphisms (SNPs). Sixteen STRs from horses foaled from 2010 to 2015 and their sires and dams ( $n = 50\,621$ ) were utilized to examine allelic richness ( $A_r$ ), expected heterozygosity ( $H_E$ ), observed heterozygosity ( $H_O$ ), unbiased heterozygosity ( $H_U$ ), inbreeding coefficient ( $F_{IS}$ ), and fixation index ( $F_{ST}$ ). These analyses found that trotting and pacing sires were less genetically diverse than dams ( $H_E P_{Bonferroni} = 0.029$  and  $6.3 \times 10^{-5}$ , respectively) and their offspring ( $A_r P_{Bonferroni} = 0.034$  and  $6.9 \times 10^{-6}$ , respectively), and pacing offspring were significantly less diverse than their dams ( $H_E P_{Bonferroni} = 2 \times 10^{-3}$ ). Inbreeding coefficients for trotters ( $F_{IS} = -0.014$ ) and pacers ( $F_{IS} = -0.012$ ) suggest that breeding practices have maintained diversity. Moderate levels of genetic differentiation ( $0.066 < F_{ST} < 0.11$ ) were found between pacing and trotting groups. Additionally, 10 of the most prolific trotting sires and their male offspring ( $n = 84$ ) were genotyped on the 670K Axiom Equine HD Array.  $H_O$  values higher than  $H_E$  ( $P < 0.001$ ), low inbreeding coefficients (mean  $F = -0.064$ ), and mean  $F_{ROH} = 21\%$  indicate relatively high levels of diversity in this cohort, further supporting the STR data. However, in contrast,  $H_O$  values were higher for trotting sires (0.41) than their offspring (0.36). This observation warrants further monitoring of diversity over time. These data provide an updated foundation of diversity indices for further, long-term analysis in the breed.

**Subject:** Genetic diversity

**Keywords:** STRs, SNPs, equine, inbreeding, population genetics

The American Standardbred was developed during the 19th century as a harness racing horse breed that competes at one of 2 different gaits: the trot (a symmetrical 2-beat, diagonally opposed gait) or the pace (a symmetrical 2-beat, lateral gait). Founding horses of the American

Standardbred include individuals of Trotter, Thoroughbred, Norfolk Trotter, Morgan, Barb, Canadian Pacer, and other imported bloodlines (Wallace 1871). Generations of assortative mating separated the American Standardbred into 2 distinct subpopulations based on

gait: trotters and pacers. In 1973, the governing body of American Standardbreds, the United States Trotting Association (USTA), closed the studbook (which includes both pacing and trotting horses), restricting registration to offspring of already registered Standardbreds (Adelman 1981; Cothran et al. 1986).

Closing a studbook limits the introduction of new alleles into the gene pool and thus can impact genetic diversity and the health of animals. For example, in Standardbreds, 26% of the 1987 foal crop (6046 foals) were sired by 2% of the sires (60 horses) (King 1992). A continued reduction in diversity has the chance to increase the frequency of genetic disorders. Among those reported for the breed are atrial fibrillation, tarsal osteochondrosis, and recurrent exertional rhabdomyolysis (Lykkjen et al. 2014; Norton et al. 2016; McCoy et al. 2016, 2019; Kraus et al. 2017, 2018).

In addition to reducing genetic diversity, inbreeding has been found to be correlated with reproductive failure in a few horse breeds with high inbreeding levels. For example, in Friesian horses, higher levels of inbreeding were correlated with an increase in the rate of retained placenta (Sevinga et al. 2004). Thus, evaluating genetic diversity is essential to monitor trends within the breed and to develop genomic resources to assist in maintaining diversity and with which to study heritable diseases and reproductive success.

In the 1980s, inbreeding was analyzed in American Standardbreds using pedigree analysis of 14 generations as well as short tandem repeats (STRs). Together, these studies showed higher levels of inbreeding in trotters when compared with pacers (MacCluer et al. 1983; Cothran et al. 1984, 1986; Cothran et al. 1987). A statistically significant association between higher expected heterozygosity and higher foaling and conception rates was also noted in pacers (MacCluer et al. 1983; Cothran et al. 1986). Furthermore, the mean observed total heterozygosity in both trotters and pacers was found to be statistically significantly decreasing over time, more so in trotters than pacers, based on an investigation of over 18 000 trotters and pacers using blood markers (King 1992). The author hypothesized that such a decrease might be attributed to the closed breeding structure and the significant impact that a single popular sire can have on a breed.

Due to the reported decrease in heterozygosity, the closure of the studbook, and a concern for potential health risks due to inbreeding depression, the USTA imposed a studbook limit of 140 mares covered per year for all trotting stallions debuting in 2009. Similarly, a studbook limit of 160 mares covered per year was imposed on pacing sires who debuted in 2009, decreasing to 150 in 2010 and 140 for those who debuted in 2011 or later. It is unknown how studbook restrictions have affected diversity, or if they have mitigated further loss of heterozygosity. To date, the only single-nucleotide polymorphism (SNP)-based study to investigate diversity in Standardbreds utilized data from the 50K SNP array to compare diversity indices of only 15 American Standardbreds (7 trotters and 8 pacers) and 25 Norwegian Standardbreds (trotters), in addition to 35 other breeds. Unlike Norwegian Standardbreds, American Standardbreds were found to have significant excess homozygosity ( $F_{IS}$ ), which could be attributed to increased inbreeding in the individuals sampled or substructure based on gait (Petersen et al. 2013).

Given the previous studies from the 1980s, which highlighted differences in inbreeding of trotting and pacing populations when analyzing STRs, excess homozygosity observed in a small sample of American Standardbreds using SNP data, and the concern for the impact of book size on diversity indices, we aim to evaluate genetic diversity in the American Standardbred using both STRs and SNPs after the book size limit was imposed. We also aim to evaluate an

extensive data set comprising 50 621 American Standardbreds to further investigate whether pacers are more genetically diverse than trotters (MacCluer et al. 1983). Additionally, our goal was to assess the feasibility of using genotypes from the 670k SNP array on a small subset of trotting individuals as an alternative to the well-established approach of evaluating genetic diversity using STR data. Finally, we hypothesize that dams are maintaining genetic diversity in American Standardbreds.

## Methods

### STR Analysis

#### Data Collection

To establish baseline values of genetic diversity indices to monitor trends over time and the potential effects of the 2009 studbook cap, we evaluated the breeding stock and first 6 foal crops after the rule was imposed (2010–2015). Genotyping records for all foals who's sires bred for at least 4 consecutive seasons from 2010 to 2015 (“offspring”), as well as their sires (“sires”), and dams (“dams”), were genotyped by Bureau Veritas Laboratories as part of the routine parentage verification process for registration with the USTA and data provided for use in this study. Horses used in this study were categorized by gait phenotype as either trotter or pacer. Phenotyping was performed by the USTA based on performance records.

#### Data Analysis

Horses were grouped according to gait (trotters or pacers) and designated subpopulations (sires, dams, and offspring) (Table 1). Records from 191 pacing dams, 77 trotting dams, 244 pacing offspring, and 40 trotting offspring were removed from the analysis due to duplicate samples and incomplete records, leaving 50 621 horses in our analyses. To compare the impact of book size (i.e., the number of mares bred by a stallion per year) within and between gait types, the sires and offspring were divided into 3 groups based on both the total number of offspring for the years under investigation and sample sizes within each group. Groupings were divided based on approximately 33% of the total offspring and categorized as “high-book,” “mid-book,” and “low-book,” accordingly. The highest producing sires, those that cumulatively sired 40% of trotters and 33% of pacers over the 6 years, were classified as “high-book sires.” Sires that cumulatively sired 28% of trotters and 34% of pacers were classified as “mid-book sires,” and the remaining, lowest producing sires were classified as “low-book sires” (Table 1). Offspring were classified by sire book size in addition to year of birth (Table 1).

Sixteen STRs were analyzed: AHT4, AHT5, ASB17, ASB2, ASB23, CA425, HMS1, HMS2, HMS3, HMS6, HMS7, HTG10, HTG4, HTG7, LEX33, and VHL20. These markers are routinely used for equine parentage testing and genetic diversity studies as they are located on 12 different chromosomes and are not in linkage disequilibrium (LD) (Sereno et al. 2008; van de Goor et al. 2011; Khanshour et al. 2013). STR data were formatted for analysis with Microsatellite Analyzer 4.05 (MSA) (Dieringer and Schlötterer 2003). For each marker, MSA was used to calculate allelic richness, expected heterozygosity, observed heterozygosity, unbiased heterozygosity, inbreeding coefficient, and pairwise fixation index using 20 000 permutations ( $F_{ST}$ ) (Nei 1978; Weir & Cockerham 1984). Mean allelic richness ( $A_r$ ), mean expected heterozygosity ( $H_E$ ), mean observed heterozygosity ( $H_O$ ), mean unbiased heterozygosity ( $H_U$ ), and mean inbreeding coefficient ( $F_{IS}$ ) were then calculated by averaging the values of each locus.  $T$ -test comparisons of  $H_E$ ,

**Table 1.** Composition of subpopulations and groups included in the analysis. The percentage is the size of the group when compared with the respective subpopulation (bold). Trotters: high-book sires ranged from an average of 71–107 offspring per year ( $\bar{\mu} = 78$ ), mid-book sires ranged from an average of 22–67 offspring per year ( $\bar{\mu} = 37$ ), low-book sires ranged from an average of 0.33–21 offspring per year ( $\bar{\mu} = 5.9$ ). Pacers: high-book sires ranged from an average of 82–136 offspring per year ( $\bar{\mu} = 102$ ), mid-book sires ranged from an average of 35–73 offspring per year ( $\bar{\mu} = 52$ ), low-book sires ranged from an average of 0.67–34 offspring per year ( $\bar{\mu} = 9.3$ )

Population		Number of trotters (percent of group type)	Number of pacers (percent of group type)
Group	Subgroup		
All		22 731	27 890
Dams		<b>7314</b>	<b>9260</b>
Sires		<b>171</b>	<b>140</b>
	<i>High-Book Sires</i>	13 (7.6%)	10 (7.1%)
	<i>Mid-Book Sires</i>	19 (11%)	20 (14%)
	<i>Low-Book Sires</i>	139 (81%)	110 (79%)
Offspring		<b>15 246</b>	<b>18 490</b>
	<i>Offspring of High-Book Sires</i>	6105 (40%)	6078 (33%)
	<i>Offspring of Mid-Book Sires</i>	4206 (28%)	6259 (34%)
	<i>Offspring of Low-Book Sires</i>	4935 (32%)	6153 (33%)
	<i>2010 Offspring</i>	2728 (18%)	3284 (18%)
	<i>2011 Offspring</i>	2883 (19%)	3497 (19%)
	<i>2012 Offspring</i>	2814 (19%)	3337 (18%)
	<i>2013 Offspring</i>	2544 (17%)	2956 (16%)
	<i>2014 Offspring</i>	2287 (15%)	2886 (16%)
	<i>2015 Offspring</i>	1990 (13%)	2530 (14%)

$H_O$ ,  $F_{IS}$ , and  $A_r$ , between and within populations, as well as linear models (lm) of offspring foaled from 2010 to 2015, were calculated using R (R Core Team 2020). A Bonferroni-corrected significance level of  $<0.05$  was used to determine the significance of all  $t$ -tests ( $P_{\text{Bonferroni}} = 1 - (1 - \alpha)^n$ ).

## SNP Analysis

### Data Collection and Genotyping

The total number of foals sired from 2010 to 2015 was used to identify 10 of the most prolific trotting sires from the study period, excluding full-siblings and father-son pairs. These 10 sires contributed to a combined 31% of the total foal crop from 2011 to 2015. To evaluate the representative diversity of these high-book stallions and their male offspring, 6–10 offspring per sire were included in the analysis, based on sample availability, for a total of 96 individuals. Genomic DNA was isolated from mane or tail hair follicles collected at the time of registration using the Qiagen Genra Puregene Blood Kit as previously described (Mack et al. 2017). Genotyping was conducted at Geneseeq Inc. (Lincoln, NE) using the 670K Axiom Equine HD Array. Upon initial quality control, 44 samples had a genotyping rate between 94% and 97%, and 50 samples had a genotyping rate  $> 97\%$ . Two samples failed initial quality control limits and were removed from downstream analyses.

### SNP-Based Diversity Analysis

SNP pruning was conducted using PLINK v.1.90 (Purcell et al. 2007; Chang et al. 2015; [https://www.cog-genomics.org/plink/1.9/general\\_usage#cite](https://www.cog-genomics.org/plink/1.9/general_usage#cite)). Only autosomal SNPs were evaluated. SNPs with minor allele frequency less than 0.05, genotyping rate  $< 95\%$ , and in strong LD ( $>0.8$ ) were removed (`--maf 0.05 --geno 0.05 --indep-pairwise 50 5 0.8`). The pruned data set included 111 688 SNPs. Using the pruned SNP set, expected heterozygosity ( $H_E$ ), observed heterozygosity ( $H_O$ ), and method-of-moments inbreeding coefficient estimates ( $F$ ) were calculated using the `--het` function in PLINK v.1.90 (Purcell et al. 2007). Pairwise comparisons between groups were

performed using Mann–Whitney  $U$ -tests. Principal component analysis (PCA) was also conducted in PLINK v.190 (`--pca`), and results were plotted using the R package `ggplot2` (Wickham 2016). PLINK files were converted to Arlequin format using PGDSpider version 2.1.1.5 (Lischer & Excoffier 2012), and Arlequin version 3.5.2.2 (Excoffier et al. 2007) was used to calculate pairwise  $F_{ST}$  values between sire families, as well as between sires and their offspring. Significance testing was conducted using 20 000 permutations and  $\alpha = 0.05$ .

### Runs of Homozygosity and Inbreeding Estimates

Genomic inbreeding was also calculated for a subset of 50 individuals with the highest quality SNP data (genotyping rate  $> 97\%$ ), which was comprised of offspring only. Runs of homozygosity (ROH)  $> 1.0$  Mb in length were identified using the `--homozyg` command in PLINK without prior MAF or LD pruning (Meyermans et al. 2020). A total of 407 091 autosomal SNPs were used for analysis using the following parameters: `--homozyg`; `--homozyg-density 25`; `--homozyg-gap 1000`; `--homozyg-kb 1000`; `--homozyg-snp 30`; `--homozyg-window-het 1`; `--homozyg-window-missing 2`; `--homozyg-window-snp 30`. Then, the  $F_{ROH}$  statistic (McQuillan et al. 2008) was calculated for each individual using the formula:  $F_{ROH} = L_{ROH}/L_{AUTO}$ , where  $L_{ROH}$  corresponds to the total length of all ROH  $> 1.0$  Mb per animal, and  $L_{AUTO}$  is the length of the equine autosomal genome (2 474 930 kb) ([https://www.ncbi.nlm.nih.gov/genome/145?genome\\_assembly\\_id=22878](https://www.ncbi.nlm.nih.gov/genome/145?genome_assembly_id=22878)).

## Results

### STR Analysis

#### Allelic Richness ( $A_r$ )

In both trotting and pacing subpopulations, high-book sires and mid-book sires had significantly lower  $A_r$  than low-book sires (trotters  $A_r = 4.4$  and  $5.0$ ;  $t$ -test:  $P_{\text{Bonferroni}} = 1.6 \times 10^{-5}$  and  $9.3 \times 10^{-3}$ ; pacers  $A_r = 4.0$  and  $4.7$ ;  $t$ -test:  $P_{\text{Bonferroni}} < 0.001$  and  $0.017$ ) (Table

**Table 2.** Measures of diversity in American Standardbreds calculated using 16 STRs. Presented are allelic richness ( $A_i$ ), expected, unbiased, and observed heterozygosity ( $H_E$ ,  $H_U$ , and  $H_O$ ), inbreeding coefficient ( $F_{IS}$ ), and pairwise fixation index ( $F_{ST}$ ) values of trotting and pacing Standardbreds foaled from 2010 to 2015 and their sires and dams. Statistically significant comparisons are denoted by corresponding letters\*.

Population Group Subgroup	Trotters						Pacers						Pairwise $F_{ST}$ of Trotters and Pacers
	$A_i$	$H_E$	$H_U$	$H_O$	$F_{IS}$	$A_i$	$H_E$	$H_U$	$H_O$	$F_{IS}$			
All	8.63	0.67	0.66	0.68	-0.014	9.1	0.67	0.67	0.68	-0.012	0.076		
Dams	abc	ab	ab	‡, a	ab	abc	abcdeghijkl	0.68	‡, abcdefghi	-0.017	0.066		
Sires	clmnopqrstu	b	0.66	0.68	-0.041	6.0	a	0.66	0.68	-0.029	0.092		
High Book Sires	bdefg	0.64	0.62	0.68	-0.095	4.0	bdefg	0.64	0.72	-0.11	0.11		
Mid Book Sires	ahijk	0.67	0.66	0.74	-0.12	4.7	ahijk	0.62	0.65	-0.041	0.076		
Low Book Sires	abdefghijk	ac	0.65	0.67	-0.029	5.9	abcdeghijk	0.66	0.68	-0.026	0.093		
Offspring	gkl	0.66	0.66	0.67	-0.015	8.9	gkl	0.66	0.67	-0.012	0.082		
Offspring of High Book Sires	dio	0.65	0.64	0.67	-0.036	7.9	dio	0.65	0.67	-0.023	0.099		
Offspring of Mid Book Sires	ehn	0.66	0.66	0.67	-0.011	7.8	ehn	0.65	0.66	-0.015	0.080		
Offspring of Low Book Sires	fjm	0.67	0.67	0.67	-0.0056	7.9	fjm	0.67	0.68	-0.011	0.077		
2010 Offspring	p	0.66	0.66	0.66	-0.0064	7.7	p	0.67	0.68	-0.016	0.080		
2011 Offspring	q	0.66	0.66	0.66	-0.0049	7.4	q	0.66	0.67	-0.014	0.082		
2012 Offspring	r	0.66	0.66	0.67	-0.0052	7.9	r	0.66	0.67	-0.0089	0.083		
2013 Offspring	s	0.66	0.66	0.66	-0.0041	7.5	s	0.66	0.67	-0.010	0.081		
2014 Offspring	t	0.66	0.66	0.66	-0.0079	7.4	t	0.66	0.67	-0.011	0.083		
2015 Offspring	u	0.66	0.66	0.66	-0.0100	7.4	u	0.66	0.67	-0.012	0.084		

\* Unique letters in the upper right corner of each cell denote all statistically significant pairwise comparisons within a column based on a t-test and  $P_{Bonferroni} < 0.05$ . For example, when looking at  $A_i$  of trotters, “a” denotes that trotting Dams, Mid Book Sires, and Low Book Sires, are each statistically significantly different in  $A_i$  from one another. Additionally, “u” in the same column denotes that trotting Sires and 2015 Offspring have statistically significantly different  $A_i$ . ‡ signifies when  $H_O$  was statistically significantly different from  $H_E$ .

2 and Supplementary Table 1). Sires as a whole have lower  $A_r$  when compared with dams and offspring (trotters  $A_r = 5.9, 8.1, 8.2$ ;  $t$ -test:  $P_{\text{Bonferroni}} = 4.8 \times 10^{-3}$  and 0.034; pacers  $A_r = 6.0, 8.4, 8.9$ ;  $t$ -test:  $P_{\text{Bonferroni}} = 2.4 \times 10^{-3}$  and  $6.9 \times 10^{-6}$ ). Even when parsed by book size, offspring have higher  $A_r$  than their respective sire groups ( $A_r$ —high-book sires and their offspring; trotters = 4.4 and 7.6, pacers = 4.0 and 7.9, mid-book sires and their offspring; trotters = 5.0 and 7.6, pacers = 4.7 and 7.8, low-book sires and their offspring; trotters = 5.9 and 7.4, pacers = 5.9 and 7.9; all  $P_{\text{Bonferroni}} < 0.02$ ).  $A_r$  of offspring did not differ when parsed by stallion book size or offspring year of birth ( $t$ -tests:  $P_{\text{Bonferroni}} > 0.05$ ).  $A_r$  did not differ when comparing respective groups of trotters to pacers, including the comparison of all trotters when compared with all pacers ( $t$ -test:  $P > 0.05$ ) (Table 2 and Supplementary Table 1).

#### Expected Heterozygosity ( $H_E$ )

Trotting and pacing dams had higher average  $H_E$  when compared with sires ( $H_E$  trotters = 0.67 and 0.65;  $H_E$  pacers = 0.68 and 0.66;  $t$ -test:  $P_{\text{Bonferroni}} = 0.03$  and  $<0.001$ , respectively). Pacing dams had statistically significantly higher  $H_E$  than that of offspring ( $H_E = 0.68$  and 0.66, respectively;  $t$ -test:  $P_{\text{Bonferroni}} = 0.002$ ) (Table 2 and Supplementary Table 1). Offspring of low-book sires have significantly higher  $H_E$  than their sire group ( $H_E$  trotters = 0.67 and 0.65;  $H_E$  pacers = 0.67 and 0.66;  $t$ -test: trotters,  $P_{\text{Bonferroni}} = 0.046$ , pacers  $P_{\text{Bonferroni}} = 0.024$ ). No significant difference in  $H_E$  was found when comparing trotters and pacers, both as a whole and by group ( $t$ -test:  $P_{\text{Bonferroni}} > 0.05$ ) (Table 2 and Supplementary Table 1).

#### Unbiased Heterozygosity ( $H_U$ )

Nei's unbiased heterozygosity ( $H_U$ ) (Nei 1978) was also calculated when parsing the entire sire population by book size in order to account for differences in the number of individuals among groups. These data corroborate what was identified using  $H_E$ , with only one additional statistically significant difference between groups:  $H_U$  of pacing mid-book sires was statistically significantly different than that of dams ( $H_U = 0.62$  and 0.68, respectively,  $t$ -test:  $P_{\text{Bonferroni}} = 0.040$ ) (Table 2 and Supplementary Table 1).

#### Observed Heterozygosity ( $H_O$ )

$H_O$  of pacing dams was significantly higher than that of offspring as a whole and offspring of mid- and low-book sires ( $H_O = 0.69, 0.67, 0.66$ , and 0.68;  $t$ -test:  $P_{\text{Bonferroni}} < 0.001, 3.8 \times 10^{-3}, 0.012$ , respectively). In trotters, only  $H_O$  of dams was significantly higher than that of offspring of low-book sires ( $H_O = 0.68$  and 0.67, respectively;  $t$ -test:  $P_{\text{Bonferroni}} = 3.2 \times 10^{-3}$ ). No other comparisons of  $H_O$  were significantly different (Table 2 and Supplementary Table 1).

#### $H_O$ and $H_E$

$H_O$  was statistically significantly higher than  $H_E$  in both trotting and pacing horses as a whole, dams, total offspring, and offspring of high-book sires. It was also statistically significantly higher in pacing offspring of mid- and low-book sires, pacing offspring foaled in 2010, 2011, and 2015, and trotting mid-book sires (Table 2 and Supplementary Table 1,  $t$ -test:  $P_{\text{Bonferroni}} > 0.05$ ).

#### Inbreeding Coefficients ( $F_{IS}$ )

Estimated inbreeding coefficients ( $F_{IS}$ ) below zero were observed for all groups (Table 2). In trotters, offspring of high-book sires were found to have lower  $F_{IS}$  than offspring of mid-book sires, offspring of low-book sires, and dams ( $P_{\text{Bonferroni}} = 3.3 \times 10^{-5}, 3.7 \times 10^{-4}$ , and

$4.5 \times 10^{-3}$ , respectively) (Table 2 and Supplementary Table 1). In trotters, dams had higher  $F_{IS}$  than mid-book sires ( $P_{\text{Bonferroni}} = 0.045$ ). In pacers, only offspring of high-book sires had significantly lower  $F_{IS}$  than offspring of low-book sires ( $P_{\text{Bonferroni}} = 0.03$ ). No significant differences in  $F_{IS}$  were observed when comparing respective trotters and pacers as a whole or by respective groups ( $t$ -test:  $P > 0.05$ ) (Supplementary Table 1).

#### Fixation Index ( $F_{ST}$ )

Moderate differentiation ( $0.05 < F_{ST} < 0.15$ ) (Wright 1978) in pairwise  $F_{ST}$  was found when comparing respective groups of trotters to pacers (Table 1 and Supplementary Table 2). Dams had the lowest pairwise  $F_{ST}$  between respective groups ( $F_{ST} = 0.066$ ). High-book sires had the highest pairwise  $F_{ST}$  between respective groups ( $F_{ST} = 0.11$ ), with their offspring not far behind ( $F_{ST} = 0.099$ ). Within the trotting and pacing subpopulations, pairwise  $F_{ST}$  was below 0.05 in all intergroup comparisons, indicating that little to no genetic differentiation occurred within gait subgroups (Supplementary Table 2).

#### Longitudinal Analysis

Pacing dams were found to have statistically significantly higher  $H_E$  and  $H_O$  than their respective offspring when grouped by year ( $P_{\text{Bonferroni}} < 0.029$ ), while pacing and trotting sires had lower  $A_r$  than offspring grouped by year ( $P_{\text{Bonferroni}} < 0.031$ ) (Supplementary Table 2). Additional analysis of offspring parsed by year observed  $A_r$ ,  $H_E$ , and  $H_O$  to be experiencing a downward trend across the 6 years examined (Figure 1). Although still below zero,  $F_{IS}$  of trotting offspring experienced an overall decrease across the 6 years, but increased in pacing offspring. Across this same period, pairwise  $F_{ST}$  between offspring of trotters and pacers increased, indicating continued differentiation between the groups. Although there were no statistically significant changes in any of the metrics examined across the 6 years investigated ( $P > 0.05$ ) (Figure 1), these data support the need for further examination of future generations. The different trends in  $F_{IS}$  of both subpopulations should be explored by SNP analysis to better investigate these trends and potential causes of differentiation.

#### SNP Analysis

##### Genomic Diversity

Genome-wide observed heterozygosity ( $H_O$ ) values ranged from 0.31 to 0.51, with a mean individual heterozygosity of 0.36 (Supplementary Table 3). This value was significantly higher than the expected heterozygosity ( $H_E$ ) for the study cohort ( $P < 0.001$ ). Additionally, mean observed heterozygosity was significantly higher for sires (0.41) when compared to their offspring (0.36;  $P < 0.001$ ). No significant pairwise differences in heterozygosity levels were detected between half-sibling families.

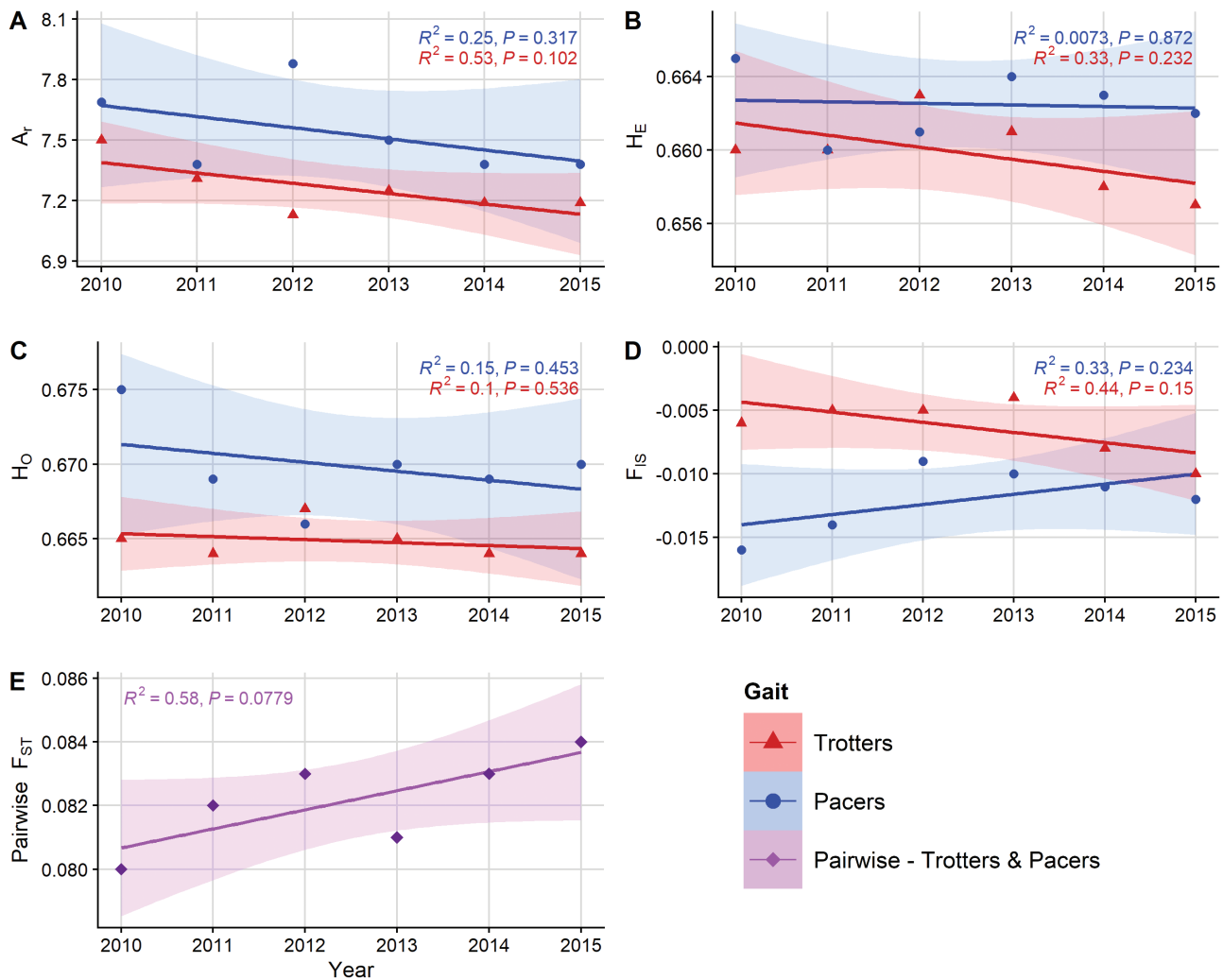
##### Inbreeding Coefficients ( $F$ )

Individual inbreeding coefficients ( $F$ ) based on observed and expected heterozygosity showed low estimates of inbreeding and an excess of heterozygosity in the study cohort, ranging from  $-0.48$  to  $0.086$  (mean =  $-0.064$ ) (Supplementary Table 3). Sires were found to be significantly less inbred than their offspring, as expected, based on mean  $F$  ( $-0.18$  and  $-0.051$ , respectively;  $P < 0.001$ ), but no significant pairwise differences in inbreeding levels were found between half-sibling families.

##### ROH Identification and $F_{ROH}$ Estimation

A total of 5756 ROH above 1 Mb in length were identified in the subset of 50 Standardbreds with high-quality SNP genotyping





**Figure 1.** Trends in diversity measures of offspring foaled from 2010 to 2015.  $A_r$ ,  $H_c$ , and  $H_o$  trended downward, suggesting ongoing loss of diversity within the breed.  $F_{is}$  differed between groups, increasing in pacers and decreasing in trotters. Pairwise  $F_{ST}$  trended upwards, suggesting ongoing differentiation between the 2 subpopulations.

data. The average number of ROH segments per individual was  $115 \pm 1.8$ , with values ranging from 82 to 154. The longest ROH identified for an individual spanned 51 Mb of ECA16, or 58% of the total chromosomal length. The average sum of all ROH lengths per animal was 522 Mb (mean ROH length = 4.5 Mb; median ROH length = 2.7 Mb) (Table 3). The number of segments < 2 Mb in length represented 66% of all ROH, whereas segments greater than 8.0 Mb, indicating recent inbreeding, corresponded to 15% of the total (Figure 2, Supplementary Table 4). Genomic inbreeding estimates using the proportion of the genome covered by ROH ( $F_{ROH}$ ) for this cohort ranged from 12% to 27% (mean  $F_{ROH} = 21\%$ ) (Table 3).

### Principal Component Analysis

Relationships among individuals belonging to different sire families were also investigated using PCA (Figure 3). PC 1 and 2 explain 19% of the genetic variation within this sample set. Most families cluster tightly, while some are dispersed across one PC (e.g., family 1) or both PCs (such as family 5). As expected, offspring of half-sibling families that shared a grandsire (indicated by “a” and “b”) grouped in the same cluster of the PCA. The sample set comprising the most

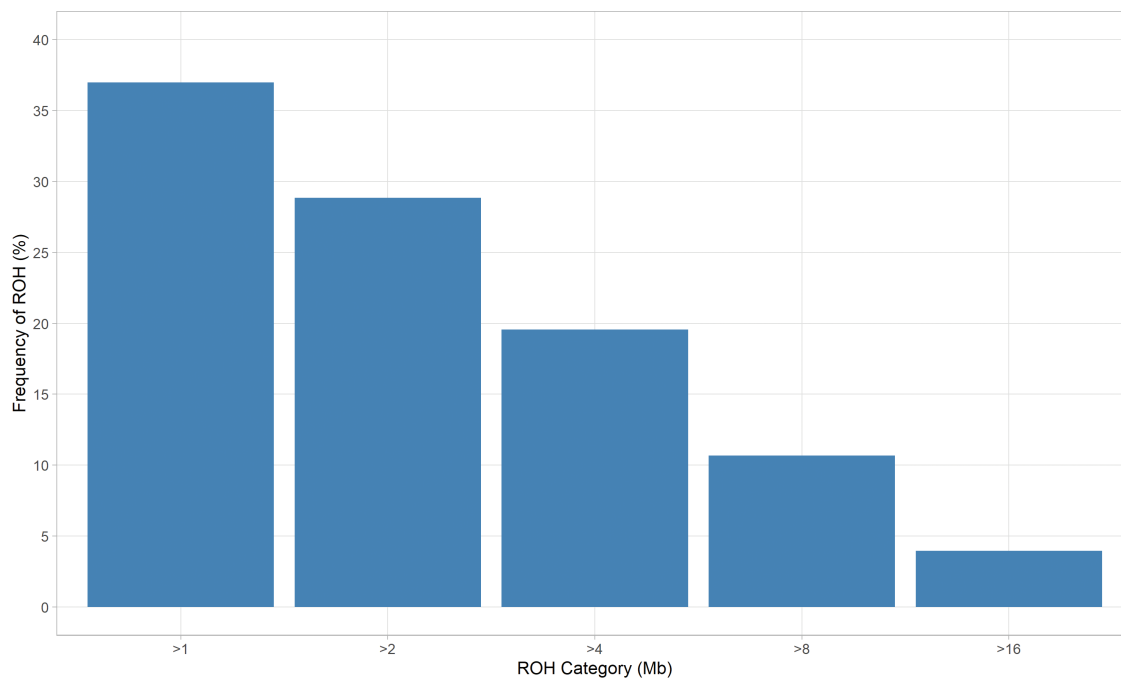
**Table 3.** Runs of homozygosity (ROH) statistics per individual. Mean (Mb), standard error (SE), standard deviation (SD), min (Mb), max (Mb),  $S_{ROH}$  (mean genome length covered by ROH),  $N_{ROH}$  (mean number of ROH),  $L_{ROH}$  (mean ROH length in Mb), and  $F_{ROH}$  (inbreeding coefficient) in a subset of 50 Standardbreds with high-quality SNP genotyping data

Total sample	Mean	SE	SD	Minimum	Maximum
$S_{ROH}$	522	10	73	307	676
$N_{ROH}$	115	1.8	13	82	154
$L_{ROH}$	4.6	0.09	0.67	3.2	6.2
$F_{ROH}$	0.21	0.004	0.03	0.12	0.27

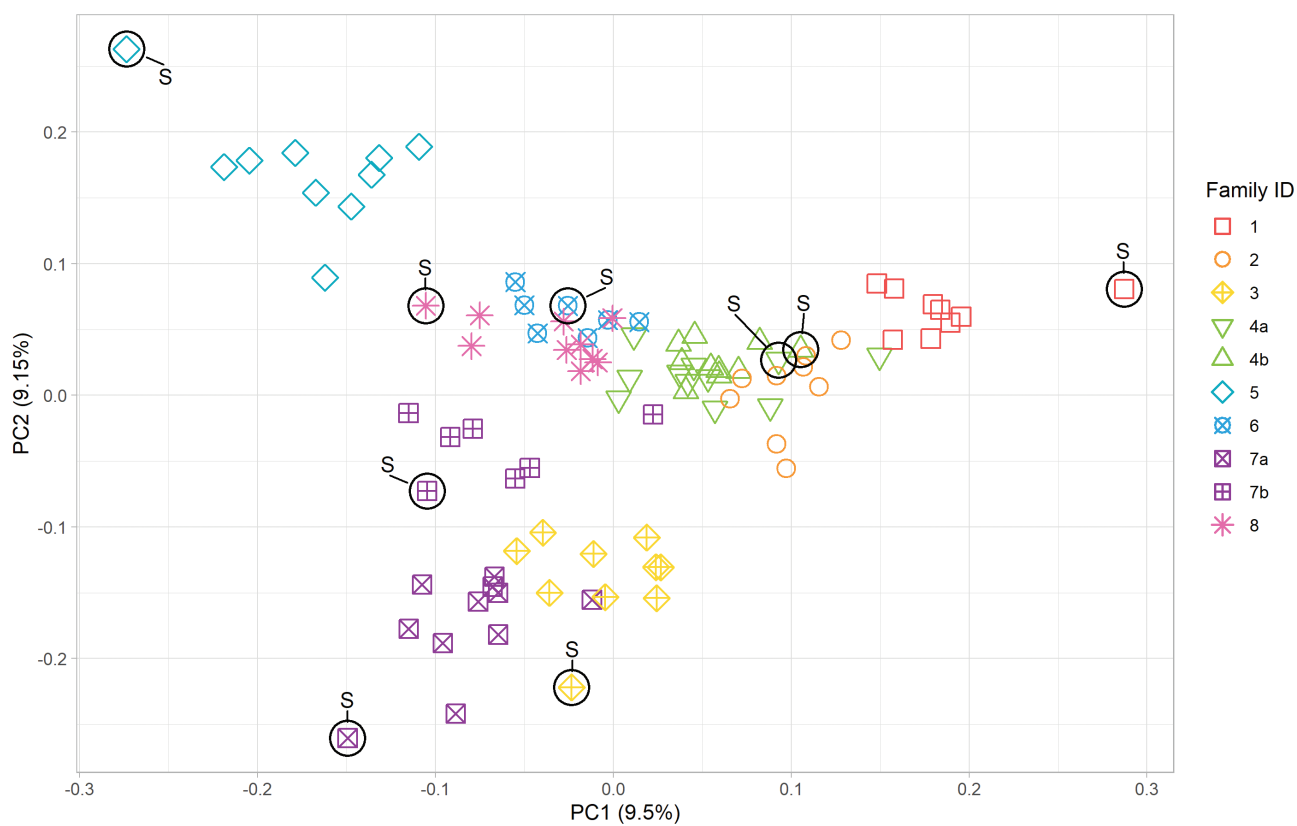
likely highest inbred portion of the trotting subpopulation shows a central cluster of 5 families (Family 2, 4a, 4b, 6, 7b, 8, Figure 3).

### $F_{ST}$

All pairwise  $F_{ST}$  values calculated between sire families were significant ( $P < 0.001$ ) using 20 000 permutations. Families 4b and 8 showed the lowest level of genetic differentiation ( $F_{ST} = 0.082$ ), whereas the greatest divergence was found between families 6 and 7a



**Figure 2.** Frequency of runs of homozygosity (ROH) based on length categories (in Mb) identified in the cohort of 50 Standardbreds with highest quality genotyping data.



**Figure 3.** PCA of trotting Standardbreds included in the SNP analysis ( $n = 94$ ). Sire families are denoted by number, and families who share a grandsire are denoted as "a" and "b." The sire of each family is denoted with an S, except for the sire of family 2, who was excluded due to poor quality data. Numbers in parentheses in the axes titles show the percentage variance explained by each PC.

( $F_{ST} = 0.13$ ). As expected, pairwise  $F_{ST}$  values between families that shared the same grandsire (offspring within families 4a and 4b; 7a and 7b) were among the lowest observed in the dataset ( $F_{ST} = 0.096$  for both) (Supplementary Table 5).

## Discussion

In this study, genotyping records of 16 STRs in 50 621 American Standardbreds and genome-wide SNP genotypes from a subset of 94 trotting individuals were used to investigate genetic diversity metrics in the breed across 6 years from 2010 to 2015. Additionally, STRs were utilized to analyze the contribution of dams to within-breed diversity as well as diversity between pacers and trotters. The application of these markers is widespread in population genetic studies due to their robustness to discern genetic differences between individuals using a limited number of highly polymorphic loci and low-quality input DNA. Furthermore, because STR genotypes are readily available for every registered Standardbred, this approach allows for an unbiased investigation of diversity in the entire population in a reproducible and cost-effective way. In contrast, high-density SNP genotyping is rapidly evolving as an alternative strategy to investigate more global genetic diversity, albeit with increased costs.

Analyses utilizing STRs suggested a lack of allelic richness ( $A_r$ ) in Standardbred sires when compared with offspring and dams, indicating lower diversity in the sire population. When parsed by percentage contribution of offspring, high-book and mid-book sires demonstrated lower  $A_r$  than low-book sires. While it is possible that this difference can be attributed to a smaller number of horses in the high-book sire ( $n = 13$  trotters and 10 pacers) and mid-book sire ( $n = 19$  trotters and 20 pacers) groups, these may represent true differences in diversity among the groups, and this diversity metric should continue to be monitored to make comparisons across generations.

$H_E$  of dams was significantly higher than sires and pacing offspring, while offspring as a whole were not significantly different from sires. However, offspring of low-book sires had significantly increased  $H_E$  when compared with their own sire group. A statistically significant difference in  $H_E$  between the offspring of high- and mid-book sires and their respective sire groups was not detected. This may be due to the short time period examined or small group size, which only includes 23 and 39 high- and mid-book sires, respectively.

Although grouping sires based on book size led to small population sizes in the high- and mid-book groups, sires with at least four consecutive breeding seasons during this period were included in the analysis and thus these data represent nearly the entire population from 2010 to 2015. Nei's unbiased heterozygosity ( $H_U$ ) was calculated to correct for sample size differences among the groups and results for  $H_U$  mirror that of  $H_E$ , with the exception of statistically significantly lower  $H_U$  in pacing mid-book sires compared to pacing dams, which was not detected with  $H_E$ . These findings suggest additional data may be necessary to determine the extent of substantial genetic differentiation between the high- and mid-book sire groups compared with dams.

The relatively small number of markers evaluated by the STR analysis could have also hindered the ability to identify small differences in  $H_E$  between sire and offspring groups. In support of this, SNP analyses utilizing genotypes from over 111 000 SNPs in 10 of the most prolific trotting sires and their offspring did detect a difference in genome-wide diversity between these groups. However, contrary to the STR results, the SNP-based analysis found sires to have significantly higher  $H_E$  and  $H_O$  when compared with their offspring.

This could be caused by sampling bias in the offspring evaluated and should be validated by testing additional offspring as well as also examining mid-book sires and their offspring with a genome-wide SNP-based analysis.

Significantly higher  $H_O$  when compared with  $H_E$  estimates were detected using both STRs and SNPs, suggesting diversity at the time of book size cap was being maintained in American Standardbreds. Furthermore, other studies comparing SNP-based estimates of heterozygosity to STR results have shown that STRs can yield inflated heterozygosity values as a consequence of ascertainment bias as it pertains to the number of genomic regions under investigation, which is caused by the utilization of a limited number of highly polymorphic markers (Fischer et al. 2017). Overall, our genomic data in trotters support what was observed in the STR data and provide justification for additional analysis using historical STR data with which to make comparisons both before and after the book size cap gaining a broader perspective of diversity in the American Standardbred. Furthermore, high levels of diversity observed by higher values of  $H_O$  when compared with  $H_E$  found in both STR and SNP data are supported by negative inbreeding coefficients, with both metrics indicating that in the years studied, the breeding practices utilized have not negatively impacted diversity in the population.

Additionally, our STR analysis found no significant change in  $H_E$  or any other measured metric across the 6 years of offspring examined. Although, all values except for  $F_{IS}$  in pacers demonstrate a downward trend, thus showing a potential increase in inbreeding levels. Although not significant, these findings suggest that an increase in the rate of inbreeding may have been detected if a larger timespan was investigated. In support of this, from 2010 to 2015, a decreasing number of sires contributed to an increasing percentage of the annual foal crop. In comparison to a previous report stating that 60 sires (2%) were responsible for 26% of the 1987 foal crop (King 1992), at most 15 sires (4.8% of sires) were responsible for the same 26% of annual foal crop studied, from 2010 to 2015 (13, 15, 15, 14, 12, and 10 sires, respectively). Meanwhile, the top 60 sires were responsible for two-thirds of the offspring studied from 2010 to 2015. This increase in the percent of offspring produced by top sires is concerning because continued utilization of a small sire pool can lead to increased inbreeding practices and therefore leaves the breed vulnerable to recessive genetic diseases that could cause significant economic impact (Adams et al. 2016). It is also possible that a statistically significant difference was not observed in genetic metrics evaluated over the 6-year time frame because the studbook cap of 140 mares per year stabilized any genetic loss that may have occurred in generations prior. Future studies comparing these data to both previous and future generations will help to gain a better understanding of the impact of inbreeding and book size limits within this closed population.

The higher genetic diversity found in dams compared with other groups (as measured by  $A_r$ ,  $H_O$ ,  $H_U$ , and  $H_E$ ) may indicate that dams are indeed maintaining the genetic diversity of offspring and counteracting the decreasing number of sires contributing to a larger percentage of the foal crop. If the broodmare population experiences a substantial decrease in the number of individuals or a significant loss of diversity due to shifts in breeding practices, the diversity of the American Standardbred could be threatened.

When comparing subpopulations, the STR analysis found no significant differences in the amount of inbreeding between the trotting and pacing American Standardbreds. This differs from the results of previous pedigree studies, which did support differences in inbreeding of trotters and pacers (MacCluer et al. 1983; Cothran et al. 1984, 1986;



). Nonetheless, the moderate  $F_{ST}$  values we found indicate that pacing and trotting groups are indeed differentiating from each other, and more so each year. This shows that although the 2 subpopulations have similar levels of genetic diversity, they are 2 moderately different populations. This differentiation was expected as there is minimal interbreeding between trotting and pacing lines, with gene flow most likely to occur from trotters to pacers (Cothran et al. 1987). However, the extent of this gene flow has yet to be investigated. Within-breed diversity indices calculated using SNP data in this study were higher than those previously reported for a cohort of 40 Standardbreds from the USA and Norway (Petersen et al. 2013). In that study, the mean individual inbreeding coefficient ( $F$ ) and expected heterozygosity ( $H_E$ ) were 0.13 and 0.30, respectively; here, we report a mean  $F$  of  $-0.064$  and  $H_E = 0.34$ . As demonstrated by these indices, higher diversity levels in our Standardbred cohort might be attributed to greater sample size and SNP density. Furthermore, mean  $F_{ROH}$  calculated for a subset of 50 Standardbreds in this study (21%) was similar to that found for 55 Thoroughbred horses born in 1994 ( $F_{ROH} = 21\%$ ), after which point it increased yearly until 2017 ( $F_{ROH} = 23\%$ ) (McGivney et al. 2020). Given that the Thoroughbred is also a closed studbook, it is likely that homozygosity levels in the Standardbred will experience similar increases and loss in diversity over the next 20 years unless carefully monitored and considered in breeding decisions. Furthermore, ROH patterns obtained for this cohort of trotting Standardbreds constitute the foundation for subsequent studies aimed at managing inbreeding rates and identifying genomic signatures of artificial selection for traits of economic importance.

These data constitute a snapshot of diversity trends in American Standardbreds over a 6-year period and illustrate the value in utilizing available STR data to assist in monitoring population-level trends in diversity. Furthermore, the SNP data corroborated STR data from the high-book trotting stallion lines investigated and further illustrates the value in supplementing STR data with high-density SNP data for deeper evaluation. Future SNP-based studies using larger populations of trotting and pacing Standardbreds are required to explore genome-wide diversity in the breed. Together, our data provide a basis of breed-wide genetic diversity in American Standardbreds and serve as a reference point for future studies.

## Supplementary Material

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

## Funding

This work was supported by the United States Trotting Association.

## Acknowledgments

The authors gratefully acknowledge the assistance of Bruce Brinkerhoff Anne Chunko, and T.C. Lane from the United States Trotting Horse Association, as well as the staff at the Bureau Veritas Laboratory for providing samples, pedigree, and STR data that made this study possible. We also graciously acknowledge Rayann Eaves for her technical assistance. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Rebecca R. Bellone, Felipe Avila, and Elizabeth Esdaile are affiliated with the UC Davis Veterinary Genetics Laboratory, a laboratory offering parentage and diagnostic DNA tests in horses and other species.

## Data Availability

These data will be made available upon reasonable request to the corresponding author.

## References

- Adams HA, Sonstegard TS, VanRaden PM, Null DJ, Van Tassell CP, Larkin DM, Lewin HA. 2016. Identification of a nonsense mutation in APAF1 that is likely causal for a decrease in reproductive efficiency in Holstein dairy cattle. *J Dairy Sci.* 99(8):6693–6701.
- Adelman ML. 1981. The first modern sport in America: harness racing in New York City, 1825–1870. *J Sport Hist.* 8(1):5–32.
- Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience.* 4:7.
- Cothran EG, MacCluer JW, Weitkamp LR, Guttormsen SA. 1986. Genetic variability, inbreeding, and reproductive performance in standardbred horses. *Zoo Biol.* 5(2): 191–201.
- Cothran, EG, MacCluer, JW, Weitkamp, LR, Pfennig, DW, Boyce, AJ. 1984. Inbreeding and reproductive performance in standardbred horses. *J Hered.* 75(3): 220–224.
- Cothran EG, MacCluer JW, Weitkamp LR, Bailey E. 1987. Genetic differentiation associated with gait within American standardbred horses. *Anim Genet.* 18:285–296.
- Dieringer, D, Schlötterer, C. 2003. Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Mol Ecol Notes.* 3(1): 167–169.
- Excoffier L, Laval G, Schneider S. 2007. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online.* 1:47–50.
- Fischer, MC, Rellstab, C, Leuzinger, M, Roumet, M, Gugerli, F, Shimizu, KK, Holderegger, R, et al. 2017. Estimating genomic diversity and population differentiation—an empirical comparison of microsatellite and SNP variation in *Arabidopsis halleri*. *BMC Genomics.* 18(1):69.
- Khanshour, A, Conant, E, Juras, R, Cothran, EG. 2013. Microsatellite analysis of genetic diversity and population structure of Arabian horse populations. *J Hered.* 104(3):386–398.
- King, JA. 1992. *Changes in heterozygosity through time in American Standardbred and American Saddlebred Horses (1960–1990)* (Master thesis). Denton (TX): University of North Texas.
- Kraus, M, Physick-Sheard, P, W, Brito, L, F, Sargolzaei, M, Schenkel, F, S. 2018. Marginal ancestral contributions to atrial fibrillation in the Standardbred racehorse: comparison of cases and controls. *PLoS One.* 13(5):e0197137.
- Kraus, M, Physick-Sheard, PW, Brito, LF, Schenkel, FS. 2017. Estimates of heritability of atrial fibrillation in the Standardbred racehorse. *Equine Vet J.* 49(6):718–722.
- Lischer HEL, Excoffier L. 2012. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics.* 28(2):298–299.
- Lykkjen, S, Olsen, HF, Dolvik, NI, Grøndahl, AM, Røed, KH, Klemetsdal, G. 2014. Heritability estimates of tarsocrural osteochondrosis and palmar/plantar first phalanx osteochondral fragments in Standardbred trotters. *Equine Vet J.* 46(1):32–37.
- MacCluer, JW, Boyce, AJ, Dyke, B, Weitkamp, LR, Pfennig, DW, Parsons, CJ. 1983. Inbreeding and pedigree structure in Standardbred horses. *J Hered.* 74(6):394–399.
- Mack M, Kowalski E, Grahn R, Bras D, Penedo MCT, Bellone R. 2017. Two variants in SLC24A5 are associated with “Tiger-Eye” iris pigmentation in Puerto Rican Paso Fino horses. *G3 (Bethesda, MD).* 7(8):2799–2806.
- McCoy AM, Beeson SK, Splan RK, Lykkjen S, Ralston SL, Mickelson JR, McCue ME. 2016. Identification and validation of risk loci for osteochondrosis in Standardbreds. *BMC Genomics.* 17:41.
- McCoy AM, Norton EM, Kemper AM, Beeson SK, Mickelson JR, McCue ME. 2019. SNP-based heritability and genetic architecture of tarsal osteochondrosis in North American Standardbred horses. *Anim Genet.* 50(1):78–81.
- McGivney, BA, Han, H, Corduff, LR, Katz, LM, Tozaki, T, MacHugh, DE, Hill, EW. 2020. Genomic inbreeding trends, influential sire lines and selection in the global Thoroughbred horse population. *Sci Rep.* 10(1):466.

- McQuillan, R, Leutenegger, A-L, Abdel-Rahman, R, Franklin, C. S, Pericic, M, Barac-Lauc, L, Smolej-Narancic, N, *et al.* 2008. Runs of homozygosity in European populations. *Am J Hum Genet.* 83(3):359–372.
- Meyermans R, Gorssen W, Buys N, Janssens S. 2020. How to study runs of homozygosity using PLINK? A guide for analyzing medium density SNP data in livestock and pet species. *BMC Genomics.* 21(1):94.
- Nei M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics.* 89(3):583–590.
- Norton EM, Mickelson JR, Binns MM, Blott SC, Caputo P, Isgren CM, McCoy AM, *et al.* 2016. Heritability of recurrent exertional rhabdomyolysis in Standardbred and Thoroughbred racehorses derived from SNP genotyping data. *J Hered.* 107(6):537–543.
- Petersen JL, Mickelson JR, Cothran EG, Andersson LS, Axelsson J, Bailey E, Bannasch D, *et al.* 2013. Genetic diversity in the modern horse illustrated from genome-wide SNP data. *PLoS One.* 8(1):e54997.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, *et al.* 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 81(3):559–575.
- R Core Team. 2020. *R: a language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Available from: <https://www.R-project.org/>
- Sereno FTP de S, Sereno JRB, Vega-Pla JL, Kelly L, Bermejo JVD. 2008. Genetic diversity of Brazilian Pantaneiro horse and relationships among horse breeds. *Pesqui Agropecu Brasil.* 43(5):595–604.
- Sevinga M, Vrijenhoek T, Hesselinks JW, Barkema HW, Groen AF. 2004. Effect of inbreeding on the incidence of retained placenta in Friesian horses. *J Anim Sci.* 82:982–986.
- van de Goor LHP, van Haeringen, WA, Lenstra JA. 2011. Population studies of 17 equine STR for forensic and phylogenetic analysis. *Anim Genet.* 42(6):627–633.
- Wallace JH. 1871. *Wallace's American Trotting Register.* Vol. 1. New York: Geo. E. Woodward, Orange Judd & Co.
- Weir BS, Cockerham CC. 1984. Estimating *F*-statistics for the analysis of population structure. *Evolution.* 38:1358–1370.
- Wickham H. 2016. *ggplot2—Elegant Graphics for Data Analysis.* 2nd ed. Cham (Switzerland): Springer International Publishing.
- Wright S. 1978. *Evolution and the Genetics of Populations.* Vol. 4. Chicago: University of Chicago Press.