Animal (2012), **6:1**, pp 19–30 © The Animal Consortium 2011
doi:10.1017/S1751731111001212

animal

Genetic diversity and admixture among Canadian, Mountain and Moorland and Nordic pony populations

J. M. Prystupa[1,2†], R. Juras[3], E. G. Cothran[3], F. C. Buchanan[1] and Y. Plante[1,2]

[1]*Department of Animal and Poultry Science, University of Saskatchewan, Saskatoon, Saskatchewan S7N 5A8, Canada;* [2]*Agriculture & Agri-Food Canada, Saskatoon, Saskatchewan S7N 0X2, Canada;* [3]*College of Veterinary Medicine & Biomedical Sciences, Texas A & M University, College Station, Texas 77848-4458, USA*

(Received 17 December 2010; Accepted 19 May 2011; First published online 26 July 2011)

As part of the requirements of the Convention on Biological Diversity, Canada has been investigating the genetic diversity of its native equine and pony populations. Along with examining four indigenous Canadian equine populations (Canadian horse, Lac La Croix pony, Newfoundland pony and Sable Island population), another 10 Mountain and Moorland, three Nordic, four horse and two feral equine populations (thought to have influenced some pony breeds) were also investigated. In total, 821 individuals were genotyped at 38 microsatellite loci. Results of the analysis of molecular variance indicated that 13.3% of genetic diversity was explained by breed differences, whereas 84.6% and 2.1% of diversity came from within and among individuals, respectively. The average effective number of alleles and allelic richness was the lowest in the Eriskay (2.51 and 3.98) and Lac La Croix (2.83 and 4.01) populations, whereas it was highest in the New Forest (4.31 and 6.01) and Welsh (4.33 and 5.87) breeds, followed closely by the Newfoundland-CDN (4.23 and 5.86) population. Expected heterozygosities varied from 0.61 in the Lac La Croix to 0.74 in the Welsh and in Newfoundland. Observed heterozygosities ranged from 0.57 in the Exmoor and 0.58 in the Sable Island herd to 0.77 in the Kerry Bog and 0.76 in the New Forest breeds. Structure and admixture analyses revealed that the most likely number of clusters was 21, although some substructure was also observed when $K = 16$*, compared with the 24 predefined populations. Information gathered from this study should be combined with other available phenotypic and pedigree data to develop, or amend, a suitable conservation strategy for all populations examined.*

Keywords: genetics, horse, Canada, microsatellite markers, conservation

Implications

In 2005, the Convention on Biological Diversity provided the first legally binding agreement for over 180 countries to examine genetic diversity and to develop optimal conservation strategies for their farm animal genetic resources (Food and Agriculture Organization (FAO), 2007). In order for Canada to fulfill its obligation, populations of various livestock species within its borders must be properly assessed, with regard to genetic diversity (FAO, 2007). The assessment of genetic diversity and admixture among Canadian, Mountain and Moorland and Nordic populations was performed to help to increase the knowledge and understanding of the Canadian equine genetic resources.

Introduction

The Food and Agriculture Organization (FAO; 2007) has recently reported that approximately 700 horse and pony (*Equus caballus*) populations are found worldwide. Of these populations, approximately 181 are listed to be at some risk of extinction, and another 272 are of unknown status (FAO, 2007). Many of the at-risk populations are pony breeds (Equus Surivial Trust, 2008; Rare Breeds Canada, 2009). Ponies are often phenotypically defined as indiviudals standing under 147 cm (14.2 hh) at the whiters, although exceptions to this rule exist, including the Haflinger and Caspian (classified as a coldblood and miniature horse breed, respectively; Lynghaug, 2009). Some of the most well-known groups of pony breeds include the Mountain and Moorland and Nordic populations. The Mountian and Moorland breeds originated from the British Isles and include the Connemara, Dale, Dartmoor, Eriskay, Exmoor, Fell, Highland, Kerry Bog, New Forest, Shetland and Welsh. The Nordic breeds also include the Shetland along with the Norwegian Fjord and Icelandic.

Canada is home to two pony populations – the Newfoundland and the Lac La Croix – along with one horse breed, the Canadian and one feral equine population located on Sable Island. All Canadian equine populations share hardiness, soundness, versatility and small population size;

† E-mail: jmp479@mail.usask.ca

they are also maintained *in situ*. There is interest in understanding what breeds contributed to the development of the Newfoundland, as early management of these ponies involved releasing individuals to run feral until they were required for work. The Lac La Croix pony is a hardy breed traditionally thought to originate from crossing Canadian horses with Spanish mustangs (Lynghaug, 2009). This breed suffered a severe bottleneck in 1977, when the population was reduced to four females, which were then crossed with Spanish mustangs in an attempt to save the population (Lynghaug, 2009). The Canadian horse, Canada's heritage equine breed, traces its origin to the first shipment of French horses, which arrived in Canada in 1665. The Canadian horse became very popular both within and across borders and significantly influenced the development of many of the American trotting and pacing breeds including the Standardbred, Missouri Fox Trotter, Tennessee Walking horse, Morgan, and Saddlebred (Lynghaug, 2009). Lastly, Sable Island is home to a horse population that has had no human interference since 1960 (Plante *et al*., 2007). Before 1960, it was not uncommon for horse breeders to release stallions on the island to breed with the feral population and then to collect the offspring and stallions the following year (Nova Scotia Museum of Nautural History, 2001). This practice makes it possible that the Sable Island population could have influenced other Canadian equine populations.

Microsatellites are simple tandem repeats and are often used in equine diversity studies because of their wide availability, simple Mendelian inheritance, wide coverage of the equine genome and selective neutrality (Cañon *et al*., 2000; Aberle *et al*., 2004; Solis *et al*., 2005). These characteristics make microsatellite loci particularly useful for estimating genetic diversity and characterizing breeds, which are necessary for developing good and viable conservation strategies (Achmann *et al*., 2004; Glowatzki-Mullis *et al*., 2006; Eggert *et al*., 2010). Recently, Leroy *et al*. (2009) have reported that pony breeds tended to cluster distinctly and separately from horse breeds using 11 microsatellites to estimate genetic diversity in 34 French equine populations. European pony breeds (Connemara, Camargue, Landais, New Forest, Poney Français de Selle, Pottok and Welsh) clustered closely to Nordic pony breeds (Shetland, Icelandic and Fjord), but separately from both warmblood and coldblood horse breeds (Leroy *et al*., 2009). Despite this, large studies involving the estimation of genetic diversity of solely pony breeds have not been performed, and studies available today tend to focus only on a few breeds or compare a large number of horse breeds with only a few pony breeds; this may not be the best way to measure the genetic diversity in pony breeds (Plante *et al*., 2007; Behl *et al*., 2008; Leroy *et al*., 2009).

In order to estimate the genetic diversity and phylogenetic relationships of the Mountain and Moorland, Nordic and native Canadian equine populations, this study aimed to investigate 15 pony breeds, three feral equine populations and five horse breeds (thought to have influenced the development of some pony breeds and/or Canadian equine populations) using 38 microsatellite loci.

Material and methods

Sampling and genetic analysis

In total, 821 randomly selected individuals from 24 populations were examined. These populations consisted of feral (Grand Turk, Sable Island and Saint-Pierre et Miquelon); horse (Canadian, Caspian, Clydesdale, Haflinger, Mongolian Domestic (serving only as an out group for phylogenetic reconstruction) and Standardbred); and pony populations. Individual pony breeds were assigned to one of the following categories: Canadian pony (Lac La Croix and Newfoundland (with two separate herd samples: Canadian (CDN) and Saskatchewan (SK)), Mountain and Moorland (Connemara, Dale, Dartmoor, Eriskay, Exmoor, Fell, Highland, Kerry Bog, New Forest and Welsh) or Nordic (Fjord, Icelandic and Shetland). Most samples were collected from North American (Canada and the United States) sources, with the exception of a few pony breeds, feral populations and the Mongolian Domestic samples. All feral populations and Mongolian Domestic samples were collected from their respective native regions. In addition, the Eriskay, New Forest and Exmoor samples were collected in the United Kingdom, whereas the Shetland samples came from both the United Kingdom and North America.

DNA templates were prepared from hair follicles or blood samples following a modified procedure by Troy *et al*. (2001). When blood samples were used, additional steps to lyse and remove the erythrocytes were incorporated. A lysis buffer containing 0.32 M Sucrose, 10 mM Tris-HCl, pH 8.0, 50 mM KCl and 0.5% Tween was mixed with 200 μl aliquots of whole fresh or frozen–thawed blood. The leukocytes were pelleted and the supernatant was aspirated. Each sample was washed twice with 400 μl of the same solution before DNA extraction.

Microsatellite loci were chosen based on a previous study by Glowatzki-Mullis *et al*. (2006) and the MoDaD microsatellite marker recommendations (Hoffmann *et al*., 2004). In total, 38 microsatellite loci were genotyped (AHT4, AHT5, AHT31, ASB2, ASB17, ASB23, ASB43, CA425, COR7, COR22, COR69, COR71, HMS1, HMS2, HMS3, HMS5, HMS6, HMS7, HMS45, HTG03, HTG04, HTG06, HTG07, HTG10, I-18, LEX33, LEX34, LEX54, LEX78, TKY301, TKY325, TKY333, TKY337, TKY341, TKY343, TKY344, UM32 and VHL20) by PCR using 1 μl of DNA template and commercially available kits and protocols (Amplitaq Gold; Applied Biosytems, Foster City, CA, USA or Qiagen Multiplexing Kit; Qiagen Inc., Burlington, Ontario, Canada). Details regarding these loci can be found in the Supplementary Table 1. Allele frequency tables for the populations examined in this study are available upon request.

The microsatellite amplicons were purified using Agencourt® AMPure® (Agencourt Bioscience Corporation/Beckman Coulter Company, Mississauga, Ontario, Canada) using the recommended protocol. Samples (0.8 μl of purified PCR product, 1 μl of 600-Liz size standard and 8.2 μl of Hi-Di Formamide (Applied Biosystems) were denatured for 5 min at 95°C, quenched on ice for 2 min and loaded onto a Genetic Analyzer 3130xl (Applied Biosystems) equipped with a 50-cm array and filled with POP7 polymer. Genotypes

were determined using GeneMapper® version 3.0 software (Applied Biosystems).

Statistical analysis

Deviations from the Hardy–Weinberg equilibrium (HWE) were estimated using MICRO-CHECKER (Hull, Yorkshire, UK; Van Oosterhout *et al.*, 2004). Linkage disequilibrium and overall F-statistics per locus were estimated using GENEPOP on the web, available at http://genepop.curtin.edu.au/ (Raymond and Rousset, 1995; Rousset, 2008). GENALEX version 6 (Peakall and Smouse, 2006) was used to estimate the average (N_a) and average effective (N_e) number of alleles, and inbreeding coefficient (F_{IS}). ARLEQUIN version 3.5 (Excoffier and Lischer, 2010) was used to compute the analysis of molecular variance (AMOVA), and observed (H_o) and expected (H_e) heterozygosities. Allelic richness (AR) was estimated using FSTAT (Goudet, 2001).

Microsatellite Analyzer (A-1210 Wien, Austria; Dieringer and Schlötterer, 2003) was used to estimate and bootstrap (1000 times) three genetic distances: Nei's unbiased (Nei, 1972), Nei's (1983) and the proportion of shared alleles (POSA; Bowcock *et al.*, 1994). In addition, Reynolds distance (Reynolds *et al.*, 1983) was also estimated and bootstrapped 1000 times using PHYLIP version 3.66 (Felsenstein, 1989 to 2006). Trees were created using the neighbor-joining method (Saitou and Nei, 1987), combined using CONSENUS (as implemented in PHYLIP version 3.66) and drawn using SplitsTree4 (Huson, 1998; Huson and Bryant, 2006). A POSA individual phylogenetic tree was also estimated using Microsatellite Analyzer and drawn using Mega4 (Tamura *et al.*, 2007).

Individuals were assigned to predefined populations using the 'leave-one-out option' and the maximum likelihood method (Paetkau *et al.*, 1995; Paetkau *et al.*, 2004) as implemented in GENALEX version 6. Structure version 2.1 (Pritchard *et al.*, 2000), with the parameter settings: K10 to 30, 50 000 burn-in, 100 000 Markov chain Monte Carlo (reaching equilibrium) with 10 replicates, was used to assign individuals to inferred clusters. Structure Harvester (Earl, 2009) was used to estimate the optimal number of clusters using the Evanno *et al.*'s (2005) approach. Clumpp version 1.1.2 (Jakobsson and Rosenberg, 2007) and Distruct version 1.1 (Rosenberg, 2004) were used to cluster the 10 independent runs into a single figure.

Contributions of each population to overall genetic diversity was also assessed using marginal loss (Weitzman, 1992 and 1993) and Nei's (1983) genetic distance as implemented in WEITZPRO (D'Arnoldi *et al.*, 1998). In addition, the loss in genetic diversity from the data set was also assessed using the Petit *et al.*'s (1998) approach and molecular kinship analyses as proposed by Caballero and Toro (2002) and implemented in MOLKIN version 2 (Gutiérrez *et al.*, 2005).

Results and discussion

In total, 821 individuals were examined across 24 populations. The analysis of the microsatellites revealed the presence of 468 alleles. This number is comparable with other studies that reported a total of 404 alleles while using 50 microsatellite loci and investigating seven equine populations (Glowatzki-Mullis *et al.*, 2006). The observation of more alleles with fewer microsatellite loci tested is likely a direct result of the increased number of populations sampled. The linkage disequilibrium analyses revealed varying results from complete equilibrium of all loci in the Newfoundland-SK and the Saint-Pierre et Miquelon population to as high as seven pairs of significantly linked loci ($P \leqslant 0.05$) in the Sable Island herd. As there were no pairs of loci consistently in linkage disequilibrium across populations, all markers were included in the analyses.

All populations were tested for deviation from the HWE, a test commonly utilized to check for null alleles that can result in excess homozygosity and an underestimation of the genetic diversity present (DeAssis *et al.*, 2009). Although, there was no specific evidence for the presence of null alleles in any of the loci tested, deviations from the HWE varied from zero loci in the Canadian, Highland, New Forest, Shetland, Standardbred and Welsh populations to as high as 13 loci in the Sable Island population ($P \leqslant 0.05$). Deviations from HWE have been reported previously in small populations (such as the Hucul, Icelandic, Jaca Navarra and Pottok), or populations that are highly inbred, and may explain the deviations seen in the Sable Island herd (Solis *et al.*, 2005; DeAssis *et al.*, 2009; Leroy *et al.*, 2009). The excess homozygosity observed could also be a result of genetic drift, nonrandom mating, substructure or small effective population size. Many of the breeds within the Canadian, Mountain and Moorland and Nordic populations have relatively small population sizes, resulting in an increase of homozygosity seen within a breed, and therefore explaining some of the deviations from the HWE observed. Specifically, the Lac La Croix was bottlenecked to four females in 1977; according to historical records, the Saint-Pierre et Miquelon population arose from approximately 75 individuals left on the islands in the 1700s (Lynghaug, 2009). Published research has also found that different maternal harems found on Sable Island show different levels of inbreeding, which may be reflective of the resource distribution (Lucas *et al.*, 2009). Lastly, the Newfoundland pony samples were treated as two separate groups (increasing the populations tested for this study to 25) as a result of a Wahlund effect detected by MICRO-CHECKER.

A summary of the population statistics can be found in Table 1. Individual locus N_a varied from five in the microsatellite HMS5 to 23 in TKY333 with an average of 12.3 alleles per locus. The number of individuals per population analyzed varied from 11 in the Kerry Bog to 60 in the Lac La Croix and Sable Island herds. The N_a observed in each population varied from 4.89 in the Eriskay to 7.84 in the Newfoundland-CDN population, and was followed closely by the Welsh with a value of 7.58. The low value for the Eriskay is likely due to the bottleneck experienced within the breed in the 1970s when it was reduced to one breeding stallion (Eriskay Pony Mother Stud Book Society (EPMSBS), 2010). The Welsh and Newfoundland likely have higher N_a values as a result of the great diversity seen within the Welsh sections and a partially opened registry maintained in the past in the

Table 1 *A summary of the basic statistics per population including sample size (*N*), average number of alleles (*N_a*), effective number of alleles (*N_e*), allelic richness (AR), observed heterozygosity (*H_o*), expected heterozygosity (*H_e*) and inbreeding coefficient (*F_{IS}*)*

Breed/populations	*N*	N_a	N_e	AR	H_o	H_e	F_{IS}
Canadian pony breeds							
Lac La Croix	60	5.18 ± 0.29	2.83 ± 0.16	4.01	0.623 ± 0.140	0.608 ± 0.132	−0.0227
Newfoundland – CDN	53	7.84 ± 0.39	4.23 ± 0.28	5.86	0.693 ± 0.106	0.726 ± 0.104	0.0430
Newfoundland – SK	19	5.63 ± 0.31	3.42 ± 0.20	5.01	0.672 ± 0.188	0.662 ± 0.142	−0.0094
Mountain and Moorland breeds							
Connemara	37	6.16 ± 0.26	3.75 ± 0.19	5.17	0.728 ± 0.114	0.705 ± 0.101	−0.0338
Dale	25	5.03 ± 0.28	3.15 ± 0.21	4.44	0.642 ± 0.169	0.629 ± 0.148	−0.0181
Dartmoor	25	5.95 ± 0.29	3.50 ± 0.19	5.07	0.677 ± 0.170	0.676 ± 0.128	−0.0004
Eriskay	27	4.89 ± 0.24	2.51 ± 0.10	3.98	0.599 ± 0.143	0.579 ± 0.104	−0.0329
Exmoor	25	5.16 ± 0.33	2.90 ± 0.16	4.34	0.569 ± 0.169	0.617 ± 0.129	0.0883
Fell	25	6.11 ± 0.30	3.65 ± 0.22	5.17	0.656 ± 0.158	0.684 ± 0.138	0.0414
Highland	25	5.63 ± 0.26	3.45 ± 0.17	4.87	0.678 ± 0.088	0.684 ± 0.094	0.0004
Kerry Bog	11	4.95 ± 0.21	3.55 ± 0.18	4.95	0.770 ± 0.151	0.689 ± 0.107	−0.1215
New Forest	26	7.16 ± 0.31	4.31 ± 0.23	6.01	0.762 ± 0.119	0.743 ± 0.088	−0.0261
Welsh	48	7.58 ± 0.41	4.33 ± 0.24	5.87	0.731 ± 0.092	0.744 ± 0.082	0.0153
Feral populations							
Grand Turk	17	5.11 ± 0.25	3.31 ± 0.19	4.69	0.675 ± 0.147	0.661 ± 0.120	−0.0252
Sable Island	60	6.39 ± 0.40	3.31 ± 0.19	4.62	0.578 ± 0.182	0.658 ± 0.135	0.1291
Saint-Pierre et Miquelon	29	6.55 ± 0.34	3.83 ± 0.24	5.40	0.711 ± 0.148	0.697 ± 0.120	−0.0213
Horse breeds							
Canadian	25	6.03 ± 0.30	3.68 ± 0.21	5.16	0.688 ± 0.144	0.688 ± 0.133	−0.0047
Caspian	25	5.84 ± 0.25	3.61 ± 0.19	5.10	0.691 ± 0.123	0.693 ± 0.107	−0.0029
Clydesdale	50	5.74 ± 0.30	2.97 ± 0.16	4.36	0.652 ± 0.122	0.623 ± 0.121	−0.0487
Haflinger	25	5.34 ± 0.26	3.28 ± 0.17	4.66	0.639 ± 0.159	0.655 ± 0.104	0.0320
Mongolian	35	8.79 ± 0.46	4.85 ± 0.23	6.64	0.756 ± 0.123	0.773 ± 0.081	0.0240
Standardbred	22	5.55 ± 0.30	3.25 ± 0.15	4.72	0.661 ± 0.163	0.659 ± 0.124	0.0016
Nordic breeds							
Fjord	50	6.05 ± 0.34	3.36 ± 0.18	4.74	0.661 ± 0.184	0.654 ± 0.171	0.0092
Icelandic	49	6.63 ± 0.34	3.54 ± 0.22	5.04	0.656 ± 0.138	0.678 ± 0.121	0.0336
Shetland	28	5.47 ± 0.25	3.23 ± 0.16	4.67	0.658 ± 0.146	0.653 ± 0.1221	−0.0025

Newfoundland pony breed. AR was calculated by standardizing the populations to a sample size of 11 individuals, and results were similar to N_a with 3.98 for the Eriskay, 5.87 for the Welsh and 5.86 for the Newfoundland-CDN population.

Following the same trend, N_e was the highest in the Welsh and New Forest breeds with values of 4.33 and 4.31, respectively. The Eriskay continued to show the lowest value at 2.51. The overall estimate for H_o and H_e, along with N_a, N_e and F indices, for each locus can be found in the Supplementary Table 2. Within the populations, the H_e varied from 0.579 in the Eriskay to 0.744 in the Welsh, followed closely by the New Forest with a value of 0.743. The H_o was highest at 0.770 in the Kerry Bog and lowest in the Exmoor at 0.569. These results are consistent with other published studies, in which fewer loci or a smaller sample size were used, for several populations including the Connemara, Fjord, Haflinger, Newfoundland, Standardbred and Sable Island (Luís *et al*., 2007; Plante *et al*., 2007; Leroy *et al*., 2009). In contrast, others have also found slightly higher values in the Connemara, Exmoor, Fell, Icelandic, Shetland, New Forest and Caspian (Cothran, 2004; Luís *et al*., 2007; Leroy *et al*., 2009). F_{IS} varied from 0.1291 in the Sable Island and 0.0883 in the Exmoor to −0.1215 in the Kerry Bog, indicating that there is a high deficiency in heterozygotes in the Sable Island and Exmoor populations, whereas there is an excess of heterozygotes in the Kerry Bog population. The high amounts of inbreeding seen in the Sable Island and Exmoor are likely a result of the small population sizes and feral and semi-feral status, respectively (Plante *et al*., 2007; Lynghaug, 2009). The combination of these two factors can lead to high levels of inbreeding and an increase in homozygosity. Results of the Kerry Bog showing an excess of heterozygotes may not reflect an accurate picture of the breed, but could be a result of the small number of samples obtained. All other populations were not significantly inbred or out crossed (<5%).

The AMOVA revealed expected results with 84.60% of genetic variance arising from within individuals, 2.07% among individuals and 13.33% occurring as a result of genetic differences among populations. These values are similar to previously reported data in which 10% to 17% and 83% to 90% of genetic diversity could be explained by breed differences and differences within individuals, respectively (Aberle *et al*., 2004; Glowatzki-Mullis *et al*., 2006; Plante *et al*., 2007).

Individual assignment tests based on the maximum likelihood methods to predefined populations found that 98% of individuals were correctly assigned to their respective group (Tables 2 and 3). Of those that could not be properly assigned include one Kerry Bog, assigned to the Welsh; two Welsh, assigned to the Icelandic and New Forest groups; six Newfoundland-CDN, four assigned to Newfoundland-SK, one to the Welsh group and one to the Saint-Pierre et Miquelon population; and one Newfoundland-SK assigned to the Saint-Pierre et Miquelon population. It is not surprising that some animals that are not actually members of this breed were assigned to the Welsh group. Many breeds, especially the above listed pony breeds, were influenced at one time or another by one or more of the four Welsh sections (Lynghaug, 2009). One Caspian was also assigned to the Welsh group, and it is believed that this sample was either mislabeled or that the inidividual was not a purebred as these two breeds are not closely related.

All phylogenetic reconstruction using genetic distances estimated showed similar topologies with low bootstrap values indicating that these breeds have likely diverged recently. Figure 1 illustrates the POSA approach that produced the phylogenetic tree making the most biological sense. The Lac La Croix and the Eriskay were not consistently placed throughout the different phylogenetic trees and is likely due to the bottleneck experienced in these breeds and subsequent hybridization with other breeds (Lynghaug, 2009; EPMSBS, 2010). An alternative explanation for the inconsistent placement of the Lac La Croix may be a result of not having Mustang samples from the area where admixture could have occurred in the past. In general, there appears to be two main groups observed in all the phylogenetic trees. The first is the Mountain and Moorland breeds group together, as expected, and share common ancestry. An interesting observation is that the Canadian horse, Grand Turk, Clydesdale and Standardbred also appear in this group. The Grand Turk and Standardbred populations were consistently paired together in all phylogenetic trees. It was originally thought that the Grand Turk population may have significantly influenced the development of Canadian equine populations due to the trade and movement flow of horses in the past; however, the phylogenetic trees do not support this opinion. The Clydesdale, also thought to have influenced the Canadian equine breeds, does appear to be distantly related to the Canadian in the POSA tree only. The relationships among horse breeds shown in this study are limited because only breeds that were expected to be closely related to the breeds of interest were examined. Thus, some horse breeds examined here may be more closely related to breeds not examined than to those in this study.

Within the Mountain and Moorland groups, several breeds also consistently shared several similarities among the phylogenetic trees. Specifically, the Dartmoor and Exmoor breeds paired together and appear to share common ancestry with the Kerry Bog ponies. This is not surprising as the Dartmoor and Exmoor breeds both developed in the South-West region of England (Lynghaug, 2009). In addition, the Welsh and Connemara breeds also consistently paired together with great confidence, supporting previously published data, and share common ancestry with New Forest breed (Leroy *et al.*, 2009). Lastly, as previously reported, the Fell and Dale also paired together in all phylogenetic trees and share common ancestry with the Highland breed (Cothran, 2004).

More interestingly, the second broad group observed in all phylogenetic trees is the combination of the Nordic breeds, Canadian pony breeds and Sable Island and Saint-Pierre et Miquelon populations. Although the Nordic breeds always grouped together, they also share common ancestry with the feral populations and Canadian pony breeds (Plante *et al.*, 2007). The consistent grouping of the Saint Pierre et Miquelon population and Newfoundland breed may indicate that these populations share recent common ancestry. The Sable Island population is likely a more distant ancestor of both the Saint Pierre et Miquelon and Newfoundland populations.

Figure 2 illustrates the individual unrooted phylogenetic tree created using POSA as an estimator of genetic distance. This tree illustrates that most of the pony breeds are generally distinct from each other, and individuals from the same breed tend to cluster together. A few breeds within the tree were observed to have overlapping clusters including the two Newfoundland herds (CDN and SK) and the Connemara, Eriskay and Welsh breeds. The overlapping clusters of the Newfoundland populations are not unexpected as these populations represent the same breed. The overlap and splitting of the Welsh breed across the Eriskay and Connemara clusters may be a result of individuals from the Welsh breed representing the different sections (types). Different sections within the Welsh breed may have influenced the Connemara and Eriskay breeds in the past and therefore may explain why these breeds cluster together.

The Bayesian analysis provided a unique way to investigate the individuals within this study and cluster them together without previous breed information. Structure Harvester determined that the most likely *K*-value was 21, although some substructure was observed at $K = 16$. When $K = 16$ (Figure 3; Table 2), there appeared to be a strong admixture, varying from one to many breeds, and some breeds grouped into clusters together. This was the case with the Fell and Dale, Welsh and Connemara, Canadian and Standardbred, Dartmoor and Haflinger and Newfoundland and Saint-Pierre et Miquelon populations. In the recent past, some breed organizations feared that their pony breeds (and Canadian populations) may have been crossed with Standardbred horses in an effort to increase certain traits within their breed. The clustering of the Standardbred and the Canadian horse when $K = 16$ is not surprising given the historical influence of the Canadian on the development of the Standardbred (Lynghaug, 2009). In addition, this cluster appeared to have some admixture with the New Forest, Highland, Fell, Dale, Haflinger, Welsh and Kerry Bog populations, which also shared admixture among each other. The Caspian, interestingly, appeared to share common ancestry with the Dartmoor and Haflinger and may be a result of the

Table 2 *The results from the population assignment (based on maximum likelihood methods) and the Bayesian analyses*

Population	Sample (*N*)	Assignment to self (%)	Assignment to inferred clusters															
			I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI
Canadian	25	100.0	0.031	0.016	0.015	0.003	0.012	0.016	0.008	0.033	**0.593**	0.091	0.050	0.039	0.007	0.006	0.032	0.050
Caspian	25	96.0	0.092	0.006	0.006	0.003	0.006	0.004	0.009	0.006	0.006	0.008	0.029	**0.780**	0.021	0.013	0.005	0.006
Clydesdale	50	100.0	0.006	0.005	0.004	0.003	0.003	0.002	0.004	0.006	0.005	0.007	0.006	0.005	**0.931**	0.005	0.004	0.004
Connemara	37	100.0	0.005	0.007	0.015	0.004	0.006	0.010	0.006	0.005	0.005	0.007	**0.898**	0.006	0.004	0.007	0.010	0.004
Dale	25	100.0	0.096	0.006	0.005	0.003	0.010	0.005	0.003	**0.612**	0.091	0.097	0.011	0.006	0.027	0.006	0.008	0.014
Dartmoor	25	100.0	0.177	0.280	0.008	0.003	0.009	0.020	0.004	0.007	0.008	0.007	0.009	**0.454**	0.005	0.004	0.004	0.003
Eriskay	27	100.0	0.004	0.008	0.005	0.002	0.007	0.003	0.004	0.004	0.006	0.008	0.006	0.004	0.003	0.003	0.005	**0.928**
Exmoor	25	100.0	0.004	0.007	0.003	0.006	0.006	**0.932**	0.002	0.004	0.007	0.004	0.005	0.007	0.003	0.002	0.004	0.004
Fell	25	100.0	0.193	0.005	0.005	0.007	0.007	0.004	0.003	**0.715**	0.013	0.016	0.004	0.006	0.004	0.004	0.006	0.010
Fjord	50	100.0	0.005	0.005	**0.937**	0.003	0.005	0.003	0.004	0.008	0.003	0.006	0.004	0.004	0.003	0.005	0.003	0.003
Grand Turk	17	100.0	0.004	**0.730**	0.005	0.004	0.100	0.010	0.003	0.006	0.009	0.095	0.008	0.005	0.007	0.006	0.004	0.006
Haflinger	25	100.0	**0.904**	0.008	0.003	0.006	0.004	0.003	0.003	0.007	0.005	0.007	0.008	0.007	0.013	0.007	0.004	0.011
Highland	25	100.0	0.006	0.005	0.004	0.003	0.004	0.007	0.002	**0.549**	0.189	0.191	0.006	0.005	0.005	0.015	0.004	0.005
Icelandic	49	100.0	0.005	0.005	0.011	0.005	0.008	0.003	0.005	0.006	0.004	0.007	0.006	0.006	0.004	**0.915**	0.006	0.004
Kerry Bog	11	90.9	0.025	0.108	0.010	0.007	0.057	0.055	0.004	0.050	0.086	**0.145**	0.125	0.063	0.006	0.042	**0.145**	0.072
Lac La Croix	60	100.0	0.003	0.003	0.002	0.007	0.003	0.006	**0.953**	0.002	0.002	0.004	0.002	0.003	0.003	0.003	0.002	0.002
Mongolian	35	100.0	0.092	0.264	0.013	0.008	0.008	0.005	0.007	0.005	0.007	**0.522**	0.013	0.015	0.011	0.008	0.017	0.006
New Forest	26	100.0	0.034	0.088	0.036	0.009	0.053	0.031	0.020	0.049	0.155	0.108	**0.182**	0.078	0.022	0.020	0.079	0.037
Newfoundland – CDN	53	88.7	0.008	0.018	0.012	0.008	0.017	0.005	0.006	0.007	0.015	0.015	0.030	0.012	0.005	0.020	**0.814**	0.009
Newfoundland – SK	19	94.7	0.004	0.004	0.009	0.014	0.017	0.007	0.004	0.006	0.004	0.005	0.004	0.003	0.004	0.003	**0.908**	0.005
Sable Island	60	100.0	0.003	0.003	0.005	**0.946**	0.005	0.004	0.004	0.003	0.003	0.004	0.003	0.004	0.002	0.004	0.005	0.003
Saint-Pierre et Miquelon	29	100.0	0.014	0.008	0.012	0.005	0.112	0.013	0.016	0.010	0.018	0.015	0.026	0.018	0.031	0.012	**0.680**	0.010
Shetland	28	100.0	0.005	0.018	0.009	0.005	**0.731**	0.010	0.005	0.005	0.016	0.006	0.008	0.013	0.004	0.152	0.006	0.007
Standardbred	22	100.0	0.003	0.010	0.003	0.002	0.009	0.007	0.005	0.004	**0.925**	0.005	0.006	0.004	0.003	0.004	0.005	0.005
Welsh	48	95.8	0.017	0.013	0.016	0.008	0.009	0.011	0.013	0.022	0.031	0.086	**0.673**	0.022	0.018	0.014	0.027	0.019

The values under columns I to XVI are from population *Q*-matrices that show distribution of *Q*-values in identified clusters. Bold values identify the clusters with highest *Q*-values (proportion of genotype membership). Shaded areas represent populations which are indigenous to Canada.

Table 3 *The results from the population assignment (based on maximum likelihood methods) and the Bayesian analyses*

Population	Sample (*M*)	Assignment to self (%)	Assignment to inferred clusters																				
			I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	XVII	XVIII	XIX	XX	XXI
Canadian	25	100.0	0.002	0.003	0.003	0.005	**0.858**	0.010	0.007	0.012	0.007	0.008	0.006	0.007	0.007	0.005	0.005	0.008	0.012	0.010	0.010	0.007	0.008
Caspian	25	96.0	0.003	0.008	0.005	0.006	0.006	0.008	0.022	0.008	0.004	0.006	0.008	0.003	0.003	**0.864**	0.017	0.005	0.004	0.006	0.005	0.005	0.004
Clydesdale	50	100.0	0.002	0.005	0.006	0.003	0.004	0.006	0.004	0.005	0.004	0.003	0.006	0.002	0.003	0.004	**0.915**	0.005	0.004	0.005	0.003	0.004	0.005
Connemara	37	100.0	0.004	0.006	0.008	0.004	0.005	0.011	**0.779**	0.094	0.005	0.005	0.004	0.010	0.014	0.006	0.005	0.004	0.004	0.009	0.012	0.007	0.004
Dale	25	100.0	0.003	0.005	0.003	0.003	0.016	0.009	0.098	**0.354**	0.007	0.008	0.008	0.005	0.007	0.004	0.027	0.169	0.169	0.009	0.005	0.006	0.085
Dartmoor	25	100.0	0.006	0.004	0.006	0.004	0.012	0.008	0.006	0.007	0.005	0.006	0.007	0.013	0.004	0.013	0.006	0.005	0.003	0.007	0.007	0.006	**0.865**
Eriskay	27	100.0	0.002	0.003	0.005	0.003	0.003	0.007	0.004	0.004	0.006	0.006	0.003	0.003	0.005	0.003	0.003	0.005	**0.910**	0.009	0.005	0.007	0.004
Exmoor	25	100.0	0.006	0.002	0.004	0.002	0.004	0.095	0.004	0.004	0.006	0.006	0.003	**0.826**	0.003	0.003	0.003	0.003	0.004	0.004	0.003	0.004	0.012
Fell	25	100.0	0.007	0.004	0.017	0.003	0.016	0.003	0.010	**0.356**	0.005	0.008	0.025	0.004	0.006	0.007	0.004	0.344	0.161	0.005	0.006	0.003	0.007
Fjord	50	100.0	0.003	0.004	0.004	0.004	0.003	0.003	0.004	0.004	0.002	0.005	0.005	0.003	0.003	0.003	0.003	0.009	0.003	0.003	**0.922**	0.004	0.006
Grand Turk	17	100.0	0.003	0.007	0.005	0.002	0.004	0.005	0.005	0.005	0.007	0.007	0.003	0.006	0.003	0.004	0.004	0.005	0.003	0.006	0.005	**0.906**	0.006
Haflinger	25	100.0	0.004	0.005	0.004	0.003	0.003	0.004	0.006	0.005	0.003	0.003	**0.907**	0.002	0.003	0.003	0.009	0.005	0.007	0.004	0.003	0.003	0.013
Highland	25	100.0	0.003	0.013	0.004	0.002	0.011	0.006	0.005	0.007	0.005	0.004	0.005	0.006	0.004	0.004	0.004	**0.888**	0.006	0.011	0.003	0.004	0.005
Icelandic	49	100.0	0.004	**0.897**	0.007	0.004	0.005	0.004	0.006	0.005	0.003	0.008	0.004	0.003	0.005	0.005	0.004	0.006	0.004	0.005	0.011	0.004	0.005
Kerry Bog	11	90.9	0.004	0.007	0.005	0.003	0.009	0.046	0.006	0.031	0.008	0.009	0.006	0.010	0.008	0.008	0.004	0.015	0.009	**0.787**	0.004	0.008	0.015
Lac La Croix	60	100.0	0.006	0.003	0.004	**0.944**	0.002	0.002	0.002	0.003	0.002	0.003	0.002	0.005	0.002	0.003	0.003	0.002	0.002	0.003	0.002	0.002	0.003
Mongolian	35	100.0	0.007	0.007	**0.834**	0.006	0.005	0.010	0.007	0.008	0.006	0.005	0.017	0.006	0.009	0.014	0.008	0.004	0.005	0.012	0.012	0.012	0.006
New Forest	26	100.0	0.007	0.009	0.038	0.017	0.027	0.077	0.028	**0.191**	0.056	0.032	0.013	0.022	0.029	0.032	0.018	0.036	0.026	0.255	0.023	0.030	0.039
Newfoundland – CDN	53	88.7	0.008	0.017	0.012	0.006	0.017	0.025	0.017	0.029	0.007	0.010	0.005	0.028	**0.739**	0.008	0.005	0.006	0.007	0.022	0.011	0.013	0.012
Newfoundland – SK	19	94.7	0.014	0.003	0.005	0.004	0.009	0.024	0.004	0.016	0.002	0.002	0.003	0.037	**0.836**	0.003	0.004	0.005	0.005	0.007	0.010	0.004	0.003
Sable Island	60	100.0	**0.936**	0.004	0.004	0.003	0.003	0.003	0.003	0.003	0.002	0.005	0.003	0.004	0.004	0.003	0.002	0.003	0.003	0.003	0.004	0.003	0.002
Saint-Pierre et Miquelon	29	100.0	0.004	0.009	0.013	0.014	0.018	0.183	0.015	0.098	0.010	0.019	0.010	0.092	**0.432**	0.016	0.023	0.007	0.008	0.008	0.011	0.004	0.007
Shetland	28	100.0	0.004	0.007	0.006	0.003	0.004	0.004	0.007	0.006	0.009	**0.888**	0.003	0.005	0.004	0.005	0.003	0.003	0.005	0.008	0.005	0.007	0.014
Standardbred	22	100.0	0.002	0.005	0.003	0.005	0.006	0.008	0.004	0.005	**0.898**	0.010	0.003	0.005	0.006	0.004	0.003	0.005	0.005	0.007	0.003	0.007	0.006
Welsh	48	95.8	0.007	0.011	0.008	0.008	0.025	**0.479**	0.041	0.045	0.016	0.005	0.009	0.010	0.014	0.013	0.013	0.011	0.014	0.248	0.011	0.006	0.008

The values under columns I to XXI are from population *Q*-matrices that show distribution of *Q*-values in identified clusters. Bold values identify the clusters with highest *Q*-values (proportion of genotype membership). Shaded areas represent populations which are indigenous to Canada.

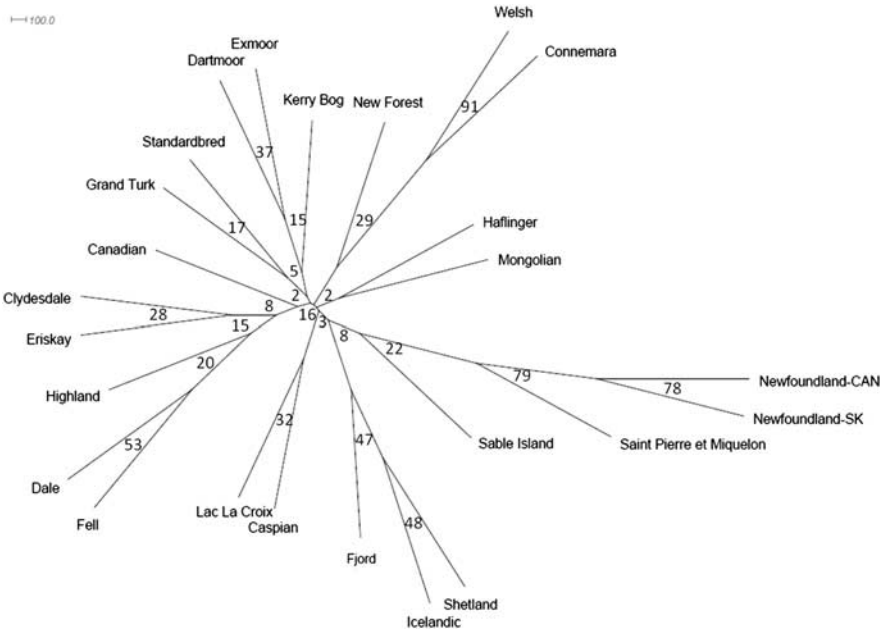

Figure 1 The unrooted population phylogenetic tree created using the proportion of shared alleles; individual confidence values on each branch are also included.

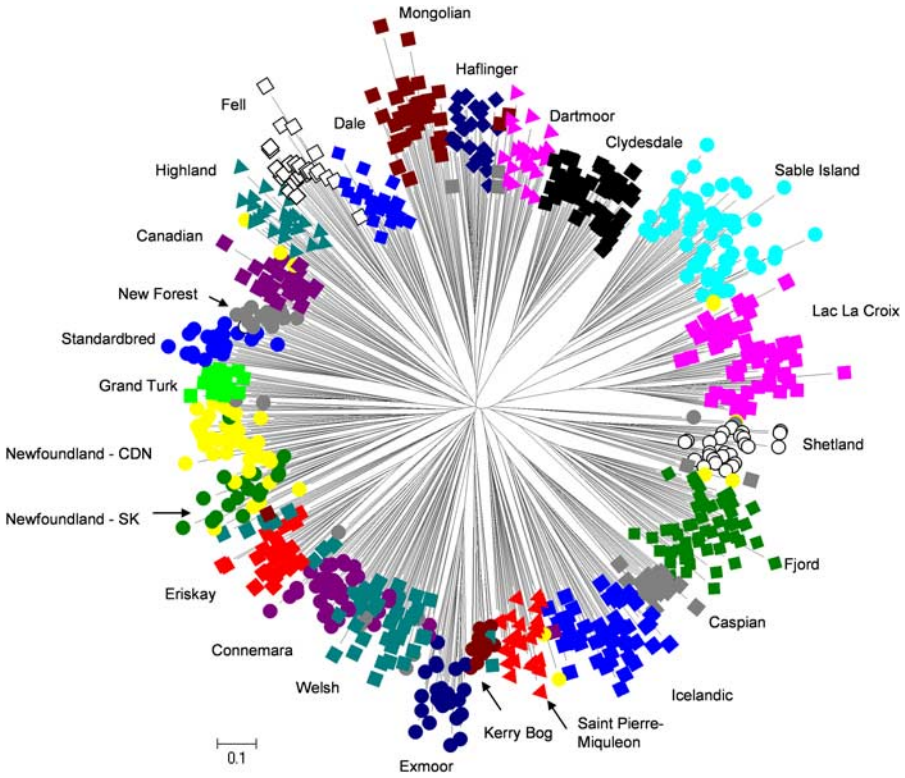

Figure 2 The unrooted individual phylogenetic tree created using a proportion of shared allele pairwise difference matrix.

known influence from the Arabian on all of these breeds (Lynghaug, 2009). In addition, the grouping of the Connemara and Welsh supports the close phylogenetic relationship observed in all four genetic distances estimated in this study.

When $K = 21$ (Figure 4; Table 3), the Canadian and Standardbred, Dartmoor and Haflinger and Welsh and Connemara, separated out into breed-specific clusters, and the strong admixture observed when $K = 16$ between some of the pony populations was no longer seen. Some pony

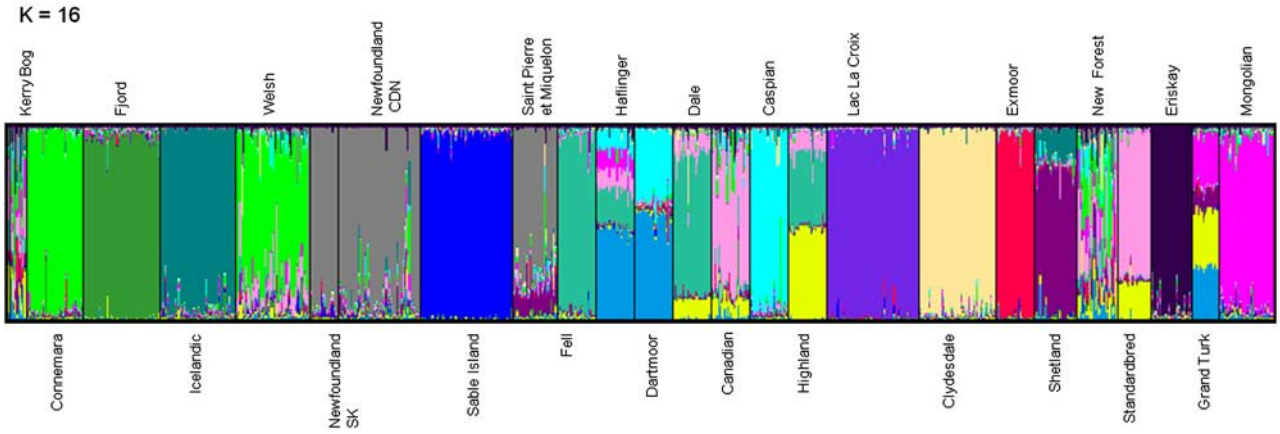

Figure 3 The consensus band for the 10 independent structure runs at $K = 16$; each single line represents one individual.

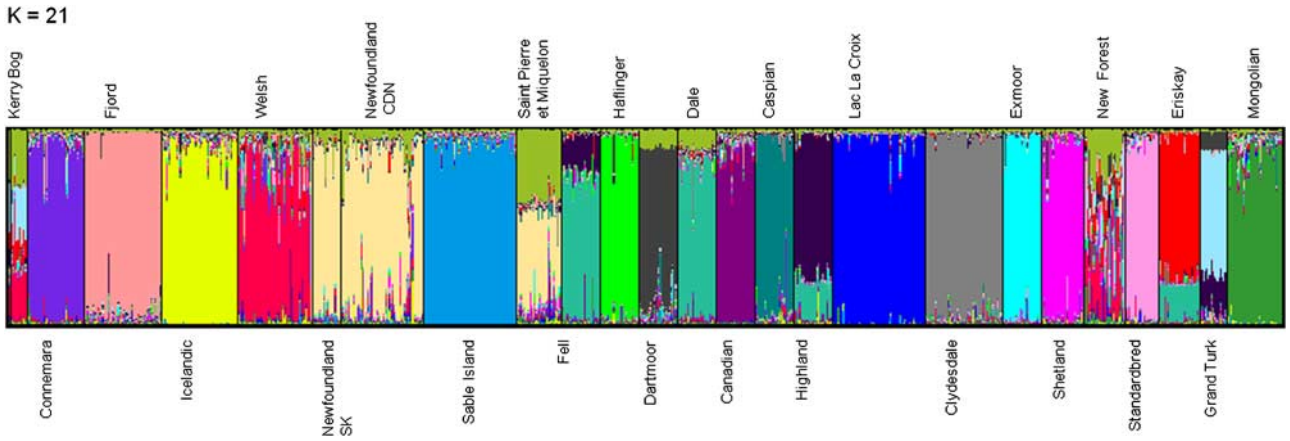

Figure 4 The consensus band for the 10 independent structure runs at $K = 21$; each single line represents one individual.

breeds, most notably the New Forest and Kerry Bog, still did not form their own individual clusters but rather appeared to be a mixture of breeds. A possible explanation for this may be that the New Forest breed was frequently bred with other Mountain and Moorland breeds until the 1930s as a way of improving the breed (Lynghaug, 2009). The Kerry Bog pony suffered a severe herd reduction to only 20 animals in the 1990s, and as a result the breed was also likely crossed with other Mountain and Moorland breeds in order to conserve the population (McGahern *et al.*, 2006). The Newfoundland populations and the Saint-Pierre et Miquelon herd remain in a single cluster, indicating that there is likely strong admixture between these two populations or that they share very recent common ancestry. The Fell and Dale also continued to share a single cluster, which is not unexpected as the breeds developed in the same region (Lynghaug, 2009). The Grand Turk population, also thought to have influenced the Canadian populations, appeared to have admixture from both horse and pony breeds; however, none of them were populations native to Canada. Interestingly, populations such as the Sable Island, Icelandic, Fjord, Lac La Croix, Clydesdale, Exmoor and Eriskay are quite distinct from the other populations examined as these breeds segregated out even at early *K*-values, which has also been reported previously for the Sable Island and Icelandic populations when comparing them with horse breeds (Glowatzki-Mullis *et al.*, 2006; Plante *et al.*, 2007).

The contribution to overall genetic diversity (Table 4) revealed varying results for the three different approaches when trying to assess the priority of the breeds included in this study. Results using the Weitzman approach (Weitzman, 1992 and 1993) favored the Lac La Croix pony as having the greatest marginal loss (6.57%) of all the populations tested when removed from the data set and was followed closely by the Exmoor (6.06%) and Eriskay (5.66%) breeds. In contrast, the Welsh population was found to have the least marginal loss (1.65%), followed closely by the New Forest (2.18%) and Newfoundland-CDN population (2.23%). The Petit *et al.*'s (1998) approach followed similar trends to the Weitzman approach with the greatest loss of diversity occurring when the Eriskay (1.89%), Lac La Croix (1.84%) and Exmoor (1.62%) were removed from the data set. The populations that contributed the least amount of diversity to the data set were the New Forest (0.43%), Newfoundland-SK (0.53%) and Welsh (0.54%). In contrast to both the previous mentioned analyses, the Caballero and Toro's (2002)

Table 4 *Results of the contribution to genetic diversity of each population within the data set using three different approaches: Weitzman's, Petit* et al.*'s and Caballero and Toro*

Breed/populations	Marginal loss (%)[1]	Internal diversity (%)[2]	Gain/loss (%)[3]
Canadian pony breeds			
Lac La Croix	6.57	−1.84	−0.08
Newfoundland – CDN	2.23	−1.14	+0.02
Newfoundland – SK	4.01	−0.53	−0.02
Mountain and Moorland breeds			
Connemara	3.55	−1.04	+0.01
Dale	3.20	−1.56	+0.14
Dartmoor	3.98	−1.10	−0.05
Eriskay	5.66	−1.89	+0.07
Exmoor	6.06	−1.62	−0.10
Fell	3.33	−1.02	−0.16
Highland	4.43	−1.23	−0.11
Kerry Bog	4.47	−1.19	−0.01
New Forest	2.18	−0.43	+0.01
Welsh	1.65	−0.54	−0.02
Feral populations			
Grand Turk	4.71	−1.37	−0.02
Sable Island	4.77	−1.39	−0.22
Saint-Pierre et Miquelon	2.89	−0.85	+0.06
Horse breeds			
Canadian	3.20	−1.03	+0.05
Caspian	3.35	−1.08	0.00
Clydesdale	4.21	−1.61	+0.10
Haflinger	4.18	−1.40	−0.05
Mongolian	3.35	0.04	−0.33
Standardbred	4.62	−0.04	−1.35
Nordic breeds			
Fjord	4.73	−1.33	−0.02
Icelandic	3.48	−1.12	−0.01
Shetland	3.76	−1.40	+0.03

[1]Data calculated using Weitzman's approach (Weitzman, 1992 and 1993).
[2]Data calculated using Petit *et al.*'s approach (Petit *et al.*, 1998).
[3]Data calculated using Caballero and Toro's approach (Caballero and Toro, 2002).

approach favored the Sable Island population as having the greatest loss of diversity (−0.22%) when removed from the other populations examined, followed closely by the Fell (−0.16%) and Highland pony (−0.11%). Surprisingly, the removal of the Dale, Eriskay and Canadian actually increased the diversity observed within the data set examined with values of 0.14, 0.07 and 0.05%, respectively. A previous study using fewer microsatellites also has found that the removal of the Sable Island herd from their data set also greatly reduced the amount of diversity observed when comparing the feral herd with horse breeds (Plante *et al.*, 2007).

Conclusions

The Canadian, Mountain and Moorland and Nordic populations significantly contribute to the genetic diversity within the species. In addition, populations such as the Fjord, Icelandic, Sable Island, Eriskay, Exmoor and Lac La Croix appear to segregate out quickly with low *K*-values, indicating that they are likely more distinct from other breeds examined. Interestingly enough, the hypothesis that horse breeds such as the Clydesdale and Standardbred, originally thought to have influenced Canadian pony populations, does not appear to be supported from the data gathered in this study; however, the Standardbred did appear to show strong admixture, or common ancestry, as expected, with the Canadian horse when $K = 16$; also, the Clydesdale did appear to be distantly related in the POSA phylogenetic tree. One of the most surprising findings in this study was the close relation of the Newfoundland and Saint-Pierre et Miquelon populations, as well as the lack of influence of the Grand Turk on any of the native Canadian equine populations. The exact priority of the populations examined for the distribution of resources varies between the different approaches estimated, but in general the Lac La Croix and Sable Island populations, along with the Eriskay, Exmoor, Fell and Highland breeds appear to be the top priorities. Further study will have to be conducted to determine the relationship of the Lac La Croix among both horse and pony breeds, as this breed was not consistently placed in the phylogenetic trees using the four genetic distances estimated.

This study provides the crucial first step in the breed characterization of several Canadian, Mountain and Moorland and Nordic populations, which should be combined in the future with other molecular information such as mitochondrial DNA sequence data, and/or currently available genealogical and phenotypic data to produce an optimal and effective conservation strategy for all breeds examined.

Acknowledgments

The authors wish to acknowledge the significant contributions of the many breed societies, Zoe Lucas with the Sable Island Green Horse Society, Patricia Morris from the Newfoundland Pony Society and Maxxam Analytics (Ontario, Canada) for providing valuable support and samples. This study was funded by the Agriculture and Agri-Food Canada.

References

Aberle KS, Hamann H, Drögemüller C and 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. Animal Genetics 35, 270–277.

Achmann R, Curik I, Dovc P, Kavar T, Bodo I, Habe F, Marti E, Sölkner J and Brem G 2004. Microsatellite diversity, population subdivision and gene flow in the Lipizzan horse. Animal Genetics 35, 285–292.

Behl R, Behl J, Gupta N and Gupta SC 2008. Evaluation of microsatellite genotyping based individual assignment in five Indian horse breeds. Indian Journal of Animal Sciences 78, 384–387.

Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR and Cavalli-Sforza LL 1994. High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368, 455–457.

Caballero A and Toro MA 2002. Analysis of genetic diversity for the management of conserved subdivided populations. Conservation Genetics 3, 289–299.

Cañon J, Checa ML, Carleos C, Vega-Pla JL, Vallejo M and Dunner S 2000. The genetic structure of Spanish Celtic horse breeds inferred from microsatellite data. Animal Genetics 31, 39–48.

Cothran EG 2004. Genetic analysis of the Lac La Croix pony. Retrieved February 24, 2011, from http://www.laclacroixindianpony.com/pdfs/cothranDNAreport.pdf

D'Arnoldi CT, Foulley JL and Ollivier L 1998. An overview of the Weitzman approach to diversity. Un apercu sur l'approche de la diversite selon Weitzman 30, 149–161.

DeAssis JB, DeLaat DM, Peixoto MG, Bergmann JA, Fonseca CG and Carvalho MR 2009. Genetic diversity and population structure in Brazilian Mangalarga Marchador horses. Genetics and Molecular Research 8, 1519–1524.

Dieringer D and Schlötterer C 2003. Microsatellite Analyser (MSA): a platform independent analysis tool for large microsatellite data sets. Molecular Ecology Notes 3, 167–169.

Earl DA 2009. Structure Harvester. Retrieved February 24, 2011, from http://taylor0.biology.ucla.edu/struct_harvest/

Eggert LS, Powell DM, Ballou JD, Malo AF, Turner A, Kumer J, Zimmerman C, Fleischer RC and Maldonado JE 2010. Pedigrees and the study of the wild horse population of assateague island national seashore. Journal of Wildlife Management 74, 963–973.

Eriskay Pony Mother Stud Book Society 2010. Our Stallions1. Retrieved February 24, 2011, from http://www.eriskaypony.org.uk/

Equus Survival Trust 2008. Equine conservation list. Retrieved February 24, 2011, from http://www.equus-survival-trust.org/

Evanno G, Regnaut S and Goudet J 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Molecular Ecology 14, 2611–2620.

Excoffier L and Lischer HEL 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10, 564–567.

Food and Agriculture Organization (FAO) 2007. The State of the World's Animal Genetic Resources for Food and Agriculture (ed. Rischkowsky B and Pilling D). FAO, Rome.

Felsenstein J 1989–2006. 'Phylip'(phylogeny inference package) *v.* 3.66. Retrieved February 24, 2011, from http://evolution.genetics.washington.edu/phylip/getme.html

Glowatzki-Mullis ML, Muntwyler J, Pfister W, Marti E, Rieder S, Poncet PA and Gaillard C 2006. Genetic diversity among horse populations with a special focus on the Franches–Montagnes breed. Animal Genetics 37, 33–39.

Goudet J 2001. FSTAT: a program to estimate and test gene diversities and fixation indices, version 2.9.3. Retrieved February 24, 2011, from http://www2.unil.ch/popgen/softwares/fstat.htm

Gutiérrez JP, Royo LJ, Álvarez I and Goyache F 2005. MolKin v2.0: a computer program for genetic analysis of populations using molecular coancestry information. Journal of Heredity 96, 718–721.

Hoffmann I, Ajmone Marsan P, Barker SF, Cothran EG, Hanotte O, Lenstra JA, Milan D, Weigend S and Simianer H 2004. New MoDaD marker sets to be used in diversity studies for the major farm animal species: recommendations of a joint ISAG/FAO working group. Proceedings of 29th International Conference on Animal Genetics, Tokyo, Japan, 107pp.

Huson DH 1998. SplitsTree: Analyzing and visualizing evolutionary data. Bioinformatics 14, 68–73.

Huson DH and Bryant D 2006. Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution 23, 254–267.

Jakobsson M and 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.

Leroy G, Callède L, Verrier E, Mériaux JC, Ricard A, Danchin-Burge C and Rognon X 2009. Genetic diversity of a large set of horse breeds raised in France assessed by microsatellite polymorphism. Genetics Selection Evolution 41, 5.

Lucas ZL, McLoughlin PD, Coltman DW and Barber C 2009. Multiscale analysis reveals restricted gene flow and a linear gradient in heterozygosity for an island population of feral horses. Canadian Journal of Zoology 87, 310–316.

Luís C, Juras R, Oom MM and Cothran EG 2007. Genetic diversity and relationships of Portuguese and other horse breeds based on protein and microsatellite loci variation. Animal Genetics 38, 20–27.

Lynghaug F 2009. The official horse breeds standards guide: the complete guide to the standards of all North American equine breed associations. Voyageur Press, Minneapolis, MN, USA.

McGahern AM, Edwards CJ, Bower MA, Heffernan A, Park SDE, Brophy PO, Bradley DG, MacHugh DE and Hill EW 2006. Mitochondrial DNA sequence diversity in extant Irish horse populations and in ancient horses. Animal Genetics 37, 498–502.

Nei M 1972. Genetic distance between populations. American Naturalist 106, 283–292.

Nei M, Tajima F and Tateno Y 1983. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. Journal of Molecular Evolution 19, 153–170.

Nova Scotia Museum of Natural History 2001. Sable Island. Retrieved February 24, 2011, from http://museum.gov.ns.ca/mnh/nature/sableisland/english_en/nature_na/horses_ho/index_ho.htm

Paetkau D, Calvert W, Stirling I and Strobeck C 1995. Microsatellite analysis of population structure in Canadian polar bears. Molecular Ecology 4, 347–354.

Paetkau D, Slade R, Burden M and Estoup A 2004. Genetic assignment methods for the direct, real-time estimation of migration rate: A simulation-based exploration of accuracy and power. Molecular Ecology 13, 55–65.

Peakall R and Smouse PE 2006. GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6, 288–295.

Petit RJ, Mousadik AEL and Pons O 1998. Identifying populations for conservation on the basis of genetic markers. Conservation Biology 12, 844–855.

Plante Y, Vega-Pla JL, Lucas Z, Colling D, De March B and Buchanan F 2007. Genetic diversity in a feral horse population from Sable Island, Canada. Journal of Heredity 98, 594–602.

Pritchard JK, Stephens M and Donnelly P 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945–959.

Rare Breeds Canada 2009. Horse breeds. Retrieved February 24, 2011 from http://www.rarebreedscanada.ca/horsebreeds.htm

Raymond M and Rousset F 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. Journal of Heredity 86, 248–249.

Reynolds J, Weir BS and Cockerham CC 1983. Estimation of the coancestry coefficient: Basis for a short-term genetic distance. Genetics 105, 767–779.

Rosenberg NA 2004. DISTRUCT: a program for the graphical display of population structure. Molecular Ecology Notes 4, 137–138.

Rousset F 2008. Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Molecular Ecology Resources 8, 103–106.

Saitou N and Nei M 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular biology and evolution 4, 406–425.

Solis A, Jugo BM, Mériaux JC, Iriondo M, Mazón LI, Aguirre AI, Vicario A and Estomba A 2005. Genetic diversity within and among four south European native horse breeds based on microsatellite DNA analysis: implications for conservation. Journal of Heredity 96, 670–678.

Tamura K, Dudley J, Nei M and Kumar S 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24, 1596–1599.

Troy CS, MacHugh DE, Bailey JF, Magee DA, Loftus RT, Cunningham P, Chamberlain AT, Sykes BC and Bradley DG 2001. Genetic evidence for near-eastern origins of European cattle. Nature 410, 1088–1091.

Van Oosterhout C, Hutchinson WF, Wills DPM and Shipley P 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes 4, 535–538.

Weitzman ML 1992. On Diversity. The Quarterly Journal of Economics 107, 363–405.

Weitzman ML 1993. What to preserve? An application of diversity theory to crane conservation. The Quarterly Journal of Economics 108, 157–183.