

# The Legend of the Canadian Horse: Genetic Diversity and Breed Origin

ANAS KHANSHOUR, RYTIS JURAS, RICK BLACKBURN, AND E. GUS COTHAN

From the Department of Veterinary Integrative Biosciences, College of Veterinary Medicine and Biomedical Science, Texas A&M University, College Station, TX 77843-4458 (Khanshour, Juras, and Cothran); Department of Integrated Studies in Education, Faculty of Education, McGill University, Montreal, Québec (Blackburn).

\*Address correspondence to E. Gus Cothran at the address above, or e-mail: [GCothran@cvm.tamu.edu](mailto:GCothran@cvm.tamu.edu)

Data deposited at Dryad: doi:<http://dx.doi.org/doi:10.5061/dryad.tt47r>

## Abstract

The Canadian breed of horse invokes a fascinating chapter of North American history and as such it is now a heritage breed and the national horse of Canada. The aims of this study were to determine the level of genetic diversity in the Canadian, investigate the possible foundation breeds and the role it had in the development of the US horse breeds, such as Morgan Horse. We tested a total of 981 horses by using 15 microsatellite markers. We found that Canadian horses have high values of genetic diversity indices and show no evidence of a serious loss of genetic diversity and the inbreeding coefficient was not significantly different from zero. Belgian, Percheron, Breton and Dales Pony, unlike the light French horses, may have common ancestries with the Canadian and could be important founders. However, the Shire and Clydesdale influenced the Canadian to a lesser extent than French and Belgian draft breeds. Furthermore, our finding indicated that there was no evidence of a clear relationship between Canadian and Oriental or Iberian breeds. Also, the Canadian likely contributed to the early development of the Morgan. Finally, these findings support the ancient legends of the Canadian Horse as North America's first equine breed and the foundation bloodstock to many American breeds and may help in the management and breeding program of this outstanding breed in North America.

**Key words:** Equine, Genetic Variation, Heritage Breed, Microsatellite, Population Genetics, Phylogenetic

**Subject areas:** Conservation genetics and biodiversity

The legend of the Canadian horse is a fascinating chapter of North American history. It is the story of North America's first equine breed that was the foundation bloodstock to many American breeds like the American Saddlebred, Standardbred, (possibly) Appaloosa, Northern Plains Mustang, and the Morgan. The Canadian traces its ancestry to horses sent from Louis XIV's royal stables to the colony of New France in the mid-17th century (Jones 1947). In spite of the colony's harsh climate and the absence of forage, shelter and pasture, the foundation herd of less than a 100 animals flourished to 30 000 by 1784 (Gendron 1993). The Canadian earned the nickname, "the little iron horse," for its storied feats of strength and endurance, regularly outclassing more muscular and heavy boned horses (Langelier 1920). The Canadian studbook was established in 1889 and is the oldest active horse breed registry in North America (Gendron 2010). In 1998 and 2002, respectively, Quebec and Canada officially recognized the Canadian as a heritage breed and the national horse for its intrinsic role in forming the 2 nations. Today the breed languishes in relative obscurity on the brink of extinction with a confirmed

population of 2456 pure Canadian horses and is classified as threatened by The Livestock Conservancy.

Comprehensive information about genetic diversity and population structure is highly important to draw the essential outlines for any appropriate conservation and sustainable management programs (Notter 1999). Also, it is essential to understand the evolution, domestication, and demographic history of breeds to support breeding programs.

Microsatellite markers have been widely used to investigate genetic structure and diversity of different horse breeds (Bigi and Perrotta 2012; Khanshour et al. 2013a) including rare horse breeds (Bodó et al. 2005; Luís et al. 2007). Stability, ease and accuracy of genotyping these co-dominant markers, together with their wide-spread distribution across the genome make microsatellite loci an attractive source of information for genetic diversity and population structure (Goldstein and Schlötterer 1999).

The allocation of resources for the conservation of heritage breeds must consider not only the genetic value of an endangered breed, but also its sociopolitical and cultural significance. Therefore, the aims of this study were to determine the

level of genetic diversity within the Canadian breed, to investigate the possible foundation breeds for the Canadian breed by comparison to a large number of contemporary breeds, and to investigate any role of the Canadian in the development of US horses, especially the Morgan Horse breed.

## Materials and Methods

We tested a total of 981 horses. One hundred and 8 of them were Canadian horses collected from different places in Canada to cover all 8 stallion bloodlines of the Canadian breed as documented by the Canadian Horse Breeders Association. All 6 maternal lineages described by (Prystupa et al. 2012a) were also sampled. Only horses registered in the Canadian Horse Breed Registry as 100% Canadian were included. The other horse samples were from animals that were registered with a breed registry and collected from different countries worldwide to represent 18 horse breeds, See [Supplementary Table 1](#) online. These breeds were expected to be linked to the Canadian based upon the historical background of the North American horses. The numbers of horses tested and breed abbreviations are listed in [Table 1](#).

Total DNA was extracted from hair follicle samples using the PUREGENE® DNA purification kit following the manufacturer's protocol (QIAGEN, MA, USA). We tested 15 microsatellite markers (AHT4, AHT5, ASB2, ASB17, ASB23, HMS6, HTG6, HMS7, HMS3, HMS2, HTG4, VHL20, HTG7, and HTG10, and LEX33) distributed on 14 chromosomes of the horse genome. Information about the loci location, primer sequences and allele sizes can be found in (Khanshour et al. 2013b). Microsatellite genotyping was performed following the method described by (Juras et al. 2003). In fulfillment of data archiving guidelines (Baker 2013), we have deposited all genotype data for the 981 horses tested in this study with Dryad. The validity of each marker was tested by calculation of the deviation from Hardy Weinberg Equilibrium (HWE) in each population by Chi-Square test implemented in GENEALEX 6.5 (Peakall and Smouse 2012). For all markers together, the Fisher's method implemented in GENPOP 3.4 (Raymond and Rousset 2001) was used to detect the significant deviation of a population from HWE after Bonferroni correction. GENPOP was also used to calculate the pair-wise  $F_{ST}$  among populations. The significant presence ( $P < 0.05$ ) of null alleles was detected and adjusted using Microchecker software (Van Oosterhout et al. 2004) according to Brookfield's approach (Brookfield 1996).

We calculated genetic diversity indices for each breed using GENEALEX 6.5. These included the average number of alleles per population ( $N_a$ ), the effective number of alleles per population ( $A_e$ ), the proportion of rare alleles per population ( $R_a$ ) defined as those with an allele frequency  $< 0.05$ , the number of private alleles ( $P_a$ ), observed ( $H_O$ ), and expected ( $H_E$ ) heterozygosity. The inbreeding coefficient ( $F_{IS}$ ) was also calculated using 10 000 bootstraps as implemented in GENETIX 4.05 (Belkhir et al. 1996–2004).

The genetic relationship of the Canadian breed to the other breeds was investigated by the Majority-rule consensus

of restricted maximum likelihood (RML) trees. We calculated the chord distances generated from 1000 bootstrapped allele frequency datasets using PHYLIP 3.69 package (Felsenstein 1989–2006). Trees were visualized by MEGA6 (Tamura et al. 2013). The Przewalski Horse population ( $N = 44$ ) was used as an out-group.

Presentation of the genetic relationships among breeds was carried out using factorial correspondence analysis (FCA) as implemented by GENETIX 4.05. Also we used the STRUCTURE 2.3.3 software (Pritchard et al. 2000) to study the admixture pattern among the Canadian and the other breeds with burn-in value of 20 000 and 100 000 Markov Chain Monte Carlo (MCMC) repetitions. Runs for each value of  $K$  ( $K = 2–21$ ) were repeated 10 times. The Przewalski Horse population ( $N = 44$ ) was used as an out-group. The software CLUMP (Jakobsson and Rosenberg 2007) and DISTRICT (Rosenberg 2004) were used to align multiple replicates and display the results, respectively. The best number of clusters was determined based upon  $\Delta K$  value (Evanno et al. 2005) using Structure Harvester application (Earl and vonHoldt 2011).

## Results

All loci tested were found to be polymorphic and all did not show stable significant deviation from HWE in all breeds tested ( $P < 0.05$ ). The Canadian breed was in Hardy Weinberg Disequilibrium (HWD) in 4 loci ASB17, HMS2, HTG7, and HMS7 with the possibility of null alleles for the HMS7 locus. The Canadian was not the only breed with HWD in these loci as other breeds showed a similar deviation, [Supplementary Table 2](#) online. However, considering all 15 loci together, no significant deviation from HWE was found in the Canadian using the overall Fisher's test and Bonferroni correction.

The genetic diversity measures within breeds showed that the highest genetic diversity was in Morgan Horse and Selle Français breeds ([Table 1](#)). Conversely, the Clydesdale population showed the lowest genetic diversity. For the Canadian breed, a high genetic diversity associated with a high number of total effective and rare alleles was found.  $F_{IS}$  was not significantly different from 0 at the 0.05 level in the Canadian breed. Pairwise  $F_{ST}$  values between the Canadian and the other breeds ranged from 0.074 for the Morgan Horse to 0.171 for the French Trotter ([Table 2](#)). Pairwise  $F_{ST}$  values among all breeds are shown in [Supplementary Table 3](#) online.

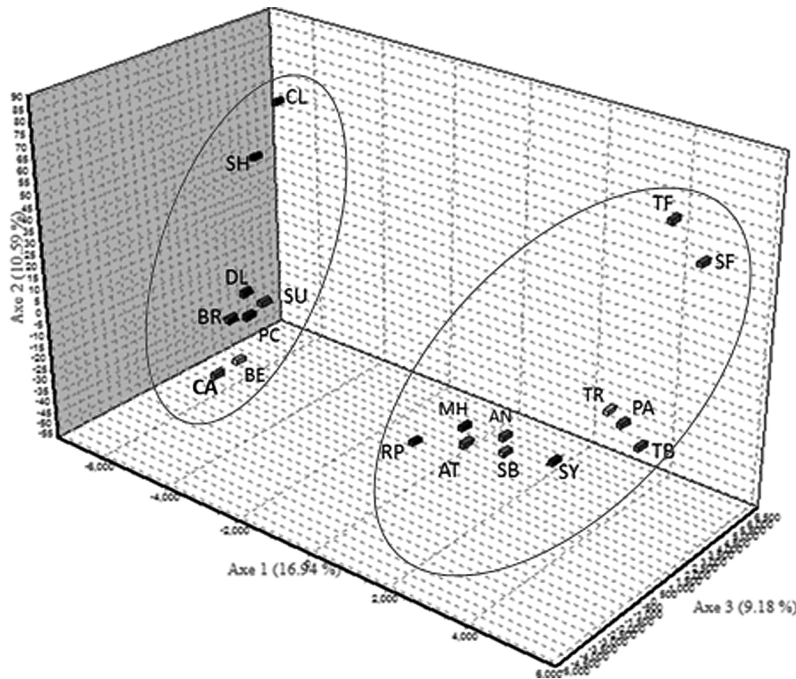
The FCA for 19 breeds indicated 2 groups along the first axis that explained the highest level of variation (16.94%; [Figure 1](#)). The Canadian horses clustered with the draft horses (Belgian, Breton, Percheron, Suffolk, Shire, and Clydesdale breeds) and Dales Pony. Similar results were also found in the RML tree ([Supplementary Figure 1](#) online) where the Canadian was clustered again with these 8 breeds (Cluster II: Canadian, Belgian, Breton, Percheron, Suffolk, Shire, Clydesdale, and Dales Pony). A closer view of the relationships among the 8 breeds in Cluster II was magnified by re-running FCA for individuals as shown in [Figure 2](#) where Canadian horses overlapped with Belgians and a few individuals of Dales Pony (cluster I). Canadian and Belgian

**Table 1** The genetic diversity measures within breeds

Populations (abbreviation)	N	Na (SE)	Ae (SE)	Pa	Ra	HO (SE)	HE (SE)	$F_{IS}$
Andalusian (AN)	30	5.67 (0.33)	3.76 (0.26)	1	0.27	0.70 (0.02)	0.71 (0.03)	0.029
Akhal Teke (AT)	80	6.47 (0.38)	3.87 (0.34)	0	0.35	0.73 (0.04)	0.71 (0.03)	-0.031
Belgian (BE)	26	6.00 (0.43)	3.62 (0.36)	0	0.25	0.67 (0.04)	0.67 (0.04)	0.021
Breton (BR)	26	5.13 (0.39)	3.41 (0.30)	0	0.20	0.73 (0.06)	0.66 (0.05)	-0.093*
Canadian (CA)	108	6.73 (0.44)	3.85 (0.36)	1	0.37	0.71 (0.04)	0.70 (0.03)	0.001
Clydesdale (CL)	33	4.47 (0.22)	2.57 (0.17)	0	0.28	0.61 (0.03)	0.59 (0.03)	-0.018
Dales Pony (DL)	85	6.13 (0.34)	3.63 (0.30)	0	0.33	0.71 (0.03)	0.69 (0.03)	-0.033*
Morgan Horse (MH)	48	7.20 (0.54)	4.28 (0.36)	2	0.36	0.75 (0.03)	0.74 (0.03)	-0.006
Standardbred Pacer (PA)	26	5.13 (0.27)	3.13 (0.20)	0	0.17	0.72 (0.03)	0.66 (0.02)	-0.062*
Percheron (PC)	35	6.47 (0.50)	3.60 (0.35)	0	0.35	0.68 (0.03)	0.69 (0.03)	0.016
Puerto Rican Paso Fino (RP)	67	7.07 (0.37)	3.50 (0.26)	2	0.38	0.70 (0.02)	0.69 (0.02)	-0.007
American Saddlebred (SB)	56	6.67 (0.51)	3.66 (0.28)	3	0.41	0.70 (0.04)	0.70 (0.03)	0.007
Selle Français (SF)	28	6.40 (0.32)	4.07 (0.26)	3	0.24	0.75 (0.03)	0.74 (0.02)	0.007
Shire (SH)	32	5.40 (0.29)	3.12 (0.22)	0	0.33	0.69 (0.04)	0.65 (0.03)	-0.033*
Suffolk (SU)	50	5.47 (0.41)	3.80 (0.31)	0	0.19	0.73 (0.04)	0.70 (0.03)	-0.034*
Syrian (SY)	69	6.93 (0.44)	3.72 (0.21)	3	0.45	0.71 (0.03)	0.72 (0.02)	0.013
Thoroughbred (TB)	70	5.13(0.31)	3.39(0.25)	0	0.20	0.69(0.03)	0.68(0.03)	-0.001
French Trotter (TF)	39	6.00(0.32)	3.41(0.29)	1	0.32	0.67(0.04)	0.67(0.03)	0.009
Standardbred Trotter (TR)	29	5.33(0.29)	3.36(0.27)	0	0.25	0.69(0.04)	0.67(0.03)	-0.005
Przewalski (PZ)	44	4.60(0.24)	3.04(0.18)	6	0.17	0.66(0.03)	0.65(0.02)	-0.002
Average	49	5.92(0.09)	3.54(0.065)	1.1	0.29	0.70(0.01)	0.68(0.01)	-0.023

Sample size (N), average number of alleles per locus per population (Na), average number of effective alleles per locus per population (Ae), number of private alleles (Pa), proportion of rare alleles (Ra), observed heterozygosity (HO), and expected heterozygosity (HE). All means are combined with its standard error (SE) and  $F_{IS}$  values are combined with the significant status.

\*Values different from 0 at  $P < 0.0$ .



**Figure 1.** The Factorial Corresponding Analysis of 19 breeds. Each breed was represented by its center of gravity point into 3D plot. Axis 1 accounts for 16.94% of the variation. Breed abbreviations are as follows: Andalusian (AN), Akhal Teke (AT), Belgian (BE), Breton (BR), Canadian (CA), Clydesdale (CL), Dales Pony (DL), Morgan Horse (MH), Standardbred Pacer (PA), Percheron (PC), Puerto Rican Paso Fino (RP), American Saddlebred (SB), Selle Français (SF), Shire (SH), Suffolk (SU), Syrian (SY), Thoroughbred (TB), French Trotter (TF), and Standardbred Trotter (TR).

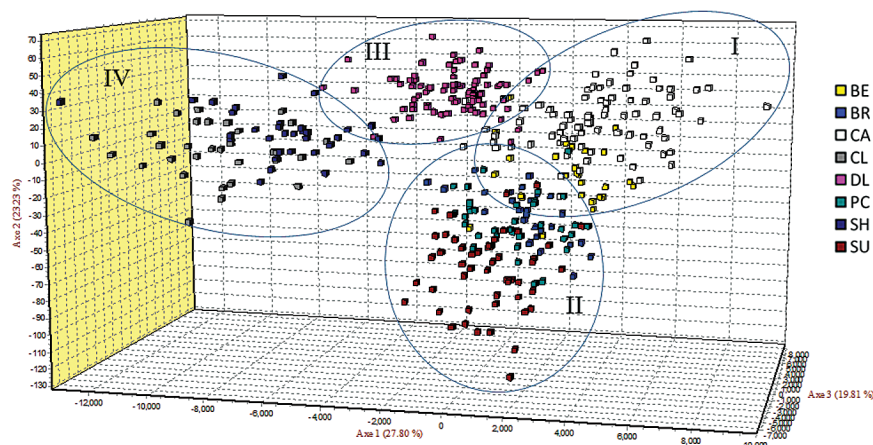


were more close to Breton, Percheron, Dales Pony, and Suffolk (cluster II) than Shire and Clydesdale breeds. Breton, Percheron, Suffolk, and some Breton and Canadian overlapped closely (cluster III). The Shire and Clydesdale breeds overlapped in a separate cluster (cluster IV).

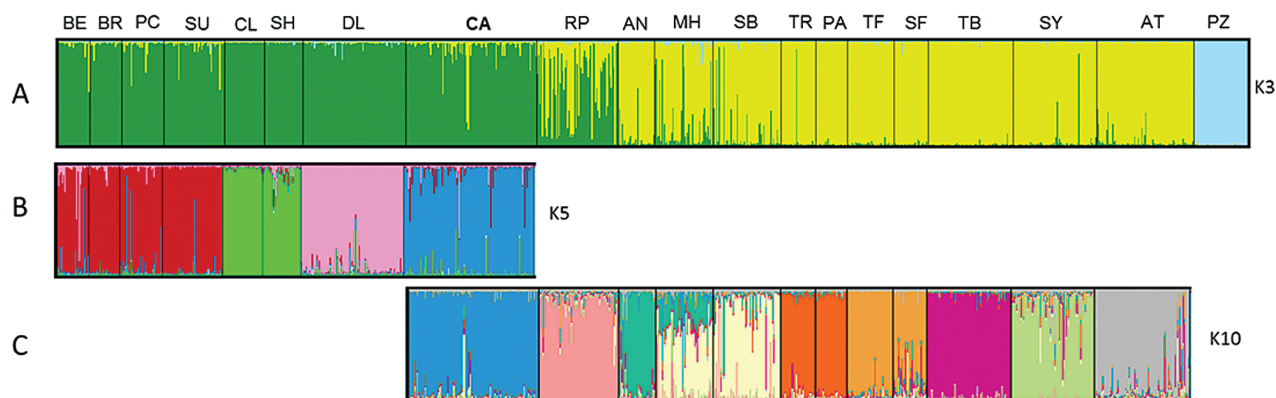
The STRUCTURE analysis of 19 breeds using  $K = 21$  showed that the highest  $\Delta K$  was at  $K = 3$  (Supplementary Figure 2 online) where 2 clusters and the out-group were formed, Figure 3A. Canadian horses along with other draft horses BE, BR, PC, SU, CL, and SH and Dales Pony formed the first cluster (8 breeds). The second cluster represented all sport and riding horse breeds (12 breeds). However, RP, MH, SB, and AN breeds showed some similarity to the horses of the first cluster. Therefore, we reran STRUCTURE for each cluster separately with the Canadian for better individual assignments in order to see which breed is more related to the Canadian.

The STRUCTURE analysis of 8 breeds showed that the highest  $\Delta K$  was at  $K = 5$  (Supplementary Figure 3 online) where 5 clusters were formed (Figure 3B). The Canadian started to form a separate cluster at  $K = 4$  and were isolated from other breeds at  $K = 5$  (Supplementary Figure 4 online). The proportion of individuals assigned into each cluster based upon the Q value at the best  $\Delta K$  showed that cluster 4, the blue color in Figure 3B, consisted of 90.5 % Canadian. The highest contribution that came from other breeds was 8.2% for the PC breed. CL population contributed only 0.006 % in this cluster (Table 3). Only 3.9% of Canadian contributed to cluster 1, the red color, that also contained high proportions of BR, SU, PC, and BE horses, Table 3 and Figure 3B.

The analysis of 12 breeds showed that the highest  $\Delta K$  was at  $K = 10$ , Supplementary Figure 5 online and Figure 3C. The Canadian horses were assigned into a separate cluster



**Figure 2.** The Factorial Corresponding Analysis of 8 breeds. Each individual was plotted into 3D plot. Axis 1 accounts for 27.8% of the variation. Breed abbreviations are as follows: Belgian (BE), Breton (BR), Canadian (CA), Clydesdale (CL), Dales Pony (DL), Percheron (PC), Shire (SH), and Suffolk (SU).



**Figure 3.** Clustering assignment obtained by STRUCTURE. (A) The analysis of 19 breeds plus the out-group at  $K = 3$ . (B) The analysis of 8 breeds at  $K = 5$ . (C) The analysis of 12 breeds at  $K = 10$ . Breed abbreviations are as follows: Andalusian (AN), Akhal Teke (AT), Belgian (BE), Breton (BR), Canadian (CA), Clydesdale (CL), Dales Pony (DL), Morgan Horse (MH), Standardbred Pacer (PA), Percheron (PC), Puerto Rican Paso Fino (RP), American Saddlebred (SB), Selle Français (SF), Shire (SH), Suffolk (SU), Syrian (SY), Thoroughbred (TB), French Trotter (TF), Standardbred Trotter (TR), and Przewalski (PZ).

starting from  $K = 5$ , [Supplementary Figure 6](#) online. As shown in [Table 4](#), the highest proportion of the Canadian horses (89.8%) was in cluster 1, the blue color in [Figure 3C](#). The other contributions in this cluster came from Morgan Horses with 7.4% of the individuals. All contributions of other breeds into this cluster were less than 2%. On the other hand, the Canadian breed contributed 11.2% of its individuals to other clusters, noticeably in cluster 4 where 50 % of MH and 74.1% of SB were assigned.

## Discussion

We described the genetic diversity and population structure in the Canadian breed and their relationship to other horse breeds. Although the Canadian was thought to have been bred in isolation for a long time, their genetic diversity represented by expected heterozygosity is high and similar to that of other domestic horses ([Plante et al. 2007](#); [van de Goor et al. 2011](#); [Conant et al. 2012](#)). The genetic diversity in the Canadian was higher than any horse breeds reported as endangered ([Bodó et al. 2005](#)). The  $F_{IS}$  value was not significantly different from zero suggesting that this breed has not suffered a recent loss of genetic diversity due to inbreeding despite a relatively small population size compared with other breeds. Similar results for the genetic variability of the Canadian was reported by [Behara et al. \(1998\)](#) using a similar set of markers and [Prystupa et al. \(2012b\)](#) who used 38 microsatellite markers. The genetic diversity, shown as the expected heterozygosity (HE), of the Canadian reported 13 years ago was 0.69 with a total number of 7632 registered horses compared to 0.7 in this study with a total number of 13 616 registered horses ([Canadian Livestock Records Corporation 2012](#)). This high level of genetic variability, supported by our finding of the high number of both total and effective number of alleles, indicates a wide base of genetic diversity of the Canadian breed through its recent history.

Regulations for Canadian horse registration have restricted introgressive gene flow since 1889 when the Canadian government established the first pedigree records ([Gendron 2010](#)). Therefore, the relatively high  $R_a$  in the Canadian is not likely due to recent genetic introgression from other horse breeds but is more likely part of the evolutionary history of the breed. The high value for  $R_a$  also could be one consequence of the isolation of this breed and represents alleles that are in the process of being lost. The presence of rare alleles was reported in some other primitive horse populations ([Leroy et al. 2009](#)).

The FCA and STRUCTURE analyses in the current study showed the Canadian as a genetically uniform breed

confirming this breed as a homogenous, interbreeding population ([Figures 2](#) and [3](#)). Although we tested horses from different areas from Canada, these horses came from a closed breed and represent a single gene pool.

FCA, RML tree and STRUCTURE placed the Canadian with the heavy draft breeds in 1 cluster ([Figures 1](#) and [3A](#)) and this result also was supported by the moderate  $F_{ST}$  value found between Canadian and the heavy draft breeds ([Table 2](#)). This finding agrees with other works that found a strong relationship between the Canadian and draft horses ([Plante et al. 2007](#); [Prystupa et al. 2012a](#); [Prystupa et al. 2012b](#)). The clear relationship between the Canadian and French draft horses—Breton and Percheron, as shown in [Figures 1, 2](#) and [3B](#), confirmed a likely common ancestry for these horses. The old horses in Canada were known as French-Canadian horses before officially called the Canadian Horse ([Pelletier 1943](#)). The old name came from the historical fact that horses of French origin, mainly from the royal stud of Louis XV, formed the basis of the French-Canadian horses. French horses imported into Canada during the 17th century came from Normandy and Brittany ([Jones 1947](#); [Gendron 2010](#)). Percheron and Breton are well known breeds from the Normandy and Brittany regions, respectively ([Hendricks 1995](#)). These 2 French breeds were reported as very genetically similar horses ([Leroy et al. 2009](#)). However, the Percheron looked more similar to the Canadian than did the Breton as STRUCTURE showed in [Figure 3B](#), and could be important in the foundation of the Canadian breed. It was reported that the Percheron resembles the French-Canadian more closely than any other breed in the world ([Jones 1947](#)). On the other hand, the Canadian clustered away from other French breeds, Selle Français and French Trotter, showing that Canadian and light French horses are genetically

**Table 3** The proportion of individuals assigned into each cluster resulted from STRUCTURE analysis at  $K = 5$  for 8 breeds

Breeds	1	2	3	4	5
CA	0.039	0.027	0.025	<b>0.905</b>	0.004
PC	<b>0.869</b>	0.023	0.020	0.082	0.006
BE	<b>0.805</b>	0.011	0.104	0.075	0.005
SU	<b>0.948</b>	0.013	0.008	0.027	0.004
SH	0.026	<b>0.908</b>	0.043	0.016	0.007
DL	0.017	0.023	<b>0.938</b>	0.016	0.006
BR	<b>0.943</b>	0.015	0.027	0.011	0.004
CL	0.007	<b>0.979</b>	0.006	0.006	0.002

Breed abbreviations are as follows: Canadian (CA), Percheron (PC), Belgian (BE), Suffolk (SU), Shire (SH), Dales Pony (DL), Breton (BR), and Clydesdale (CL). The highest  $Q$  values are in bold.

**Table 2** Pairwise  $F_{ST}$  between Canadian and other breeds

Pop	MH	DL	PC	BE	SY	AT	SU	AN	SB	SH	RP	SF	BR	TB	TR	PA	CL	TF
CA	<b>0.074</b>	<b>0.085</b>	<b>0.086</b>	<b>0.095</b>	<b>0.103</b>	<b>0.107</b>	<b>0.109</b>	<b>0.111</b>	<b>0.113</b>	<b>0.117</b>	<b>0.121</b>	<b>0.121</b>	<b>0.123</b>	<b>0.131</b>	<b>0.143</b>	0.154	0.165	0.171

Breed abbreviations are as follows: Canadian (CA), Morgan Horse (MH), Dales Pony (DL), Percheron (PC), Belgian (BE), Syrian (SY), Akhal Teke (AT), Suffolk (SU), Andalusian (AN), American Saddlebred (SB), Shire (SH), Puerto Rican Paso Fino (RP), Selle Français (SF), Breton (BR), Thoroughbred (TB), Standardbred Trotter (TR), Standardbred Pacer (PA), Clydesdale (CL), and French Trotter (TF). Values showed moderate differentiation are in bold.

**Table 4** The proportion of individuals assigned into each cluster resulted from STRUCTURE analysis at  $K = 10$  for 12 breeds

Breeds	1	2	3	4	5	6	7	8	9	10
CA	<b>0.898</b>	0.012	0.010	0.016	0.013	0.008	0.014	0.011	0.012	0.005
MH	0.074	0.058	0.216	<b>0.503</b>	0.031	0.021	0.032	0.038	0.018	0.009
AT	0.018	0.013	0.013	0.016	0.020	0.015	0.042	0.028	<b>0.832</b>	0.004
AN	0.017	0.025	<b>0.829</b>	0.015	0.014	0.008	0.034	0.022	0.031	0.005
SF	0.017	0.014	0.039	0.026	0.027	<b>0.738</b>	0.082	0.031	0.020	0.007
SB	0.013	0.055	0.052	<b>0.741</b>	0.029	0.014	0.036	0.015	0.041	0.005
TB	0.011	0.008	0.007	0.007	0.009	0.008	<b>0.926</b>	0.012	0.007	0.004
SY	0.011	0.010	0.018	0.018	0.023	0.021	0.037	<b>0.828</b>	0.029	0.005
RP	0.010	<b>0.883</b>	0.020	0.023	0.014	0.012	0.010	0.014	0.012	0.004
TF	0.008	0.005	0.007	0.008	0.016	<b>0.927</b>	0.009	0.007	0.007	0.006
TR	0.006	0.011	0.013	0.021	<b>0.905</b>	0.007	0.019	0.010	0.006	0.003
PA	0.006	0.005	0.008	0.009	<b>0.921</b>	0.009	0.015	0.012	0.012	0.003

Breed abbreviations are as follows: Canadian (CA), Morgan Horse (MH), Akhal Teke (AT), Andalusian (AN), Selle Français (SF), American Saddlebred (SB), Thoroughbred (TB), Syrian (SY), Puerto Rican Paso Fino (RP), French Trotter (TF), Standardbred Trotter (TR), and Standardbred Pacer (PA). The highest  $Q$  values are in bold.

different compared with other tested breeds (Figure 3C). Pairwise  $F_{st}$  between Canadian and Selle Français was moderate (0.121) but it was high for the French Trotter (0.171). French Trotters have been reported before as genetically distinct breeds from the Breton and Percheron breeds (Leroy et al. 2009; Petersen et al. 2013).

The Belgian breed was also very close to the Canadian breed (Figure 1) and individuals clearly overlapped as shown in Figure 2. STRUCTURE also showed that the Belgian clustered with other heavy horses and could have admixed with the Canadian. The Dales Pony was close to Belgian (Figure 3B) with moderate differentiation based upon the pairwise  $F_{st}$  (Supplementary Table 3 online). The strong relationship between Belgian and Percheron was reported before (Petersen et al. 2013) and, in fact, all the heavy horse breeds are genetically close (Aberle et al. 2004; Iwańczyk et al. 2006). The Belgian breed is a very ancient breed in Western Europe (Hendricks 1995). However, 36.5% of the registered Belgian horses are from unknown origins (Pirault et al. 2013). Therefore, the admixture pattern and the high similarity among Western European horses are expected. Regardless of the origin of the Belgian breed, it could be one of the candidate founders for the Canadian breed along with Percheon and Dales Pony.

The relationship between the Canadian breed and the British draft horses revealed by the Factorial Corresponding Analysis, pairwise  $F_{st}$ , and STRUCTURE (Figures 1, 2, 3A and B; Table 2) was not same with Shire and Clydesdale as compared to the Suffolk. Shire and Clydesdale formed one cluster, as reported in other studies (Petersen et al. 2013), but notably showed less relationship with the Canadian than did the Suffolk (Figures 1 and 2; Table 2). Horses that came with the English settlement were part of the horse breed establishment in Canada and the Northern US states (Howard 1965). The power and strength of the British horses such as Shire and Suffolk impressed Canadian farmers. Therefore, they used them to produce horses capable not only of general draught work but of supplying power required for pioneer cultivation on the high plains (Moore-Colyer 2000).

Similarly, the Clydesdale, from Scotland, found their way to Canada to meet the needs of Scottish settlers. Our results showed that Shire and Clydesdale influenced the Canadian breed less than French and Belgian draft horses did. This could be because various British exporters sent many unsound, unsuitable and even impotent British stallions to the North American market before 1900 (Moore-Colyer 2000). These poor quality stallions came to be unfavorably compared with horses of Percheron and Belgian types whose breeders made sure that only the very best specimens reach North America at a reasonable price. Another reason that could be of major significance is that the draft British horses, except Suffolk, have long hairs on the lower part of their legs (known as the feather) that was considered as unsuitable feature for the needs of most Canadian horse breeders (Moore-Colyer 2000). Furthermore, the reconstitution of the Canadian breed pedigree records that happened in 1907 followed a severe selective inspection where only perfectly healthy horses with no features of large drought breeds, such as Shire and Clydesdale, was accepted (Gendron 2010). Our findings indicate that the Dales Pony had a moderate level of differentiation based upon pair-wise  $F_{st}$  and it is a possible contributor to the Canadian breed. This finding is supported by the fact that all breeding efforts after 1907 were centered to keep the small size of the Canadian (Gendron 2010).

The STRUCTURE analysis showed that Arabian, Akhal Teke, and Andalusian breeds were distinct from the Canadian at early  $K$  ( $K = 5$ ). Similar results came from the genetic distance trees and the Factorial Corresponding Analysis where the Canadian breed was in a different clade than all Arabian, Akhal Teke, and Iberian breeds. Therefore, our finding indicated no evidence of a clear relationship between Canadian and those horses. The Canadian was thought to have Arabian or Akhal Teke blood through French founders such as Percheron (Jones 1947). Yet, the Percheron themselves, as well as Belgian and Breton, showed no relationship with either the Andalusian or Arabian breeds in this study as well as other studies (Leroy et al. 2009; Petersen et al. 2013).



The Morgan Horse showed similarity with the Canadian based upon pair-wise *F<sub>st</sub>* (Table 2) and STRUCTURE analysis (Supplementary Figure 6 online) where they were the more closely related to the Canadian than any other sport and riding horse breeds (Figure 3C; Tables 2, 4). Such as results were previously reported (Behara et al. 1998). Our finding about the Canadian and Morgan breeds could be explained by 2 scenarios. First, Morgan Horses were used to recover the Canadian after different waves of the Canadian breed contraction in number and quality which happened after 1900. The second scenario is that the Canadian influenced the development of Morgan Horse at some level. The first scenario is rejected because although the Morgan Horse was among the breeds that fought for the honor of recovering the Canadian between 1913 and 1923, all the attempts of these crossbreeding projects were abandoned (Gendron 2010). The second scenario is more likely. The Morgan breed was established around 1790 from a single foundation stallion, named Figure and then called Morgan, and mares with no information of their origin or pedigree (Hendricks 1995). Around that time, purebred and mongrel French-Canadian horses crossed the borders from Lower Canada to the northern states of US, such as Vermont, through trade with farmers and American dealers (Jones 1947). Mares that resulted from the intermixing of French-Canadian with American horses on the United States side evidentially contributed to the development of the Morgan horse (Jones 1947; Lynghaug 2009). Furthermore, the presence of the French-Canadian genetic materials in Morgan horses was indicated by specific characteristics such as the excellent legs and feet, heavy crimped mane and tail (Herbert et al. 1871). Interestingly, it was mentioned in The National Live-stock Journal in 1881 that H. Bexon, a Vermonter, has said: “there is a terrible fear the Morgan horse would be found to have some French blood in him” (Jones 1947).

In conclusion, although the Canadian breed has experienced significant difficulties throughout their history, they have not suffered a serious loss of genetic diversity due to inbreeding despite now being considered as a rare breed. Belgian, Percheron, Breton and Dales Pony, unlike the light French horses, may have common ancestries with the Canadian breed and could be important founders. Shire and Clydesdale influenced the Canadian breed less than French and Belgian draft horses. Also, our finding indicated that there was no evidence of a clear relationship between Canadian and Oriental or Iberian breeds. The Canadian evidentially contributed to the early development of the Morgan horses and many other American breeds. Finally, this study provided in depth investigation of the genetic history of North America's first equine breed. These findings may help in the population management and breeding program of the Canadian and provided some answers related to the ancient legends of the Canadian Horse.

## Supplementary Material

Supplementary material can be found at <http://www.jhered.oxfordjournals.org/>.

## Funding

Animal Genetic lab at Texas A&M University.

## Acknowledgments

We would like to thank all the horse breeders who provided us their horse samples. Many thanks to Andre Auclair (Executive Director, Quebec Federation of Heritage Breeders) and Gerard Lambert (President, National Syndicate of Canadian Horse Breeders) for their help in selecting Canadian samples. Laurent Schibler of UNCEIA, France, provided samples of the French light breeds.

## References

- Aberle KS, Hamann H, Drögemüller C, Distl O. 2004. Genetic diversity in German draught horse breeds compared with a group of primitive, riding and wild horses by means of microsatellite DNA markers. *Anim Genet.* 35:270–277.
- Baker CS. 2013. Journal of heredity adopts joint data archiving policy. *J Hered.* 104:1.
- Behara AMP, Colling DT, Cothran EG, Gibson JP. 1998. Genetic relationships between horse breeds based on microsatellite data: application for live-stock conservation. In: *6th world congress on genetics applied to livestock production*. Armidale (Australia): University of New England. p. 119–122.
- Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F. 1996–2004. GENETIX 4.05, Windows TM software for population genetics. In: *Laboratoire genome, populations, interactions*. Montpellier (France): Université Montpellier II.
- Bigi D, Perrotta G. 2012. Genetic structure and differentiation of the Italian catria horse. *J Hered.* 103:134–139.
- Bodó I, Alderson L, Langlois B. 2005. *Conservation genetics of endangered horse breeds*. Wageningen (the Netherlands): Wageningen Academic Publishers. p. 111.
- Brookfield JF. 1996. A simple new method for estimating null allele frequency from heterozygote deficiency. *Mol Ecol.* 5:453–455.
- Canadian Livestock Records Corporation. 2012. Annual Statistics. [cited 2013 December 10]. Available from: <http://www.clrc.ca/index2010.shtml>.
- Conant EK, Juras R, Cothran EG. 2012. A microsatellite analysis of five Colonial Spanish horse populations of the southeastern United States. *Anim Genet.* 43:53–62.
- Earl DA, vonHoldt BM. 2011. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* 4:359–361.
- Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol.* 14:2611–2620.
- Felsenstein, J. 1989–2006. PHYLIP (phylogeny inference package). [cited March 10 2013]. Available from: <http://evolution.genetics.washington.edu/phylip.html>.
- Gendron M. 1993. Le cheval canadien. Québec, Les Éditions du Septentrion. *Revue d'histoire de l'Amérique française.* 46:501–503.
- Gendron M. 2010. *A brief history of the Canadian horse*. Granby (QC): Société d'histoire de la Haute-Yamaska (SHHY).
- Goldstein D, Schlötterer C. 1999. *Microsatellites: evolution and applications*. New York: Oxford University Press.
- Hendricks BL. 1995. *International Encyclopedia of Horse Breeds*. Norman (OK): University of Oklahoma Press.
- Herbert HW, Bruce SD, Bruce BG. 1871. *Frank forester's horse and horseman-ship of the United States and British provinces of North America*. New York: Geo. E. Woodward.

- Howard RW. 1965. *The horse in America*. Chicago, NY: Follett Publishing Company.
- Iwańczyk E, Juras R, Cholewiński G, Cothran EG. 2006. Genetic structure and phylogenetic relationships of the Polish Heavy horse. *J Appl Genet*. 47:353–359.
- Jakobsson M, Rosenberg NA. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*. 23:1801–1806.
- Jones RL. 1947. *The old French-Canadian horse: its history in Canada and the United States*. Toronto, Ontario: University of Toronto Press.
- Juras R, Cothran EG, Klimas R. 2003. Genetic analysis of three Lithuanian native horse breeds. *Acta Agr. Scand. A* 53:180–185.
- Khanshour A, Conant E, Juras R, Cothran EG. 2013a. Microsatellite analysis of genetic diversity and population structure of Arabian horse populations. *J Hered*. 104:386–398.
- Khanshour AM, Juras R, Cothran EG. 2013b. Microsatellite analysis of genetic variability in Waler horses from Australia. *Aust. J. Zool*. 61:357–365.
- Langelier G. 1920. *The French Canadian Horse: "He never gives out, it does not matter what he is at."*. Bulletin 95 ed. Ottawa (Canada): Ministry of Agriculture. p. 11.
- Leroy G, Callède L, Verrier E, Mériaux JC, Ricard A, Danchin-Burge C, Rognon X. 2009. Genetic diversity of a large set of horse breeds raised in France assessed by microsatellite polymorphism. *Genet Sel Evol*. 41:5.
- Luis C, Cothran EG, Oom Mdo M. 2007. Inbreeding and genetic structure in the endangered Sorraia horse breed: implications for its conservation and management. *J Hered*. 98:232–237.
- Lyngshaug F. 2009. *The official horse breeds standards guide: the complete guide to the standards of all North American Equine Breed Associations*. Minneapolis, MN: MBI Publishing Company.
- Moore-Colyer RJ. 2000. Aspects of the trade in British pedigree draught horses with the United States and Canada, c1850–1920. *Agric. Hist. Rev*. 48:42–59.
- Notter DR. 1999. The importance of genetic populations diversity in live-stock populations of the future. *J. Anim. Sci*. 77:61–69.
- Peakall R, Smouse PE. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*. 28:2537–2539.
- Pelletier JR. 1943. *Twenty-one years' work (1919–1940) for the improvement of the Canadian horse breed, as carried on at the St. Joachim horse farm, Quebec. Pub. by authority of the Hon. James G. Gardiner, Minister of agriculture*. p. 5.
- Petersen JL, Mickelson JR, Cothran EG, Andersson LS, Axelsson J, Bailey E, Bannasch D, Binns MM, Borges AS, Brama P, et al. 2013. Genetic diversity in the modern horse illustrated from genome-wide SNP data. *PLoS One*. 8:e54997.
- Pirault P, Danvy S, Verrier E, Leroy G. 2013. Genetic structure and gene flows within horses: a genealogical study at the french population scale. *PLoS One*. 8:e61544.
- Plante Y, Vega-Pla JL, Lucas Z, Colling D, de March B, Buchanan F. 2007. Genetic diversity in a feral horse population from Sable Island, Canada. *J Hered*. 98:594–602.
- Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics*. 155:945–959.
- Prystupa JM, Hind P, Cothran EG, Plante Y. 2012a. Maternal lineages in native Canadian equine populations and their relationship to the Nordic and Mountain and Moorland pony breeds. *J Hered*. 103:380–390.
- Prystupa JM, Juras R, Cothran EG, Buchanan FC, Plante Y. 2012b. Genetic diversity and admixture among Canadian, Mountain and Moorland and Nordic pony populations. *Animal*. 6:19–30.
- Raymond M, Rousset F. 2001. GENEPOP Update of the version described in Raymond, M. e Rousset, F (1995)-GENEPOP: population genetics software for exact tests and ecumenicism. [online]. <http://wbiomed.curtin.edu.au/genepop/2001>.
- Rosenberg NA. 2004. DISTRUCT: a program for the graphical display of population structure. *Mol. Ecol. Notes*. 4:137–138.
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol*. 30:2725–2729.
- van de Goor LHP, van Haeringen WA, Lenstra JA. 2011. Population studies of 17 equine STR for forensic and phylogenetic analysis. *Anim. Genet*. 42:627–633.
- Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes*. 4:535–538.

Received June 18, 2014; First decision August 11, 2014;  
Accepted October 10, 2014

Corresponding Editor: Ernest Bailey