Genetic Diversity in Canadian, Mountain and Moorland, and Nordic Pony Populations

A Thesis Submitted to the
College of Graduate Studies and Research
in Partial Fulfilment of the Requirements
for the Degree of Master of Science
in the Department of Animal and Poultry Science
University of Saskatchewan
Saskatoon, SK
Canada

Jaclyn Mercedes Prystupa

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of the University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by Dr. Yves Plante, who supervised my thesis, or, in his absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make further use of material in this thesis in whole or in part should be addressed to:

Head of the Department of Animal and Poultry Science

University of Saskatchewan

51 Campus Drive

Saskatoon, Saskatchewan S7N 5A8

ABSTRACT

The legally binding international declaration of the Convention on Biological Diversity (signed by over 180 countries) recently acknowledged the importance of conserving genetic diversity within livestock species. This study aimed to help Canada assess molecular diversity in its horse and pony (*Equus ferus caballus)* genetic resources. Here, 24 populations were examined, with special focus on the native Canadian, Mountain and Moorland, and Nordic pony populations, using two well accepted molecular tools. Additional horse breeds and feral populations were also included in this project as some may have influenced the development of the three equine groups of interest. Altogether, 821 individuals were genotyped at 38 microsatellite loci, and 280 individuals were sequenced using a 421 base pair portion of the mitochondrial displacement Hypervariable Region I.

Results from the microsatellite analyses indicated that 13.33% of genetic diversity arose from breed differences, whereas 84.60% and 2.07% of diversity arose from within and among individuals respectively. The New Forest and Welsh breeds were found to be the most diverse while having the highest average effective number of alleles and allelic richness (4.31 and 6.01; 4.33 and 5.87 respectively). The Eriskay and Lac La Croix breeds were found to have the lowest average effective number of alleles and allelic richness (2.51 and 3.98; 2.83 and 4.01 respectively). Expected heterozygosities were lowest in the Lac La Croix (0.61) and highest in the Welsh and New Forest (0.74) breeds, whereas observed heterozygosities were highest in the Kerry Bog (0.77) and lowest in the Exmoor (0.57) breeds. The genetic structure and admixture analyses suggested that the most probable number of unique genetic clusters was 21 as opposed to the 24 predefined populations.

Results from the mitochondrial sequence data revealed that there were 36 informative sites producing 62 haplotypes, 20 of which were previously unreported. The Connemara was found to have the highest haplotype diversity of the pony breeds (0.89); however, the Highland pony was found to have the highest nucleotide diversity and pairwise difference (0.16 and 6.73 respectively). In contrast, the Fell pony had the lowest haplotype diversity (0.22), and the feral Sable Island population had the lowest nucleotide diversity and pairwise difference (0.01 and 0.29 respectively). Multiple phylogenetic trees were reconstructed and produced similar topologies. In general, the Mountain and Moorland and Nordic breeds were spread among the clades, whereas native Canadian populations were most frequent in the D and E clades. Interestingly, a large portion of ponies were found within the rare E clade as opposed to horses.

Information gathered from this project can be incorporated with other available data into pre-existing conservation/breeding programs currently managed by the various breed societies to ensure that the most optimal and sustainable strategies are in place.

ACKNOWLEDGMENTS

Ever since I was a little girl I have been fascinated with the horse. Thankfully, my parents fed my interest about horses with not only books, but also riding lessons, and a horse of my very own. With that said, I must admit that as an avid equine enthusiast, I was thrilled to be able to work on a graduate project which focused on ponies. This project was definitely a labour of love, and I would like to take the time to acknowledge and thank a few key people for their help and support, as without them this project would not have been possible. First off, I would like to thank my supervisor, Dr. Yves Plante who guided me throughout all stages of this thesis, as well as introduced me to the topic of conservation genetics. I would also like to acknowledge and thank the members of my committee including: Dr. Fiona Buchanan, Dr. Janet Hill, Dr. Bernard Laarveld, and Dr. Phil McLoughlin who provided valuable support and input. I would also like to acknowledge the many breed societies, Dr. José Andrés, Dr. Gus Cothran, Dr. Don Hamiliton, Pamela Hind, Dr. Rytis Juras, Zoe Lucas, Liz Mackenzie, Patricia Morris, and Maxxam Analytics (Ontario, Canada), who provided valuable support and/or samples. In addition, I would like to thank the Canadian Animal Genetic Resources program of Agriculture and Agri-Food Canada for providing the funding for this project, as well as the University of Saskatchewan for providing conference funding.

Last but not least, I would like to thank my family, including my mom, my dad and Barrington, who not only supported me throughout my schooling, but also listened to countless practice presentations and rough drafts.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

AFLP = Amplified fragment length polymorphism

AKBPS = American Kerry Bog Pony Society

AMOVA = Analysis of Molecular Variance

AnGR = Animal Genetic Resources

AR = Allelic richness

BLAST = Basic Local Alignment Search Tool

CIPRES = Cyberinfrastructure for Phylogenetic Research

cm = centimetre

D-Loop = Displacement loop

DNA = Deoxyribonucleic acid

EPMSBS = Eriskay Pony Mother Stud Book Society

EST = Equus Survival Trust

FAO = Food and Agriculture Organization

GPAAR = Global Plan of Action for Animal Resources

GTR = General Time Reversal

HCl = Hydrogen chloride

H_E = Expected heterozygosity

hh = Hands high

H_O = Observed heterozygosity

HVR = Hyper Variable Region

HWE = Hardy-Weinberg Equilibrium

IISD = International Institute for Sustainable Development

KCl = Potassium chloride

MCMC = Markov Chain Monte Carlo

MtDNA = Mitochondrial DNA

MJ = Median-Joining

M = Molar

mM = milliMolar

N = Number of individuals sampled

N_a = Average number of alleles

NCBI = National Center for Biotechnology Information

N_e = Average number of effective alleles

NSMNH = Nova Scotia Museum of Natural History

PCR= Polymerase Chain Reaction

RAPD = Random amplified polymorphism DNA

RBC = Rare Breeds Canada

RFLP = Restricted fragment length polymorphism

SD = Standard Deviation

SNP = Single nucleotide polymorphism

TWEEN = Polyoxyethylene (20) sorbitan monolaurate

µl = microlitre

1.0 General Introduction

Livestock are an essential part of the global agricultural industry. Animals provide not only meat, fibre, and milk; but also, in many parts of the world, a source of power and transportation for agriculture and industry (Food and Agriculture Originization (FAO) 2007). Careful selection of sires and dams, and increased knowledge of nutrition has allowed for improved efficiency in production; however, this has not come without cost. A growing trend in many sectors within the livestock industry is a decrease in the number of breeds utilized. This alarming trend has resulted in the loss of some breeds and genetic diversity within species, as well as an increase in disease, decrease in fertility, and a lack of adaptability to changing environments and/or shifting consumer demands (FAO 2007). In order to prevent the loss of even more breeds, and to help hedge against changing environments and consumer demands, over 180 countries have agreed to catalogue, characterize, and assess the genetic diversity in their farm animal breeds for the purpose of establishing, or amending, conservation strategies for all livestock species within their borders (FAO 2007). Conservation involves preserving diversity (both phenotypic and genotypic) within and among breeds, and is important for both rare and popular breeds alike (Gizaw *et al.* 2008). In order to develop a feasible and sustainable conservation program one must first characterize the populations of interest (Glowatzki-Mullis *et al.* 2006; Plante *et al.* 2007). Once these breeds have been properly characterized and diversity estimated, various organizations (both governmental and non-governmental) can use the information to help prioritize and develop (or amend) breeding strategies.

Equines (*Equus ferus caballus*) are just one of the species important to the livestock sector. Though they are restricted to mostly a recreational role within developed countries, they still provide a vital function in agriculture and food production in many other parts of the world (FAO 2007). Within Canada there are several breeds of both horses and ponies engaged in many different equine related activities. Of all the equine breeds recognized in Canada, only three, the Newfoundland and the Lac La Croix pony, and the Canadian (or Canadien) horse are actually native breeds. In addition, Canada is also home to one native feral equine population located off the East coast of Canada on Sable Island. There is interest in understanding which breeds have contributed to the development of the Canadian populations, especially the Lac La Croix, Newfoundland, and feral Sable Island populations, which in the past (or present in the case of the Sable Island population), were often left to roam and breed when they were not being used (Pickeral 2001; Plante *et al.* 2007; Lynghaug 2009). Historical records indicate that these populations likely developed from crossing Mustangs with Canadian horses, Mountain and Moorland breeds, and Nordic breeds respectively.

The term pony is used to describe equines standing 14.2 hands high (hh; 147 cm) at the withers or less. Many pony breeds have relatively small population sizes, especially in North America, and are considered to be at some risk of extinction, with several pony breeds being placed on conservation lists including: Rare Breeds Canada (RBC), American Livestock Breeds Conservancy, Equus Survival Trust (EST), and Rare Breeds Survival Trust (EST 2008; FAO 2007; Lynghaug 2009).

Genetically speaking, pony breeds have been found to cluster separately from horse breeds (Leroy *et al.* 2009). Despite this, there is very little information published investigating the genetic diversity among pony breeds. The information that is available focuses on relatively few pony breeds, and/or compares them with a large number of horse breeds (Kaminski & Duncan 1981; Cunningham *et al.* 2001; Vilà *et al.* 2001; Jansen *et al.* 2002; Bjørnstad *et al.* 2003; Aberle & Distl 2004; Royo *et al.* 2005; Glowatzki-Mullis *et al.* 2006; McGahern *et al.* 2006; Aberle *et al.* 2007; Plante *et al.* 2007; Leroy *et al.* 2009). The purpose of this research project was to investigate the genetic diversity and admixture among native Canadian, Mountain and Moorland, and Nordic equine populations, using both microsatellites and mitochondrial DNA (mtDNA) sequence data.

2.0 Literature Review

2.1. Conservation genetics

In order to fully understand the concept of conservation genetics it is critical to understand the conservation theory. The foundation of conservation genetics is that the longevity of a breed is directly related to the diversity observed within the breed (Aranguren-Mendez *et al.* 2002; Luís *et al.* 2007a). Therefore, maintaining and maximizing diversity in farm animal breeds should be the core goal in livestock production, selection, and conservation, as it is important for genetic progress, and allows for the ability to adapt to different breeding strategies (Morais *et al.* 2005). Establishing an optimal conservation strategy is breed dependent and involves many steps, the first beginning with breed characterization (Glowatzki-Mullis *et al.* 2006; Plante *et al.* 2007). Breed characterization involves not only classifying the physical and production aspects of a breed (currently recorded by many non-governmental agencies), but also an estimation of genetic diversity (FAO 2007). The estimation of genetic diversity provides important information regarding the admixture and possible genetic erosion among and within breeds or populations (Aranguren-Mendez *et al.* 2002; Morais *et al.* 2005; Valera *et al.* 2005). Genetic erosion can be defined as the loss of rare alleles within a population which results in increased homozygosity, and therefore a reduction in the genetic diversity observed (Van Treuren *et al.* 1991; Ouborg & Treuren 1995).

In 1992, 'Agenda 21' was a global plan of action elected at all stages of government to combat the increasing loss of breeds within species (and genetic diversity), as well as environmental degradation due to human activities (FAO 2007). Canada, and 178 other countries, adopted the non-legally binding guidelines and goal of increasing sustainable

agriculture practices to meet current and future production and consumer demands (FAO 2007). 'Agenda 21' set forth the following guidelines for all countries that adopted the agreement:

> '"a) draw up breed conservation plans for endangered populations, including semen/embryo collection and storage, farm-based conservation of indigenous stock and *in situ* conservation, b) plan and initiate breed development strategies, and c) select indigenous populations on the basis of regional importance and genetic uniqueness, for a ten-year programme, followed by selection of an additional cohort of indigenous breeds for development"' (FAO 2007).

The 'Rio Earth Summit' (the '1992 United Nations Conference on Environment and Development'), coincided with 'Agenda 21' and presented the 'Convention on Biological Diversity Declaration' which was the first legally binding framework regarding genetic diversity, and was originally accepted, signed, and adopted by 150 countries, including Canada (FAO 2007). The convention also addressed sustainable resources, "equitable sharing of the benefits arising from the utilization of genetic resources", and the right of individual states to determine their own strategy for genetic resource management (FAO 2007). Since the declaration was originally presented another 38 countries have also joined and signed the legally binding agreement (FAO 2007).

In 2007, the 'International Technical Conference on Animal Genetic Resources for Food and Agriculture' was held in Interlaken, Switzerland and was the "first intergovernmental conference to focus exclusively on Animal Genetic Resources (AnGR)" (International Institute for Sustainable Development (IISD) 2007). This conference addressed two documents; the 'Global Plan of Action for Animal Resources' (GPAAR) and the 'Interlaken Declaration', originally brought forth by the' Commission on Genetic

Resources for Food and Agriculture' (IISD 2007). The 'GPAAR' is made up of four distinct priority areas (IISD 2007). Priority area (1) focuses on the "Characterization, Inventory and Monitoring of Trends and Associated Risks" of AnGR including the need for the development of "international technical standards and protocols" and the establishment of individual "country-based early warning and response systems" to monitor AnGR (IISD 2007). Priority area (3) focuses on the conservation of AnGR and the development and implementation of conservation policies (both national and international) as well as conservation strategies (both *in situ* and *ex situ*) using "technical standards for conservation" (IISD 2007).

The 'Interlaken Declaration' acknowledged "the essential roles and values of AnGR for food and agriculture, in particular their contribution to food security for present and future generations" as well as confirmed the acceptance of the 'GPAAR' by all delegates of the 'International Technical Conference on Animal Genetic Resources for Food and Agriculture' (IISD 2007).

Conservation programs vary between countries, species and breeds, although in general there are two main approaches. The best known, and likely the more economical approach, is called the *in situ* method (FAO 2007). This method is advantageous because the animal is housed on the ground where the breed originated and can be used for its original purpose. The alternative method (*ex situ*) involves keeping individuals outside of where the breed originated (*in vivo*), or *in vitro* which is achieved by storing tissue cryogenically (allowing for storage for extended periods of time; Talle *et al.* 2005; FAO 2007). Optimal conservation programs should be developed using a balance of both methods.

Equine breeds are commonly conserved using both approaches. Artificial insemination has been used in both horse and pony breeds alike (American Kerry Bog Pony Society (AKBPS) 2008; DeAssis *et al.* 2009). On the other hand, equine oocyte collection, use, and storage is a difficult and tedious task, and results in fewer high quality oocytes available to freeze "compared to other large domestic species" (Lagutina *et al.* 2005; Galli *et al.* 2007). The exact ratio of animals on the ground to tissues stored is dependent on the availability of *in vitro* resources, as well as the diversity and/or distinctiveness within a breed, and the risk of extinction for that breed (Petit *et al.* 1998; FAO 2007). Developing countries may not have sufficient financial resources to store tissue cryogenically and therefore will likely only develop conservation strategies using an *in situ* approach. Alternatively, developed countries will likely use both methods, and may even come to an agreement with developing countries to store some of their genetic resources cryogenically. Due to limited financial resources, breeds that are distinct, have a unique trait, or have a small population size, (and therefore a higher risk of extinction), may be prioritized higher than other breeds for the development of a conservation strategy (Petit *et al.* 1998; Aranguren-Mendez *et al.* 2002; Behl *et al.* 2007; FAO 2007).

2.2. Equine breeds and populations

There are approximately 700 different equine breeds found around the world (FAO 2007). Each breed was developed in a specific region and environment to fulfill a need, such as power, transportation, food, sport or war (Pickeral 2001). Equines were frequently imported into new regions to facilitate empire expansion, trade, or colonization, and often resulted in the displacement of some of the local breeds with similar roles (FAO 2007). Mechanization and the modernizations of agriculture and industrial practices had a severe

impact on all equine breeds as machines took the place of both horses and ponies alike, resulting in a decrease in their numbers, breeds, and genetic diversity (FAO 2007).

In developing countries equine breeds are conserved as a result of the necessity for a particular need (usually in agriculture); however, in developed countries this is not always the case. The three main reasons equines are conserved in developed countries are for cultural/historical, recreational, and agricultural/forestry reasons (Luís *et al.* 2007a). A breed that fulfills a cultural or historical need is most likely to be conserved in its country of origin using an *in situ* method (Behara 2000). Recently, equines have been bred in all shape, sizes and dispositions to fulfill recreational roles in various disciplines. Lastly, equines are conserved in developed countries for their traditional role in agriculture and forestry. Some breeds are still utilized for meat or milk production, whereas others are utilized to move stock in the fields or at auctions as they can manoeuvre quickly. The following sections discuss all breeds relevant to this thesis. The breeds have been classified into four main categories: Canadian, feral, Mountain and Moorland, and Nordic populations.

2.2.1. Canadian populations

Canadian populations likely arose from crossing various types of equines originally brought to the New World with the colonial settlers. The four native equine populations developed in Canada are: the Canadian horse, the Lac La Croix pony, the Newfoundland pony, and the feral Sable Island population. With the exception of the Sable Island population, these breeds were developed for versatility and to have kind dispositions. In addition, all native Canadian equine populations also have good feet and legs, and are well acclimated to Canada's harsh weather and long winters.

2.2.1.1. The Canadian horse

The Canadian, Canada's national heritage horse, originates from a shipment of horses sent from France to Canada in 1665 (Lynghaug 2009). Canadian horses were selectively bred to have dark coats and to stand between 14 and 16 hh, (Lynghaug 2009). The Canadian horse was included in this study for two main reasons: (1) for completeness of this study, and (2) because Canadian horses may have influenced the development of the other Canadian populations, or vice versa. RBC (2009) lists the Canadian horse as 'at risk' with only 151 to 500 new female registrations each year. Additional information regarding the influence of the Canadian horse on other North American breeds can be found in Chapter 3 of this thesis.

2.2.1.2. The Lac La Croix pony

The Lac La Croix pony was developed in the North American Great Lakes region, bordering the United States (Minnesota) and Canada (Ontario), and dates back to approximately the 1870s (Lynghaug 2009). Generally speaking, the Lac La Croix Pony originated from crosses between Spanish Mustangs and Canadian horses. The breed stands between 12 and 14.2 hh, and can be found in all coat colours except white, cream, or broken patterns (Lynghaug 2009). In 1977, this breed suffered a severe bottleneck with a reduction to only four mares, which were bred to Spanish mustangs (Behara 2000). These four mares were left to run feral on private property until the late 1990s, and today there are approximately 250 individuals, with about 75 listed as breeding animals registered (Lynghaug 2009). RBC (2009) lists the Lac La Croix as 'critical' with only 1 to 15 new female registrations each year.

2.2.1.3. The Newfoundland pony

The Newfoundland pony is believed to have been developed from crossing several of the Mountain and Mooreland pony breeds; although the exact breeds are not known, as traditional breeding management involved releasing ponies to run feral when not required for work (Lynghaug 2009). The Newfoundland pony stands between 11 and 14.2 hh, has a thick mane and tail, and is found in all colours with the exception of broken patterns (Lynghaug 2009). In the past, Newfoundland ponies were typically used to plough fields and carry everything from firewood to fish. In the 1970s, the Newfoundland pony population reached an estimated 10,000 animals; however, their numbers plummeted in the 1980s to fewer than 100 individuals, as a result of the mechanization of rural community industries, and many thousands of ponies being sold for meat (Lynghaug 2009). Over the last few decades the Newfoundland pony numbers increased to approximately 400 animals, and RBC (2009) still lists the breed as 'critical'.

2.2.1.4. The Sable Island population

Although not recognized as a breed, the feral equine population located on Sable Island (off the East coast of Canada) has been identified as a unique and valuable genetic resource (Plante *et al.* 2007). Since 1960, the feral herd has been protected under the Canada Shipping Act hence preventing any human interference (Plante *et al.* 2007). Prior to 1960, horse owners frequently left stallions on the island to breed with the herd, and then collected the offspring and stallions (Nova Scotia Museum of Natural History (NSMNH) 2001). This practice makes it possible that the Sable Island population could share ancestry with other Canadian breeds. In general, Sable Island equines are hardy, stand around 14 hh, are dark

coloured, and live in small harems on the island (Pickeral 2001). In 2008, there were 375 individuals surveyed on the island, although the total population fluctuates between about 250 and 500 individuals per year, depending on environmental conditions (Plante *et al.* 2007; Lucas *et al.* 2009; Personal Coummication with Dr. Phil McLoughlin, University of Saskatchewan, 2011). RBC (2009) lists the population as 'critical'.

2.2.2. Feral equine populations

The following sections discuss two additional feral populations that may have influenced the development of the native Canadian equine populations.

2.2.2.1. The Grand Turk equine population

In the past, Turks and Caicos (part of the West Indies) was an extremely important player in the salt raking industry (Burleson & Burleson 2007). Salt was an important food preservative and was imported in large quantities to North America. The salt industry required horses, ponies, and donkeys to move the valuable resource to ships, and therefore to market. As the salt raking industry diminished many equines were released and left to roam on the islands (Burleson & Burleson 2007). Due to the large trade flow that existed in the past between Turks and Caicos and North America there is a possibility that equines could also have been traded, and therefore they may share common ancestry with some North American populations/breeds.

2.2.2.2. The Saint-Pierre et Miquelon population

Located approximately 10 kilometres off the coast of Green Island (Newfoundland and Labrador) lie the two small French islands of Saint-Pierre et Miquelon. These islands are home to a semi-feral horse population historically thought to have arisen from approximately 75 individuals left on the island (Personal Communication with Patricia Morris, Newfoundland Pony Society, 2010). Due to the close proximity of the French islands to Canada there is a possibility that these equines may also share common ancestry with the Canadian populations.

2.2.3. The Mountain and Moorland breeds

The British Isles are home to a unique group of native ponies known as the Mountain and Moorland breeds. The Mountain and Moorland group originally consisted of nine breeds (Connemara, Dale, Dartmoor, Exmoor, Fell, Highland, New Forest, Shetland and Welsh), and are thought to have influenced the development of many breeds, including the possibility of Canadian ones (Lynghaug 2009). In the recent past, two other breeds; the Eriskay and the Kerry Bog have also been loosely incorporated into the Mountain and Moorland pony category. All the Mountain and Moorland breeds are strong, versatile, and surefooted with excellent dispositions. In the 16th century, a law was passed by King Henry VIII to prevent the use of "nags of small stature" for breeding purposes; and also prohibited the "use of any horse under 15 hh", as a result, many ponies from all Mountain and Moorland breeds were destroyed (Lynghaug 2009). Queen Elizabeth I later revoked this law. The World Wars along with the mechanization of agriculture, forestry, mining, and industry, put further strain on many of these breeds. The following sections describe each of the Mountain and Moorland pony breeds in further detail, including their presence in North America.

2.2.3.1. The Connemara pony

The Connemara originated in Ireland, and was first brought to North America in 1951 with the importation of two stallions and four mares (Lynghaug 2009). This breed is known to excel as a performance pony. Connemara ponies are typically grey (although can be found in any coat colour) and stand between 13 and 14.2 hh (Pickeral 2001). In 1957, there were approximately 155 ponies in the United States, and today there are approximately 4,000 purebred registered Connemara ponies in North America (Lynghaug 2009).

2.2.3.2. The Dale pony

Developed in the Northern part of England, the Dale stands between 14 and 14.2 hh, and is predominately found in black; however, chestnut, bay, and roan coat colours are also accepted (Pickeral 2001). The Dale pony has an interesting past, it was frequently crossed with Clydesdales to produce vanners (small draft horses), and was also conscripted during both World Wars by the British Army to move supplies and ammunition (Lynghaug 2009). Many ponies died during these wars, and due to mechanization, as well as crossbreeding, there was a severe decrease in Dale numbers (Lynghaug 2009). Today, there are approximately 2,000 Dale ponies worldwide, and this has resulted in the breed being classified as 'endangered' by Rare Breeds Survival Trust (Lynghaug 2009). This breed was only recently (recorded to be) imported to Canada (1991), although some likely came with the first settlers, and today there are approximately 250 individuals in North America (Lynghaug 2009). Rare Breeds Canada (2009) lists the Dale as 'critical'.

2.2.3.3. The Dartmoor pony

The Dartmoor pony originated in the South part of England, and is found in the same coat colours as the Dale, however is slightly more refined and smaller standing at approximately 12.2 hh (Pickeral 2001). This breed unlike other Mountain and Moorland pony breeds was developed primarily for riding purposes, and was severely impacted by the laws imposed by King Henry VIII (Lynghaug 2009). The Dartmoor numbers increased to some extent when it became known as a great mount for polo, although their numbers were reduced again during the World Wars as they were frequently used as a source of meat. In addition, the traditional method of breeding involving releasing Dartmoor ponies on the moors resulted in crossbreds, as the moors were considered common land, and various other horse and pony breeds were also released on to the land (Lynghaug 2009). Today, most Dartmoor ponies are bred on private land, and although they are still listed as 'vulnerable' by Equus Survival Trust, there are now between 5,000-7,000 breeding females worldwide. The Dartmoor pony was first introduced into North America in the 1930s, and today there are still relatively few ponies with approximately 350 individuals residing in the United States (Lynghaug 2009). Rare Breeds Canada (2009) lists the Dartmoor as 'critical'.

2.2.3.4. The Eriskay pony

Considered to be one of the newer members of the Mountain and Moorland breeds, the Eriskay pony was developed on the Western Isle of Eriskay (Scotland). This breed stands between 12 and 13.2 hh, and can only be found in grey, black or bay coat colours (Eriskay Pony Mother Stud Book Society (EPMSBS) 2010). This breed suffered a severe population reduction and bottleneck in the 1970s when the breed was reduced to approximately 20

individuals, with one purebred stallion (EPMSBS 2010). Today, the breed population has increased to approximately 420 individuals; however, the breed is still listed as ‘critical’ by Rare Breeds Survival Trust (EPMSBS 2010). Currently, there are no Eriskay ponies recorded in North America.

2.2.3.5. The Exmoor pony

Considered to be the oldest of the Mountain and Moorland breeds, the Exmoor is a small bay, brown or dun breed with no white markings, and standing between 11.2 and 12.3 hh (Lynghaug 2009). Like the Dartmoor, the Exmoor was also allowed to run feral on the moors, as well as the in the Royal Forest (Exmoor) until 1818, when the deed to the land was sold (Lynghaug 2009). Before the land changed hands, the seller took approximately 30 ponies with him in order to conserve the breed (Lynghaug 2009). Over the next decades, the Exmoor breed suffered severe herd reductions as a result of crossbreeding, and being used as a food source and mount during the World Wars (Lynghaug 2009). Today, the breed is listed as ‘endangered’ by Rare Breeds Survival Trust and there are only 2,000 Exmoors found worldwide; of that, only about 100 are found in their natural habitat on the moors, and another 100 are found in North America (Lynghaug 2009). Rare Breeds Canada (2009) lists the Exmoor as ‘critical’.

2.2.3.6. The Fell pony

The Fell pony is similar to the Dale, although is slightly smaller, standing under 14 hh and are thought to originate from the same ancestral stock (Pickeral 2001). The Fell developed in the mountains between England and Scotland, where forage resources are limiting, and therefore limited the height of equines found in this region (Lynghaug 2009).

Today, there are approximately 6,000 Fell ponies worldwide, with 500 breeding females and approximately 350 individuals in North America (Lynghaug 2009). Rare Breeds Canada (2009) lists the Fell as 'critical'.

2.2.3.7. The Highland pony

The Highland pony is a stout breed, standing between 13 and 14.2 hh, and was originally developed in Scotland (Pickeral 2001). This breed was used not only used for agricultural purposes but also as a mount during war (Lynghaug 2009). Highlands can be found in several colours including: dun, grey, chestnut, bay, and black; additional zebra and dorsal stripes are also found within the breed. Highlands were originally brought to North America by the settlers, and were also recently imported into the United States in the 1960s (Lynghaug 2009). Today, the Highland is listed as 'vulnerable' by Equus Survival Trust and are extremely rare in North America with a population size of only 70 individuals in 2008 (Lynghaug 2009).

2.2.3.8.The Kerry Bog pony

The Kerry Bog pony is a small (10 to 12 hh), compact pony of various colours and patterns with a thick coat to withstand severe environmental conditions. These ponies developed in Ireland and were commonly utilized for pack purposes. At the turn of the 20th century the population suffered a severe herd reduction to only 20 animals within Ireland (McGahern *et al.* 2006). Kerry Bog ponies first arrived in North America in 2003 when four mares and two stallions were imported into the United States (AKBPS 2008).

2.2.3.9. The New Forest pony

The New Forest pony was named after the forest where the breed originated. Developed as a performance breed, many New Forest mares were frequently bred to stallions from the various other Mountain and Moorland breeds to improve performance right up until the 1930s. New Forest ponies were first introduced into Canada in 1951 with the importation of 22 individuals from England (Lynghaug 2009). The population has since increased to approximately 110 pure and part-bred registered animals in North America, although it is believed that there is likely another 300 that are not registered or registered in other registries, and currently residing in Canada or the United States (Personal Communication with Priscilla Hawkyard, New Forest Pony Society of North America, 2011).

2.2.3.10. The Shetland pony

One of the most well-known and smallest pony breeds in the world is the Shetland. This breed stands no more than 10.2 hh, and can be found in many different colours with the exception of broken patterns (Lynghaug 2009). Currently in North America, there are two types of Shetlands, although both types developed from the same ancestral stock. The first is known as the American Shetland, and is more refined then the second, the United Kingdom Shetland, as a result of backcrossing with other breeds (Lynghaug 2009). The United Kingdom Shetland developed in Scotland and is considered the traditional type. The Shetland gained tremendous popularity in the 1840s when the breed excelled as 'pit ponies' as a result of a ban on the use of children and women in the mining industries in England. The mining industries exclusively used stallions for this work, and demand for stout Shetland stallions put a strain on the breed (Lynghaug 2009). Mechanization eventually replaced the need for 'pit ponies', and the Shetland began a new role in driving and as a child's mount.

Shetlands were first brought to North America in 1884, and were used almost exclusively in the mining industries. As occurred in England, mechanization rendered the 'pit pony' role obsolete, and the Shetland pony found new roles in riding, driving, and the development of the American Shetland (Lynghaug 2009). Rare Breeds Canada (2009) lists the United Kingdom Shetland as 'critical'.

2.2.3.11. The Welsh pony

The popular Welsh pony was originally developed in Wales. This breed was utilized traditionally for draft and driving purposes, although can be found in many different disciplines. Over the years, four sections (types) developed, in which all coat colours are permitted with the exception of broken patterns (Pickeral 2001). The first type is the Welsh Mountain pony, standing no higher than 12 hh (Lynghaug 2009). Section 'B' stands at 13.2 hh or less, and is defined as more refined compared to section 'A'. The final two sections are cobs (small stout horses), and can be differentiated by the 'C' section being heavier set and with a height restriction of 13.2 hh, in comparison to the lighter and taller section 'D' with no height restriction (Pickeral 2001). Welsh ponies were originally imported into North America in the 1880s, and as of 2007, there were 48,000 individuals registered among the four sections in United States (Lynghaug 2009). Rare Breeds Canada (2009) lists the Welsh pony and Cob as 'vulnerable' with only 51 to 150 new female registrations each year.

2.2.4. The Nordic breeds

Three breeds in this study are considered to fall under the category of the Nordic breeds, including the: Norwegian Fjord, Icelandic, and Shetland. The Nordic breeds phenotypically fall under the height category of pony (14.2 hh or less at the withers); although many owners and enthusiasts insist that the Norwegian Fjord and Icelandic are in fact horses. Regardless of the classification, these breeds share a small stout frame. Although the Shetland was discussed in the previous section, the remaining two will be discussed in detail in the following sections.

2.2.4.1. The Icelandic

The Icelandic breed developed from equines present in Iceland prior to approximately 1,000 A.D, at which time the government banned the importation of equines (Lynghaug 2009). The Icelandic is a small equine standing between 12 and 14 hh, and is found in all coat colours except appaloosa patterns (Lynghaug 2009). Intriguingly, this breed has a very unique feature, five natural gaits (Caoline 1998). In addition to the walk, trot, and gallop, these equines also perform the pace (where the two legs on the same side move at the same time), and the '*tőlt*' (also known as the running walk), characterized as a four beat gait (Lynghaug 2009). Individuals from the Icelandic breed were likely brought to North America with the first colonial settlers.

2.2.4.2. The Norwegian Fjord

Developed in Norway, the Fjord stands no more than 14 hh and always appears in a dun coat colour with the mane having both black and cream strands of hair (Pickeral 2001). Phenotypically speaking, this breed closely resembles the Asian Wild Horse, and may be the ancestor of many draft breeds. Fjords were first seen in the United States in 1888, when a single colt was imported, although most breeding stock found currently was not imported until the 1950s (Lynghaug 2009).

2.3. Genetic diversity in equines

Genetic diversity can lead to the development of distinctive traits, disease resistance, and climate adaptability within breeds (FAO 2007). Distinctive traits have been found in breeds such as the Bashkir Curly which has a curly and hypoallergenic coat compared to other horse and pony breeds, and the Icelandic with its additional tőlt gait (Lynghaug 2009). Resistance to diseases have been found in the Kuda Padi and Bajau (local Southeast Asian equine breeds), Pantaneiro (a local Brazilian horse breed), and the Pottok (a local French/Spanish pony breed), which have been reported to be resistant to internal parasites, Equine Infectious Anaemia (mosquito born) and Piroplasmosis (tick born) respectively (FAO 2007; Domestic Animal Diversity Information System 2011). Climate adaptations have been found in many breeds including the Canadian ones with thick coats for Canadian winters, and Arabians and Caspians with thin skin to dissipate heat in desert environments (FAO 2007).

2.4. Molecular methods to determine genetic diversity

Several molecular tools and techniques have been used over time to estimate genetic diversity, relationships and the origin of equine breeds (FAO 2007). These techniques allow for the estimation of basic population parameters, aid in the development of conservation strategies, and characterization of genetic traits of interest (Talle *et al.* 2005; FAO 2007). Initially studies focused on the differences in the electrophoretic mobility of selected proteins and blood types (Sandberg & Cothran 2000). Unfortunately these methods were subject to high cost (for tissue storage), and were affected by allele recombination, chimeric animals, and environmental and developmental factors (Sandberg & Cothran 2000; FAO 2007). Blood typing and protein analyses were replaced by DNA based methods because they provide increased confidence in results (Talle *et al.* 2005). In addition, DNA based methods allow for the use of a smaller amount of source material, increased storage capacities, and decreased cost (FAO 2007). They are also less likely to be affected by environmental and developmental factors (FAO 2007). Equine diversity studies have frequently involved the use of many types of DNA based methods including: Restriction Fragment Length Polymorphisms (RFLP), Random Amplified Polymorphic DNA (RAPD), Amplified Fragment Length Polymorphisms (AFLP), microsatellites, mtDNA sequencing, and Single Nucleotide Polymorphisms (SNP; Wang *et al.* 1994; Apostolidis *et al.* 2001; Rieder *et al.* 2001; Cothran *et al.* 2005; Do Egito *et al.* 2007; Plante *et al.* 2007; Haase *et al.* 2008).

RFLPs were one of the first DNA based techniques used and required both high quality DNA and a large volume of source material (Schlotterer 2004). They were later combined with the Polymerase Chain Reaction (PCR; reducing the source material required), or replaced by other molecular techniques, such as RAPDs and AFLPs (Wang *et al.* 1994;

Ishida *et al.* 1996; Ajmone-Marsan *et al.* 1997). Both RAPDs and AFLPs are simple, quick, inexpensive, are used frequently where little genetic information is available and for DNA fingerprinting, but are not useful when trying to distinguish between heterozygous and homozygous individuals, therefore these techniques are not routinely utilized in equine genetic diversity studies (Schlotterer 2004; Ajmone-Marsan *et al.* 1997; Do Egito *et al.* 2007; FAO 2007). SNPs involve single changes in nucleotide bases throughout the genome that can then be measured by various laboratory techniques (Boulding *et al.* 2008). SNPs are frequently used to map and understand their effects on phenotypic traits (such as coat colour) in equines (Rieder *et al.* 2001; Haase *et al.* 2008), and although they can be automated and results transferred between laboratories, they are also biallelic, and relatively expensive to assess and genotype (Talle *et al.* 2005; FAO 2007). Today, most equine genetic diversity studies utilize microsatellites and/or mtDNA sequence data, and are described in detail in the following sections. Each of these methods has its own strengths, and therefore they are often utilized in conjunction with each other (Aberle & Distl 2004; Pérez-Gutiérrez *et al.* 2008).

2.4.1. Microsatellites

Currently, microsatellite markers are the most widely accepted molecular tool to estimate genetic diversity within breeds (Aranguren-Mendez *et al.* 2002; Morais *et al.* 2005; Glowatzki-Mullis *et al.* 2006; Plante *et al.* 2007). Microsatellites are simple tandem repeats that are spaced throughout the genome (Cañon *et al.* 2000; Aberle *et al.* 2004; Solis *et al.* 2005). The availability of horse microsatellite maps and amplification primer sequences, high mutation rates, selective neutrality, and simple Mendelian inheritance makes them ideal markers to measure genetic diversity, verify parentage, and to reconstruct the phylogenetic

relationships of both horse and pony breeds (Aranguren-Mendez *et al.* 2002; Achmann *et al.* 2004; Penedo *et al.* 2005; Glowatzki-Mullis *et al.* 2006; FAO 2007; Plante *et al.* 2007; Pérez-Gutiérrez *et al.* 2008; Eggert *et al.* 2010). The repeating segments vary in length and this is thought to be as a result of slipping by the DNA polymerase during replication (Binns *et al.* 1995). Microsatellite markers are also hypervariable, meaning that they have multiple alleles of intermediate frequencies (FAO 2007). The use of microsatellite loci in equine genetic diversity studies will be discussed in detail in Chapter 3 of this thesis.

2.4.2. MtDNA Sequence Data

MtDNA is a maternally inherited small circular genome (16,660 base pairs) generally thought to lack recombination (Xu & Árnason 1994). Within the mtDNA there are two regions, known as the Hyper Variable Regions (HVRI and HVRII) of the displacement loop (D-Loop) and the Cytochrome-B, that have been shown to be highly informative to estimate genetic diversity within closely related populations along with assessing breed origin and time of domestication (Vilà *et al.* 2001; Jansen *et al.* 2002; Aranguren-Mendez *et al.* 2004; Cozzi *et al.* 2004; Cothran *et al.* 2005; Pérez-Gutiérrez *et al.* 2008). Though there are two HVR, the HVRI tends to be slightly more informative than HVRII (Kornienko & Vodolazhsky 2010). The use of mtDNA sequence data in equine genetic diversity studies will be discussed in detail in Chapter 4 of this thesis.

2.5. Genetic diversity within the equine populations of interest

2.5.1. Canadian equine populations

The genetic diversity of Canadian populations has been studied to a limited extent. Behara (2000) compared the Canadian populations to other horse breeds using 25 microsatellite loci. This approach was useful in determining the phylogenetic relationship of the Canadian horse, found to be closely related to the French Trotter and Morgan horse, but not particularly useful in determining the phylogenetic relationship of the two remaining native Canadian pony breeds (Behara 2000). Cothran (2004) assessed the genetic diversity in 43 Lac La Croix ponies, using blood typing and 12 microsatellite loci, and concluded that the genetic variation in the Lac La Croix was slightly lower, with exception to the Friesian, than other breeds tested. Lastly, Plante *et al*. (2007) examined all four Canadian populations in comparison to other breeds using 12 to 16 microsatellite loci, and determined that the feral Sable Island herd was genetically distinct, but less diverse then the other breeds tested. In addition, it was observed that the Canadian horse was closely related to the Belgian and Percheron (draught breeds); whereas the Sable Island population was closely related to the Nordic breeds (Plante *et al.* 2007). MtDNA sequence data has not been examined in native Canadian equine populations.

2.5.2. Feral populations

Currently, and to the best of my knowledge, there is no published research regarding the genetic diversity of the Grand Turk or Saint-Pierre et Miquelon populations.

2.5.3. The Mountain and Moorland breeds

Information regarding the genetic diversity of the Mountain and Moorland breeds is also limited. To date, I am unaware of any information published regarding the estimation of genetic diversity and breed characterization encompassing all Mountain and Moorland pony breeds, although some individual Mountain and Moorland pony breeds have been investigated and compared to other equine breeds.

Cothran's (2004) report regarding the genetic diversity of the Lac La Croix pony also included the Fell, Shetland and Welsh pony breeds along with seven other horse breeds using 12 microsatellite loci, blood typing, and protein analyses. In this particular study it was observed that when using the 12 microsatellite loci the Shetland pony was the most diverse breed tested with an effective number of alleles (N_e; defined as the minimum number of alleles required to explain the diversity observed within a population) of 5.30 alleles, whereas the Fell pony had only a mid ranged N_e value of 4.03 (Cothran 2004). In contrast, the Shetland and Welsh pony breeds had mid range N_e values of 2.21 and 2.13 respectively when blood typing and protein analyses were used to estimate genetic diversity (Cothran 2004). A restricted maximum likelihood phylogenetic tree was also included in this report and included the Dale, Exmoor, Fell, Highland, Shetland, and Welsh pony breeds, although they were only briefly mentioned when concluding that the Lac La Croix pony likely came from Iberian descent (Cothran 2004).

The New Forest pony was also examined using blood typing and was compared to two other 'pony' breeds, the Camargue and Haflinger, and two horse breeds, the Arabian and the Thoroughbred (Kaminski & Duncan 1981). These authors noted that there were unique hemotypes in approximately half of the three pony breeds which was significantly higher

than the percent of unique hemotypes observed in the horse breeds, 10% in the Thoroughbred, and 18% in the Arabian (Kaminski & Duncan 1981). Cunningham *et al.* (2001) examined the genetic diversity of the Thoroughbred and included the Shetland pony in their study. These authors observed that the Shetland pony was the second least diverse breed following the Thoroughbred (Cunningham *et al.* 2001).

Aberle *et al.* (2004) included the Exmoor in their study in comparison to several draught breeds, and concluded that the breed was less diverse compared to more recently domesticated breeds, but more diverse than other primitive breeds, and shared common ancestry with the Sorraia (another primitive pony breed).

More recently Leroy *et al.* (2009) focused on a number of equine breeds (including the Connemara, New Forest, Shetland and Welsh pony breeds) found in France using 11 microsatellite loci, and observed that the European pony breeds formed a unique genetic cluster separate from Nordic, warmblood and coldblood breeds. These authors also observed deviations from the Hardy-Weinberg Equilibrium (HWE) for the Connemara and New Forest pony breeds, likely a result of their small population size. The Welsh and New Forest pony breeds were observed to be very diverse compared to other breeds tested, likely a result of the four sections, and partially open registry maintained in the past (Leroy *et al.* 2009; Lynghaug 2009).

The genetic diversity within the Mountain and Moorland pony breeds has also been investigated using mtDNA sequence data, and was first mentioned by Vilà *et al.* (2001) when the authors examined a large number of *Equus ferus caballus* samples and observed that all equine breeds formed six general clades. Included in this study were the Connemara and Exmoor pony breeds; however, they were not specifically discussed as the focus was on

when equines were first domesticated (Vilà *et al.* 2001). Jansen *et al.* (2002) further investigated maternal lineages in equine breeds and included the Exmoor, Fell, Highland, Shetland and Welsh pony breeds in their study. These authors concluded that the Northern European pony breeds formed a unique cluster, C1, that dated back approximately 1500 years (Jansen *et al.* 2002).

Royo *et al.* (2005) investigated the origin of Iberian breeds and included the Exmoor pony in their study. These authors concluded that most Exmoors were not strongly related to the Southern Iberian breeds, but they did share a few haplotypes with the Northern Iberian breeds (Royo *et al.* 2005). Most Exmoor ponies also formed their own cluster, likely the C1 cluster previously reported (Jansen *et al.* 2002; Royo *et al.* 2005). This information was later further supported by Aberle *et al.* (2007) when the authors confirmed that Exmoor ponies tended to fit into a single 'B' clade. McGahern *et al.* (2006) investigated the origins and genetic diversity of all native Irish equine populations (Irish Draught, Kerry Bog, and Connemara) and found that the majority of these breeds fit in the D clade (of Iberian descent) as defined by Jansen *et al.* (2002), with the exception the Kerry Bog which fit in a rare clade E and C, 31% and 26 %, of the time respectively.

2.5.4. The Nordic breeds

The genetic diversity of the Nordic equine breeds has been studied using both microsatellite loci and mtDNA sequence data. In addition to the Mountain and Moorland breeds studied by Leroy *et al.* (2009), these authors also included the Nordic breeds which also formed their own unique cluster close to the European pony breeds. Plante *et al.* (2007) also included the Nordic breeds while investigating the genetic diversity of the feral Sable

Island population. The Nordic breeds were found to be moderately diverse while having a few breed specific alleles (two in the Fjord, one in the Icelandic), and N_e values of 3.55 and 4.03 in the Fjord and Icelandic respectively (Plante *et al.* 2007). Bjørnstad *et al.* (2003) examined the possible influence of the Mongolian Domestic horse on the Norwegian breeds using 26 microsatellite loci, and found that the Icelandic and the Norwegian Fjord likely shared common ancestry dating back approximately 1000 years ago. The Icelandic was also found to be more closely related to the Mongolian Domestic horse compared to the Norwegian Fjord (Bjørnstad *et al.* 2003). Glowatzki-Mullis *et al.* (2006) included the Icelandic breed while investigating the Franches-Montagnes (a local Swiss horse breed) using 50 microsatellite loci. The Icelandic was found to have one breed specific allele and was moderately diverse when compared to the horse breeds included in the study; however, it was very distinct compared to the other breeds as it segregated out early when using a Bayesian approach (Glowatzki-Mullis *et al.* 2006).

MtDNA sequence data has also been used to study the origin and diversity of the Nordic breeds. Jasen *et al.* (2002) concluded that the Nordic breeds shared two main clades, the first being the C1 clade, common to the Northern European pony breeds and the second, a rare E clade unique to Shetland, Icelandic, and Fjord breeds.

2.6. Conclusions

Recent international interest in conservation genetics has resulted in many countries characterizing their genetic resources, including their livestock species (FAO 2007). Current methods for the evaluation of genetic diversity rely on the variation at the DNA level. The use of microsatellite markers along with mtDNA sequence data in equines are widely

accepted, and are often used together to help portray the most accurate picture of breeds or populations of interest. The molecular technique used is only one part of a complicated equation when estimating the genetic diversity in livestock species (FAO 2007). Collecting representative samples and choosing the proper models to fit the population for data analyses is also vital to help prioritize breeds for conservation.

Equines have clearly shaped human civilization and humans have also significantly impacted equine evolution; and although horse breeds have been extensively studied this is not the case with ponies (Kaminski & Duncan 1981; Cunningham *et al.* 2001; Vilà *et al.* 2001; Jansen *et al.* 2002; Bjørnstad *et al.* 2003; Aberle & Distl 2004; Royo *et al.* 2005; Glowatzki-Mullis *et al.* 2006; McGahern *et al.* 2006; Aberle *et al.* 2007; Plante *et al.* 2007; Leroy *et al.* 2009). The development of local breeds as a result of society needs and the local environmental conditions has resulted in the creation of many unique breeds; some with distinct traits, disease resistance or climate/environmental adaptations (FAO 2007). Breeds can also vary both at the genotypic and the phenotypic levels. Currently, in developed countries equines are preserved for their importance to recreation, culture/history and to some extent, their importance to agriculture (Luís *et al.* 2007a). Estimating genetic diversity is just the first step in developing optimal conservation strategy, and it should be noted that conservation strategies will likely not succeed unless both the state governments and the public partake in the process (Gizaw *et al.* 2008).

2.6.1. Objectives

There are two main objectives driving this project. The first is to estimate the genetic diversity of the native Canadian, Mountain and Moorland, and Nordic pony breeds using microsatellite loci (Chapter 3) and mtDNA sequence data (Chapter 4). The second is to reconstruct the ancestry of the native Canadian populations, and examine the possible influence of the Mountain and Moorland, Nordic, and historically important horse and feral populations on the development of the native Canadian equine populations using microsatellite loci and mtDNA sequence data.

3.0 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] & Y. Plante[1, 2]

[1]Dept. Animal & Poultry Science, University of Saskatchewan, Saskatoon, Saskatchewan, Canada; [2]Agriculture & Agri-Food Canada (AAFC), Saskatoon, Saskatchewan, Canada; [3]College of Veterinary Medicine & Biomedical Sciences, Texas A & M University, College Station, Texas, United States of America

Corresponding author: Jaclyn M. Prystupa. E-mail: jmp479@mail.usask.ca

3.1. Implications

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 (FAO 2007). In order for Canada to fulfill its obligation, populations of various livestock species within its borders must be properly assessed, with regards to genetic diversity (FAO 2007). The assessment of genetic diversity and admixture among Canadian, Mountain and Moorland and Nordic pony populations was performed to help increase the knowledge and understanding of the Canadian equine genetic resources.

* Accepted in 'Animal' May 19, 2011 (manuscript number ANIMAL-10-10860R2).
Contribution of authors: J. M. Prystupa provided genotypes for 22 to 26 microsatellite loci per population, performed the statistical analyses with the exception of the contribution to genetic diversity within the dataset, and wrote the manuscript. E.G. Cothran facilitated the collection of samples and provided critical review, along with the analyses for the contribution to genetic diversity within the dataset. R. Juras provided genotypes for 12 to 16 microsatellite loci for all samples provided by E.G. Cothran. F. C Buchanan provided samples and critical review. Y. Plante provided funding, samples, critical review, and insight with statistical analyses.

3.2. Introduction

The FAO (2007) recently reported that there are close to 700 horse and pony populations 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 (EST 2008; RBC 2009). Ponies are often phenotypically defined as individuals standing under 14.2hh at the withers, 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, small population size, and are maintained *in situ*. There is interest in understanding what breeds contributed to the development of the Newfoundland, as the 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 origins to the first shipment of French

horses that 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 which has had no human interference since 1960 (Plante *et al.* 2007). Prior to 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 (NSMNH 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, hypervariability, 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) reported that pony breeds tended to cluster distinctly and separately from horse breeds using 11 microsatellite loci 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 the 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 solely of pony breeds have not been performed, and studies

available today tend to only focus on a few breeds, or compare a large number of horse breeds to only a few pony breeds, and this may not be the best way to evaluate and compare the genetic diversity in pony breeds (Plante *et al.* 2007; Behl *et al.* 2007; Leroy *et al.* 2009).

In order to estimate the genetic diversity and phylogenetic relationships of 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 influence the development of some pony breeds and/or native Canadian equine populations) using 38 microsatellite loci.

3.3. Materials and Methods

3.3.1. 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; CDN and 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 Mongolian Domestic samples. All feral populations and Mongolian Domestic horse samples were collected from their

respective native regions. In addition, the Eriskay, New Forest, and Exmoor pony samples were collected in the United Kingdom, while the Shetland pony 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: 0.32M Sucrose, 10mM Tris-HCl, pH 8.0, 50mM KCl, and 0.5% TWEEN was mixed with 200µl aliquots of whole fresh or frozen-thawed blood. The leukocytes were pelleted and the supernatant was aspirated. Each sample was washed twice with 400µl of the same solution prior to DNA extraction.

Microsatellite loci were chosen based on a previous study by Glowatzki-Mullis *et al.* (2006) and the FAO's Measurement of Domestic Animal Diversity 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 1µl of DNA template and commercially available kits and protocols (Amplitaq Gold, Applied Biosytems, Foster City, California, USA or Qiagen Multiplexing (Qiagen Inc., Burlington, Ontario, Canada). Details regarding these loci can be found in Appendix A.

The microsatellite amplicons were purified using Agencourt® AMPure® (Agencourt Bioscience Corporation/ Beckman Coulter Company, Mississauga, Ontario, Canada) using the recommended protocol. Samples (0.8µl of purified PCR product, 1µl of 600-Liz size

standard and 8.2µl of Hi-Di Formamide (Applied Biosystems)) were denatured for 5 minutes at 95ºC, quenched on ice for 2 minutes, 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).

3.3.2. Statistical Analysis

Deviations from HWE were estimated using MICRO-CHECKER (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 1995; Rousset 2008). GENALEX v. 6 (Peakall & Smouse 2006) was used to estimate the average (N_a), effective (N_e) number of alleles, and F_{IS}. ARLEQUIN v. 3.5 (Excoffier & Lischer 2010) was used to compute the analysis of molecular variance (AMOVA), and observed (H_O) and expected (H_E) heterozygosities. Allelic richness was estimated using FSTAT (Goudet 2001).

MICROSATELLITE ANALYZER (Dieringer & Schlötterer 2003) was used to estimate and bootstrap (1000 times) three genetic distances: Nei's standard (Nei 1972, 1978), 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 v. 3.66 (Felsenstein 1989-2006). Trees were created using the neighbor-joining method (Saitou & Nei 1987), combined using CONSENUS (as implemented in PHYLIP v. 3.66), and drawn using SPLITSTREE4 (Huson 1998; Huson & 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 v. 6. STRUCTURE v. 2.1 (Pritchard *et al.* 2000), with the parameter settings: K 10 to 30; 50,000 burn-in; 100,000 MCMC (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.* (2005) approach. CLUMPP (Jakobsson & Rosenberg 2007) and DISTRUCT (Rosenberg 2004) were used to cluster the 10 independent runs into a single figure.

A contribution of each population to overall genetic diversity was assessed using marginal loss (Weitzman 1992, 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 dataset was also assessed using the Petit *et al.* (1998) approach and molecular kinship analyses as proposed by Caballero and Toro (2002) and implemented in MOLKIN v. 2 (Gutiérrez *et al.* 2005).

3.4. Results and Discussion

In total, 821 individuals were examined across 24 populations. The analysis of the microsatellites revealed the presence of 468 alleles in the 38 microsatellite loci examined. This number is comparable to other work 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 7 pairs of significantly linked loci ($P \leq 0.05$) 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 which 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 ($P \leq 0.05$). Deviations from HWE have been reported previously in small populations (such as the Hucul, Icelandic, Jaca Navarra and Pottok), or populations which 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 also could be a result of genetic drift, non-random 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, therefore also explaining some of the deviations from the HWE observed. Specifically, the Lac La Croix pony was bottlenecked to four females in 1977; and according to historical records, the Saint-Pierre et Miquelon population arose from approximately 75 individuals left on the islands in the 1700s (Personal Communication with Patricia Morris, Newfoundland Pony Society; Lynghaug 2009). Published research has 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 3.4.1. 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 pony and Sable Island herd. 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 purebred breeding stallion (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 Newfoundland pony breed. Allelic richness 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 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 Appendix B. Within 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.* 2007b; 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 breeds (Cothran 2004; Luís *et al.* 2007b; 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 pony 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 (less than 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).

Table 3.4.1 A summary of the basic statistics per population (± Standard deviation (SD)) 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 F_{IS}.

Breed/ Populations	*N*	$N_a \pm SD$	$N_e \pm SD$	*AR*	$H_o \pm SD$	$H_e \pm SD$	F_{IS}
Canadian pony							
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							
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							
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

Individual assignment tests based on the maximum likelihood methods to predefined populations found that 98% of individuals were correctly assigned to their respective group (Table 3.4.2). Of those that could not be properly assigned: one Kerry Bog pony, assigned to the Welsh group; two Welsh ponies, assigned to the Icelandic and New Forest groups; six Newfoundland-CDN ponies, four assigned to Newfoundland-SK group, one to the Welsh group and one to the Saint-Pierre et Miquelon population; and one Newfoundland-SK pony assigned to the Saint-Pierre et Miquelon population. It is not surprising that some animals were assigned to the Welsh pony breed which are not actually members of that breed. 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 breed, and it is believed that this sample was either mislabelled or that the inidividual was not a purebred since these two breeds are not closely related.

All phylogenetic reconstruction using the genetic distances estimated showed similar topologies with low bootstrap values indicating that these breeds have likely diverged recently. Figure 3.4.1 illustrates the POSA approach which produced the phylogenetic tree making the most biological sense. The Lac La Croix and the Eriskay pony breeds 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 appear to be two main groups observed in all the phylogenetic trees. The first is the Mountain and Moorland pony breeds which group together, as expected, and shared common ancestry. An interesting observation

Table 3.4.2 The individual assignment test results in percent based on a maximum likelihood method to predefined populations/breeds.

Population	Sample (n)	Assignment to self (%)
Caspian	25	96.0
Clydesdale	50	100.0
Connemara	37	100.0
Dale	25	100.0
Dartmoor	25	100.0
Eriskay	27	100.0
Exmoor	25	100.0
Fell	25	100.0
Fjord	50	100.0
Grand Turk	17	100.0
Haflinger	25	100.0
Highland	25	100.0
Icelandic	49	100.0
Kerry Bog	11	90.9
Lac La Croix	60	100.0
Mongolian	35	100.0
New Forest	26	100.0
Newfoundland- CDN	53	88.7
Newfoundland- SK	19	94.7
Sable Island	60	100.0
Saint-Pierre et Miquelon	29	100.0
Shetland	28	100.0
Standardbred	22	100.0
Welsh	48	95.8

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 expected to be

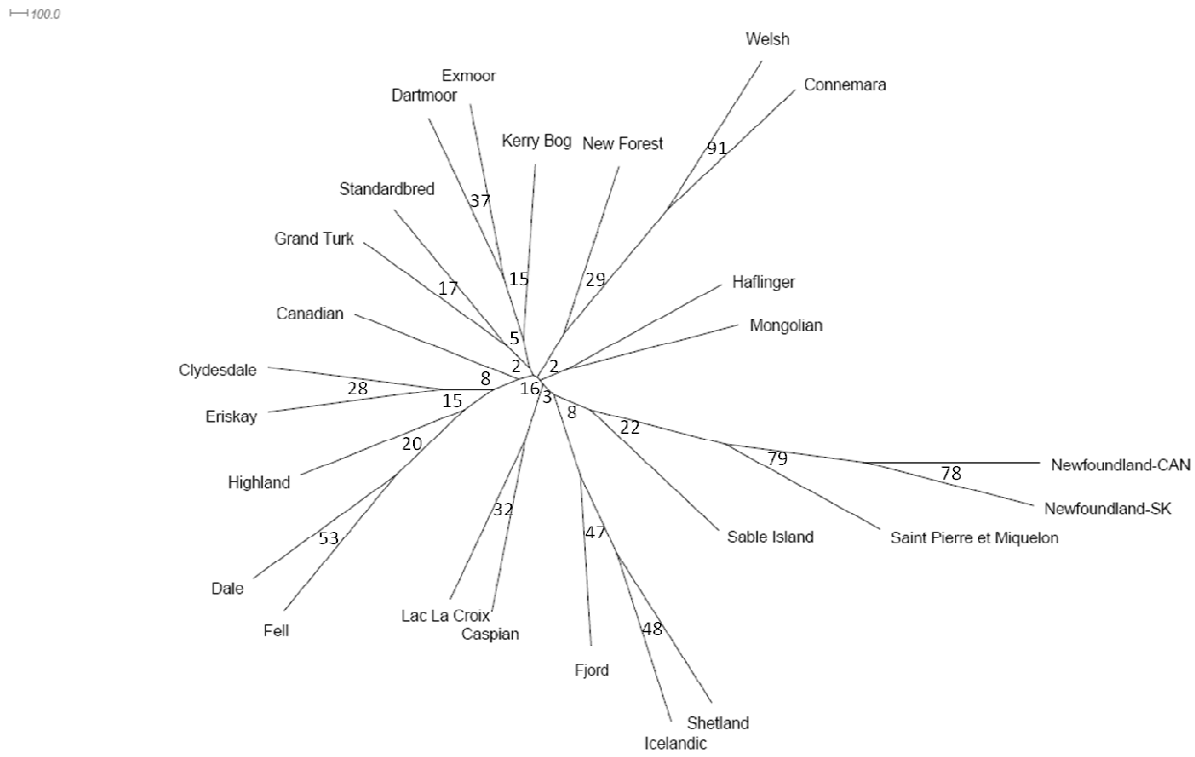

Figure 3.4.1 The un-rooted population phylogenetic tree created using a POSA approach. Individual confidence values on each branch are also included.

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 group several breeds also consistently shared several similarities among the phylogenetic networks. 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 the 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 networks is the combination of the Nordic breeds, Canadian pony breeds, Sable Island and Saint-Pierre et Miquelon populations. Though the Nordic breeds always grouped together, they also share common ancestry with two 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 3.4.2 illustrates the individual un-rooted 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 Connemara, Eriskay and Welsh pony 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. 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.

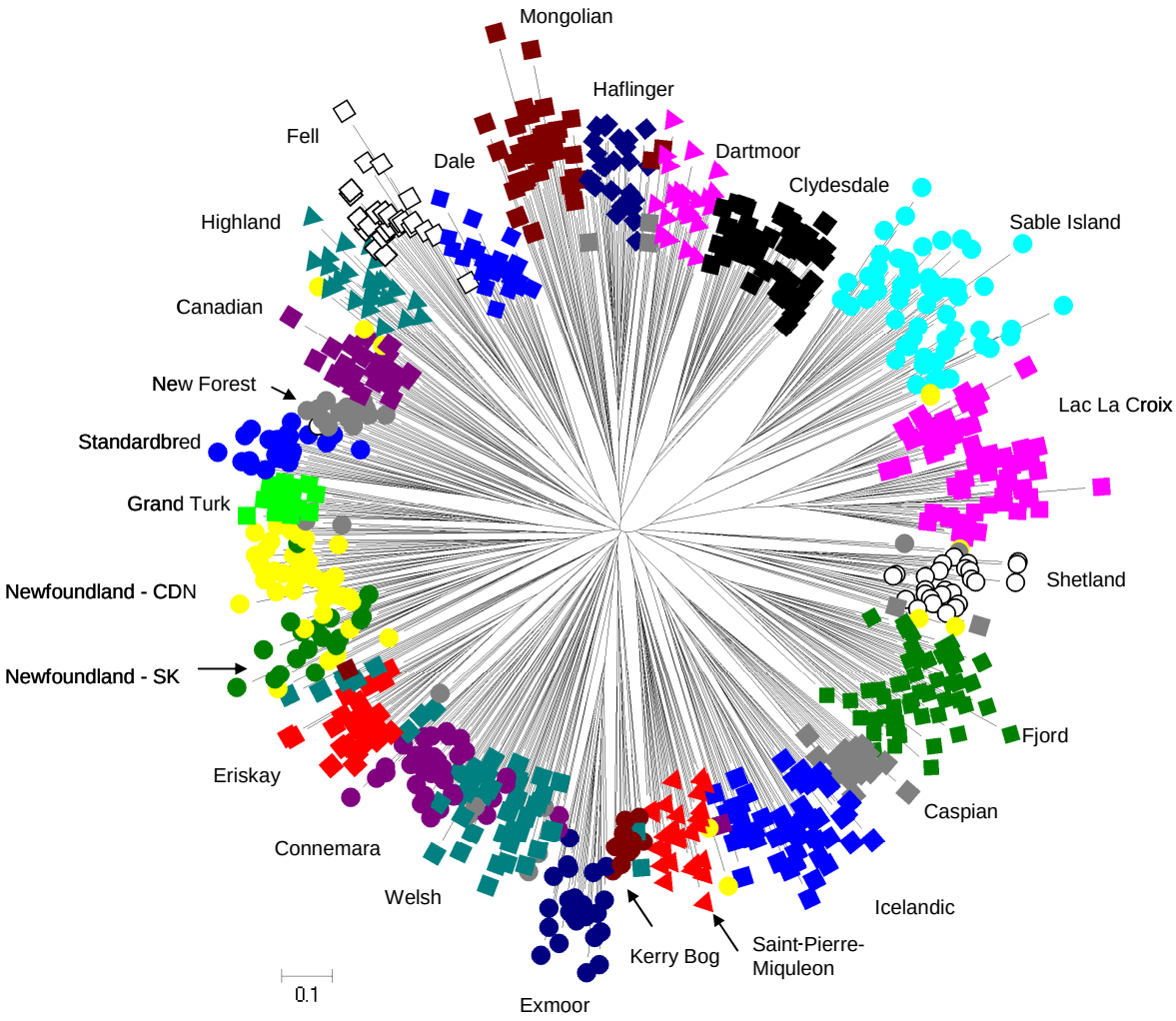

Figure 3.4.2 The un-rooted individual phylogenetic tree created using a POSA pairwise difference matrix**.**

The Bayesian analysis provided a different way to investigate the individuals within this study and cluster them together without prior breed information. STRUCTURE HARVESTER determined that the most likely K value was 21, although some substructure was observed at K = 16. When K =16 (Table 3.4.3; Figure 3.4.3) there appeared to be a strong admixture, varying from one to many breeds, and some breeds grouped into clusters together.

Table 3.4.3 The Bayesian analyses (for K=16). The values under columns I-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).

Population	Assignment to inferred clusters															
	I	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**	**IX**	**X**	**XI**	**XII**	**XIII**	**XIV**	**XV**	**XVI**
Canadian	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	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	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	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	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	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	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	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	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	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	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	**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	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	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	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	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	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	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	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	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	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	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	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	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	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

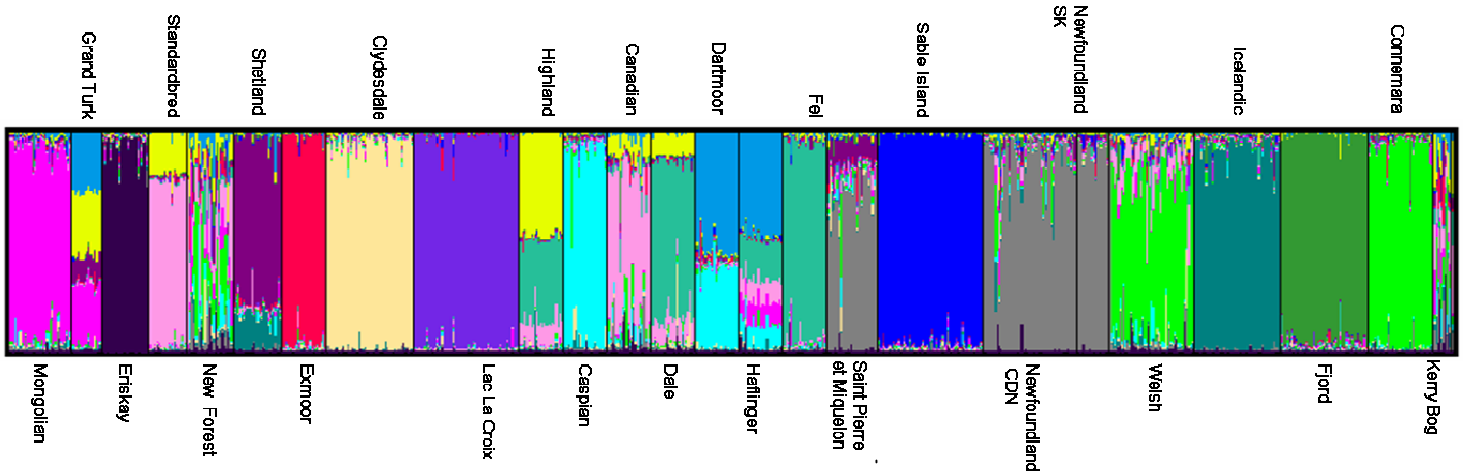

Figure 3.4.3 The consensus band for the 10 independent structure runs when K = 16, each single line represents one individual. Individual and populations with multiple colours indicate a genotypes matching and sharing admixtures with other populations respectively.

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 amongst each other. The Caspian, interestingly, appeared to share common ancestry with the Dartmoor and Haflinger and may be a result of the 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 (Table 3.4.4; Figure 3.4.4) 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 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 pony 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

Table 3.4.4 The results from the Bayesian analyses (for K=21). The values under columns I-XI 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).

Population	Assignment to inferred clusters										
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI
Canadian	0.002	0.003	0.003	0.005	**0.858**	0.010	0.007	0.012	0.007	0.008	0.006
Caspian	0.003	0.008	0.005	0.006	0.006	0.008	0.022	0.008	0.004	0.006	0.008
Clydesdale	0.002	0.005	0.006	0.003	0.004	0.006	0.004	0.005	0.004	0.003	0.006
Connemara	0.004	0.006	0.008	0.004	0.005	0.011	**0.779**	0.094	0.005	0.005	0.004
Dale	0.003	0.005	0.003	0.003	0.016	0.009	0.098	**0.354**	0.007	0.008	0.008
Dartmoor	0.006	0.004	0.006	0.004	0.012	0.008	0.006	0.007	0.005	0.006	0.007
Eriskay	0.002	0.003	0.005	0.003	0.003	0.007	0.004	0.004	0.006	0.006	0.003
Exmoor	0.006	0.002	0.004	0.002	0.004	0.095	0.004	0.004	0.006	0.006	0.003
Fell	0.007	0.004	0.017	0.003	0.016	0.003	0.010	**0.356**	0.005	0.008	0.025
Fjord	0.003	0.004	0.004	0.004	0.003	0.003	0.004	0.004	0.002	0.005	0.005
Grand Turk	0.003	0.007	0.005	0.002	0.004	0.005	0.005	0.005	0.007	0.007	0.003
Haflinger	0.004	0.005	0.004	0.003	0.003	0.004	0.006	0.005	0.003	0.003	**0.907**
Highland	0.003	0.013	0.004	0.002	0.011	0.006	0.005	0.007	0.005	0.004	0.005
Icelandic	0.004	**0.897**	0.007	0.004	0.005	0.004	0.006	0.005	0.003	0.008	0.004
Kerry Bog	0.004	0.007	0.005	0.003	0.009	0.046	0.006	0.031	0.008	0.009	0.006
Lac La Croix	0.006	0.003	0.004	**0.944**	0.002	0.002	0.002	0.003	0.002	0.003	0.002
Mongolian	0.007	0.007	**0.834**	0.006	0.005	0.010	0.007	0.008	0.006	0.005	0.017
New Forest	0.007	0.009	0.038	0.017	0.027	0.077	0.028	**0.191**	0.056	0.032	0.013
Newfoundland-CDN	0.008	0.017	0.012	0.006	0.017	0.025	0.017	0.029	0.007	0.010	0.005
Newfoundland-SK	0.014	0.003	0.005	0.004	0.009	0.024	0.004	0.016	0.002	0.002	0.003
Sable Island	**0.936**	0.004	0.004	0.003	0.003	0.003	0.003	0.003	0.002	0.005	0.003
Saint-Pierre et Miquelon	0.004	0.009	0.013	0.014	0.018	0.183	0.015	0.098	0.010	0.019	0.010
Shetland	0.004	0.007	0.006	0.003	0.004	0.004	0.007	0.006	0.009	**0.888**	0.003
Standardbred	0.002	0.005	0.003	0.005	0.006	0.008	0.004	0.005	**0.898**	0.010	0.003
Welsh	0.007	0.011	0.008	0.008	0.025	**0.479**	0.041	0.045	0.016	0.005	0.009

Table 3.4.4 The results from the Bayesian analyses (for K=21) continued. The values under columns XI I-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).

	Assignment to inferred clusters									
Population	**XII**	**XIII**	**XIV**	**XV**	**XVI**	**XVII**	**XVIII**	**XIX**	**XX**	**XXI**
Canadian	0.007	0.007	0.005	0.005	0.008	0.012	0.010	0.010	0.007	0.008
Caspian	0.003	0.003	**0.864**	0.017	0.005	0.004	0.006	0.005	0.005	0.004
Clydesdale	0.002	0.003	0.004	**0.915**	0.005	0.004	0.005	0.003	0.004	0.005
Connemara	0.010	0.014	0.006	0.005	0.004	0.004	0.009	0.012	0.007	0.004
Dale	0.005	0.007	0.004	0.027	0.169	0.169	0.009	0.005	0.006	0.085
Dartmoor	0.013	0.004	0.013	0.006	0.005	0.003	0.007	0.007	0.006	**0.865**
Eriskay	0.003	0.005	0.003	0.003	0.005	**0.910**	0.009	0.005	0.007	0.004
Exmoor	**0.826**	0.003	0.003	0.003	0.003	0.004	0.004	0.003	0.004	0.012
Fell	0.004	0.006	0.007	0.004	0.344	0.161	0.005	0.006	0.003	0.007
Fjord	0.003	0.003	0.003	0.003	0.009	0.003	0.003	**0.922**	0.004	0.006
Grand Turk	0.006	0.003	0.004	0.004	0.005	0.003	0.006	0.005	**0.906**	0.006
Haflinger	0.002	0.003	0.003	0.009	0.005	0.007	0.004	0.003	0.003	0.013
Highland	0.006	0.004	0.004	0.004	**0.888**	0.006	0.011	0.003	0.004	0.005
Icelandic	0.003	0.005	0.005	0.004	0.006	0.004	0.005	0.011	0.004	0.005
Kerry Bog	0.010	0.008	0.008	0.004	0.015	0.009	**0.787**	0.004	0.008	0.015
Lac La Croix	0.005	0.002	0.003	0.003	0.002	0.002	0.003	0.002	0.002	0.003
Mongolian	0.006	0.009	0.014	0.008	0.004	0.005	0.012	0.012	0.012	0.006
New Forest	0.022	0.029	0.032	0.018	0.036	0.026	0.255	0.023	0.030	0.039
Newfoundland- CDN	0.028	0.739	0.008	0.005	0.006	0.007	0.022	0.011	0.013	0.012
Newfoundland- SK	0.037	**0.836**	0.003	0.004	0.005	0.005	0.007	0.010	0.004	0.003
Sable Island	0.004	0.004	0.003	0.002	0.003	0.003	0.003	0.004	0.003	0.002
Saint-Pierre et Miquelon	0.092	0.432	0.016	0.023	0.007	0.008	0.008	0.011	0.004	0.007
Shetland	0.005	0.004	0.005	0.003	0.003	0.005	0.008	0.005	0.007	0.014
Standardbred	0.005	0.006	0.004	0.003	0.005	0.005	0.007	0.003	0.007	0.006
Welsh	0.010	0.014	0.013	0.013	0.011	0.014	0.248	0.011	0.006	0.008

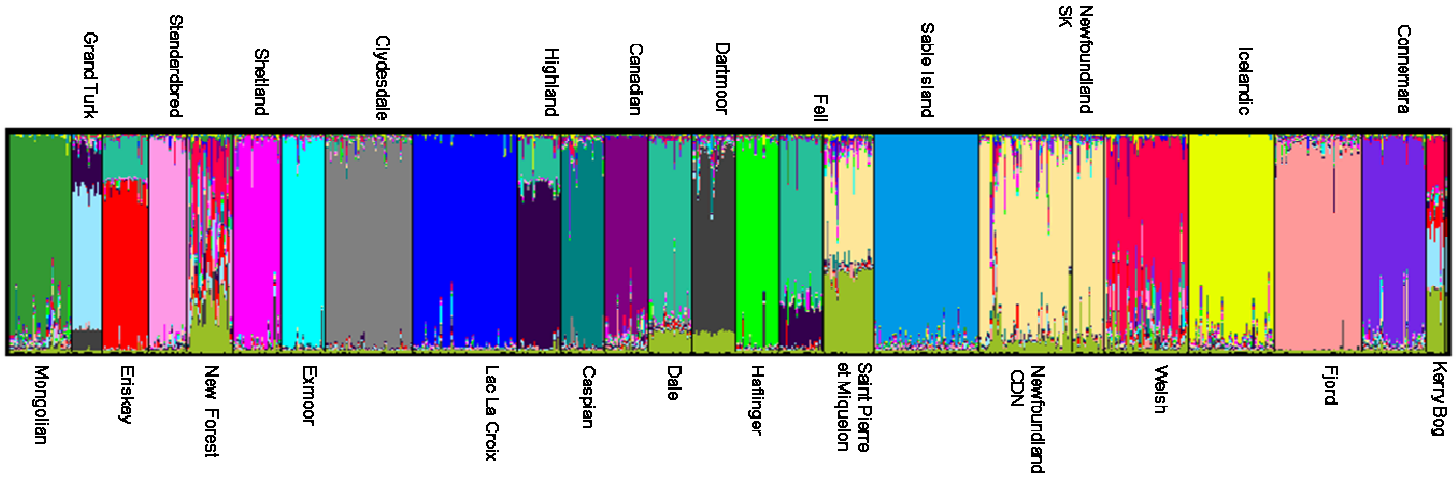

Figure 3.4.4 The consensus band for the 10 independent structure runs when K = 21, each single line represents one individual. Individual and populations with multiple colours indicate a genotypes matching and sharing admixtures with other populations respectively.

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 a 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 to horse breeds (Glowatzki-Mullis *et al.* 2006; Plante *et al.* 2007).

The contribution to overall genetic diversity (Table 3.4.5) 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, 1993) favoured the Lac La Croix pony as having the greatest marginal loss (6.57%) of all the populations tested when removed from the dataset, and was followed closely by the Exmoor (6.06%) and Eriskay (5.66%) breeds. In contrast, the Welsh pony 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.* (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 dataset. The populations that

Table 3.4.5 Results of the contribution to genetic diversity of each population in percent within the dataset using three different approaches: Weitzman (Weitzman 1992, 1993), Petit *et al.* (1998), and Caballero and Toro (2002).

Breed/ Populations	Weitzman Marginal Loss (%)	Petit *et al.* (1998) Internal Diversity (%)	Caballero and Toro (2002) Gain/Loss (%)
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

contributed the least amount of diversity to the dataset 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 (2002) approach favoured 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 dataset examined with values of 0.14, 0.07, and 0.05% respectively. A previous study using fewer microsatellite loci also has found the removal of the Sable Island herd from their dataset to also greatly reduce the amount of diversity observed when comparing the feral herd to horse breeds (Plante *et al.* 2007).

3.5. 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; as well 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 population 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 work will have to be done 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 mtDNA sequence data, and/or currently available genealogical and phenotypic data to produce an optimal and effective conservation strategy for all breeds examined.

4.0 Maternal lineages in Canadian, Mountain and Moorland, and Nordic pony populations[†]

J. M. Prystupa[1, 2], E. G. Cothran[3] & Y. Plante[1, 2]
[1]Dept. Animal & Poultry Science, University of Saskatchewan, Saskatoon, Saskatchewan, Canada; [2]Agriculture & Agri-Food Canada (AAFC), Saskatoon, Saskatchewan, Canada; [3]College of Veterinary Medicine & Biomedical Sciences, Texas A & M University, College Station, Texas, United States of America

Corresponding author: Jaclyn M. Prystupa. E-mail: jmp479@mail.usask.ca

4.1. Introduction

Viable conservation strategies are important for both domestic and wild species alike. One of the first, and crucial, steps in the development of any conservation strategy is to properly assess and characterize populations of interest at the phenotypic and genotypic levels (Glowatzki-Mullis *et al.* 2006; Plante *et al.* 2007; Gizaw *et al.* 2008). Different molecular tools and techniques have been utilized to assess the genetic diversity present within a species, including mtDNA sequence data (Cothran *et al.* 2005; Luís *et al.* 2006; Pérez-Gutiérrez *et al.* 2008). In particular, the HVR of the D-Loop, have been found to be useful when estimating genetic diversity and phylogenetic relationships among closely related populations (Aranguren-Méndez *et al.* 2001; Cozzi *et al.* 2004; Cothran *et al.* 2005; McGahern *et al.* 2006).

Previous studies examining the origin of *Equus ferus caballus* using mtDNA sequence data have reported a large number of ancestral haplogroups (17 to 19) belonging to

[†] Submitted to the 'Journal of Heredity' (manuscript number JOH-2011-068) in March 2011. Contribution of authors: J.M. Prystupa performed all laboratory work and statistical analyses. E.G. Cothran provided samples and critical review. Y. Plante provided funding, samples and provided insight with statistical analyses.

six or seven main clades (Vilà *et al.* 2001; Jansen *et al.* 2002; Cieslak *et al.* 2010). This is significantly higher than in other farm animal species, including: cattle (Jia *et al.* 2010), sheep (Pedrosa *et al.* 2005), goats (Naderi *et al.* 2007), and donkeys (Beja-Pereira *et al.* 2004). Possible explainations for the differences observed among the species is likely a result of the domestication process for each species. *Equus ferus caballus* is thought to have multiple origins of domestication (both in geography and time) along with a strong bias for the recuriment of mares over stallions from the wild (Vilà *et al.* 2001; Jansen *et al.* 2002; Wallner *et al.* 2003; Cieslak *et al.* 2010). A recent study has concluded that 39 of 87 ancestral haplotypes found in various ancient horse remains (dating from 12,000 BC to 1,000 AD) are still present in modern day horses and ponies (Cieslak *et al.* 2010). In addition, new haplotypes have since diverged (Kavar *et al.* 1999; Bowling *et al.* 2000).

Ponies are generally classified as any member of *Equus ferus caballus* which stands under 14.2 hh at the withers, although there are some horse breeds such as the Caspian and Haflinger that also stand under this height restriction (Pickeral 2001). In general, pony breeds are well known for their strength, hardiness, endurance, surefootedness, versatility and unique disposition. Within the classification of pony there are two well-known groups: the Mountain and Moorland breeds (Connemara, Dale, Dartmoor, Eriskay, Exmoor, Fell, Highland, Kerry Bog, New Forest, Shetland and Welsh), and the Nordic breeds (Fjord, Icelandic and Shetland). The Shetland is generally thought to link these two groups together. The Mountain and Moorland breeds originated in the British Isles, and as a result of multiple population reductions and bottlenecks experience throughout history, several of these breeds are listed as endangered by various conservation organizations (EST 2008; Lynghaug 2009;

RBC 2009). The Nordic breeds originated from a few different countries. In addition to the Shetland originating in the British Isles, the Norwegian Fjord and Icelandic arose from which they are named.

Canada is also home to two different pony populations, the Lac La Croix and the Newfoundland, which are maintained *in-situ* and are also listed as endangered by a few conservation groups (EST 2008; Lynghaug 2009; RBC 2009). Currently, the breed societies associated with these populations are trying to register them under the Canadian Animal Pedigree Act. The Newfoundland pony originated from the province of Newfoundland and Labrador, and herds of ponies were frequently left to roam feral when not required as a source of labour (Lynghaug 2009). Due to this early management strategy the exact breeds used to develop the Newfoundland pony are unknown, but are generally thought to originate from a number of the Mountain and Moorland breeds (Lynghaug 2009). The Lac La Croix pony originated from the Great Lakes area (between the Canada – United States of America international border), and is believed to have arisen from crossing Canadian horses (breed) with Mustangs. This population was reduced to four females in 1977 (Behara 2000). In order to save the population these mares were bred with Spanish Mustangs and were left to run feral on private property until the late 1990s (Lynghaug 2009). A possible link in the development of the Canadian pony populations may be from the feral Sable Island population (located off the East coast of Canada). The Sable Island horse population has been protected from human interference since 1960, but prior to that individual horse owners would leave stallions on the island to breed with the mares (Plante *et al.* 2007). The progeny would then be collected the following year and brought back to the mainland where they would be offered for sale (NSMNH 2001).

The declaration of the Convention on Biological Diversity served as an important cornerstone regarding the importance of genetic diversity within livestock breeds, as well as provided the first legal binding framework regarding estimating the genetic diversity within farm animal species in various countries around the world (FAO 2007). This study examines the genetic diversity and possible admixture among native Canadian, Nordic and Mountain and Moorland pony breeds, along with six horse breeds and three feral populations (which may have influenced the development of some of the pony breeds) using a 421 base pair section of the mtDNA D-Loop. Results from this study will not only serve as a partial fulfillment of the assessment of genetic diversity within equine populations in Canada, but also will aid the individual breed societies to amend their breeding programs to obtain the best possible conservation strategy.

4.2. Materials and Methods

4.2.1. Samples

In total, 280 (blood or hair) samples were randomly collected from 24 different populations. Six horse breeds (Caspian, Canadian, Haflinger, Mongolian, Clydesdale, and Standardbred) where included, along with three feral populations (Sable Island, Saint-Pierre et Miquelon, and Grand Turk) to investigate their possible influence on the development of the pony breeds within North America. Three broad groups of pony breeds were focused on in this study including the Mountain and Moorland (Connemara, Dale, Dartmoor, Eriskay, Exmoor, Fell, Highland, Kerry Bog, New Forest, and Welsh), Nordic (Fjord, Icelandic,

Shetland), and Canadian (Lac La Croix and Newfoundland). Each breed/population consisted of approximately 10 randomly selected ponies or horses when DNA sample template quality permitted.

4.2.2. DNA Extraction

DNA was extracted from either whole blood or hair follicles using the Troy *et al.* (2001) protocol. When whole blood samples were used two additional steps were included to lyse and remove the red blood cells. Small aliquots (200µl) of blood, fresh or frozen-thawed, were treated with 200µl of 0.32M Sucrose, 10mM Tris-HCl, pH 8.0, 50mM KCl, and 0.5% TWEEN. Samples were then spun to pellet the white blood cells and the supernatant was carefully removed. Each sample was then washed twice with 400µl of the same solution before extracting the DNA. Samples of inferior quality for sequencing were re-extracted from hair follicles (if available) using a DNeasy Blood & Tissue Kit (Qiagen Inc.) and a user-developed protocol titled the: "Purification of DNA from nails, hair, or feathers using DNeasy Blood & Tissue Kit-DY04" available on the Qiagen website (http://www.qiagen.com/literature/render.aspx?id=519).

4.2.3. PCR and Sequencing

A 421 base pair section of the mitochondrial D-Loop was amplified using PCR and previously published primers by Cothran *et al*. (2005). PCR reactions were made using Platinum®Taq polymerase (Invitrogen Canada Inc., Burling, Ontario) and the manufacturer's

recommendations. PCR amplicons were purified using Exonuclease-I and Shrimp Alkaline Phosphatase (Fermentas Canada Inc., Burling, Ontario) according to the manufacturer's specifications.

Sequencing reactions were made using BigDye™ Terminator Cycle Sequencing Kits v. 3.1 (Applied Biosystems) and the forward primer according to the manufacturer's protocols. Sequencing reactions were purified using Agencourt® CleanSEQ® (Agencourt Bioscience Corporation/ Beckman Coulter Company) using the recommended protocol. Purified sequencing products were loaded on to a Genetic Analyzer 3130xl (Applied Biosystems) equipped with a 50 cm capillary array and POP7 polymer. Sequences were examined using SEQMAN (Swindell and Plasterer 1997).

Sequences from the populations included in this study were entered into the National Center for Biotechnology Information (NCBI) Genbank database available at: http://www.ncbi.nlm.nih.gov/, and were assigned accession numbers HQ592784 to HQ593063. Additional accessions from 22 horse, pony and feral populations were also downloaded where DNA samples were not available. Details regarding all the accessions spanning the 46 populations can be found in Appendix C. In addition, accessions fitting into the major equines clades originally presented by Vila *et al.* (2001) were also included in this study. Exact accession numbers were chosen based on a previously published study (Pérez-Gutiérrez *et al.* 2008) and can also be found in Appendix C.

4.2.4. Statistical Analysis

All sequences were aligned to the NCBI accession X79547 (Xu and Árnason 1994), and shortened to 257 base pairs to allow for maximum sample size, using CLUSTALW (Thompson *et al*. 1994), as implemented in MEGA4 (Tamura *et al.* 2007). A BLAST search using the NCBI database was used to determine any previously unreported haplotypes. Several genetic diversity parameters were used for the 24 populations of interest, without the consideration of gaps, including the number of variable sites, number of haplotypes, and haplotype diversity using several of Nei's (1987) equations as implemented by DNASP v. 5 (Librado & Rozas 2009). Pairwise comparisons and nucleotide diversity were also estimated using ARLEQUIN v. 3.5 (Excoffier & Lischer 2010).

A median-joining (MJ) network showing the haplotypes proportionate to frequency was drawn, from all available sequences, using NETWORK 4.5 (Bandelt *et al.* 1999; Polzin & Daneshmand 2003) while down weighting "mutational hotspots" as recommended by Jansen *et al.* (2002) and Cieslak *et al* .(2010). Three additional statistical approaches (maximum parsimony, maximum likelihood and Bayesian) were also used to further investigate the phylogenetic relationships among the equine populations (horse, pony and feral) using programs available from the Cyberinfrastructure for Phylogenetic Research (CIPRES) Science Gateway available at: http://www.phylo.org/ (Miller *et al*. 2009). All programs were run with default CIPRES settings unless otherwise specified. Maximum parsimony was estimated using PAUPRAT (Sikes & Lewis 2001) with 200 iterations, 5 levels (5, 10, 15, 20, and 25%) of perturbed data, and 15 independent runs. Maximum likelihood was estimated using the General Time Reversible (GTR) model (Tavare 1986), with and without an estimate of the proportion of invariable sites (GTRGramma+1), as

implemented in the RAXML-HPC black box (Stamatakis 2006). The Bayesian approach was estimated using MR. BAYES (Huelsenbeck & Ronquist 2001) using the parameter settings of 2 runs, 4 chains, heating parameter of 0.2, sample frequency every 100 generations, burn-in fraction of 0.25, and 1 million generations (the number of generations were increased until the "average standard deviation of split frequencies" fell below 0.05) using a 4 by 4 nuclear model and Nst=6 setting in the CIPRES gateway. All phylogenetic consensus trees (rooted to Przewalski's horse; *Equus ferus przewalskii*; accession AFO72994) were depicted using MEGA4 (Tamura *et al.* 2007).

4.3. Results

In total, 61 haplotypes from 36 polymorphic sites were found in 23 of the 24 populations examined in this study. Unfortunately, due to poor DNA sample template quality only five New Forest samples could be sequenced and therefore were removed to prevent skewing of the analysis when trying to estimate the genetic diversity parameters (Dataset A). The New Forest samples were re-entered into the dataset in order to perform the BLAST search on all haplotypes arising from the 24 populations. This new dataset (Dataset B) revealed 62 haplotypes, and a subsequent BLAST search revealed that 20 of the 62 haplotypes found were not previously reported. The exact nuclear base substitutions and positions for the 20 unreported haplotypes (using Dataset B) can be found in Table 4.3.1. The unreported haplotypes included two breed specific haplotypes; one in the Icelandic and one in the Highland, along with one haplotype found only in pony breeds (Dale and Newfoundland). Two breed specific lineages were also confirmed during the BLAST search in the Shetland and Caspian breeds (Hill *et al.* 2002; Flannery & Cothran 2003).

Table 4.3.1 The summary of the 20 unreported haplotypes with specific nucleotide substitutions and positions within the mtDNA genome listed within the dataset

Haplotype	15494	15495	15496	15519	15521	15534	15538	15542	15546	15585	15595	15596	15597	15600	15601	15602	15603	15604	15635	15649	15650	15657	15659	15666	15703	15709	15718	15720	15726	15728	15737	15740
(X79547)	T	T	A	C	G	C	A	C	C	G	A	A	A	G	T	C	T	G	C	A	A	T	T	G	T	C	C	G	G	T	T	A
Hap_2	C	C	G	.	.	T	.	.	.	.	.	.	.	.	.	T	C	.	.	G	.	C	.	.	.	.	.	A	.	.	.	.
Hap_14	.	C	.	.	.	.	.	T	.	.	.	.	.	.	.	T	.	.	.	.	G	.	.	A	.	.	.	A	.	.	.	.
Hap_17	.	C	.	.	A	.	.	.	.	.	.	.	.	.	.	T	.	A	.	.	.	.	.	.	C	.	.	A	A	.	.	G
Hap_23	.	C	.	.	.	.	.	.	.	.	.	G	.	.	.	T	.	A	T	.	.	.	.	.	C	.	.	A	.	.	.	G
Hap_25	C	C	G	.	.	T	.	.	T	.	.	.	.	.	.	T	C	.	.	G	.	.	.	.	.	.	.	A	.	.	.	.
Hap_26	.	C	.	.	.	.	.	.	.	A	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	A	.	.	.	.
Hap_29	.	C	G	.	.	T	.	.	.	.	.	.	.	.	.	T	C	.	.	G	.	.	.	.	.	.	.	A	.	.	.	.
Hap_32	.	C	.	.	A	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.
Hap_33	.	C	.	.	A	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	T	A	.	.	C	.
Hap_35	.	C	.	.	.	.	.	.	.	.	.	G	.	.	.	T	.	A	.	.	.	.	.	.	C	.	.	A	.	.	.	G
Hap_39	.	C	G	.	.	T	.	.	.	.	.	.	.	.	.	T	C	.	.	G	.	C	.	.	.	.	.	A	.	.	.	.
Hap_41	C	C	C	.	.	T	.	.	.	A	.	.	.	.	.	T	C	A	.	G	.	.	.	.	.	.	.	A	.	.	.	.
Hap_43	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	A	.	.	G	.	.	.	C	.	.	A	A	.	.	G
Hap_46	C	C	G	.	.	T	.	.	.	A	.	.	.	.	.	T	C	A	.	.	.	.	.	.	.	.	.	A	.	.	.	.
Hap_49	C	C	G	.	.	T	.	.	.	A	.	.	.	.	.	T	C	A	.	G	.	.	.	.	.	.	.	A	.	C	.	.
Hap_51	.	C	.	T	.	.	G	.	.	.	.	.	G	.	.	T	.	.	.	.	G	.	.	.	.	T	.	A	.	.	.	.
Hap_54	C	C	G	.	.	T	.	.	.	.	.	.	.	.	.	.	C	.	.	G	.	.	.	.	.	.	T	A	.	.	.	.
Hap_56	.	C	.	.	.	.	G	.	.	A	.	G	.	.	.	T	.	.	.	.	G	.	.	.	.	T	.	A	.	.	.	.
Hap_60	.	C	.	.	.	.	.	.	.	.	.	.	.	A	C	T	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.
Hap_61	.	C	.	.	.	.	G	.	.	.	G	.	.	.	.	T	.	.	.	.	G	.	.	.	.	T	.	A	.	.	.	.

A summary of the results from the genetic diversity (Dataset A) estimates and haplotypes found (Dataset B) in each breed/population are presented in Table 4.3.2 and Table 4.3.3 respectively. Variable sites within the defined breeds and populations varied from only 3 in the Sable Island population to as high as 22 in the Newfoundland pony and the Canadian horse. The highest number of haplotypes (10) was observed in the Newfoundland pony, whereas the lowest number of haplotypes (2) and haplotype diversity (0.22 ± 0.17) was found in the Fell pony. The Sable Island population had the lowest nucleotide diversity (0.007 ± 0.009) and pairwise difference (0.286 ± 0.317). In contrast, the pony breeds with the highest

Table 4.3.2 The summary of the genetic diversity estimates for the 24 populations of interest including: sample size (n), number of variable sites, number of haplotypes, haplotype diversity (± SD), nucleotide diversity (± SD), and pairwise difference (± SD).

Breed/Population	n	Variable sites	# of Haplotypes	Haplotype Diversity ± SD	Nucleotide Diversity ± SD	Pairwise Difference ±SD
Native Canadian						
Canadian	12	22	6	0.85 ± 0.07	0.157 ± 0.090	6.455 ± 3.287
Lac La Croix	10	11	3	0.60 ± 0.13	0.107 ± 0.065	4.400 ± 2.372
Newfoundland	17	22	10	0.88 ± 0.06	0.114 ± 0.007	4.691 ± 2.417
Sable Island	21	3	4	0.29 ± 0.12	0.007 ± 0.009	0.286 ± 0.317
Mountain & Moorland						
Connemara	12	18	8	0.89 ± 0.08	0.160 ± 0.009	6.545 ± 3.328
Dale	12	14	5	0.58 ± 0.16	0.100 ± 0.006	3.893 ± 2.100
Dartmoor	12	10	3	0.44 ± 0.16	0.089 ± 0.005	3.636 ± 1.980
Eriskay	12	12	3	0.59 ± 0.11	0.106 ± 0.064	4.363 ± 2.319
Exmoor	11	18	6	0.84 ± 0.09	0.162 ± 0.094	6.654 ± 3.403
Fell	9	4	2	0.22 ± 0.17	0.022 ± 0.020	0.889 ± 0.681
Highland	11	15	6	0.84 ± 0.09	0.164 ± 0.095	6.727 ± 3.437
Kerry Bog	10	11	5	0.80 ± 0.10	0.096 ± 0.006	3.956 ± 2.162
New Forest	5	*Due to the small number of samples that could be sequenced from this breed it was removed from the dataset before analysis*				
Welsh	10	15	7	0.88 ± 0.11	0.007 ± 0.009	5.511 ± 2.895
Nordic						
Fjord	11	14	3	0.62 ± 0.10	0.156 ± 0.090	6.400 ± 3.285
Icelandic	12	15	7	0.88 ± 0.08	0.151 ± 0.087	6.181 ± 3.161
Shetland	17	17	6	0.80 ± 0.06	0.140 ± 0.079	5.735 ± 2.889
Feral						
Grand Turk	11	6	3	0.56 ± 0.13	0.046 ± 0.026	1.418 ± 0.936
Saint-Pierre & Miquelon	11	17	6	0.84 ± 0.09	0.150 ± 0.079	6.145 ± 3.166
Horse						
Caspian	10	15	8	0.96 ± 0.06	0.154 ± 0.091	6.333 ± 3.281
Clydesdale	10	16	7	0.93 ± 0.06	0.134 ± 0.080	5.511 ± 2.895
Haflinger	11	19	6	0.87 ± 0.07	0.154 ± 0.089	6.328 ± 3.251
Mongolian	11	15	8	0.95 ± 0.05	0.144 ± 0.084	5.891 ± 3.048
Standardbred	12	15	6	0.82 ± 0.10	0.109 ± 0.065	4.364 ± 2.319

Table 4.3.3 The summary of the haplotypes found within the 24 populations of interest. Bolded = new haplotypes, Underlined = breed specific.

Population/Breed	Haplotypes (n)
Native Canadian	
Canadian	Hap1 (1);Hap3 (3); Hap5 (1); Hap9 (2); Hap10 (1); Hap11 (4)
Lac La Croix	Hap3 (6); Hap20 (3); Hap27 (1)
Newfoundland	**Hap2 (1)**; Hap3 (4); Hap4 (5); Hap5 (1); Hap6 (1); Hap13 (1); Hap18 (1); Hap19 (1); **Hap39 (1)**; Hap57 (1)
Sable Island	Hap11 (18); Hap40 (1); **Hap41 (1)**; **Hap46 (1)**
Mountain & Moorland	
Connemara	Hap1 (4); Hap3 (1); Hap11 (2); Hap27 (1); Hap28 (1) **Hap29 (1);** Hap30 (1); **Hap35 (1)**
Dale	**Hap2 (1)**; Hap3 (1); Hap7 (8); Hap8 (1); Hap9 (1)
Dartmoor	Hap1 (9);Hap3 (1); Hap6 (2)
Eriskay	Hap9 (4); Hap20 (7); Hap59 (1)
Exmoor	Hap1 (1); Hap5 (3); Hap8 (4); Hap10 (1); **Hap51 (1)**; Hap52 (1)
Fell	**Hap26 (1)**; Hap27 (8)
Highland	Hap3 (1); Hap4 (1); Hap9 (1); Hap16 (3); **Hap17 (4)**; Hap18 (1)
Kerry Bog	Hap7 (4); Hap19 (3); Hap20 (1); Hap21 (1); Hap27 (1)
New Forest	Hap4 (1); Hap 18 (1); Hap 50 (1); Hap55 (1); **Hap56 (1)**
Welsh	Hap1 (1); Hap4 (1); Hap18 (4); Hap19 (1); Hap36 (1); Hap37 (1); Hap38 (1)
Nordic	
Fjord	Hap11 (6); Hap31 (4); Hap60 (1)
Icelandic	Hap3 (1); Hap4 (1); Hap11 (2); Hap19 (1); **Hap32 (1); Hap33 (2)**; Hap34 (4)
Shetland	Hap10 (1); Hap16 (4); Hap20 (6); Hap34 (4); Hap53 (1); **Hap54 (1)**
Feral	
Grand Turk	Hap20 (1); Hap30(7); Hap50 (3)
Saint-Pierre & Miquleon	Hap3 (3); Hap13 (1); Hap18 (1); Hap19 (1); Hap42 (4); **Hap43 (1)**
Horse	
Caspian	Hap3 (1); Hap12 (2); Hap13 (2); **Hap14 (1)**; Hap15 (1); Hap24 (1); Hap44 (1); Hap45(1)
Clydesdale	Hap10 (2); Hap11(1); Hap13 (2); Hap18 (1); Hap19 (2); **Hap61(1)**; **Hap62 (1)**
Haflinger	Hap18 (3); Hap19 (2); Hap22 (3); **Hap23 (1)** ;Hap24 (1); **Hap25 (1)**
Mongolian	Hap3 (2); Hap11 (1); Hap19 (1); Hap28 (2); Hap47 (1); Hap48 (2); **Hap49 (1);** Hap50 (1)
Standardbred	Hap11 (1); Hap13 (1); Hap19 (2); Hap55 (5)**;** Hap57 (1); **Hap58 (2)**

values belonged to the Mountain and Moorland group, specifically the Connemara pony with a haplotype diversity of 0.89 ± 0.08; and the Highland pony with values of 0.164 ± 0.095 and 6.727 ± 3.437 for nucleotide diversity and pairwise difference, respectively.

All statistical approaches used in this study produced similar topologies amongst the haplotypes. The phylogenetic tree created using the Bayesian approach (Figure 4.3.1) made the most biological sense and produced a comparable, although much more confident, phylogenetic tree than either of the maximum likelihood trees. In addition, the maximum parsimony phylogenetic trees and MJ network also had strong similarities despite the fear that the various levels of perturbed data would cause a dampening of the data.

Although all clades described by Vilà *et al*. (2001) were identified within the Bayesian, maximum parsimony and maximum likelihood trees and MJ network depicted, the individual haplogroups, as described by Jansen *et al.* (2002), within the B, D, and F clades could not be differentiated with any confidence. When the "mutational hotspots" were taken into account the number of haplotypes decreased from 112 to 83. The MJ network (Figure 4.3.2) illustrates the frequency of each haplogroup along with the proportion of the Nordic, Mountain and Moorland, and Canadian populations in comparison to other horse, pony and feral populations present within the haplogroups.

Canadian populations appeared most frequently in the D clade, specifically the D1 and D3 haplogroups which combined together when the additional "mutational hotspot" reported by Cieslak *et al*. (2010) was accounted for. This combined haplogroup incorporated 89% of the feral Sable Island population, 60% of the Lac La Croix ponies, 58% of the Canadian horses, and 53% of the Newfoundland ponies. An additional 12% of the Newfoundland ponies were also found in the D2 haplogroup. A large proportion of the Lac La Croix ponies (30%) were also found in the E haplogroup. The remaining native Canadian

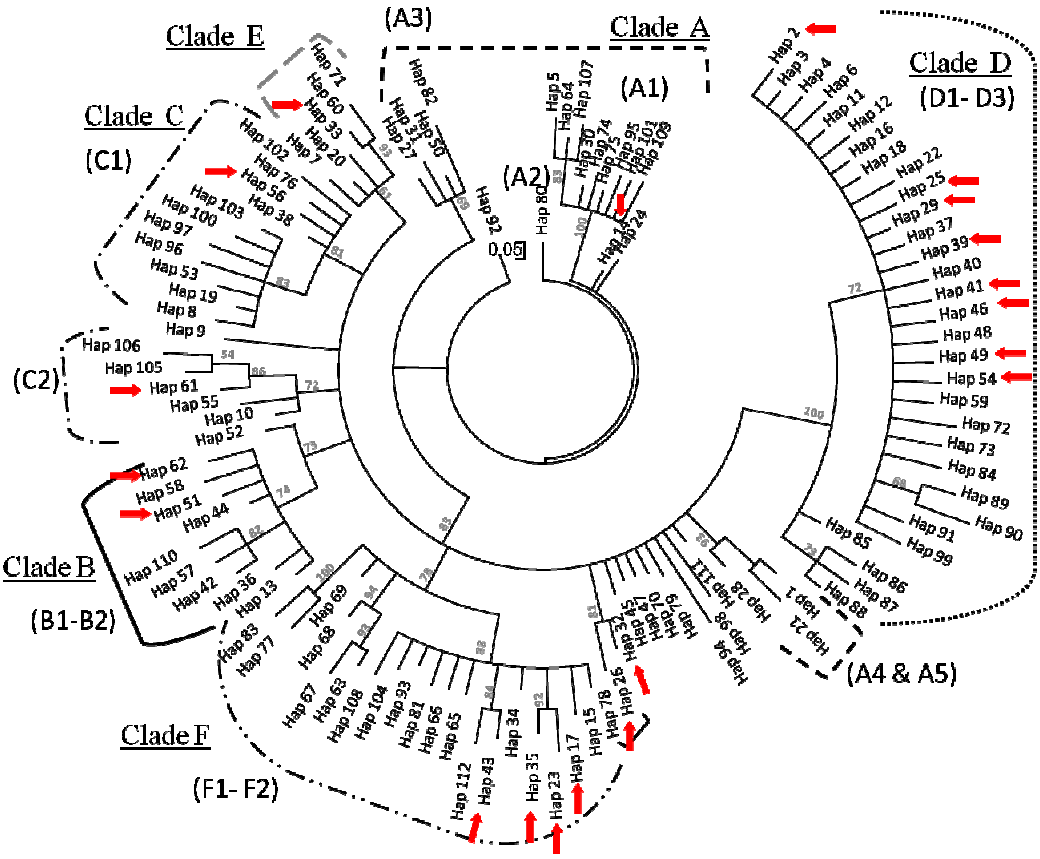

Figure 4.3.1 The phylogenetic tree created using a Bayesian approach of the 112 haplotypes observed in this study rooted to the Prezwalski haplotype (A2). Gray numbers on individual branches indicates the posterior probabilities associated with them. Red arrows show previously unreported haplotypes**.**

population samples were spread throughout the network including 10% of the Lac La Croix ponies in the A3 haplogroup, 6% of Newfoundland ponies in the B1 haplogroups, and 8% of the Canadian horses in the A5 and C2 haplogroups.

The Mountain and Moorland breeds were observed to be spread across the clades. The majority of the Fell ponies (88%) were found in the A3 haplogroup along with 10% of the Kerry Bog and 8% of the Connemara. On the other hand, 75% of the Dartmoor, 33% of

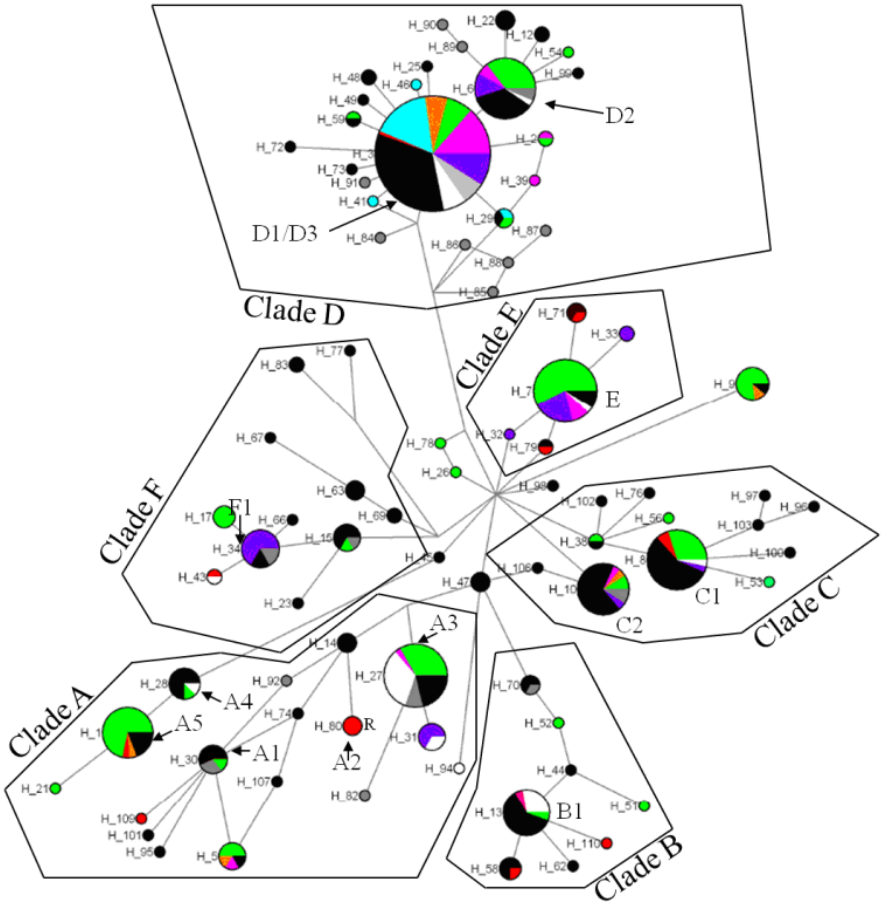

Figure 4.3.2 The MJ network representing all 46 equine populations and taking into account the "mutational hotspots" with the recommendations of previous research (Jansen *et al*. 2002; Cieslak *et al.* 2010). The major equine clades originally presented by Vila *et al.* (2001) along with the common haplogroups identified by Jansen *et al*. (2002) are labelled with circles proportionate to the frequency of haplotypes. Colours represent different equine groups; horse breeds = black, reference samples = red, feral populations = white, Mountain and Moorland breeds = green; Nordic breeds = purple; Canadian pony breeds = pink; Canadian feral populations = blue; Canadian horse populations = orange; all other pony populations =grey. "R" = Przewalski's horse.

the Connemara, 10% of the Welsh, and 9% of the Exmoor samples were found in the A5 haplogroup. A number of different breeds were also found in the C1 haplogroup including: Exmoor (36%), Kerry Bog (30%), Welsh (10%), and Dale (8%). An additional 20% of the

New Forest, 10% of the Welsh, and 9% of the Exmoor samples were found in the C2 haplogroup. Samples found in the D2 haplogroup included Welsh (40%), Highland (36%), New Forest (20%), Dartmoor (17%) and Connemara (8%). In comparison, the D1/D3 haplogroup included Connemara (25%), New Forest (20%), Welsh (20%), Highland (18%), Dartmoor (8%), and Dale (8%) breeds. A substantial proportion of the Dale (66%), Eriskay (58%), and Kerry Bog (50%) ponies were also found in the E haplogroup. The Mountain and Moorland breeds were also found to a lesser extent in the A1, A4, and B1 haplogroups which contained 8% of the Connemara, 8% of the Connemara, and 10% of the Welsh samples respectively.

The Nordic pony breeds were distributed across all clades with the exception of the B clade. The majority of the Nordic samples fell into the D clade, specifically 75% of the Fjord and 33% of the Icelandic ponies which fell into the D1/D3 haplogroup. In addition, 35% of the Shetland samples fell in the D2 haplogroup. A large proportion of the Nordic breeds (24% of the Shetland and 33% of the Icelandic) were also found to belong to the F1 haplogroup. The E haplogroup was also prominent within this group and included the 35 % of the Shetland, 8% Icelandic, and 8% of the Fjord. A very small percentage of the Icelandic (9%) and Shetland (6%) were also found in the C1 and C2 haplogroups respectively.

Several of the other horse, pony and feral populations included in this study were also spread throughout the clades. The A1 group consisted mostly of the Middle Eastern breeds such as the Arabian. All Przewalski's horse samples fell into the A2 subgroup. In addition, some haplogroups were dominated by specific populations such as the Grand Turk population which was found in the A3 haplogroup 91% of the time. Other haplogroups such as the A4 and A5 encompassed many breeds including the Mongolian Domestic, Mustang,

Noriker, Irish Draught, Pantanerio, Thoroughbred and Belgian. The B1 haplogroup consisted of individuals from various breeds and populations including 45% of the Saint-Pierre et Miquelon population, 20% of the Clydesdales, 20% of the Capsians, and 17% of the Standardbreds. The C1 haplogroup also included 20% of the Clydesdales, 18% of the Haflingers, 17% of the Standardbreds, along with 9% of the Mongolian Domestic and 9% of the Saint-Pierre et Miquelon population. Additional Mustang, Friesian and Cleveland Bay samples were also found in this haplogroup. The C2 haplogroup included 42% of Standardbred and 30% of the Clydesdale samples, 9% of the Saint-Pierre et Miquelon population, along with individuals from the Campolina breed. A large number of horse breeds from Europe, South America, and North America fell into the D clade. A small percentage of horse breeds fell into the E haplogroup including one Thoroughbred and one Akhal-Teke along with 9% of the Grand Turk population. The F clade contained mostly Arabians, Akhal-Tekes,Caspains along with Garrano and Asturcon ponies.

4.4. Discussion

The MJ network (Figure 4.3.2) revealed a strong admixture among horse, pony, and feral populations, and supports Yang *et al.*'s (2002) suggestion that haplotypes are not useful for breed assignment. From the genetic diversity parameters estimated for the 24 populations included in this study it can be said that the overall genetic diversity collectively observed within the pony breeds is comparable to that found within the horse breeds. Although some individual pony breeds may be slightly less diverse than some individual horse breeds, they may have haplotypes which are not frequently found in horse breeds and thus contribute to the overall genetic diversity within *Equus ferus caballus*.

The Nordic breeds in particular were found to share several haplotypes amongst each other, reemphasizing previous data about the close relationship of these breeds using microsatellites (Plante *et al.* 2007; Leroy *et al.* 2009). In addition, our present work was able to find more haplotypes than previously reported for the Icelandic and Exmoor pony breeds, in which only 2 and 3 haplotypes were previously reported respectively (Aberle *et al.* 2007). The low bootstrap values for the maximum likelihood trees have been reported previously in the German draft horse breeds (Aberle *et al.* 2007). The low bootstrap values among the various haplogroups within the B, D, and F clades in the phylogenetic trees may be a result of the crossing of individuals to produce new breeds, or for breed improvement. Vila *et al*. (2001) suggested that the high numbers of haplotypes without having breed specific lineages may be a result of breeding management or trade, as most breeds have been founded with a few stallions being bred to mares of various origins to develop new breeds types. Alternatively, Cieslak *et al.* (2010) recently reported that the large number of haplotypes observed today is a result of ancestral variability as opposed to the breeding management and flow of horses in the past.

Cieslak *et al.* (2010) found that the most prominent ancient haplotype in modern day horses and ponies is the D1and D3 haplogroup as reported by Jansen *et al.* (2002). This haplogroup has been found in specimens dating back to the Iron Age (1200-322BC) in both Asia and Europe (Cieslak et al. 2010). In addition, the D2 haplotype was the third most common ancient haplotype found today, and dates back as far as the Bronze Age (3200-1200BC). Other studies have also reported similar findings in which haplotypes found in the

D clade, as described by Vilà *et al*. (2001), was also commonly found in equines of British or Iberian descent (McGahern *et al.* 2006; Lei *et al.* 2009). Our findings therefore suggest that Canadian equine populations likely arose from British and Iberian stocks.

The most interesting finding in this study was the large proportion of pony breeds (Canadian, Mountain and Moorland, and Nordic) found in the E haplogroup. This haplogroup is closely related to another haplogroup, G1 as described by Cieslak *et al.* (2010), and was first found in Europe in specimens dating back as far as the Copper Age (3500-2300 BC; Cieslak *et al.* 2010). A large portion of haplogroups related to this rare haplogroup have since gone extinct, or are found very infrequently in modern day horses and ponies (Cieslak *et al.* 2010). The E haplogroup as described by Jansen *et al.* (2002), found today dates as far back to the Bronze Age in Siberia and Kazakhstan; and the Iron Age in China, Mongolia and Korea (Cieslak *et al.* 2010). Previous studies have also reported this haplogroup to be rare worldwide, appearing in only 3% in living equines, although very apparent (31%) in Kerry Bog ponies (Kakoi *et al.* 2007; Lei *et al*. 2009; McGahern *et al.* 2006). This study also found a high percentage of Kerry Bog ponies belonging to this group (50%) along with Dales, Eriskay, Shetlands, Lac La Croix, Icelandic and Fjord. This finding may provide added support to ensure the development of suitable conservation strategies for the Dale, Eriskay and Lac La Croix populations which are all considered endangered (EST 2008; RBC 2009).

In addition to a large proportion of individuals belonging to the D and E clades, many others belonged to the A clade, previously reported to be commonly found in horses and ponies of both European and Asian descent (McGahern *et al.* 2006; Lei *et al.* 2009). Cieslak *et al.* (2010) concluded that the A3 as reported by Jansen *et al.* (2002), is the second most common ancient haplogroup present today. Our work found that the Mountain and Moorland

breeds shared haplotypes with both each other in this haplogroup along with the Grand Turk population. One possible explanation for this may be that ponies from Europe were likely sent to the island at one time where they either remained, or were traded to other countries including Canada. A second interesting observation was that although the Fell and Dales ponies have very similar phenotypic traits and are very similar based upon protein and microsatellite variation (Luis *et al.* 2007b); these breeds did not share overlapping haplotypes. The majority of the Fell ponies were grouped within the A clade; whereas the majority of the Dale ponies grouped into the E clade.

The C clade was previously reported to contain predominately Northern European ponies (Jansen *et al.* 2002; McGahern *et al.* 2006). The C1 haplogroup has been found in pre-domesticated horse remains dating back as far as 12,000 BC in Europe and later in Siberia and Kazakhstan; therefore it is not surprising that previous studies have found European ponies to assign to the C1 haplogroup (Jansen *et al.* 2002; Cieslak *et al.* 2010). Our study found that some horse breeds, specifically the Standardbred and Clydesdale, were just as prominent in the C1 haplogroup as some of the Mountain and Moorland breeds. Interestingly, the F1 haplogroup is reported to have first arisen in the Eastern Asian region during the Bronze Age and can still be found in local primitive breeds (Cieslak *et al.* 2010). The strong presence of Nordic breeds within this haplogroup may indicate that mares used to found the Nordic breeds originated from East Asia.

Although most of the Canadian populations were found in the D clade, the Canadian pony breeds did not share overlapping haplotypes with the Sable Island population. The most common haplotype within the Sable Island population was also found in individuals from a number of different breeds including the: Canadian, Connemara, Fjord, Icelandic,

Mongolian, Standardbred, Andulasian, Quarter Horse, Thoroughbred, Mustang, Campolina, Shire, Irish Draught and Belgian, perhaps indicating that this population arose from several different types of equines and supporting historical data (NSMNH 2001). This study supports previous research involving microsatellite loci and the relationship of the Sable Island population and the Nordic breeds (Plante *et al.* 2007). The low diversity in the Sable Island population is most likely a result of the presence of one main haplotype with small mutations of only one base occurring over time to produce a new haplotype, alternatively, the four observed haplotypes may reflect the resource distribution on Sable Island, maternal lines associated with certain areas or territories, or the fluctuation in population size from year to year (Lucas *et al.* 2009).

The Newfoundland pony shared several haplotypes with horse breeds, pony breeds (Asturcon, Connemara, Dale, Dartmoor, Garrano, Highland, Icelandic, Kerry Bog, Lac La Croix, New Forest, Pottok, and Welsh) and feral populations (Saint-Pierre et Miquelon and Mustang), perhaps a result of early management and the combination of several types of origin and ancestry (Lynghaug 2009). In the past, there was the opinion that some Newfoundland ponies were crossed with some horse breeds (particularly the Clydesdale and Standardbred breeds) in order to improve certain desired traits. This study showed that some of the Newfoundland ponies do share haplotypes dominated by both these breeds, suggesting that at one time at least some Newfoundland ponies were crossed with Standardbreds and Clydesdales, however; this was not observed by Prystupa *et al.* (2010) while looking at the genetic diversity of these ponies using microsatellite markers.

Intriguingly, three separate haplogroups were present within the Lac La Croix pony despite the severe population bottle neck of 1977, and this may be a result of the out crossing with Mustangs in order to preserve the population (Lynghaug 2009). This number of haplotypes is higher than expected, and the haplogroups observed in the Lac La Croix were those found frequently in other populations (horse, pony and feral). The Lac La Croix did share one haplotype found in common with American Mustangs, Canadian horses, Mountain and Moorland, and Nordic populations. In addition, the Lac Lac Croix also shared one haplotype (Hap_30) with the Grand Turk population, and gives support to the opinion that some Canadian equine populations could have been influenced from equines shipped to Canada from Grand Turk in the past. Alternatively, both these populations are noted to have arisen from, or were recently influenced by, Spanish stock (Mustangs and horse or pony breeds) and therefore might also explain the overlapping of haplotypes.

4.5. Conclusions

As expected, all pony groups (Canadian, Nordic and Mountain and Moorland) shared several haplotypes amongst each other, and appear to contribute greatly to the overall genetic diversity observed within the species. Haplotype assignment also revealed that both Canadian pony populations shared maternal lineages with the Saint-Pierre Miquelon population and the Canadian horse. The Sable Island population was shown to share haplotypes with the Nordic breeds in particular, along with the Mountain and Moorland pony breeds. The Newfoundland pony appears to have the highest number of maternal lines, indicating that perhaps mares of many different backgrounds were used to repopulate the breed, or that the breed contains a high level of ancestral variability. The Lac La Croix

appears to be quite diverse regardless of the bottleneck experienced in the past, and approximately a third of the ponies tested were found to contain the rare E haplotype. The Mountain and Moorland breeds also share many haplotypes amongst each other in the A through F clades. The E clade is considered to be found very infrequently throughout the world, although it was found frequently in the Dale, Kerry Bog, Eriskay and Shetlands. Information gathered from this study should be combined with all available pedigree, and molecular information to aid in the development, or amendment, of viable and sustainable conservation program.

5.0 General Discussion and Conclusions

Creating a viable and sustainable conservation program for any livestock species and breed is a complex process beginning with breed characterization (Glowatzki-Mullis *et al.* 2006; FAO 2007; Plante *et al.* 2007). Accurate breed characterization depends on measuring and recording phenotypic traits and characteristics, as well as measuring genetic diversity (FAO 2007). The primary objective of this project was to estimate the genetic diversity of native Canadian, Mountain and Moorland, and Nordic populations using microsatellite loci and mtDNA sequence data. The genetic diversity observed within these populations varied between individual breeds and molecular tools, but in general, three statements can be made.

The first is that all populations within this study are highly structured, with very little gene flow among the breeds, shown with the analyses of the microsatellite data. This is suggested with the majority of the individuals from the same breed clustering strongly together in the individual phylogenetic tree. In addition, there was also a very high percentage of correct breed assignment of individuals (98%) using the maximum likelihood methods with prior breed information. Interestingly, if the two separate herds of Newfoundland ponies were combined into a single breed, the percentage of correct breed assignment would have further increased to 99.3%. In contrast, the mtDNA sequence data could not be used for breed assignment because variation accrued before breed development (Vilà *et al.* 2001; Jansen *et al.* 2002; Cieslak *et al.* 2010).

Secondly, although there is little or no gene flow today among the Canadian, Mountain and Moorland, and Nordic breeds, there is evidence that gene flow existed in the past, observed from the analyses of both molecular markers. The low bootstrap values on the various population phylogenetic networks (produced from microsatellite data) clearly

illustrates that the populations are only recently diverged. This is not surprising given the recent bottlenecks and herd reductions experienced by many of the populations between the 1970s and 1990s, and the following introgression to rejuvenate the breeds or populations (McGahern *et al.* 2006; Lynghaug 2009; EPMSBS 2010). The structure data also supported this opinion as several breeds including the Dale, Fell, Kerry Bog, Newfoundland, New Forest, and Saint-Pierre et Miquelon populations showed admixture between one or more of the breeds. The mtDNA sequence data further confirmed this statement with the observation of multiple haplotypes found within the populations but very few breed specific lineages.

The majority of the results of both molecular tools were consistent with prior historical information. For instance, the admixture observed between the Fell and Dale can be traced to at least one Fell stallion in the 1950s which was likely out crossed into the Dale breed (Fox-Clipsham *et al.* 2009). There is no overlap of haplotypes observed between these breeds, therefore indicating that the admixture was likely a result of the flow of stallions only. In addition, the admixture seen within the Kerry Bog and New Forest breeds with other Mountain and Moorland breeds is unsurprising given the historical records. The New Forest pony was frequently noted to have been crossed with other native British pony breeds as a way to improve desired traits (Lynghaug 2009). The admixture observed from the microsatellite and mtDNA sequence data reflects the use of both stallions and mares from other Mountain and Moorland breeds. If template quality permitted, the New Forest pony would have likely also rivalled the Newfoundland pony for the highest number of haplotypes observed within a breed. The admixture seen in the Kerry Bog ponies examined in this study is likely a result of two factors. The first is that the major population reduction in the past likely resulted in some out crossing with other Mountain and Moorland breeds to increase the

population size (McGahern *et al.* 2006). In addition, the samples obtained for this study were only of North American sources, and likely explains the severe excess of heterozygotes observed from the microsatellite data. The American Kerry Bog Society (2008) has recently reported the first Kerry Bog pony born from artificial insemination in 2004, indicating that they are likely bringing in new gene flow from Ireland. Lastly, the observation of admixture between the Newfoundland and Saint-Pierre et Miquelon populations served as a surprising, although logical, finding. This admixture was confirmed by both the microsatellite and mtDNA sequence data, indicating that likely both mares and stallions were recently transported back and forth between the islands and the mainland which are located very close together.

Finally, along with the past gene flow among Canadian, Mountain and Moorland, and Nordic breeds, these populations also contribute significantly to the overall diversity within the species. Genetic diversity within a species can be achieved either by having a high variation within breeds, or alternatively, by having unique breeds which may or may not be genetically diverse (Petit *et al.* 1998). Three pony breeds were consistently found to have high genetic variation (confirmed by both markers), within this study and include the: Newfoundland, New Forest, and Welsh. Similar findings of high genetic variation observed from both microsatellite and mtDNA sequence data have been reported in other horse breeds confirmed to be recently diverged (Pérez-Gutiérrez *et al.* 2008).

Breeds can also be unique from one another as a result of genetic drift alone, a unique founding population, or some combination of the two. The Sable Island population was found to be less genetically diverse than most other breeds included in this study as well as more inbred. The lack of heterozygotes observed within the Sable Island population is likely

due to the small gene pool found on the island (Pérez-Gutiérrez *et al.* 2008; Lucas *et al.* 2009). Regardless of the low genetic diversity and high inbreeding found within the Sable Island population, there would still be a high loss in genetic diversity from the entire dataset if the feral population was removed, and supports the Plante *et al*. (2007) study which also reported similar findings. The mtDNA sequence data revealed that the Sable Island population had very few haplotypes found on the island, and all of which fell into the most common clade found in modern day horses and ponies. From these combined results it can be concluded that the Sable Island population is unique as a result of genetic drift alone. In contrast, the Dale was also only found to have moderate genetic diversity when using microsatellite loci data compared to other populations within the dataset; however, the majority of the Dales sequenced (66%) were found to be in the rare E haplogroup as reported by Jansen *et al.* (2002). The information obtained from the mtDNA sequence data suggested that the Dale breed is unique mostly as a result of the founding mares of the breed, or alternatively, the Dale mares that survived the many bottlenecks were those that fell into the E haplogroup. Unique haplotypes have also been reported to make a breed distinct (Pérez-Gutiérrez *et al*. 2008), and this haplogroup has only been found in about 3% of horses and ponies alive today (McGahern *et al.* 2006; Lei *et al.* 2009). Other pony breeds such as the Lac La Croix and Eriskay were found to be distinct as a result of genetic drift and a unique founding source. Both these populations had very low genetic diversity when examined using microsatellite data, but they segregated out at very low cluster values. In addition to being unique as a result of genetic drift, mtDNA sequence data also revealed that a large portion of them also fell into the E haplogroup.

The second stated objective of this project was to reconstruct the phylogeny of the native Canadian equine populations in reference to the feral, horse, Mountain and Moorland, and Nordic populations/breeds also incorporated in this study. Only the ancestry of two of the Canadian populations could be definitively reconstructed with the information gathered from the molecular tools. The first is that the Newfoundland pony shared very recent common ancestry with the Saint-Pierre et Miquelon population. MtDNA sequence data analysis also revealed that at least at some time in the past many Mountain and Moorland breeds, such as the Dale, Dartmoor, Highland, and New Forest were likely used to develop the breed. The Canadian, Clydesdale, and Standardbred breeds may have influenced the repopulation of the breed. The Sable Island population also shared a more distant common ancestry with both the Saint-Pierre and Miquelon and Newfoundland populations. In addition, the feral population was also confirmed to be very closely related to the Nordic breeds with the information gathered from both molecular tools and another published study (Plante *et al.* 2007). The remaining two indigenous populations (Canadian and Lac La Croix) could not be reconstructed with great confidence. Both these populations were not consistently placed in the phylogenetic trees depicted using the microsatellite data, and further studies should incorporate other horse breeds and Mustang samples to further determine the phylogenetic ancestry of these breeds. The mtDNA sequence data revealed that the Lac La Croix shared haplotypes with the Canadian horse, Mountain and Moorland breeds, Mustang, and Grand Turk population, as well as others, and supports some historical information regarding the trade flow of horses (Lynghaug 2009). The lack of a strong relationship observed between the Canadian horse and the Lac La Croix pony despite the strong historical importance is likely due to the fact that Mustangs were more recently used

to repopulate the breed. The Canadian horse was found also to share haplotypes with the Mountain and Moorland, Sable Island, Newfoundland, and draught breeds, which also supports historical data and other published work (Behara 2000; Plante *et al.* 2007; Lynghaug 2009).

The conservation of farm animal breeds has gained international legal support with many countries agreeing not only to examine, but also to develop suitable conservation strategies for breeds, especially indigenous ones, within their borders (FAO 2007). The microsatellite loci and mtDNA sequence data employed in this study represent neutral markers which are utilized in the hopes that sufficient time has elapsed between populations for unique differences to be incurred (Talle *et al*. 2005). Neutral markers are frequently utilized for estimating genetic diversity within species because of the positive association with "biologically important differences" without being impacted by selection factors (Talle *et al.* 2005). Results from this project should be incorporated with other available pedigree, census, and phenotypic characteristics to develop, or amend, individual breed conservation programs. Several of the pony breeds included in this study are also considered to be at some risk of extinction, therefore the ability to efficiently and correctly prioritize breeds for conservation resources is also vital (EST 2008; RBC 2009).

The priority of breeds for conservation is and will become an exceedingly difficult task as there are more breeds to conserve than there are financial resources (Talle *et al.* 2005). Effective conservation strategies must combine all available information to consistently prioritize the livestock breeds within countries and encompass as many factors as possible. Molecular tools can be employed to help simplify this task. For instance, the use of molecular kinship analyses as well as other approaches can provide a way to evaluate the

potential risk of removing (or losing) a breed within a dataset (Weitzman 1992, 1993; Petit *et al.* 1998; Caballaro & Torro 2002). Not all these approaches rely on the same assumptions and may even have differing thoughts for the specific criteria utilized to prioritize the breeds (marginal loss, internal diversity, and/or genetic distance), and therefore it may be particular useful to examine a dataset using multiple approaches. In addition, one must also take other factors into account when trying to assess the priority of a breed for the development of a conservation strategy such as the amount of *ex situ* genetic material stored, economic importance, unique traits, and the risk of extinction for that breed which can be affected by population size as well as the increase in the frequency of a lethal negative trait (FAO 2007). Lethal negative recessive traits have already been recorded in a couple of the Mountain and Moorland breeds. Immunodeficiency/Anaemia Syndrome originally thought to only affect Fell ponies has now been recently reported to be present in Dales (Fox-Clipsham *et al.* 2009). In addition, a conservation program is not limited to a once in a lifetime basis, and therefore populations should be monitored to ensure their success. The FAO (2007) suggests that populations should be monitored at least once per generation (which is approximately 8-10 years in horse and pony breeds), and therefore the information gathered from this project can be compared with other data in years to come.

For the specific populations in this study the most optimal conservation strategies will likely arise from the *in situ* management within each of the respective countries of origin, along with some *ex situ* germplasm cryogenically stored within government facilities. By doing this both the government and public will become part of the conservation process, ensuring not only the role of the individual breeds, but also ensuring the sustainability of the individual conservation programs developed (Gizaw *et al.* 2008).

6.0 References

Aberle K.S., Hamann H., Drögemüller C. & Distl O. (2004) Genetic diversity in German draught horse breeds compared with a group of primitive, riding and wild horses by means of microsatellite DNA markers. *Animal Genetics* **35**, 270-7.

Aberle K.S., Hamann H., Drögemüller C. & Distl O. (2007) Phylogenetic relationships of German heavy draught horse breeds inferred from mitochondrial DNA D-loop variation. *Journal of Animal Breeding and Genetics* **124**, 94-100.

Achmann R., Curik I., Dovc P., Kavar T., Bodo I., Habe F., Marti E., Sölkner J. & Brem G. (2004) Microsatellite diversity, population subdivision and gene flow in the Lipizzan horse. *Animal Genetics* **35**, 285-92.

Ajmone-Marsan P., Valentini A., Cassandro M., Vecchiotti-Antaldi G., Bertoni G. & Kuiper M. (1997) AFLP(TM) markers for DNA fingerprinting in cattle. *Animal Genetics* **28**, 418-26.

American Kerry Bog Pony Society (2008) The Survival of the The Kerry Bog Pony. Available at http://www.kerrybogpony.org/ (last accessed 14 March 2011).

Apostolidis A.P., Mamuris Z., Karkavelia E. & Alifakiotis T. (2001) Comparison of Greek breeds of horses using RAPD markers. *Journal of Animal Breeding and Genetics* **118**, 47-56.

Aranguren-Mendez J., Beja-Pereira A., Avellanet R., Dzama K. & Jordana J. (2004) Mitochondrial DNA variation and genetic relationships in Spanish donkey breeds (*Equus asinus*). *Journal of Animal Breeding and Genetics* **121**, 319-30.

Aranguren-Mendez J., Jordana J. & Gomez M. (2002) Genetic conservation of five endangered Spanish donkey breeds. *Journal of Animal Breeding and Genetics* **119**, 256-63.

Bailey E., Skow L., Bernoco D., Delvalle A., Scavone M.D.P., Bowling A.T. & Murray J.D. (2000) Equine dinucleotide repeat loci LEX071 through LEX078. *Animal Genetics* **31**, 286-7.

Behara A.M.P. (2000) Genetic diversity in Canadian horse breeds based on microsatellite analysis applications for livestock conservation University of Guelph, Guelph.

Behl R., Behl J., Gupta N. & Gupta S.C. (2007) Genetic relationships of five Indian horse breeds using microsatellite markers. *Animal* **1**, 483-8.

Beja-Pereira A., England P.R., Ferrand N., Jordan S., Bakhiet A.O., Abdalla M.A., Mashkour M., Jordana J., Taberlet P. & Luikart G. (2004) African origins of the domestic donkey. *Science* **304**, 1781.

Binns M.M., Holmes N.G., Rolliman A. & Scott A.M. (1995) The identification of polymorphicmicrosatellite loci in the horse and their use in thoroughbred parentage testing. *British Veterinary Journal* **151**, 9-15.

Bjørnstad G., Nilsen N.Ø. & Røed K.H. (2003) Genetic relationship between Mongolian and Norwegian horses? *Animal Genetics* **34**, 55-8.

Boulding E.G., Culling M., Glebe B., Berg P.R., Lien S. & Moen T. (2008) Conservation genomics of Atlantic salmon: SNPs associated with QTLs for adaptive traits in parr from four trans-Atlantic backcrosses. *Heredity* **101**, 381-91.

Bowcock A.M., Ruiz-Linares A., Tomfohrde J., Minch E., Kidd J.R. & Cavalli-Sforza L.L. (1994) High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368**, 455-7.

Bowling A.T., Del Valle A. & Bowling M. (2000) A pedigree-based study of mitochondrial D-loop DNA sequence variation among Arabian horses. *Animal Genetics* **31**, 1-7.

Breen M., Lindgren G., Binns M.M., Norman J., Irvin Z., Bell K., Sandberg K. & Ellegren H. (1997) Genetical and physical assignments of equine microsatellites - First integration of anchored markers in horse genome mapping. *Mammalian Genome* **8**, 267-73.

Burleson J. & Burleson, D. (2007) *Caribbean Insider Adventures - Turks and Caicos Islands: Find Hidden Treasures and Back-Door Secrets (Insider Adventures)*. Rampant Technpress, Kittrell.

Caballero A. & Toro M.A. (2002) Analysis of genetic diversity for the management of conserved subdivided populations. *Conservation Genetics* **3**, 289-99.

Cañon J., Checa M.L., Carleos C., Vega-Pla J.L., Vallejo M. & Dunner S. (2000) The genetic structure of Spanish Celtic horse breeds inferred from microsatellite data. *Animal Genetics* **31**, 39-48.

Caoline D. (1998) *The kingdom of the horse* Quintet Plublishing, Ontario.

Cieslak M., Pruvost M., Benecke N., Hofreiter M., Morales A., Reissmann M. & Ludwig A. (2010) Origin and history of mitochondrial DNA lineages in domestic horses. *PLoS ONE* **5**.

Coogle L., Reid R. & Bailey E. (1996) Equine dinucleotide repeat loci from LEX025 to LEX033. *Animal Genetics* **27**, 289-90.

Coogle L., Reid R. & Bailey E. (1997) Equine dinucleotide repeat loci LEX034-LEX048. *Animal Genetics* **28**, 309.

Cothran E.G., Juras R. & Macijauskiene V. (2005) Mitochondrial DNA D-loop sequence variation among 5 maternal lines of the Zemaitukai horse breed. *Genetics and Molecular Biology* **28**, 677-81.

Cothran G.E. (2004) Genetic analysis of the Lac La Croix Pony. Available at http://www.laclacroixindianpony.com/pdfs/cothranDNAreport.pdf (last accessed 14 March 2011).

Cozzi M.C., Strillacci M.G., Valiati P., Bighignoli B., Cancedda M. & Zanotti M. (2004) Mitochondrial D-loop sequence variation among Italian horse breeds. *Genetics Selection Evolution* **36**, 663-72.

Cunningham E.P., Dooley J.J., Splan R.K. & Bradley D.G. (2001) Microsatellite diversity, pedigree relatedness and the contributions of founder lineages to thoroughbred horses. *Animal Genetics* **32**, 360-4.

D'Arnoldi C.T., Foulley J.L. & Ollivier L. (1998) An overview of the Weitzman approach to diversity. *Un apercu sur l'approche de la diversite selon Weitzman* **30**, 149-61.

DeAssis J.B., DeLaat D.M., Peixoto M.G., Bergmann J.A., Fonseca C.G. & Carvalho M.R. (2009) Genetic diversity and population structure in Brazilian Mangalarga Marchador horses. *Genetics and Molecular Research* **8**, 1519-24.

Dhar A.K., Beaney,K., Koss,K. and Alcivar-Warren,A. (1998) Sequence variation in the mitochondrial D-loop region of Arabian and Belgian horses. [Cited 2011 February 3]. Available from: http://www.ncbi.nlm.nih.gov/nuccore

Dieringer D. & Schlötterer C. (2003) MICROSATELLITE ANALYSER (MSA): A platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes* **3**, 167-9.

Do Egito A.A., Fuck B.H., McManus C., Paiva S.R., Albuquerque M.D.S.M., Santos S.A., De Abreu U.G.P., Da Silva J.A., Sereno F.T.P.D.S. & Mariante A.D.S. (2007) Genetic variability of Pantaneiro horse using RAPD-PCR markers. *Revista Brasileira de Zootecnia* **36**, 799-806.

Domestic Animal Diversity Information System. (2011) Pantaneiro/ Brazil. Food and Agriculture Organization of the United Nations. Available at http://dad.fao.org/ (last accessed 1 April 2011).

Earl D.A. (2009) Structure Harvester Available at http://users.soe.ucsc.edu/~dearl/software/struct_harvest/ (last accessed 14 March 2011).

Eggert L.S., Powell D.M., Ballou J.D., Malo A.F., Turner A., Kumer J., Zimmerman C., Fleischer R.C. & Maldonado J.E. (2010) Pedigrees and the study of the wild horse population of assateague island national seashore. *Journal of Wildlife Management* **74**, 963-73.

Eggleston-Stott M.L., DelValle A., Bautista M., Dileanis S., Wictum E. & Bowling A.T. (1997) Nine equine dinucleotide repeats at microsatellite loci UCDEQ136, UCDEQ405, UCDEQ412, UCDEQ425, UCDEQ437, UCDEQ467, UCDEQ487, UCDEQ502 and UCDEQ505. *Animal Genetics* **28**, 370-1.

Ellegren H., Johansson M., Sandberg K. & Andersson L. (1992) Cloning of highly polymorphic microsatellites in the horse. *Animal Genetics* **23**, 133-42.

Eriskay Pony Mother Stud Book Society (2010) Our stallions1. Available at http://www.eriskaypony.org.uk/ (last accessed 14 March 2011).

Equus Survivial Trust (2008) Equine Conservation List. Available at http://www.equus-survival-trust.org/ (last accessed 14 March 2011).

Evanno G., Regnaut S. & Goudet J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology* **14**, 2611-20.

Excoffier L. & Lischer H.E.L. (2010) Arlequin suite (verison 3.5): A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564-7.

Food and Agriculture Organization (2007) *The State of the World's Animal Genetic Resources for Food and Agriculture* (Ed. by Pilling, Barbara, Rischkowsky, & Dafydd), Rome.

Felsenstein J. (1989-2006) 'Phylip'(phylogeny inference package) v. 3.66. Available at http://evolution.genetics.washington.edu/phylip.html (last accessed 14 March 2011).

Flannery A.R., and Cothran, E.G. (2003) Mitochondrial DNA sequence variation and the domestication pattern of the horse. Available at http://www.ncbi.nlm.nih.gov/nuccore (last accessed 14 March 2011).

Fox-Clipsham L., Swinburne J.E., Papoula-Pereira R.I., Blunden A.S., Malalana F., Knottenbelt D.C. & Carter S.D. (2009) Immunodeficiency/anaemia syndrome in a Dales pony. *Veterinary Record* **165**, 289-90.

Galli C., Colleoni S., Duchi R., Lagutina I. & Lazzari G. (2007) Developmental competence of equine oocytes and embryos obtained by in vitro procedures ranging from in vitro maturation and ICSI to embryo culture, cryopreservation and somatic cell nuclear transfer. *Animal Reproduction Science* **98**, 39-55.

Gizaw S., Komen H., Windig J.J., Hanotte O. & Arendonk J.A.M.V. (2008) Conservation priorities for Ethiopian sheep breeds combining threat status, breed merits and contributions to genetic diversity. *Genetics Selection Evolution* **40**, 433-47.

Glowatzki-Mullis M.L., Muntwyler J., Pfister W., Marti E., Rieder S., Poncet P.A. & Gaillard C. (2006) Genetic diversity among horse populations with a special focus on the Franches-Montagnes breed. *Animal Genetics* **37**, 33-9.

Godard S., Vaiman D., Oustry A., Nocart M., Bertaud M., Guzylack S., Mériaux J.C., Cribiu E.P. & Guérin G. (1997) Characterization, genetic and physical mapping analysis of 36 horse plasmid and cosmid-derived microsatellites. *Mammalian Genome* **8**, 745-50.

Goudet J. (2001) FSTAT: A program to estimate and test gene diversities and fixation indices (Version 2.9.3) Available at http://www2.unil.ch/popgen/softwares/fstat.htm (last accessed 14 March 2011).

Guérin G., Bertaud M. & Amigues Y. (1994) Characterization of seven new horse microsatellites: HMS1, HMS2, HMS3, HMS5, HMS6, HMS7 and HMS8. *Animal Genetics* **25**, 62.

Gutiérrez J.P., Royo L.J., Álvarez I. & Goyache F. (2005) MolKin v2.0: A computer program for genetic analysis of populations using molecular coancestry information. *Journal of Heredity* **96**, 718-21.

Haase B., Jude R., Brooks S.A. & Leeb T. (2008) An equine chromosome 3 inversion is associated with the tobiano spotting pattern in German horse breeds. *Animal Genetics* **39**, 306-9.

Hill E.W., Bradley D.G., Al-Barody M., Ertugrul O., Splan R.K., Zakharov I. & Cunningham E.P. (2002) History and integrity of thoroughbred dam lines revealed in equine mtDNA variation. *Animal Genetics* **33**, 287-94.

Hoffmann I., Ajmone Marsan, P., Barker, S.F., Cothran, E.G., Hanotte, O., Lenstra, J.A., 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. In: *Proceedings of 29th International Conference on Animal Genetics*, pp. 35–6. FAO, Meiji University, Tokyo, Japan.

Hopman T.J., Han E.B., Story M.R., Schug M.D., Aquadro C.F., Bowling A.T., Murray J.D., Caetano A.R. & Antczak D.F. (1999) Equine dinucleotide repeat loci COR001-COR020. *Animal Genetics* **30**, 225-6.

Huelsenbeck J.P. & Ronquist F. (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-5.

Huson D.H. (1998) SplitsTree: Analyzing and visualizing evolutionary data. *Bioinformatics* **14**, 68-73.

Huson D.H. & Bryant D. (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23**, 254-67.

International Institute for Sustainable Development. (2007) Summary of the first international technical conference on animal genetic resources for food and agriculture: 3-7 September 2007. Available at http://www.iisd.ca/vol09/enb09387e.html (last accessed 31 March 2011).

Irvin Z., Giffard J., Brandon R., Breen M. & Bell K. (1998) Equine dinucleotide repeat polymorphisms at loci ASB 21, 23, 25 and 37-43. *Animal Genetics* **29**, 67.

Ishida N., Hasegawa T., Takeda K., Sakagami M., Onishi A., Inumaru S., Komatsu M., Mukoyama H. (1994) Polymorphic sequence in the D-loop region of equine mitochondrial DNA. *Animal Genetics* **25,** 215-221.

Ishida N., Hasegawa T., Oyunsuren T. & Mukoyama H. (1996) PCR-RFLP analysis of the cytochrome b gene in horse mitochondrial DNA. *Animal Genetics* **27**, 359-63.

Jakobsson M. & Rosenberg N.A. (2007) CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**, 1801-6.

Jansen T., Forster P., Levine M.A., Oelke H., Hurles M., Renfrew C., Weber J. & Olek K. (2002) Mitochondrial DNA and the origins of the domestic horse. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 10905-10.

Jia S., Zhou Y., Lei C., Yao R., Zhang Z., Fang X. & Chen H. (2010) A new insight into cattle's maternal origin in six Asian countries. *Journal of Genetics and Genomics* **37**, 173-80.

Kakoi H., Tozaki T. & Gawahara H. (2007) Molecular analysis using mitochondrial DNA and microsatellites to infer the formation process of Japanese native horse populations. *Biochemical Genetics* **45**, 375-95.

Kaminski M. & Duncan P. (1981) Hemotypes and genetic structure of Camargue horses. *Biochemical Systematics and Ecology* **9**, 365-71.

Kavar T., Habe F., Brem G. & Dovč P. (1999) Mitochondrial D-loop sequence variation among the 16 maternal lines of the Lipizzan horse breed. *Animal Genetics* **30**, 423-30.

Kim K.I., Yang Y.H., Lee S.S., Park C., Ma R., Bouzat J.L., Lewin H.A. (1999) Phylogenetic relationships of Cheju horses to other horse breeds as determined by mtDNA D-loop sequence polymorphism. *Animal Genetics* **30,** 102-108.

Kornienko I.V. & Vodolazhsky D.I. (2010) Patterns of single nucleotide polymorphism distribution in hypervariable regions of the D-loop of human mitochondrial DNA. *Molecular Biology* **44**, 386-92.

Lagutina I., Lazzari G., Duchi R., Colleoni S., Ponderato N., Turini P., Crotti G. & Galli C. (2005) Somatic cell nuclear transfer in horses: Effect of oocyte morphology, embryo reconstruction method and donor cell type. *Reproduction* **130**, 559-67.

Lei C.Z., Su R., Bower M.A., Edwards C.J., Wang X.B., Weining S., Liu L., Xie W.M., Li F., Liu R.Y., Zhang Y.S., Zhang C.M. & Chen H. (2009) Multiple maternal origins of native modern and ancient horse populations in China. *Animal Genetics* **40**, 933-44.

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

Librado P. & Rozas J. (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-2.

Lindgren G., Sandberg K., Persson H., Marklund S., Breen M., Sandgren B., Carlsten J. & Ellegren H. (1998) A primary male autosomal linkage map of the horse genome. *Genome Research* **8**, 951-66.

Lister A.M., Kadwell,M., Kaagan,L.M. and Stanley,H.F. (1998) Ancient and modern DNA in a study of horse domestication. [Cited 2011 February 3]. Available from: http://www.ncbi.nlm.nih.gov/nuccore

Lucas Z.L., McLoughlin P.D., Coltman D.W. & 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**, 310-6.

Luís C., Bastos-Silveira C., Cothran E.G. & Oom M.D.M. (2006) Iberian origins of new world horse breeds. *Journal of Heredity* **97**, 107-13.

Luís C., Cothran E.G. & Oom M.D.M. (2007a) Inbreeding and genetic structure in the endangered Sorraia horse breed: Implications for its conservation and management. *Journal of Heredity* **98**, 232-7.

Luís C., Juras R., Oom M.M. & Cothran E.G. (2007b) Genetic diversity and relationships of Portuguese and other horse breeds based on protein and microsatellite loci variation. *Animal Genetics* **38**, 20-7.

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.

Marklund S., Ellegren H., Eriksson S., Sandberg K. & Andersson L. (1994) Parentage testing and linkage analysis in the horse using a set of highly polymorphic microsatellites. *Animal Genetics* **25**, 19-23.

Marti E., Breen M., Fischer P., Swinburne J. & Binns M.M. (1998) Six new cosmid derived and physically mapped equine dinucleotide repeat microsatellites. *Animal Genetics* **29**, 236-8.

McGahern A.M., Edwards C.J., Bower M.A., Heffernan A., Park S.D.E., Brophy P.O., Bradley D.G., MacHugh D.E. & Hill E.W. (2006) Mitochondrial DNA sequence diversity in extant Irish horse populations and in ancient horses. *Animal Genetics* **37**, 498-502.

Miller M.A., Holder, M.T., Vos, R., Midford, P.E., Liebowitz,T., Chan, L., Hoover, P., Warnow, T. (2009) The CIPRES Science Gateway. Available at http://www.phylo.org/sub_sections/portal (last accessed 14 March 2011).

Mirol P.M., Peral García P., Vega-Pla J.L., Dulout F.N. (2002) Phylogenetic relationships of Argentinean Creole horses and other South American and Spanish breeds inferred from mitochondrial DNA sequences. *Animal Genetics* **33,** 356-363.

Morais J., Oom M.M., Malta-Vacas J. & Luís C. (2005) Genetic structure of an endangered Portuguese semiferal pony breed, the Garrano. *Biochemical Genetics* **43**, 347-64.

Murphie A.M., Hopman T.J., Schug M.D., Aquadro C.F., Bowling A.T., Murray J.D., Caetano A.R. & Antczak D.F. (1999) Equine dinucleotide repeat loci COR021-COR040. *Animal Genetics* **30**, 235-7.

Naderi S., Rezaei H.R., Taberlet P., Zundel S., Rafat S.A., Naghash H.R., El-Barody M.A.A., Ertugrul O., Pompanon F., Abo-Shehada M., Ajmone P.M., Al Tarrayrah J., Angiolillo A., Baret P., Baumung R., Beja-Pereira A., Bertaglia M., Bordonaro S., Bruford M., Caloz R., Canali G., Canon J., Cappuccio I., Carta A., Cicogna M., Crepaldi P., Dalamitra S., Krugmann D., Dobi P., Popielarczyk D., Dunner S., D'Urso G., England P., Erhardt G., Prinzenberg E.M., Ibeagha-Awemu E., Strzelec E., Aziz F., Fornarelli F., Garcia D., Georgoudis A., Gesine L., Giovenzana S., Gutscher K., Hewitt G., Hoda A., Horst B., Istvan A., Juma G., Joost S., Jones S., Karetsou K., Kliambas G., Koban E., Kutita O., Lazlo F., Lenstra J.A., Ligda C., Lipsky S., Luikart G., Glowatzki M.L., Marilli M., Marletta D., Milanesi E., Negrini R., Nijman I.J., Obexer-Ruff G., Papachristoforou C., Pariset L., Pellecchia M., Peter C., Perez T., Pietrolà E., Pilla F., Niznikowski R., Roosen J., Scarpa R., Sechi T., Taylor M., Togan I., Trommetter M., Valentini A., Van Cann L.M., Vlaic A. & Wiskin L. (2007) Large-scale mitochondrial DNA analysis of the domestic goat reveals six haplogroups with high diversity. *PLoS ONE* **2**.

Nei M. (1972) Genetic distance between populations. *American Naturalist* **106**, 283-92.

Nei M. (1987) In: *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nei M., Tajima F. & Tateno Y. (1983) Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *Journal of Molecular Evolution* **19**, 153-70.

Nova Scotia Museum of Natural History (2001) Sable Island Available at http://museum.gov.ns.ca/mnh/nature/sableisland/index.htm (last accessed 14 March 2011).

Ouborg N.J. & Treuren R.V. (1995) Variation in Fitness-Related Characters Among Small and Large Populations of Salvia Pratensis. *Journal of Ecology* **83**, 369-80.

Paetkau D., Calvert W., Stirling I. & Strobeck C. (1995) Microsatellite analysis of population structure in Canadian polar bears. *Molecular Ecology* **4**, 347-54.

Paetkau D., Slade R., Burden M. & 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. & Smouse P.E. (2006) GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288-95.

Pedrosa S., Uzun M., Arranz J.J., Gutiérrez-Gil B., San Primitivo F. & Bayón Y. (2005) Evidence of three maternal lineages in near eastern sheep supporting multiple domestication events. *Proceedings of the Royal Society B: Biological Sciences* **272**, 2211-7.

Penedo M.C.T., Millon L.V., Bernoco D., Bailey E., Binns M., Cholewinski G., Ellis N., Flynn J., Gralak B., Guthrie A., Hasegawa T., Lindgren G., Lyons L.A., Røed K.H., Swinburne J.E. & Tozaki T. (2005) International equine gene mapping workshop report: A comprehensive linkage map constructed with data from new markers and by merging four mapping resources. *Cytogenetic and Genome Research* **111**, 5-15.

Pérez-Gutiérrez L.M., De La Peña A. & Arana P. (2008) Genetic analysis of the Hispano-Breton heavy horse. *Animal Genetics* **39**, 506-14.

Petit R.J., Mousadik A.E.L. & Pons O. (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology* **12**, 844-55.

Pickeral T. (2001) *The encylopedia of horses and ponies*. Parragon Publishing, Bath.

Plante Y., Vega-Pla J.L., Lucas Z., Colling D., De March B. & Buchanan F. (2007) Genetic diversity in a feral horse population from Sable Island, Canada. *Journal of Heredity* **98**, 594-602.

Polzin T. & Daneschmand S.V. (2003) On Steiner trees and minimum spanning trees in hypergraphs. *Operations Research Letters* **31**, 12-20

Pritchard J.K., Stephens M. & Donnelly P. (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-59.

Prystupa J.M., Juras, R, Cothran, E.G, Buchanan, F.C, and Plante, Y (2010) Assessment of genetic diversity and admixture in pony breeds. In: *International Society for Animal Genetics*, July 26-30. Endinburgh, U.K.

Raymond M., Rousset, F (1995) GENEPOP (Version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**, 248-9.

Rare Breeds Canada (2009) Canada's Livestock and Poultry Conservation Lists. Avaiable at http://www.rarebreedscanada.ca/2009conslist.pdf (last accessed 14 March 2011).

Reynolds J., Weir B.S. & Cockerham C.C. (1983) Estimation of the coancestry coefficient: Basis for a short-term genetic distance. *Genetics* **105**, 767-79.

Rieder S., Taourit S., Mariat D., Langlois B. & Guérin G. (2001) Mutations in the agouti (ASIP), the extension (MCIR), and the brown (TYRP1) loci and their association to coat color phenotypes in horses (*Equus caballus*). *Mammalian Genome* **12**, 450-5.

Rosenberg N.A. (2004) DISTRUCT: A program for the graphical display of population structure. *Molecular Ecology Notes* **4**, 137-8.

Rousset F. (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources* **8**, 103-6.

Royo L.J., Álvarez I., Beja-Pereira A., Molina A., Fernández I., Jordana J., Gómez E., Gutiérrez J.P. & Goyache F. (2005) The origins of Iberian horses assessed via mitochondrial DNA. *Journal of Heredity* **96**, 663-9.

Saitou N. & Nei M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution* **4**, 406-25.

Sandberg K.C. & Cothran, E. G. (2000) Blood Groups and Biochemical Polymorphisms. In: *The genetics of the horse* (ed. by Bowling A.T. & Ruvinsky A.), pp. 85-108. CABI publishing, New York.

Schlotterer C. (2004) The evolution of molecular markers [mdash] just a matter of fashion? *Nat Rev Genet* **5**, 63-9.

Sikes, D.S. & Lewis, P.O. (2001) Beta software, version 1. PAUPRat: PAUP* implementation of theparsimony ratchet. In: *Department of Ecology and Evolutionary Biology*. University of Connecticut, Storrs.

Solis A., Jugo B.M., Mériaux J.C., Iriondo M., Mazón L.I., Aguirre A.I., Vicario A. & 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**, 670-8.

Stamatakis A. (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688-90.

Swinburne J.E., Lockhart L., Aldridge V., Marti E., Breen M. & Binns M.M. (2000) Characterisation of 25 new physically mapped horse microsatellite loci: AHT24-48. *Animal Genetics* **31**, 237-8.

Swindell S.R. & Plasterer T.N. (1997) SEQMAN. Contig assembly. *Methods in molecular biology (Clifton, N.J.)* **70**, 75-89.

Talle S.B., Chenyabuga W.S., Fimland E., Syrstad O., Meuwissen T. & Klungland H. (2005) Use of DNA technologies for the conservation of animal genetic resources: A review. *Acta Agriculturae Scandinavica - Section A: Animal Science* **55**, 1-8.

Tamura K., Dudley J., Nei M. & Kumar S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24**, 1596-9.

Tallmadge R.L., Hopman T.J., Schug M.D., Aquadro C.F., Bowling A.T., Murray J.D., Caetano A.R. & Antczak D.F. (1999) Equine dinucleotide repeat loci COR061-COR080. *Animal Genetics* **30**, 462-3.

Tavare S. (1986) Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences *In American Mathematical Society: Lectures on Mathematics in the Life Sciences* **17**, 57-86.

Thompson J.D., Higgins D.G. & Gibson T.J. (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673-80.

Tozaki T., Kakoi H., Mashima S., Hirota K.I., Hasegawa T., Ishida N., Miura N., Choi-Miura N.H. & Tomita M. (2001) Population Study and Validation of Paternity Testing for Thoroughbred Horses by 15 Microsatellite Loci. *Journal of Veterinary Medical Science* **63**, 1191-7.

Troy C.S., MacHugh D.E., Bailey J.F., Magee D.A., Loftus R.T., Cunningham P., Chamberlain A.T., Sykes B.C. & Bradley D.G. (2001) Genetic evidence for Near-Eastern origins of European cattle. *Nature* **410**, 1088-91.

Valera M., Molina A., Gutiérrez J.P., Gómez J. & Goyache F. (2005) Pedigree analysis in the Andalusian horse: Population structure, genetic variability and influence of the Carthusian strain. *Livestock Production Science* **95**, 57-66.

Van Haeringen H., Bowling A.T., Stott M.L., Lenstra J.A. & Zwaagstra K.A. (1994) A highly polymorphic horse microsatellite locus: VHL20. *Animal Genetics* **25**, 207.

Van Oosterhout C., Hutchinson W.F., Wills D.P.M. & Shipley P. (2004) MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**, 535-8.

Van Treuren R., Bijlsma R., van Delden W. & Ouborg N.J. (1991) The significance of genetic erosion in the process of extinction. I. Genetic differentiation in Salvia pratensis and Scabiosa columbaria in relation to population size. *Heredity* **66**, 181-9.

Vilà C., Leonard J.A., Götherström A., Marklund S., Sandberg K., Lidén K., Wayne R.K. & Ellegren H. (2001) Widespread origins of domestic horse lineages. *Science* **291**, 474-7.

Wallner B., Brem G., Müller M. & Achmann R. (2003) Fixed nucleotide differences on the Y chromosome indicate clear divergence between *Equus przewalskii and Equus caballus*. *Animal Genetics* **34**, 453-6.

Wang W., Liu A.H., Lin S.Y., Lan H., Su B., Xie D.W. & Shi L.M. (1994) Multiple genotypes of mitochondrial DNA within a horse population from a small region in Yunnan province of China. *Biochemical Genetics* **32**, 371-8.

Weitzman M.L. (1992) On Diversity. *The Quarterly Journal of Economics* **107**, 363-405.

Weitzman M.L.(1993) What to preserve? An application of diversity theory to crane conservation. *The Quarterly Journal of Economics* **108**, 157-83.

Xu X. & Árnason U. (1994) The complete mitochondrial DNA sequence of the horse, *Equus caballus:* Extensive heteroplasmy of the control region. *Gene* **148**, 357-62.

Yang Y.H., Kim K.I., Cothran E.G. & Flannery A.R. (2002) Genetic diversity of Cheju horses (*Equus caballus)* determined by using mitochondrial DNA D-loop polymorphism. *Biochemical Genetics* **40**, 175-86.

7.0 Appendices

Appendix A A detailed summary including the multiplex (M) with annealing temperature (AT), chromosome (Ch.), locus, primer sequence (5'-3'), allele size (ISAG markers have been standardized), dye, volume of primer (Vol.), and reference for microsatellite locus utilized this study (taken and modified from Glowatzki-Mullis *et al.* (2006).

M & AT (ºC)	Ch.	Locus	Primer Sequence (5'-3')	Allele Size	Dye	Vol. (µl)	Reference
1	4	LEX33	TTTAATCAAAGGATTCAGTTG	195,	NED	0.45	Coogle *et al.* 1996
51.6			GGGACACTTTCTTTACTTTC	203 -221			
	24	AHT4	AACCGCCTGAGCAAGGAAGT	148-168	FAM-6	0.07	Binns *et al.* 1995
			CCCAGAGAGTTTACCCT				
2	2	ASB17	ACCAGTCAGGATCTCCACCG	85-127	VIC	0.19	Breen *et al.* 1997
58.0			GAGGGCGGTACCTTTGTACC				
	4	HMS6	GAAGCTGCCAGTATTCAACCATTG	153,	VIC	0.19	Guérin *et al.* 1994
			CTCCATCTTGTGAAGTGTAACTCA	159-171			
	9	HTG4	CTATCTCAGTCTTGATTGCAGGAC	123-141	FAM-6	0.07	Ellegren *et al.* 1992
			CTCCCTCCCTCCCTCTGTTCTC				
	30	VHL20	CAAGTCCTCTTACTTGAAGACTAG	89-107	FAM-6	0.25	Van Haeringen
			AACTCAGGGAGAATCTTCCTCAG				*et al.* 1994
3	4	HTG7	CCTGAAGCAGAACATCCCTCCTTG	114-130	NED	0.22	Marklund *et al.* 1994
58.0			ATAAAGTGTCTGGGCAGAGCTGCT				
	15	HTG6	CCTGCTTGGAGGCTGTGATAAGAT	82-106	VIC	0.17	Lindgren *et al*. 1998
			GTTCACTGAATGTCAAATTCTGCT				
4	3	ASB23	ACATCCTGGTCAAATCACAGTCC	181-195,	VIC	0.19	Breen *et al.* 1997
58.0			GAGGGCAGCAGGTTGGGAAGG	199-213			
	15	ASB2	CACTAAGTGTCGTTTCAGAAGG	222-224, 230-232,	FAM-6	0.17	Breen *et al.* 1997
			GCACAACTGAGTTCTCTGATAGG	236-258			

M & AT (ºC)	Ch.	Locus	Primer Sequence (5'-3')	Allele Size	Dye	Vol. (µl)	Reference
5	14	UM32	AAATGGTCAGCCTCTCCTC	143-157	PET	0.17	Swinburne *et al.*
51.6			TGTCTCTCTAGTCCCACTCCTC				2000
	16	I-18	CAACAAAGATGTTGCAAGGG	81-109	FAM	0.08	Marti *et al.* 1998
			GTGTGCCTCTTGTCTCTTAGG				
6	5	HMS5	TAGTGTATCCGTCAGAGTTCAAGG	104-112	FAM	0.12	Guérin *et al.* 1994
58.0			GCAAGGAAGTCAGACTCCTGGA				
	16	HTG3	TAACCTGGGTGCAAAGCCACCCAT	113-129	VIC	0.27	Lindgren *et al.* 1998
			GTCAGGGCCAATCTTCCTCAC				
	23	TKY301	AATGGTGGCTAATCAATGGG	144-166	VIC	0.12	Tozaki *et al.* 2001
			GTGTATGATGCCCTCATCTC				
	28	TKY333	CCTTCACTAGCCTTCAAATG	88-114	NED	0.12	Tozaki *et al.* 2001
			TTGTGTTTAGACAGTGCTGC				
7	13	COR69	AGCCACCAGTCTGTTCTCTG	263-285	NED	0.17	Tallmadge *et al.*
58.0			AATGTCCTTTGGTGGATGAAC				1999
	14	LEX78	CAAGCCATGCTGTGGAAACG	152-166	FAM-6	0.12	Bailey *et al.* 2000
			AATGTGCGCATTTAACCACTGTG				
	17	COR07	GTGTTGGATGAAGCGAATGA	154-190	NED	0.08	Hopman *et al.* 1999
			GACTTGCCTGGCTTTGAGTC				
	18	LEX54	TGCATGAGCCAATTCCTTAT	163-183	VIC	0.17	Breen *et al.* 1997
			TGGACAGATGACAGCAGTTC				
	22	COR22	AAGACGTGATGGGAAATCAA	257-269	PET	0.22	Murphie *et al.* 1999
			AGAAAGTTTTCAAATGTGCCA				
	29	ASB43	TCACTTAGTAGGGGCATGC	77-103	VIC	0.17	Irvin *et al.* 1998
			GTGTTTGTCCTTGACTCTCC				
	31	ATH31	TCTTGCTGCATTTCCTTGG	116-136	0.35µl	0.35	Swinburne *et al.* 2000
			GTGCAGGAAAAGTTTGATGACC				

M & AT (ºC)	Ch.	Locus	Primer Sequence (5'-3')	Allele Size	Dye	Vol. (µl)	Reference
8 58.0	5	LEX34	GCGGAGGTAAGAAGTGGTAG GGCCTAAGATGAGGGTGAA	245-255	PET	0.20	Coogle *et al.* 1997
	26	COR71	CTTGGGCTACAACAGGGAATA CTGCTATTTCAAACACTTGGA	190-202	FAM-6	0.20	Tallmadge *et al.* 1999
	27	HMS45	TGTTACAGGTATTGGTAAACTGTGC GGAACAAGAAGAAATCACTAATGTC	185-197	VIC	0.20	Godard *et al.* 1997
9 51.6	5	TKY344	GTGTCCATCAATGGATGAAG CTTAAGGCTAAATAATATCCC	91-107	FAM-6	0.17	Tozaki *et al.* 2001
	16	TKY341	TATCCAGTCACCCATTTTAC TTGTGTCAGTACACTCTATG	134-154	NED	0.17	Tozaki *et al.* 2001
10 58	4	TKY337	AGCAGGGTTTAATTACCGAG TAGATGCTAATGCAGCACAG	167-187	VIC	0.17	Tozaki *et al.* 2001
	28	CA425	AGCTGCCTCGTTAATTCA CTCATGTCCGCTTGTCTC	230-250	VIC	0.17	Eggleston-Stott *et al.* 1997
	29	TKY325	GGATGGAGTGAGATAATACC TGGATGAACCATGAATAGTG	172-194	NED	0.17	Tozaki *et al.* 2001
Single 51.6	1	HMS7	CAGGAAACTCTCATGTTGATACCATC TGTTGTTGAAACATACCTTGACTGT	167, 173-189	FAM-6	0.10	Guérin *et al.* 1994
Single 58	8	ATH5	ACGGACACATCCCTGCCTGC GCAGGCTAAGGAGGCTCAGC	126-146	VIC	0.19	Binns *et al.* 1995
Single 54.3	9	HMS3	CCAACTCTTTGTCACATAACAAGA CCATCCTCACTTTTTCACTTTGTT	150-152, 156-172	PET	0.13	Guérin *et al.* 1994
Single 58.0	10	HMS2	ACGGTGGCAACTGCCAAGGAAG CTTGCAGTCGAATGTGTATTAAATG	282-308	VIC	0.17	Guérin *et al.* 1994

M & AT (ºC)	Ch.	Locus	Primer Sequence (5'-3')	Allele Size	Dye	Vol. (µl)	Reference
Single	11	TKY343	TAGTCCCTATTTCTCCTGAG	138-174	NED	0.19	Tozaki *et al.* 2001
56.4			AAACCCACAGATACTCTAGA				
Single	15	HMS1	CATCACTCTTCATGTCTGCTTGG	167-193	PET	0.17	Guérin *et al.* 1994
51.6			TTGACATAAATGCTTATCCTATGGC				
Single	21	HTG10	CAATTCCCGCCCCACCCCCGGCA	89-115	NED	0.22	Marklund *et al.* 1994
58.0			TTTTTATTCTGATCTGTCACATTT				

Appendix B A overall summary of the statistics parameters estimated (average number of alleles (N_a), effective number of alleles (N_e), observed heterozygosity (H_O), expected heterozygosity (H_E), and *F*-indices for each individual locus tested when *P = 0.05*

Locus	N_a	N_e	H_o	H_e	F_{IS}	F_{ST}	F_{IT}
AHT04	6.36	3.35	0.857	0.853	0.024	0.128	0.149
AHT05	6.00	2.90	0.676	0.806	0.019	0.158	0.174
AHT31	5.12	1.86	0.686	0.773	0.041	0.190	0.224
ASB02	7.20	3.86	0.853	0.850	0.033	0.145	0.173
ASB17	8.44	4.29	0.800	0.877	0.014	0.115	0.127
ASB23	6.16	3.04	0.686	0.784	0.017	0.140	0.155
ASB43	4.88	2.13	0.886	0.795	0.012	0.106	0.116
CA425	5.60	2.30	0.353	0.562	0.062	0.147	0.200
COR07	6.84	3.13	0.824	0.835	0.042	0.102	0.140
COR22	5.12	2.28	0.714	0.792	0.082	0.131	0.202
COR69	5.76	2.46	0.882	0.808	0.012	0.149	0.159
COR71	7.24	3.28	0.857	0.834	0.014	0.100	0.113
HMS01	5.20	2.09	0.765	0.757	0.043	0.162	0.198
HMS02	5.92	2.67	0.829	0.783	-0.029	0.234	0.212
HMS03	5.72	2.92	0.686	0.838	-0.008	0.119	0.112
HMS05	3.48	1.66	0.588	0.676	0.060	0.136	0.149
HMS06	5.08	2.35	0.800	0.797	0.010	0.116	0.124
HMS07	5.28	2.53	0.686	0.761	0.050	0.080	0.126
HMS45	5.52	1.56	0.686	0.757	-0.061	0.086	0.030
HTG03	5.20	2.28	0.657	0.780	0.061	0.089	0.145
HTG04	4.72	1.89	0.486	0.523	-0.002	0.141	0.139
HTG06	4.28	1.37	0.600	0.704	0.131	0.255	0.352
HTG07	4.08	1.74	0.800	0.733	-0.006	0.135	0.130
HTG10	7.16	3.25	0.857	0.817	0.023	0.154	0.173
I-18	6.60	3.14	0.743	0.859	0.034	0.114	0.144
LEX33	6.76	3.44	0.824	0.810	-0.015	0.124	0.111
LEX34	5.04	1.80	0.686	0.713	-0.001	0.101	0.101
LEX54	6.60	2.92	0.800	0.815	0.034	0.094	0.124
LEX78	4.20	1.56	0.886	0.714	0.022	0.103	0.123
TKY301	7.12	3.32	0.853	0.871	0.047	0.147	0.187
TKY325	8.64	4.94	0.800	0.885	0.044	0.105	0.144
TKY333	9.20	4.16	0.971	0.888	0.008	0.122	0.129
TKY337	6.04	2.74	0.857	0.846	0.025	0.109	0.131
TKY341	5.92	2.38	0.824	0.784	0.025	0.106	0.128
TKY343	7.92	3.82	0.886	0.843	0.024	0.122	0.144
TKY344	6.68	2.22	0.629	0.639	0.040	0.190	0.222
UM32	5.12	2.08	0.676	0.799	0.009	0.120	0.128
VHL20	7.00	3.66	0.800	0.829	0.033	0.170	0.197

Appendix C A summary of all 46 breeds/ populations included in this study with type, accession numbers and references (taken and modified from Pérez-Gutiérrez *et al.* 2008).

Breed/ Population	Type	Accession Numbers	Reference
Akhal-Teke	Horse	AY246174-79 DQ327950, DQ327954-5, DQ327957	Flannery & Cothran 2003 McGahern *et al.*, 2006
Andulasian	Horse	AY519911-13	Royo *et al.*, 2005
Arabian	Horse	AF132568-77	Royo *et al.*, 2005
Asturcon	Pony	AY519871-80	Royo *et al.*, 2005
Belgian	Horse	AF064630-32 AY246186-94	Dhar *et al.* 1998 Flannery & Cothran 2003
Campolina	Horse	AY997139-44	Luís *et al.* 2006
Canadian	Horse	HQ592811- 19, HQ592822, HQ592932-33	Prystupa 2011
Caspian	Horse	HQ592820-21, HQ592823-27, HQ592934-36	Prystupa 2011
Cleveland Bay	Horse	AY246209-13	Flannery & Cothran 2003
Clydesdale	Horse	HQ593049-58	Prystupa 2011
Connemara	Pony	HQ592860-70, HQ592889	Prystupa 2011
Creole	Horse	AF465986-8 AF516496, AF516498, AY997128, AY997145, AY997147, AY997149	Mirol *et al.* 2002 Luís *et al.* 2006
Dale	Pony	HQ592800-10, HQ593046	Prystupa 2011
Dartmoor	Pony	HQ592790-99, HQ592930-31	Prystupa 2011
Eriskay	Pony	HQ593012-23	Prystupa 2011
Exmoor	Pony	HQ592968-78	Prystupa 2011
Fell	Pony	AF072981 HQ592856-59, HQ592942-46	Lister *et al.* 1998 Prystupa 2011
Fjord	Pony	HQ592871-80, HQ593047	Prystupa 2011
Friesian	Horse	AY246225-29	Flannery & Cothran, 2003
Garrano	Pony	AY519915-23	Royo *et al.* 2005
Grand Turk	Feral	HQ593024-34	Prystupa 2011
Haflinger	Horse	HQ592847-55, HQ592940-41	Prystupa 2011
Highland	Pony	HQ592828-37, HQ592937	Prystupa 2011
Icelandic	Pony	HQ592881-88, HQ592947-49, HQ593048	Prystupa 2011
Irish Draught	Horse	DQ327891-900	McGahern *et al.* 2006
Kerry Bog	Pony	HQ592838-46	Prystupa 2011
Lac La Croix	Pony	HQ592914-17, HQ593036-41	Prystupa 2011
Lipizzaner	Horse	AF168689-98	Kavar *et al.* 1999
Mongolian Domestic	Horse	HQ592957-67	Prystupa 2011

Breed/ Population	Type	Accession Numbers	Reference
Mustang	Feral	AF516489, AF516494, AJ413753, AJ413797, AJ413802, AY997152, AY997154, AY997178, AY997181, AY997200	Luís *et al.* 2006
New Forest	Pony	HQ592995-99	Prystupa 2011
Newfoundland	Pony	HQ592784-89, HQ592898-900, HQ592918-19, HQ592950, HQ593042-43, HQ593061-62	Prystupa 2011
Noriker	Horse	AY246248-52	Flannery & Cothran, 2003
Pantaneiro	Horse	AY997160-64, AY997199	Royo *et al.* 2005
Percheron	Horse	AB329603	Kakoi *et al.* 2007
Pottok	Pony	AY519958-67	Royo *et al.* 2005
Przewalski's horse	Root	AF072994-5	Lister *et al.* 1998
Quarter Horse	Horse	AF072980	Lister *et al.* 1998
Sable Island	Feral	HQ592901-13, HQ592951-56, HQ593035, HQ593063	Prystupa 2011
Saint-Pierre and Miquelon	Feral	HQ592920-29, HQ593044	Prystupa 2011
Shetland	Pony	HQ592939, HQ592979-94	Prystupa 2011
Shire	Horse	AF072975-6 AJ413893-900	Lister *et al.* 1998 Jansen *et al.* 2002
Standardbred	Horse	HQ593000-11	Prystupa 2011
Suffolk	Horse	AF072986	Lister *et al.* 1998
Thoroughbred	Horse	AF072990-1 AB329590, AB329593-5, AB329599 D14991, D23665-6	Lister *et al.* 1998 Kakoi *et al.* 2007 Ishida *et al.* 1994
Welsh	Pony	HQ592890-97, HQ593059-60	Prystupa 2011
Clade A	Reference	AF014409 AF072987	Lister *et al.* 1998 Kim *et al.* 1999
Clade B	Reference	AF072989-90	Lister *et al.* 1998
Clade C	Reference	AF072988, AF072992	Lister *et al.* 1998
Clade D	Reference	AF064632 D23665	Dhar *et al.* 1998 Ishida *et al.* 1994
Clade E	Reference	AF014412 D14991	Kim *et al.* 1999 Ishida *et al.* 1994
Clade F	Reference	AF014410 AF072996	Kim *et al.* 1999 Lister *et al.* 1998
Accession to align sequences	Reference	X79547	Xu & Árnason, 1994