



AUTHOR(S):

TITLE:

YEAR:

Publisher citation:

OpenAIR citation:

Publisher copyright statement:

This is the _____ version of an article originally published by _____
in _____
(ISSN _____; eISSN _____).

OpenAIR takedown statement:

Section 6 of the "Repository policy for OpenAIR @ RGU" (available from <http://www.rgu.ac.uk/staff-and-current-students/library/library-policies/repository-policies>) provides guidance on the criteria under which RGU will consider withdrawing material from OpenAIR. If you believe that this item is subject to any of these criteria, or for any other reason should not be held on OpenAIR, then please contact openair-help@rgu.ac.uk with the details of the item and the nature of your complaint.

This publication is distributed under a CC _____ license.

Comparative genetic diversity in a sample of pony breeds from the U.K. and North America: a case study in the conservation of global genetic resources

Clare L. Winton¹, Yves Plante², Pamela Hind², Robert McMahon¹, Matthew J. Hegarty¹, Neil R. McEwan¹, Mina C. G. Davies-Morel¹, Charly M. Morgan¹, Wayne Powell³ & Deborah M. Nash¹

¹IBERS, Aberystwyth University, Aberystwyth, Ceredigion SY23 3DA, U.K.

²Agriculture and Agri-Food Canada, Saskatoon, Saskatchewan S7N 5A8, Canada

³CGIAR, F-34394, Montpellier Cedex 5, France

Keywords

Conservation, founder effect, horse, microsatellite, mtDNA, phylogenetics.

Correspondence

Deborah M. Nash, IBERS, Aberystwyth University, Aberystwyth, Ceredigion SY23 3DA, U.K.

Tel: (+44) 1970 622575;

Fax: 01970 622350;

E-mail: dmn@aber.ac.uk

Funding Information

We are grateful for funding provided by an IBERS PhD Scholarship (Aberystwyth University) and a Royal Society Research Grant 2010/R2, which supported Clare Winton.

Received: 8 September 2014; Revised: 14

May 2015; Accepted: 18 May 2015

Ecology and Evolution 2015; 5(16):
3507–3522

doi: 10.1002/ece3.1562

Introduction

The maintenance of populations to preserve maximal genetic variability and future adaptation potential is a common management challenge of conservation efforts in both domestic and wild species. Every finite population risks losing variability through stochastic sampling particularly when the number of reproducing individuals is restricted, creating founder effects and bottlenecks, for example, during: reintroduction of wild animals to previous habitats, formation of ex situ captive breeding groups, and formation of domestic species (Food and Agriculture Organization, 2007; Shen et al. 2009; Vonholdt et al. 2010). Establishing subpopulations requires sufficient sampling of the founding source's genetic vari-

Abstract

Most species exist as subdivided ex situ daughter population(s) derived from a single original group of individuals. Such subdivision occurs for many reasons both natural and manmade. Traditional British and Irish pony breeds were introduced to North America (U.S.A. and Canada) within the last 150 years, and subsequently equivalent breed societies were established. We have analyzed selected U.K. and North American equivalent pony populations as a case study for understanding the relationship between putative source and derived subpopulations. Diversity was measured using mitochondrial DNA and a panel of microsatellite markers. Genetic signatures differed between the North American subpopulations according to historical management processes. Founder effect and stochastic drift was apparent, particularly pronounced in some breeds, with evidence of admixture of imported mares of different North American breeds. This demonstrates the importance of analysis of subpopulations to facilitate understanding the genetic effects of past management practices and to lead to informed future conservation strategies.

ability to be sustainable and remains representative of the origin (Vonholdt et al. 2010), and small founder numbers and/or limited lineage sampling is associated with increased rates of inbreeding and divergence from the source (Frankam et al. 2002). Maintaining gene flow between, and limiting reproductive variance within, the derived subpopulations is predicted to counteract the negative effects of genetic drift on diversity reduction and adaptive allele loss, particularly for small populations (Miller et al. 2009; Vonholdt et al. 2010). These principles apply universally. However, in many cases, the genetic consequences of current or previous management practices are not well characterized or understood. For domestic animals, divergent selection and gene flow have been occurring on a local and continental scale since

prehistoric times, yet formalized breeding organizations or societies have arisen only in the last ~300 years (Ensminger and Parker 1986; Food and Agriculture Organization, 2007). Subpopulations of recognized breeds have been introduced into countries beyond their native origin, under the influence of commercial organizations, breeders' associations, and hobby enthusiasts (Food and Agriculture Organization, 2007). These introductions have often been documented, but the degree to which daughter populations effectively preserve the genetics of the original stock is often poorly characterized. Because the dynamics of these domesticated groups can be considered as modeling a larger range of conservation issues, genetic investigations to determine how representative subpopulations are of the original stocks in relation to known recorded information can potentially provide insight into future international strategies for conservation.

Mitochondrial DNA and microsatellite markers have been utilized to discover the origins and relationships of many horse breeds (Royo *et al.* 2005; Solis *et al.* 2005; Cai *et al.* 2009; Van de Goor *et al.* 2011; Prystupa *et al.* 2012a,b) and to elucidate the domestication process (Vila *et al.* 2001; Jansen *et al.* 2002; Cieslak *et al.* 2010; Lippold *et al.* 2011; Achilli *et al.* 2012). Others characterized the concordance between pedigrees and actual genetic relationships of current Thoroughbreds and Arabs (Hill *et al.* 2002; Glazewska *et al.* 2007). However, limited data compare populations of the same breed established on different continents. Nevertheless, pony populations derived from U.K. and Irish Native breeds have been examined in diverse locations, for example, maternal origins and mitochondrial diversity of native ponies have been studied in Ireland (McGahern *et al.* 2006) and North America (Prystupa *et al.* 2012a,b). Others have incorporated subsets of British and Irish pony breeds as part of a wider dataset for microsatellite (Simple Sequence Repeat; SSR) analysis, although the sample population origins are not always stated or are from exported subpopulations (Luis *et al.* 2007; Leroy *et al.* 2009; Van de Goor *et al.* 2011). There are 11 recognized breeds of pony native to Britain and Ireland plus a distinct feral population living in the Carneddau mountain range of Snowdonia, North Wales (Winton *et al.* 2013). Prystupa *et al.* (2012a,b) analyzed samples from a comprehensive range of these mountain and moorland-type pony breeds from North America (U.S.A. and Canada) using mtDNA/SSR markers. As such, an opportunity exists to compare the genetic diversity and structure present within these subpopulations with data from the U.K. source populations. Moreover, a broad study of U.K. and North American native breeds employing comprehensive analyses provides an overarching view of international horse gene flow and divergence, rather than

solely focusing on diversity of specific subpopulations of localized breeds.

Importation of recognized horse breeds into North America started in the 1800s; some breeds have been systematically imported for over 100 years, while others became established more recently (Fig. 1). Welsh pony imports commenced in the 1880s and were first officially recorded in 1909 when six stallions and 27 mares were sent to the U.S.A. and Canada. Imports into the U.S.A. peaked at 427 in 1957, with an annual median of eight animals between 1957 and 1991, although there were also no recorded imports in some years, particularly during wartime (Davies 1993). The importance of this breed in North America is illustrated by the establishment of a formalized breed society in 1907 (Evans 2001). The earliest named shipment of Shetland ponies to the U.S.A. was 1884, and the U.S.A. breed society was established in 1888. High demand for Shetlands existed (1332 ponies imported 1900–1910) and, except for a decline during the World Wars, importation continues to the present day (Russell 1996; Hodson 1997). In contrast, the Kerry Bog Pony has only recently become established in the U.S.A. with an initial breeding herd imported in 2003 (Anon, 2013a).

This study uses mitochondrial DNA sequencing and a panel of microsatellite markers to determine whether North American daughter pony populations maintain genetic diversity while remaining representative of the U.K. source breeds despite different patterns of importation.

Materials and Methods

Sample collection and genomic DNA extraction

Studies were undertaken using animals representative of different bloodlines for each of eight British and Irish native pony populations. Populations were referred to using their breed name and those sampled from North America (the U.S.A. and Canada, as described by Prystupa *et al.* (2012b)) were prefixed with the identifier "U.S." The Welsh Pony and Cob is subdivided into four distinct sections (A, B, C, and D) defined by the animal's adult height and conformation characteristics (Winton *et al.* 2013). U.S. Welsh Pony samples were not subclassified, so both U.K. sections A and B were included, as these ponies are more common in North America than the rare Section C. The Section D is a larger animal, often technically classed as a horse. DNA samples were collected as previously described for U.K. and Irish sources by Winton *et al.* (2013) and North America by Prystupa *et al.* (2012b). Mitochondrial sequences from 39 Kerry

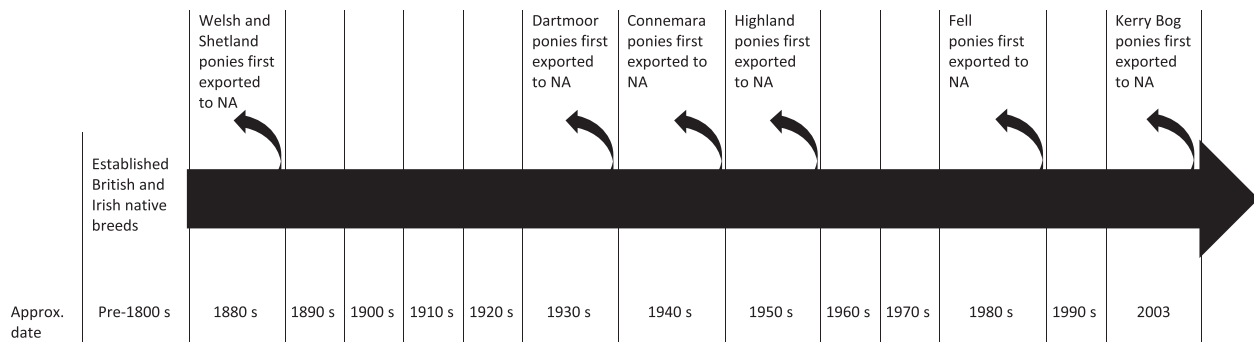


Figure 1. Timeline of breed export to North America (NA). Established British and Irish native breeds include the following: Welsh (sections A and B), Dartmoor (English), Fell (English), Highland (Scottish), Shetland (Scottish), Connemara (Irish), and Kerry Bog (Irish).

Bog Ponies sampled from Ireland (McGahern et al. 2006) were downloaded from GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>). Mitochondrial sequence data for the corresponding breeds of North American origin analyzed by Prystupa et al. (2012a) were downloaded from GenBank for analysis. A total of 410 animals were used for the mtDNA analysis and 472 for the microsatellite analysis. See Table 1 for population sample numbers. See Appendix 1 for origin of data and GenBank accession numbers.

Microsatellite genotyping

SSR genotyping of the samples were performed for 17 SSR loci using commercially available kits and protocols as described previously for the U.K., Winton et al.

(2013), and the U.S.A. for 38 markers including these 17, Prystupa et al. (2012b). PCR products were sized using a ABI 3730 or 3130xl Genetic Analyzer (Applied Biosystems, Paisley, UK) with the internal 600-Liz size standard. Genotypes were determined using Genemapper[®] Software v3.0 (Applied Biosystems).

Microsatellite analysis

To integrate U.S. and U.K. datasets, subsets of samples were exchanged between laboratories for genotyping and allele scoring. Datasets for the overlapping markers used were adjusted accordingly and standardized according to ISAG recommendations. Allele frequencies were examined for significant deviation ($P < 0.05$) from Hardy–Weinberg expectations using the exact test in Arlequin 3.5 (Excoffier

Table 1. Summary statistics for mtDNA and SSR genotyping for each U.S. and U.K. pony population.

Population	<i>N</i>	<i>H</i>	mtDNA		<i>N</i>	AR	SSR		
			<i>h</i> ± SD	π ± SD			<i>H_e</i>	<i>H_o</i>	<i>F_{IS}</i>
Welsh Section A	47	15	0.925 ± 0.016	0.0239 ± 0.0013	52	6.14	0.698	0.674	0.021
Welsh Section B	29	12	0.904 ± 0.028	0.0187 ± 0.0030	25	6.05	0.716	0.685	0.028
U.S. Welsh Pony	10	7	0.867 ± 0.107	0.0200 ± 0.0035	48	6.73	0.762	0.748	0.018
Fell	46	14	0.929 ± 0.014	0.0231 ± 0.0013	46	6.73	0.738	0.73	0.003
U.S. Fell	9	2	0.222 ± 0.166	0.0042 ± 0.003	25	5.25	0.683	0.641	0.055
Highland	46	15	0.916 ± 0.017	0.0178 ± 0.0012	40	5.93	0.707	0.722	−0.045
U.S. Highland	11	6	0.836 ± 0.089	0.0233 ± 0.0030	25	4.76	0.657	0.658	−0.006
Connemara	46	18	0.882 ± 0.032	0.0192 ± 0.0019	47	5.90	0.743	0.73	0.010
U.S. Connemara	12	6	0.818 ± 0.096	0.0243 ± 0.0040	37	5.93	0.751	0.778	−0.037
Dartmoor	40	10	0.814 ± 0.037	0.0122 ± 0.0020	39	6.08	0.694	0.685	0.003
U.S. Dartmoor	12	3	0.439 ± 0.158	0.0114 ± 0.0039	25	4.97	0.639	0.609	0.042
Shetland	39	8	0.767 ± 0.054	0.0225 ± 0.0026	35	5.60	0.664	0.657	0.005
U.S. Shetland	14	6	0.835 ± 0.057	0.0220 ± 0.0026	28	5.18	0.643	0.614	0.045
Kerry Bog Pony	39	17	0.934 ± 0.020	0.0210 ± 0.0017	NA	NA	NA	NA	NA
U.S. Kerry Bog Pony	10	5	0.800 ± 0.100	0.0151 ± 0.0028	NA	NA	NA	NA	NA

N, sample number; *H*, number of mtDNA haplotypes; *h*, mtDNA haplotype diversity; π , mtDNA nucleotide diversity; SD, standard deviation; AR, allelic richness based upon a minimum sample size of 23 individuals; *H_e*, expected heterozygosity; *H_o*, observed heterozygosity; and *F_{IS}*, inbreeding coefficient.

et al. 2005; Excoffier and Lischer 2010), with Markov chain length set to 100,000 following 10,000 dememorization steps. Markers displaying highly significant deviations in more than one population and/or containing large amounts of missing data values were removed to limit the influence of allele dropout. Loci with missing data unevenly distributed across the populations (>5% per breed) were also excluded resulting in seven loci being removed from the analysis, including two that had demonstrated null alleles in other studies (Rendo *et al.* 2012). Individual samples with poor-quality DNA with calls missing from >3 loci were also excluded.

An unbiased Bayesian approach using Markov chain Monte Carlo (MCMC) clustering of samples was conducted via the STRUCTURE v2.2.3 software (Pritchard *et al.* 2000). Parameters were set as diploid data for each individual and assessed for values of K ranging from 1 to 16. Burn-in and MCMC iteration settings were 50,000 and 100,000, respectively. Allele frequencies were treated as correlated. For each value of K , six replicate simulations were conducted. The K statistic (the second-order rate of change in log probability between successive values of K) was calculated using STRUCTURE Harvester v0.6.7 (http://taylor0.biology.ucla.edu/struct_harvest/) as per Evanno *et al.* (2005). Results from replicate runs at the optimal K were combined in CLUMPP (Jakobsson and Rosenberg 2007) and the average Q-table exported to DISTRUCT for graphical presentation (Rosenberg 2004). Population summary statistics, analyses to detect population expansion according to Harpending's raggedness index, pairwise F_{ST} comparisons, and analysis of molecular variance (AMOVA) were calculated using Arlequin 3.5 (Excoffier and Lischer 2010). To account for difference in sample size, allelic richness (AR) was calculated for each locus and population based on a minimum sample size of 23 diploid individuals using FSTAT v2.9.3.2 (Goudet 2001).

mtDNA genotyping

U.K. samples were sequenced for a 606-bp mtDNA fragment of the mitochondrial control region (MTCR as described previously; Winton *et al.* 2013). Sequence data were trimmed to a 212-bp sequence in order to align the sequence fragment shared with the U.S. samples (Prystupa *et al.* 2012a). Alignment was performed using BioEdit Sequence Alignment Editor v7.0.9.0 (Hall 1999) before being exported in phylip (.phy) format.

mtDNA analysis

A median-joining network was created to explore the relationship between haplotypes using NETWORK

(<http://www.fluxus-engineering.com>). In line with other studies, hypervariable nucleotide positions 15,585, 15,597, and 15,650 were removed (downweighted to 0) and nucleotide positions 15,604, 15,659, and 15,737 were downweighted (from 10 to 5). Nucleotide positions refer to that of the complete horse reference sequenced by Xu and Arnason (1994). The 83 reference sequences from Achilli *et al.* (2012) were downloaded to create a skeleton network of the globally defined horse haplogroups. Additionally, 208 archaeological samples (Cieslak *et al.* 2010) were obtained from GenBank to identify the distribution of ancestral diversity and its relation to modern populations. Sequence traces were checked for nucleotide call errors and either manually rectified or removed in the case of overall low-quality sequence.

Haplotype number, diversity, and nucleotide diversity were calculated for each population using DnaSP v5 (Librado and Rozas 2009). A haplotype list was generated for each dataset by DnaSP v5 based on the original sequences and used in Arlequin 3.5 (Excoffier and Lischer 2010) to calculate the average pairwise nucleotide differences between each population and to estimate number of migrants exchanged between populations. Mismatch distributions were also analyzed to detect evidence of population expansion.

To infer past population dynamics, such as changes in relative effective population size throughout time indicating evidence of genetic bottlenecks, Nexus files of aligned mtDNA sequences were prepared for each population and used to generate extended Bayesian Skyline plots by Bayesian Monte Carlo simulation of coalescent molecular trees using the phylogenetic inference software package BEAST (Bayesian Evolutionary Analysis Sampling Trees) (Drummond and Rambaut 2007; Heled and Drummond 2010; Drummond *et al.* 2012). Priors were set to the same values for all populations, with default parameter setting being used with the exception of the prior mutation rate, which was set to a constant of 0.05 mutations per site per million years. These plots provide relative comparisons of molecular evolutionary history dynamics between U.S. and U.K. breeds by estimating the effective number of females throughout a continuous time period based upon the mitochondrial molecular diversity within the current populations. As such, they are not absolute estimates of past population sizes and absolute chronological timings of events. This decision was taken because of the inferred potential for different accuracy in estimation of parameters from the discrepancy in available sample sizes from the two groups. To rule out a consistent bias of sample size in the estimated demographics, a random subset of samples with equivalent numbers of molecules was subsampled from the U.K. populations. These samples revealed that, in direct contrast to the U.S. samples,

estimates of predomestication population size were similar to those predicted using the larger U.K. population dataset, but with larger confidence intervals.

Results

Parameters of population genetic diversity

Summary statistics are presented in Table 1. Analysis of mitochondrial control region (MTCR) sequence from 410 individuals identified 46 haplotypes (DNASP v5) based on 34 variable nucleotide sites (212-bp sequence, DNASP v5). Overall haplotype diversity was high (0.949 ± 0.003) with high nucleotide diversity (0.0232 ± 0.0005). The U.S. Fell population possessed the lowest haplotype/nucleotide diversity of all populations and the lowest number of haplotypes (two). The Kerry Bog group demonstrated the highest haplotypes diversity (0.934) and the U.S. Connemara had the highest overall nucleotide diversity (0.0243), while the Connemara had the highest number of haplotypes (18). With the exception of Shetlands, U.S. populations showed lower mitochondrial diversity than U.K. equivalents, although both U.S. Connemara and U.S. Highland pony populations displayed greater nucleotide diversity values than their U.K. equivalents.

All U.S. populations had lower autosomal diversity than U.K. counterparts according to SSR allelic richness, observed heterozygosity (H_O), and inbreeding coefficients (F_{IS}), with the exception of the U.S. Welsh Pony and the U.S. Connemara (Table 1). The U.S. Welsh Pony demonstrated the highest values within the entire dataset for allelic richness and high H_O , while the U.S. Connemara

exhibited the highest H_O and low F_{IS} . For all parameters, the U.S. Dartmoor displayed evidence of the lowest genetic diversity.

Overall pony populations sampled from the U.S. showed lower levels of diversity than U.K. counterparts, particularly, the U.S. Fell and U.S. Dartmoor.

Population gene flow and structure

Significant genetic structure was found within the dataset ($P < 0.001$), with 6.91% of total variation existing between populations within U.S. and U.K. groups. There was no significant variance found between U.S. and U.K. breeds. Nonetheless, Bayesian analysis using STRUCTURE of the log likelihood ($\ln \Pr(X/K)$) of the posterior probability for a given K suggested the optimum value was $K = 2$. Delta K plots also showed a secondary peak for $K = 4$ clusters (Fig. 2A and B). At $K = 2$, clusters were divided according to geographic sample origin (U.S. and U.K.), with the exception of the Connemara. Where $K = 4$ clusters of sample origin were also evident, but the Fell and Shetland populations cluster together irrespective of geographical origin. Pairwise F_{ST} comparisons were performed based on the sum of squared allele size differences (R_{ST}) and substantiated the relationship between populations of U.K. and U.S. origin (Table 2). Breeds did not show significant continental divergence except for Dartmoors, as is reflected in high estimates of N_m (migrants) between populations of the same breed. Of particular note was the Connemara with very low divergence between populations and a correlated N_m of 105.32. Similar to the summary statistics (Table 1), Theta

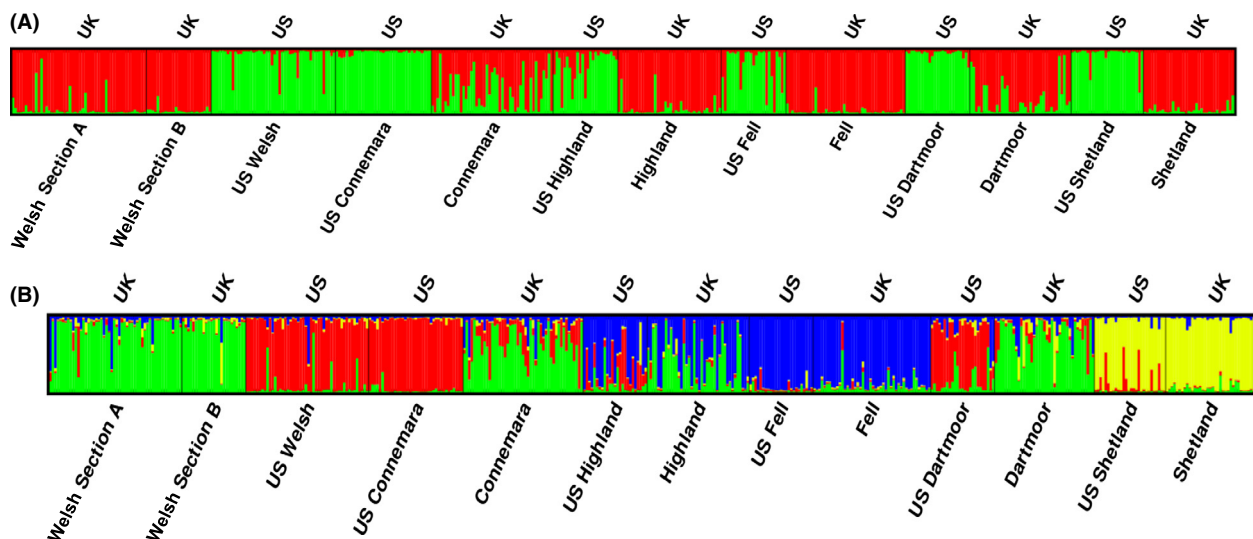


Figure 2. The results of STRUCTURE analysis for 10 SSR loci of each population for (A) $K = 2$ and (B) $K = 4$, averaged over six independent runs. STRUCTURE Harvester analysis suggests true value of K was 2 and 4.

Table 2. Pairwise F_{ST} comparisons based on sum of squared size differences (R_{ST}) and effective number of migrants for U.S. and U.K. pony populations

	Welsh Sect. A	Welsh Sect. B	U.S. Welsh	Fell	U.S. Fell	High.	U.S. High.	Conn.	U.S. Conn.	Dart.	U.S. Dart.	Shet.	U.S. Shet.
Welsh Sect. A	1.87	15.84	44.76	9.2	1.765	2.6	3.045	4.09	8.47	3.02	3.315	4.745	7.41
Welsh Sect. B	0.016	1.96	12.01	10.58	1.655	2.36	2.02	8.13	11.215	6.775	5.31	5.355	6.29
U.S. Welsh	0.006	0.02	2.25	0.037	1.505	0.099	3.05	0.069	0.038	0.101	3.15	0.077	4.8
Fell	0.026	0.023	6.535	2.08	13.26	5.755	13.9	7.68	11.82	3.83	4.29	3.29	3.34
U.S. Fell	0.124	0.131	0.142	0.019	1.81	0.081	3.67	0.089	0.081	0.139	1.13	0.153	1.195
High.	0.088	0.096	2.28	0.042	2.825	1.92	20.07	4.74	3.865	3.42	1.49	1.37	1.19
U.S. High.	0.076	0.11	0.076	0.018	0.064	0.012	1.72	0.069	0.056	0.119	0.17	0.157	1.46
Conn.	0.058	0.03	3.375	0.032	2.565	0.05	3.395	2.12	105.32	10.64	1.7	3.71	2.515
U.S. Conn.	0.029	0.022	6.27	0.021	2.855	0.061	4.255	0.002	2.18	0.028	2.46	0.044	4.725
Dart.	0.076	0.036	7.235	0.061	1.55	0.068	1.85	0.023	8.565	1.85	2.44	3.17	2.16
U.S. Dart.	0.07	0.045	0.074	0.055	0.181	0.144	1.225	0.128	0.092	0.093	1.67	0.152	1.725
Shet.	0.05	0.045	3.005	0.071	1.39	0.155	1.345	0.063	5.485	0.073	1.4	1.74	14.28
U.S. Shet.	0.033	0.038	0.05	0.07	0.173	0.173	0.146	0.09	0.05	0.104	0.127	0.017	1.68

Below diagonal: Pairwise F_{ST} based on sum of squared size differences (R_{ST}). Above diagonal: effective number of migrants (N_m) per generation, based on Slatkin, 1995. Diagonal elements: Theta H estimates $4N_e\mu = (1/(1-H_e)^2) - 1$. Abbreviated populations are defined as follows: Welsh Section A, Welsh Section B; U.S. Welsh Pony, U.S. Welsh U.S.; Highland, High.; U.S. Highland, U.S. High.; Connemara, Conn.; U.S. Connemara, U.S. Conn.; Dartmoor, Dart.; U.S. Dartmoor, U.S. Dart.; Shetland, Shet.; U.S. Shetland, U.S. Shet. All values are significantly different ($P < 0.05$) with the exception of those in cells shaded grey.

H ($4N_e\mu$) values for autosomal data (Table 2) were reduced in most U.S. populations compared to U.K. counterparts indicating smaller effective population sizes, except for the U.S. Connemara and U.S. Welsh.

Maternal-specific data are broadly consistent (Table 3), indicating more divergence among U.S. than U.K. populations (average $N_m = 4.197$ [$n = 21$] vs. $N_m = 7.972$ [$n = 28$]). Excepting the Fells and Dartmoors, transcontinental populations of the same breed did not show significant divergence in maternal genotypes as indicated by pairwise comparisons and estimated N_m (Table 3). However, the U.S. Welsh Pony, U.S. Connemara, and U.S. Highland exhibited particularly low divergence from other breeds of the U.S. and the U.K. (Table 3). Indeed, the U.S. Welsh Pony matrilineal lines show greater affinity with four U.K. breeds (Section A, Highland, Connemara and Dartmoor) than to the equivalent U.S. breeds, or indeed to that which the U.K. Section A shows to the other three U.K. breeds. U.S. Welsh Ponies showed the least distinction from, and highest number of migrants shared with, the Section B population. In contrast, the U.S. Fell pony was maternally significantly diverged from every other breed analyzed, including the U.K. Fell population. Possible admixture in some U.S. breeds, namely the U.S. Highland and U.S. Connemara, is indicated. For example, the U.S. Connemara population showed twice as many migrants exchanged with the U.K. Section A than with the equivalent U.K. Connemara.

The median-joining network (Fig. 3) displays the relationships of the maternal lines of the British, Irish, and

U.S. pony populations according to shared mitochondrial haplotype frequencies and using ancestral fossil samples to “root” the network. The U.S. Fell pony is noteworthy due to the overrepresentation of a single haplogroup: 89% of the U.S. Fells belong to haplogroup B, with only a single individual containing a haplotype found within another haplogroup (Figs. 3, 4). For some U.S. breeds, the haplogroup frequencies compared to the U.K. equivalent were similar (e.g., haplogroup L makes up 57% and 55% of the U.K. and U.S. Highland ponies respectively), but others show discrepancies (haplogroup A is present in 75% of the U.S., but only 5% of the U.K. Dartmoors, Fig. 3). In the more extreme cases, the “daughter” population contains haplogroups lacking from the U.K. source. For example, 36% of U.S. Highland samples contained the Q haplogroup that is absent from the U.K. Highlands, and similarly the haplogroup G sequences in U.S. Connemaras and haplogroup N in the U.S. Shetlands. Nonetheless, the general pattern is one of fewer haplogroups in the U.S. populations relative to the corresponding U.K. breed.

Skyline plots generated using mitochondrial sequence data illustrated that the effective population size of the U.S. breeds was lower than that of U.K. counterparts at a given time-point (x -axis) demonstrating the associated reduction in ancestral diversity as a result of a recent bottleneck in U.S. populations (Fig. 5 and Appendix 2). All populations appear to have undergone expansion, although for U.K. populations, this is estimated to have started substantially earlier than U.S. equivalents. Welsh populations did not show the same pattern as other

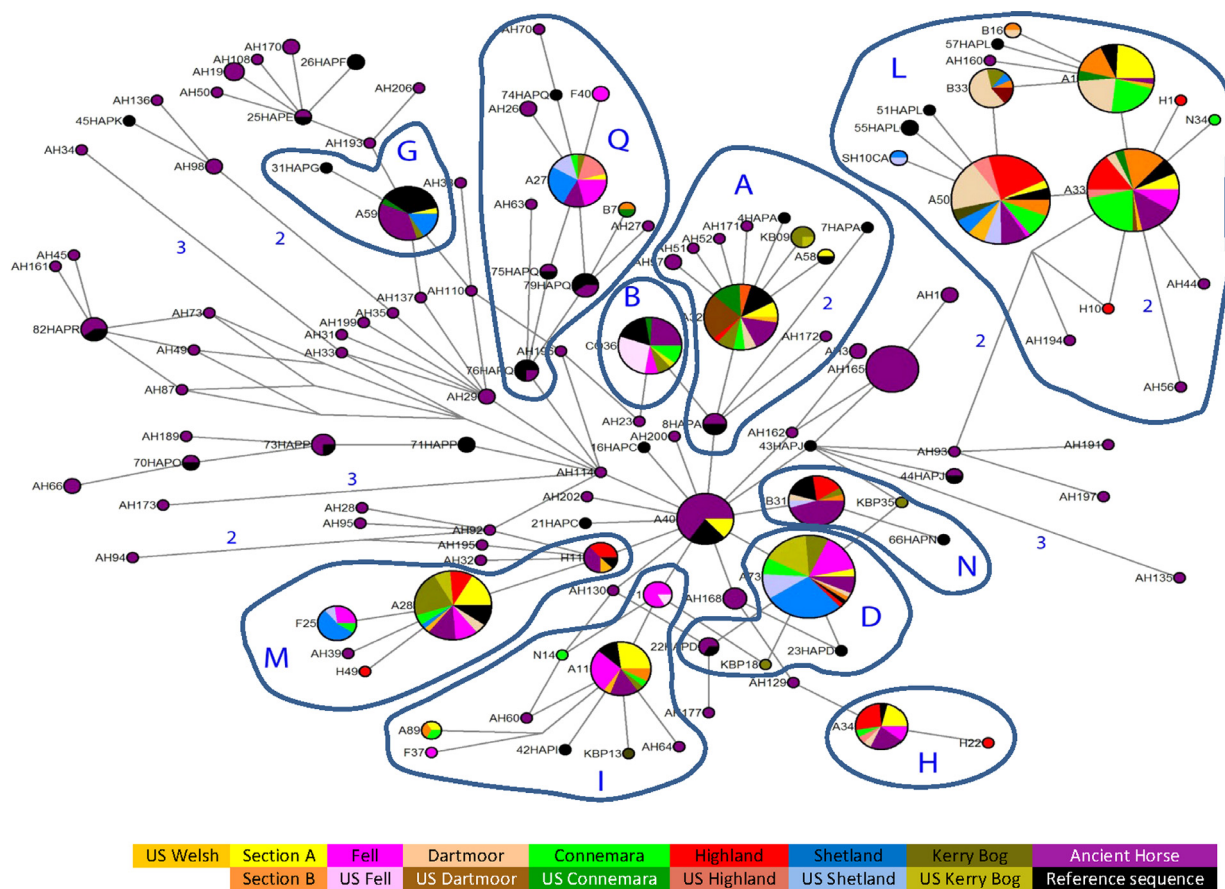


Figure 3. A median-joining network for mtDNA haplotypes displaying the relationships of the maternal lines of the British, Irish, and U.S. pony populations according to shared mitochondrial haplotype frequencies. Node size represents overall haplotype frequency with pie charts within nodes showing frequency of that haplotype by population. Reference samples representing examples of each of the major haplogroups identified by (Achilli et al. 2012) are displayed in black and labeled according to sample number and haplogroup, for example, “1HapA” Numerical values indicate the number of nucleotide changes (>1 mutation) between primary nodes. The haplotype designated as A40 is placed most ancestrally relative to the entire dataset and contains ancestral samples.

groups, whereby the Section A showed a decline in effective population size and lower values than the U.S. Welsh population. Harpending’s raggedness index test of the mitochondrial data for population expansion gave low raggedness index, indicating a unimodal distribution (Appendix 3). This generally supported the Skyline plots as unimodal distribution is found in populations having undergone recent population expansion. However, a significantly high ($P < 0.05$) raggedness index was found for the Welsh Section A, U.S. Connemara, Shetland, and U.S. Shetland population, which indicated that they existed in a multimodal distribution consistent with a stationary population for some time.

Discussion

Establishing secondary ex situ populations to protect genetic resources is a common management tool in the

conservation of a range of species (Fraser 2008; Ozer et al. 2011; Ransler et al. 2011; El Alqamy et al. 2012). For domestic livestock species, approximately 1080 breeds exist as “transboundary” populations, located internationally (DAD-IS, 2006). As an exemplar of whether a daughter population maintains the genetic diversity of its putative source once ex situ, our data suggest that pony populations sampled from the U.S. have lower levels of diversity than U.K. counterparts. However, specific exceptions exist for either mtDNA or SSR markers within certain breeds. The U.S. Fell and the U.S. Dartmoor have considerably reduced diversity for both mitochondrial and autosomal markers and seem to be descended from a low number of sires and very limited imported mares. Mitochondrial haplotype and nucleotide diversity within most of the other U.S. breeds was only slightly reduced compared to the U.K. populations. Based solely on these values, it may appear that most of the U.S. subpopulations

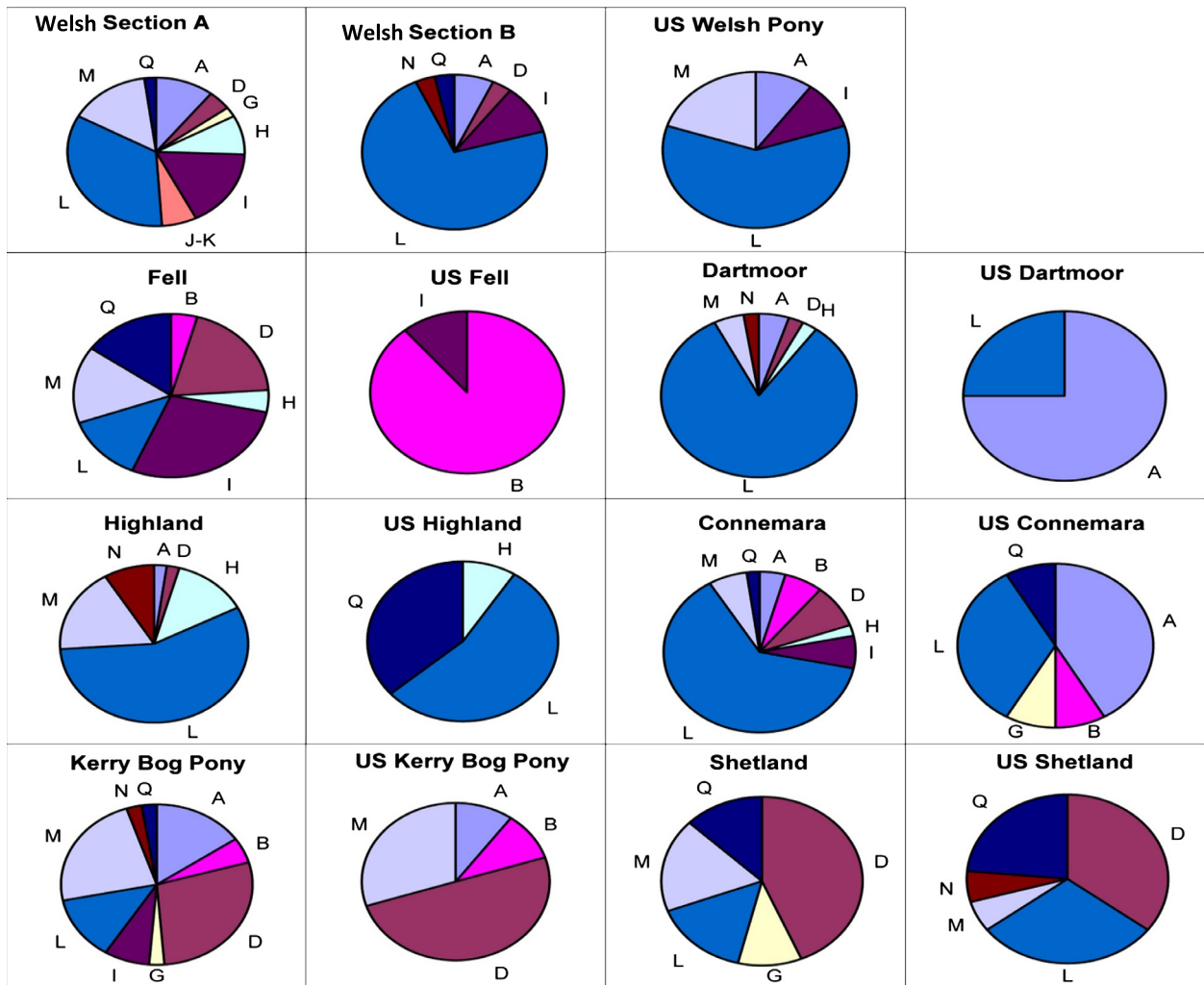


Figure 4. Haplogroup frequency (%) according to breed. Segments represent proportion of individuals belonging to labeled haplogroup. The first row identifiers “A” to “Q” represent the major worldwide mtDNA haplogroups in horses: haplogroup classification is based upon control region motifs, as described by (Achilli et al. 2012).

have maintained much of the original maternal diversity of the U.K. populations they are derived from. However, closer examination of the mitochondrial haplotype distribution and population pairwise differences reveals discrepancies.

Certain haplogroups were present in the U.S. population, but absent in the U.K. equivalent for Highlands, Connemaras, and Shetlands, and although mtDNA and SSR diversities between U.S. breeds indicate more divergence than between the U.K. breeds, as expected for populations undergoing drift, there are some increases in lineage sharing between particular populations. In particular, the Welsh Section A, U.S. Connemara, U.S. Highland, and U.S. Welsh share more haplotypes than expected from other relationships possibly due to admix-

ture. One interpretation of the intermediate probability of allocating Connemara animals to the U.S. or U.K. groupings in STRUCTURE plots is that this breed or the local derivatives have been crossed with other breeds in North America and to a lesser extent in the U.K. This may have occurred perhaps even before the respective stud books were closed (a closed stud book is one that prohibits registering animals whose parents were not themselves registered). Autosomal divergence between the U.K. and U.S. populations suggests that stochastic drift has occurred between source and daughter populations. The clustering according to breed rather than country of sample origin, evidenced by STRUCTURE analysis and pairwise F_{ST} comparisons (based on R_{ST} distance analysis), excludes this being the result of from laboratory-based genotyping

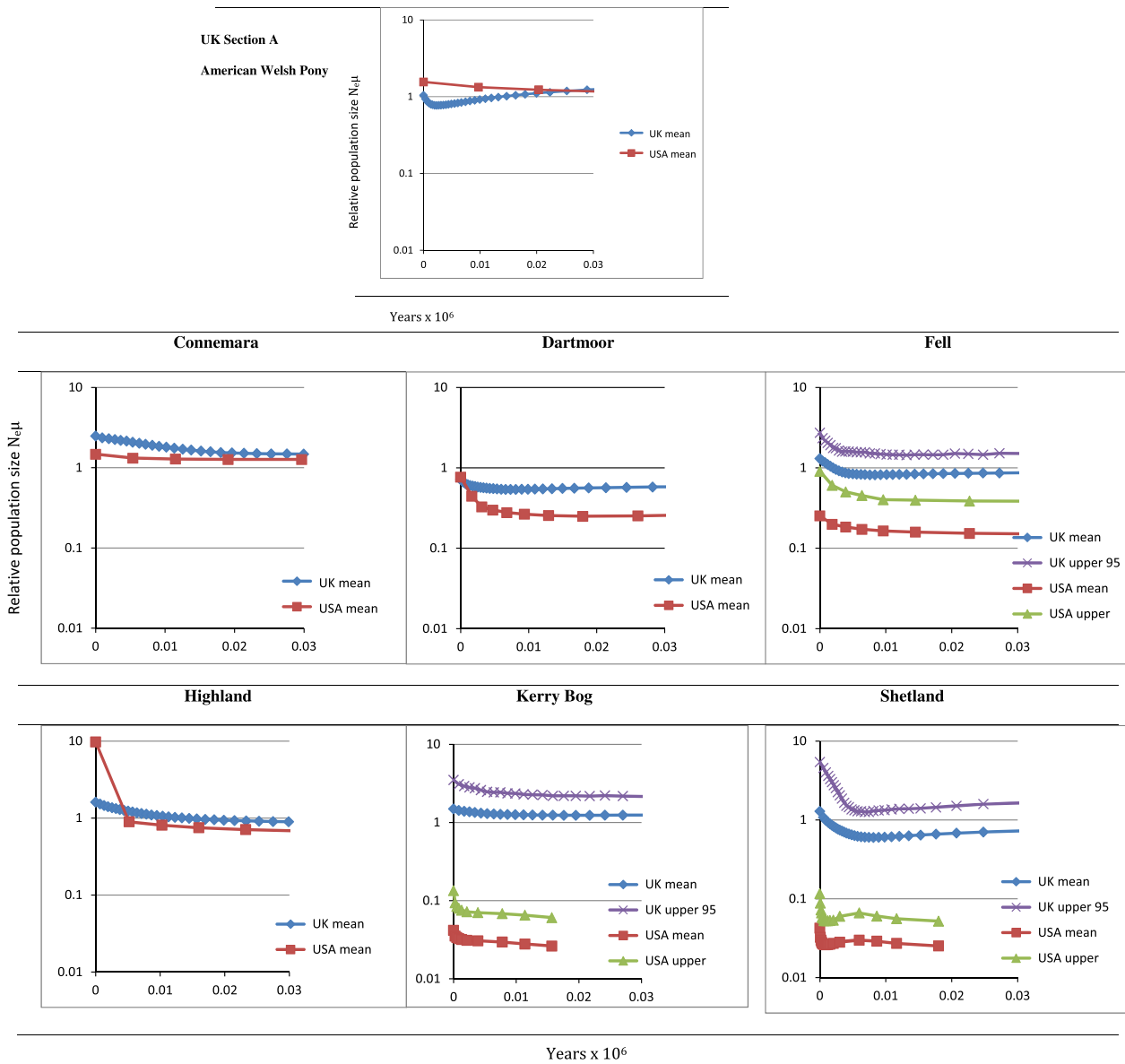


Figure 5. Extended Bayesian Skyline plots of estimated past population dynamics by Bayesian Monte Carlo simulation of coalescent molecular trees based on mitochondrial DNA data. The y-axis indicates the effective relative number of females, while the x-axis is years ($\times 10^6$) before present. The plot lines are as follows: red, median estimate; blue, mean estimate; green, lower 95% limits; and purple, upper 95% limits. Note differing axis assigned to U.S. Welsh pony.

discrepancies. Divergence between captive or wild sub-populations and source populations has been demonstrated for other species, where increasing gene flow or multiple restoration/reintroduction sources can limit the effects (Biebach and Keller 2009; Ozer *et al.* 2011; Wilson *et al.* 2012).

The Welsh breeds were an exception to the typical shared characteristics of an origin–daughter relationship observed within this study. It was not possible to identify an individual Welsh Section as the founder of the U.S.

Welsh pony. The U.S. Welsh consisted of mitochondrial signatures that were shared by both U.K. sections As and Bs, yet the association with the Section Bs was most pronounced. This female amalgamation was also reflected in the Skyline plots whereby, unlike other U.S. populations, the U.S. Welsh pony showed an overall increase and a higher effective population size at the same relative time-point compared to the Section A. This is consistent with the U.S. Welsh pony sample used here having originated from a “mixed” population of Welsh sections A and B

ancestors. As such, the U.S. Welsh population captures ancestral mitochondrial diversity from a larger population of origin than the U.K. Section A pony does alone. In effect, the bottleneck caused by breed formation was more recent in the Welsh Section As compared to the “mixed” origin of the U.S. Welsh. While cross-breeding and amalgamation seems to have occurred, particularly in the formation of the U.S. Welsh Pony, this contrasts with estimates of N_m for maternal data indicating American populations are more divergent from each other. This is therefore likely evidence of two processes. In one, founder and/or drift effects resulting in allele frequency differences (including loss of rare alleles) have driven populations in the U.S. apart and given lower estimates for previous population sizes due to loss of haplotypes passing through the bottleneck. In the other, cross-breeding or use of a common external breed (not included in the U.K. populations analyzed) as a crossing sire/dam in the U.S. has brought some of them together, but to a lesser degree than the stochastic changes that have caused the evident divergence.

As an example of this, SSR analyses demonstrated closer relationships between the U.S. Welsh and the Section A, with nearly four times the number of migrants exchanged than with the Section B, in contrast to the mtDNA patterns. It thus seems likely that U.S. Welsh ponies have a different maternal and paternal history; seemingly a product of mares from at least both sections of Welsh ponies from the U.K., while stallion introductions to the U.S. have been largely Section A. The diversity seen in the U.S. Welsh ponies today is therefore a consequence of original breed formation and subsequent repeated “migration” from one subpopulation of the source group. This highlights the importance for conservation strategies more generally to appreciate the degree of genetic variability and the level of gene flow among each subpopulation (Ciofi and Bruford 1999).

Skyline plots in the U.S. populations tend to be dominated by evidence of recent bottleneck effects, as only a limited signal of previous population changes can survive through the bottleneck to the current population. This is illustrated by mitochondrial haplogroup I, predominantly found in Section A and Fell ponies, but in only a single sample from each respective U.S. population. The branching star-shaped pattern of haplogroup I indicates these lineages may have existed within the earliest ancestral ponies of Britain and Ireland and is strongly characteristic of some of the more isolated populations (Winton *et al.* 2013). The “A40” node (Fig. 2) placed most ancestrally relative to the entire dataset is restricted to a substantial number of fossil samples and the Section A ponies. Similarly, the D haplogroup is rare worldwide (Achilli *et al.* 2012) and has been considered an old clade characteristic

of the small ponies distributed on the western fringe of Europe (McGahern *et al.* 2006; Kakoï *et al.* 2007; Cieslak *et al.* 2010; Prystupa *et al.* 2012a) as confirmed in our U.K. populations. However, other than the recently imported U.S. Kerry Bog Pony and the U.S. Shetland, this rare haplogroup has been lost from the U.S. samples.

The recently characterized Kerry Bog is a prime example of the importance of consistency between conservation management organizations or breed societies. Our results indicate more restricted matrilineal and lower diversity values in the U.S., but with no major divergence from the maternal source population. However, the recently established American Kerry Bog Pony Society accepts pied coloration not found in Ireland (Anon, 2013a); this divergence from the original breed indicating admixture is reflected in the lack of breed-specific clustering according to SSR analysis by Prystupa *et al.* (Prystupa *et al.* 2012b). While no SSR data currently exist for the U.K. Kerry Bog population, the introduction of pied-color animals will inherently change the genetic signature, causing greater divergence from the source population.

While statistically significant comparisons cannot be made for individual Skyline plots between the U.K. and U.S. populations, together they provide a comparative story of reduced population size and recent bottlenecks in the U.S. breeds. As expected, the results in Figure 5 broadly indicate more recent establishment of U.S. populations and lower levels of subsequent importation resulting in larger discrepancies between predicted population sizes of the U.K. and U.S. breeds. The major exceptions are the U.S. Welsh and Shetland breeds. The U.S. Fell and U.S. Dartmoor pony populations consistently demonstrated major reductions in diversity compared to the original U.K. populations for both autosomal and mitochondrial analysis. Not only is there a marked decrease in relative diversity values, but also there is the significant overrepresentation of U.S. individuals within mitochondrial haplogroups that are rare within the respective U.K. populations, suggestive of severe population size restriction and consequent stochastic changes including the loss of ancestral matrilineal (Lawler *et al.* 1995; Luis *et al.* 2006; Alvarez *et al.* 2012). Although this has resulted in significant mitochondrial divergence from the original U.K. populations, SSR analyses demonstrate a maintained link between autosomal allele frequencies among U.K. and U.S. Fells. Anecdotal records suggest that Dartmoor ponies, first introduced to the U.S.A. in the 1930s (Anon, 2013b,c), may have been maintained as a “pure” breed by mares derived from a single stud within the U.S., hence the divergence in autosomal and maternal markers. However, apparently extensive cross-breeding also occurred following U.S. import prior to the establishment of the Dartmoor Pony Registry of America in 1956;

this may have occurred with local U.S. ponies, contributed further to allelic frequency divergence between the U.S. and U.K. groups (Anon, 2013b). The U.K. Dartmoor has also suffered reduced diversity as a result of historic demographic restrictions. As such, it is important that the ex situ population of this breed in particular is carefully managed to maintain maximal representation of the ancestral diversity, using methods such as avoidance of mate choice limitations and sex-biased skewing in reproductive success (Goyache *et al.* 2003; Eales *et al.* 2008; Conard *et al.* 2010; Brekke *et al.* 2011). The effects of various herd management strategies and the resulting variance in gene flow and admixture, applied here to different breeds, have resulted in a range of genetic signatures within the U.S. and U.K. subpopulations and indicate that recurrent importation has maintained genetic characteristics in some pony breeds, for example, the U.K./U.S. Connemara but has been less effective in others, such as the Fells.

Conclusion

By studying ex situ subpopulations and putative source populations through the analysis of genetic diversity of ponies, we have established it is possible to identify different genetic characteristics that are compatible with known histories of those populations. We detected clear indications of founder effect and stochastic drift in most daughter populations, more pronounced in some breeds than others, and evidence of admixture of imported mares of different breeds within the U.S.A. To conserve a valuable resource and maintain sufficient adaptive genetic variability, it is important that derived populations remain representative of the original source once ex situ and that enough founder lines are present within the original or subsequently exported animals to maintain diversity levels and avoid extreme stochastic drift. It is imperative that the objectives of daughter breed societies/management groups are clear and consistent from the outset. We demonstrate that genetic analysis of subpopulations can assist in understanding the genetic effects of past management practices and inform future conservation strategies even in the absence of recorded histories.

Acknowledgments

We are grateful for funding provided by an IBERS PhD Scholarship (Aberystwyth University) and a Royal Society Research Grant 2010/R2, which supported Clare Winton.

Data Accessibility

Mitochondrial sequence data generated as part of this project are stored in FASTA format and will be uploaded

to the NCBI/GenBank nucleotide sequence repository. Other mtDNA sequences incorporated into the analysis were downloaded from this source and can be retrieved as per the relevant citations. Breed-specific Nexus files were created for use in the Skyline analysis and will be made available via the DRYAD data repository. Processed SSR data will be also be uploaded to the DRYAD data repository as a tabular spreadsheet of allele scores with individual/breed/marker identifiers. This data was converted into several formats for use in packages such as STRUCTURE; should access to each individual format be deemed necessary, they will also be stored on the DRYAD repository. Raw SSR data (chromatogram peaks) are maintained in.fsa format and can be made available on request.

Conflict of Interest

None declared.

References

- Achilli, A., A. Olivieri, P. Soares, H. Lancioni, B. Hooshiar Kashani, U. A. Perego, *et al.* 2012. Mitochondrial genomes from modern horses reveal the major haplogroups that underwent domestication. *Proc. Natl Acad. Sci. USA* 109:2449–2454.
- Alvarez, I., I. Fernandez, L. Lorenzo, L. Payeras, M. Cuervo, and F. Goyache. 2012. Founder and present maternal diversity in two endangered Spanish horse breeds assessed via pedigree and mitochondrial DNA information. *J. Anim. Breed. Genet.* 129:271–279.
- Anon. 2013a. <http://www.kerrybogpony.org/KerryBogPony.html>.
- Anon. 2013b. <http://www.dartmoorpony.com/>.
- Anon. 2013c. <http://www.albc-usa.org/cpl/dartmoor.html>.
- Biebach, I., and L. Keller. 2009. A strong genetic footprint of the re-introduction history of Alpine ibex (*Capra ibex ibex*). *Mol. Ecol.* 18:5046–5058.
- Brekke, P., P. Bennett, A. Santure, and J. Ewen. 2011. High genetic diversity in the remnant island population of hihi and the genetic consequences of re-introduction. *Mol. Ecol.* 20:29–45.
- Cai, D., Z. Tang, L. Han, C. F. Speller, D. Y. Yang, X. Ma, *et al.* 2009. Ancient DNA provides new insights into the origin of the Chinese domestic horse. *J. Archaeol. Sci.* 36:835–842.
- Cieslak, M., M. Pruvost, N. Benecke, M. Hofreiter, A. Morales, M. Reissmann, *et al.* 2010. Origin and history of mitochondrial DNA lineages in domestic horses. *PLoS ONE* 5:e15311.
- Ciofi, C., and M. W. Bruford. 1999. Genetic structure and gene flow among Komodo dragon populations inferred by microsatellite loci analysis. *Mol. Ecol.* 12(Suppl. 1):S17–S30.

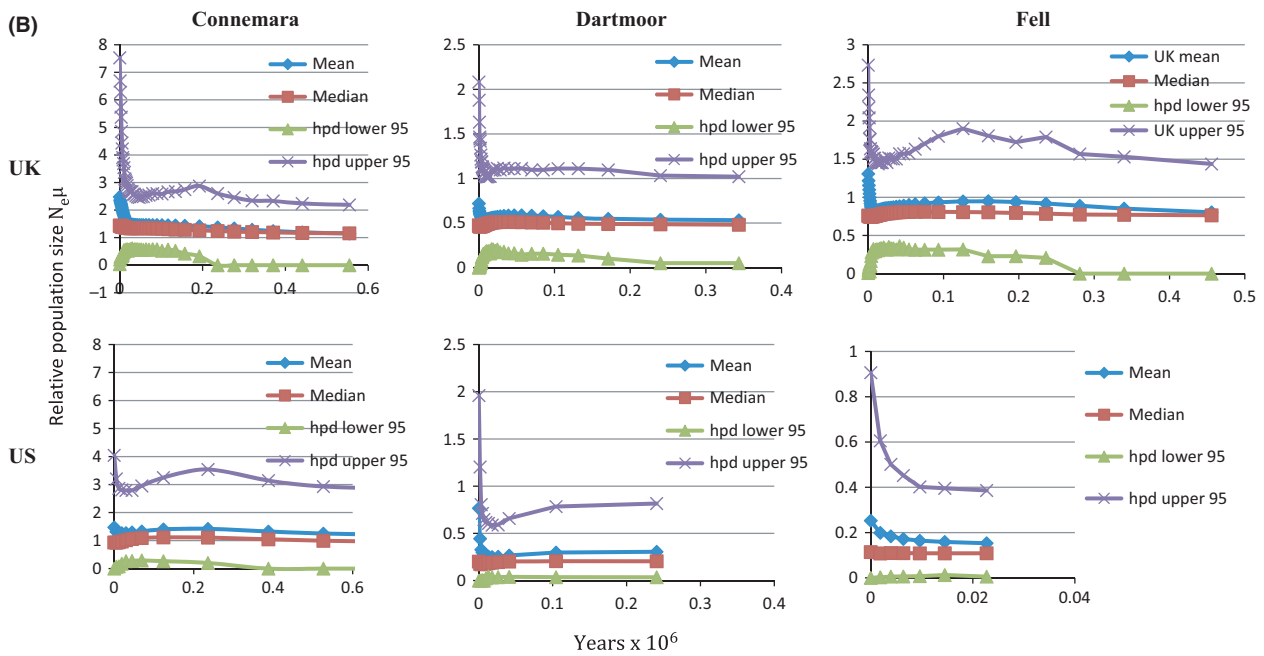
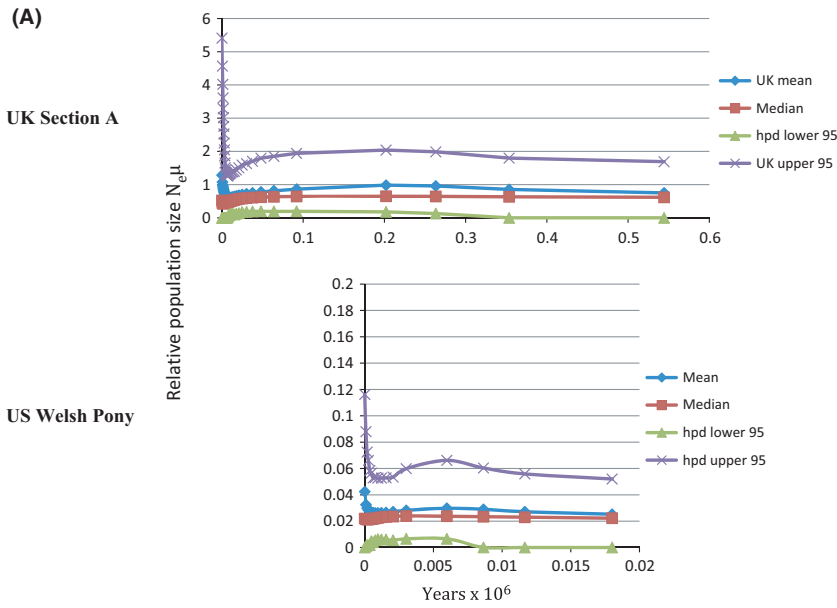
- Conard, J., M. Statham, P. Gipson, and S. Wisely. 2010. The influence of translocation strategy and management practices on the genetic variability of a re-established elk (*cervus elaphus*) population. *Restor. Ecol.* 18:85–93.
- DAD-IS. 2006. *Domestic Animal Diversity Information System (DAD-IS)*. FAO, USA. Available at: www.fao.org/dad-is/.
- Davies, W. 1993. The Welsh Mountain pony. Pp. 39–49. J.A. Allen and Company Limited, London, U.K.
- Drummond, A. J., and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214.
- Drummond, A. J., M. A. Suchard, D. Xie, and A. Rambaut. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29:1969–1973.
- Eales, J., R. Thorpe, and A. Malhotra. 2008. Weak founder effect signal in a recent introduction of Caribbean *Anolis*. *Mol. Ecol.* 17:1416–1426.
- El Alqamy, H., H. Senn, M. Roberts, R. McEwing, and R. Ogden. 2012. Genetic assessment of the Arabian oryx founder population in the Emirate of Abu Dhabi, UAE: an example of evaluating unmanaged captive stocks for reintroduction. *Conserv. Genet.* 13:79–88.
- Ensminger, M., and O. Parker. 1986. *Sheep and goat science*, 5th edn, 32 pp. Interstate Printers & Publishers, Madison.
- Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14:2611–2620.
- Evans, J. 2001. *Horses: a guide to selection, care and enjoyment*, 3rd edn. Pp. 142–146. WH Freeman and Co., New York.
- Excoffier, L., and H. E. L. Lischer. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Res.* 10:564–567.
- Excoffier, L., G. Laval, and S. Schneider. 2005. Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evol. Bioinform. Online* 1:47–50.
- Food and Agriculture Organization. 2007. *The state of the world's animal genetic resources for food and agriculture*. Communication Division – FAO, Rome. www.fao.org.
- Frankam, R., J. Ballou, and D. Briscoe. 2002. *Introduction to conservation genetics*. Cambridge University Press, Cambridge, U.K.
- Fraser, D. 2008. How well can captive breeding programs conserve biodiversity? A review of salmonids. *Evol. Appl.* 1:535–586.
- Glazewska, I., A. Wysocka, B. Gralak, R. Purs, and J. Sell. 2007. A new view on dam lines in Polish Arabian horses based on mtDNA analysis. *Genet. Sel. Evol.* 39:609–619.
- Goudet, J. 2001. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from <http://www2.unil.ch/popgen/softwares/fstat.htm>.
- Goyache, F., J. Gutiérrez, I. Fernandez, E. Gomez, I. Alvarez, J. Díez, et al. 2003. Using pedigree information to monitor genetic variability of endangered populations: the Xalda sheep breed of Asturias as an example. *J. Anim. Breed. Genet.* 120:95–105.
- Hall, T. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41:95–98.
- Heled, J., and A. J. Drummond. 2010. Bayesian inference of species trees from multilocus data. *Mol. Biol. Evol.* 27:570–580.
- Hill, E., D. Bradley, M. E. Al-Barody, R. Splan, I. Zakharov, and E. Cunningham. 2002. History and integrity of thoroughbred dam lines revealed in equine mtDNA variation. *Anim. Genet.* 33:287–294.
- Hodson, A. 1997. *The Shetland pony*. Pp. 105–107. JA Allen & Co. Ltd., London, U.K.
- Jakobsson, M., and N. A. Rosenberg. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23:1801–1806.
- Jansen, T., P. Forster, M. Levine, H. Oelkes, M. Hurles, C. Renfrew, et al. 2002. Mitochondrial DNA and the origins of the domestic horse. *Proc. Natl Acad. Sci. USA* 99:10905–10910.
- Kakoi, H., T. Tozaki, and H. Gawahara. 2007. Molecular analysis using mitochondrial DNA and microsatellites to infer the formation process of Japanese native horse populations. *Biochem. Genet.* 45:375–395.
- Lawler, S., R. Sussma, and L. Taylor. 1995. Mitochondrial DNA of the Mauritian macaques (*Macaca fascicularis*): an example of founder effect. *Am. J. Phys. Anthropol.* 96:133–141.
- Leroy, G., L. Callède, E. Verrier, J. C. Mériaux, A. Ricard, C. Danchin-Burge, et al. 2009. Genetic diversity of a large set of horse breeds raised in France assessed by microsatellite polymorphism. *Genet. Sel. Evol.* 41:5.
- Librado, P., and J. Rozas. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452.
- Lippold, S., N. Matzke, M. Reissmann, and M. Hofreiter. 2011. Whole mitochondrial genome sequencing of domestic horses reveals incorporation of extensive wild horse diversity during domestication. *Evol. Biol.* 11:328–348.
- Luis, C., C. Bastos-Silveira, J. Costa-Ferreira, E. Cothran, and M. Oom. 2006. A lost Sorraia maternal lineage found in the Lusitano horse breed. *J. Anim. Breed. Genet.* 123:399–402.
- Luis, C., R. Juras, M. Oom, and E. Cothran. 2007. Genetic diversity and relationships of Portuguese and other horse breeds based on protein and microsatellite loci variation. *Anim. Genet.* 38:20–27.
- McGahern, A., C. Edwards, M. Bower, A. Heffernan, S. D. Park, P. O. Brophy, et al. 2006. Mitochondrial DNA sequence diversity in extant Irish horse populations and in ancient horses. *Anim. Genet.* 37:498–502.
- Miller, K., N. Nelson, H. Smith, and J. Moore. 2009. How do reproductive skew and founder group size affect genetic

- diversity in reintroduced populations? *Mol. Ecol.* 18:3792–3802.
- Ozer, F., H. Gellerman, and M. Ashley. 2011. Genetic impacts of Anacapa deer mice reintroductions following rat eradication. *Mol. Ecol.* 20:3525–3539.
- Pritchard, J., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
- Prystupa, J., P. Hind, E. Cothran, and Y. Plante. 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, J., R. Juras, E. Cothran, F. Buchanan, and Y. Plante. 2012b. Genetic diversity and admixture among Canadian, Mountain and Moorland and Nordic pony populations. *Animal* 6:19–30.
- Ransler, F., T. Quinn, and S. Oyler-McCance. 2011. Genetic consequences of trumpeter swan (*Cygnus buccinator*) reintroductions. *Conserv. Genet.* 12:257–268.
- Rendo, F., M. Iriondo, C. Manzano, and A. Estomba. 2012. Effects of a 10-year conservation programme on the genetic diversity of the Pottoka pony – new clues regarding their origin. *J. Anim. Breed. Genet.* 129:234–243.
- Rosenberg, N. A. 2004. DISTRUCT: a program for the graphical display of population structure. *Mol. Ecol. Notes* 4:137–138.
- Royo, L., I. Alvarez, A. Beja-Pereira, A. Molina, I. Fernandez, J. Jordana, et al. 2005. The origins of Iberian horses assessed via mitochondrial DNA. *J. Hered.* 96:663–669.
- Russell, V. 1996. Shetland ponies. Pp. 194–196. Whittet Books Ltd., London, U.K.
- Shen, F., Z. Zhang, W. He, B. Yue, A. Zhang, L. Zhang, et al. 2009. Microsatellite variability reveals the necessity for genetic input from wild giant pandas (*Ailuropoda melanoleuca*) into the captive population. *Mol. Ecol.* 18:1061–1070.
- Solis, A., B. Jugo, J. Mériaux, M. Iriondo, L. Mazón, A. Aguirre, et al. 2005. Genetic diversity within and among four south European native horse breeds based on microsatellite DNA analysis: implications for conservation. *J. Hered.* 96:670–678.
- Van de Goor, L., W. van Haeringen, and J. Lenstra. 2011. Population studies of 17 equine STR for forensic and phylogenetic analysis. *Anim. Genet.* 42:627–633.
- Vila, C., J. Leonard, A. Gotherstrom, S. Marklund, K. Sandberg, K. Liden, et al. 2001. Widespread origins of domestic horse lineages. *Science* 291:474–477.
- Vonholdt, B., D. Stahler, E. Bangs, D. Smith, M. Jimenez, M. Mack, et al. 2010. A novel assessment of population structure and gene flow in grey wolf populations of the Northern Rocky Mountains of the United States. *Mol. Ecol.* 19:4412–4427.
- Wilson, C. D., G. E. Beatty, C. R. Bradley, H. C. Clarke, S. J. Preston, D. Roberts, et al. 2012. The importance of population genetic information in formulating ex situ conservation strategies for the freshwater pearl mussel (*Margaritifera margaritifera* L.) in Northern Ireland. *Anim. Conserv.* 15:593–602.
- Winton, C. L., M. J. Hegarty, R. McMahon, G. T. Slavov, N. R. McEwan, M. C. G. Davies Morel, et al. 2013. Genetic diversity and phylogenetic analysis of native mountain ponies of Britain and Ireland reveal a novel rare population. *Ecol. Evol.* 3:934–947.
- Xu, X., and U. Arnason. 1994. The complete mitochondrial genome of the horse, *Equus caballus*: extensive heteroplasmy of the control region. *Gene* 148:357–362.

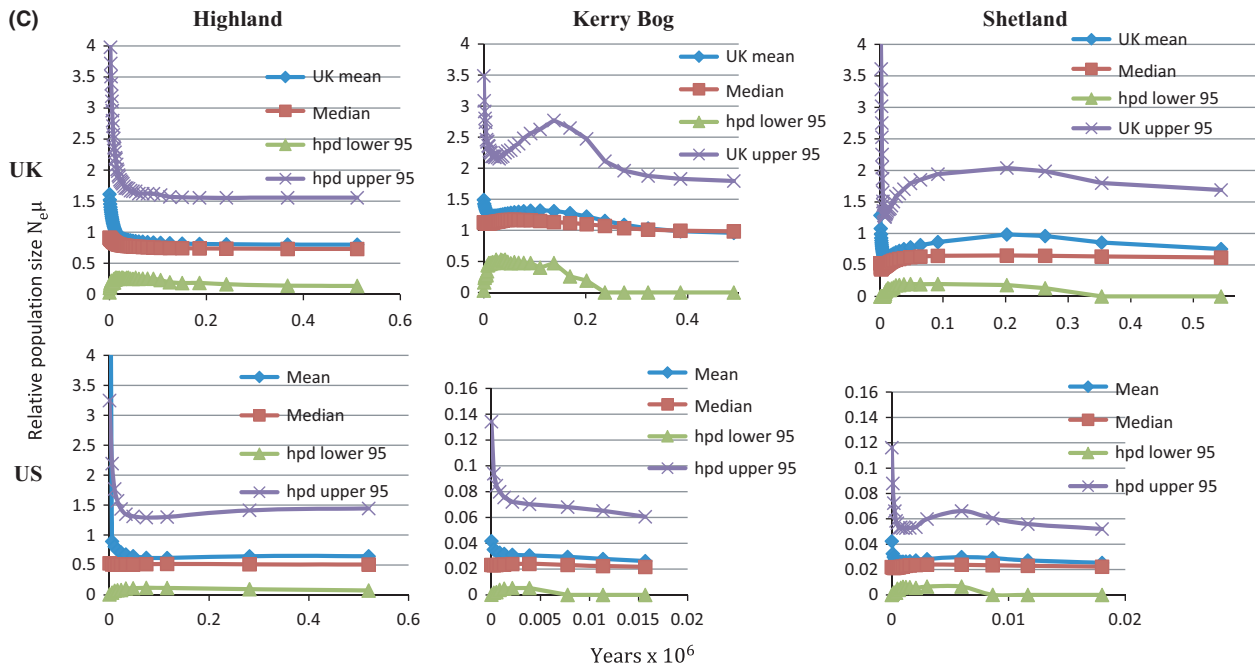
Appendix 1: Origin of data used within the study and GenBank sequences

	U.K. and Irish samples	North American samples	Kerry Bog samples	Ancient horse samples	Haplogroup identifier samples
SSR data	Original data generated for the current study and Winton et al. (2013)	Prystupa et al. (2012b)	NA	NA	NA
MtDNA data	Original data generated for the current study, accession numbers KR361761–KR362244, and Winton et al. (2013)	Prystupa et al. (2012a); GenBank accession numbers HQ592784–HQ593063	McGahern et al. (2006); GenBank accession numbers DQ327852–DQ327890	Cieslak et al. (2010); GenBank accession numbers detailed in original manuscript “Supporting Information (Table S2)”	Achilli et al. (2012); GenBank accession numbers JN398377–JN398457

Appendix 2: Supplementary figure to compliment Figure 5



Appendix 2. Continued.



Appendix 3: Harpending's raggedness index data (significant *P* values in bold)

Population	Harpending's raggedness index	<i>P</i> value (simulated raggedness \geq observed raggedness)
Welsh Section A	0.0576	0.01
Welsh Section B	0.0423	0.64
U.S. Welsh Pony	0.0790	0.54
Fell	0.0179	0.42
U.S. Fell	0.7037	0.72
Highland	0.0547	0.13
U.S. Highland	0.1867	0.11
Connemara	0.0218	0.74
U.S. Connemara	0.1782	0.05
Dartmoor	0.0591	0.63
U.S. Dartmoor	0.4277	0.98
Shetland	0.1095	0.01
U.S. Shetland	0.3075	0.02
Kerry Bog Pony	0.0264	0.24
U.S. Kerry Bog Pony	0.1753	0.09