

**GENOMIC DIVERSITY AND ORIGINS OF THE FERAL
HORSES (*Equus ferus caballus*) OF SABLE ISLAND AND THE
ALBERTA FOOTHILLS**

A Thesis submitted to the College of
Graduate and Postdoctoral Studies
In Partial Fulfillment of the Requirements
For the Degree of Master of Science
In the Department of Biology
University of Saskatchewan
Saskatoon, Saskatchewan

By

CHRISTINA TOLLETT

PERMISSION TO USE

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this 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 the professor or professors who supervised my thesis work or, in their 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 other use of material in this thesis in whole or in part should be addressed to:

Head of the Department of Biology
University of Saskatchewan
112 Science Place
Saskatoon, SK
S7N 5E2

OR

Dean
College of Graduate Studies and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldsen Building, 110 Science Place
Saskatoon, SK
S7N 5C9

ABSTRACT

The genetic structure of domestic horse populations has been studied extensively using microsatellite markers; however, this study is the first to use genomics to assess the genetic diversity levels, uniqueness, and ancestral relationships of the Sable Island and Alberta feral horse populations. Here, I use a genome-wide single nucleotide polymorphism SNP-based approach to describe the structure and relatedness of the Sable Island, Nova Scotia and Alberta Foothills feral horse populations of Canada. Investigating the genomic diversity of these populations can provide researchers with important information for the conservation of genetics and improve the understanding of microevolution in natural populations. I characterized broad patterns of genetic diversity in the feral horse populations of Sable Island and the Alberta foothills using the Illumina GGP 65K SNP array, and assessed their relationships with domestic breeds to gain insight into each population's respective origin. My analyses indicate that Sable Island is the most diverged and inbred of the two populations and relative to other horse breeds and populations, being greatly distinguished from even its closest relatives, the Mongolian horse and breeds of Nordic origin. In contrast, the Alberta Foothills population appears to be the product of extensive admixture with a strong relationship to draft breeds. Higher than anticipated inbreeding levels in Alberta further suggests the presence of breeding groups across the landscape. My study presents important information regarding the origin and genetic diversity feral horse populations in Canada, and will assist with future conservation and management of these populations.

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to my supervisors Dr. Philip D. McLoughlin and Dr. Jocelyn Poissant for affording me all the amazing opportunities that came with working in the University of Saskatchewan's Population Ecology lab. To my committee members, Dr. Neil Chilton and Dr. Emily Jenkins, thank you for your time, patience, and guidance throughout my degree, it has been appreciated beyond measure.

I am also profoundly grateful for the financial support provided by the National Science and Engineering Research Council, the Ramsay family for the Malcolm A. Ramsay award, the Alberta Conservation Association, and to the Department of Biology. I would also like to extend my gratitude to the Parks Canada Agency and Department of Fisheries and Oceans for all of their generosity and support throughout our field seasons, which has extended far beyond any expectations. In addition, I would like to thank the Wild Horses of Alberta Society (WHOAS) for their support and assistance in sample collection among other things. Furthermore, Dr. Jessica L. Peterson and Dr. Sher Hendrickson deserve my gratitude for enabling this project by providing us with access to the raw data files from previous analyses.

To the field crews, before and after me, thank you for all of your hard work, consistency, and new ideas, our projects would not have been possible without the combined efforts of years worth of students, postdocs, and professors. Perhaps, most importantly, I would like to thank all of my fellow field crewmembers for everyone's spunk, enthusiasm, and persistence. You made our long days not only bearable, but into some of my most cherished memories. It is these memories, which make me see great potential and hope in our present and future scientists.

Throughout my degree I have been provided with endless amounts of support and advice from my family and friends. I am especially grateful to my fiancé Kyle Tschritter for all of his emotional support, laughter, and midnight snack runs; I could not have done this without you. In addition I would like to thank all of my lab mates, all of who deserve recognition. Most notably though, a big thank you to Patricia Tomchuk, Charlotte Regan, Julie Colpitts and Ruth Gruel for their countless revisions, nights out, and motivational speeches. Finally, a warmest and deepest thank you to my parents who have done everything in their power to help me along the way and are truly, embarrassingly, proud of me.

TABLE OF CONTENTS

PERMISSION TO USE.....	i
ABSTRACT.....	ii
ACKNOWLEDGEMENTS	iii
LIST OF FIGURES	vii
LIST OF TABLES	ix
1. INTRODUCTION.....	1
1.1 Background	1
1.2 Horse Demographic History	3
<i>1.2.1 Domestication, Expansion and Differentiation.....</i>	<i>3</i>
<i>1.2.2 Formation of Specialized Breeds and Further Horse Adaptation</i>	<i>4</i>
1.3 Feral Horse Population of Sable Island	5
<i>1.3.1 Background and Ancestry</i>	<i>5</i>
1.4 The Alberta Foothills Feral Horse Population	6
<i>1.4.1 Background and Ancestry</i>	<i>6</i>
1.5 SNP Array	8
1.6 Objectives and Predictions.....	9
<i>1.6.1 Objectives</i>	<i>9</i>
<i>1.6.2 Sable Island Predictions.....</i>	<i>9</i>
<i>1.6.3 Alberta Foothills Predictions</i>	<i>9</i>
2. METHODS	10
2.1 Sable Island Study Site	10
2.2 Alberta Foothills Study Site.....	11
2.3 DNA samples and genotyping	13
2.5 Data Pruning	15
2.6 Within-Breed Diversity.....	15
<i>2.6.1 Proportion of Polymorphic Sites.....</i>	<i>15</i>
<i>2.6.2 Homozygosity and Inbreeding.....</i>	<i>15</i>
2.7 Among-Breed Diversity.....	16

2.7.1 Breed Divergence.....	16
2.7.2 AMOVA	16
2.8 Among-Breed Relationships.....	16
2.8.1 Principal Component Analysis.....	16
2.8.2 Bayesian Cluster Analysis.....	16
3. RESULTS	17
3.1 Within-Breed Diversity.....	17
3.1.1 Proportion of Polymorphic Sites.....	17
3.1.2 Homozygosity and Inbreeding.....	17
3.2 Among-Breed Diversity: Breed Divergence.....	21
3.2.1 Breed Divergence	21
3.2.2 AMOVA	21
3.3 Among-Breed Relationships.....	23
3.3.1 Principal component analysis	23
3.3.2 Bayesian Cluster Analysis.....	25
4. DISCUSSION	28
4.1 Within-breed Diversity	28
4.2 Among-breed Diversity	30
4.3 Among-breed Relationships.....	31
4.4 Management Implications.....	33
4.5 Conclusions.....	35
REFERENCES.....	37
APPENDIX A: PHYLOGENTIC RELATIONSHIP	49
APPENDIX B: ALTERNATE PRINCIPAL COMPONENT ANALYSIS	50
APPENDIX C: BAYESIAN CLUSTER ANALYSIS AT $K = 3$ BEFORE SABLE ISLAND WAS SEPARATED AS A SUB-POPULATION	51
APPENDIX D: CROSS-VALIDATION ERROR AS A FUNCTION OF THE K-VALUE	52
APPENDIX E: PRINCIPAL COMPONENT ANALYSES EMPHASIZING THE ALBERTA POPULATION.....	53

APPENDIX F: BAYESIAN CLUSTER ANALYSIS OF THE ALBERTA POPULATION AND ITS CONTRIBUTORS	54
APPENDIX G: INDIVIDUAL INBREEDING COEFFICIENTS BY EQUINE ZONE	55

LIST OF FIGURES

Figure 2.1: A map of Sable Island, Nova Scotia, Canada and its relative location to Halifax Nova Scotia (Cabrera et al., 2017).....	10
Figure 2.2: The 2018 feral horse survey count and equine zone divisions in Alberta, Canada (Alberta Environment and Parks, 2018).....	12
Figure 2.3: An ArcGIS map of Sable Island with the median locations plotted for the eighteen horses (of twenty-three) from the Sable Island population, which have median locations, recorded.....	14
Figure 3.1: A bar plot representing the proportion of polymorphic sites found across each populations data set, the mean is 91.82%. Sable Island has the lowest proportion of polymorphic sites (74.5%), while the Quarter Horse has the largest proportion (99.17%).....	20
Figure 3.2: A heat map of the pairwise F_{ST} values for 39 horse breeds. The red indicates a higher F_{ST} value and thus a greater divergence between populations, while the blue indicates a low F_{ST} value.....	22
Figure 3.3: A three-axis principal component analysis of 43 horse breeds, based on 37 495 autosomal SNP sites. The leading three components explain approximately 72% of the variation (PC1 42.96%, PC2 16.6%, and PC3 12.76%).....	24
Figure 3.4: The Alberta percentage admixture obtained from the ADMIXTURE simulation results at $K = 25$, while excluding an individual identified as an outlier. The “Other” is a mixture of various breeds that all occurred at a mean of $< 5\%$	25
Figure 3.5: Bayesian clustering output from ADMIXTURE depicting the results for three values of K (4,25, and 35) in 856 horses from 43 populations. ADMIXTURE identified $K = 25$ as the most likely genetic group allocation for the data.....	26
Figure A.1: TreeMix was used to create a maximum likelihood phylogeny of the 43 horse breeds included in our data set in addition to a Donkey sample to root the tree (1.13; Pickrell and Pritchard, 2012).....	48
Figure A.2: A two-axis principal component analysis of 43 horse breeds, based on 37 495 autosomal SNP sites. The leading two components explain 59.6% of the variation (PC1 42.96% and PC2 16.6%).....	49

Figure A.3: Bayesian clustering output from ADMIXTURE depicting the results for a K -value of 3 in 856 horses from 43 populations. ADMIXTURE identified $K = 25$ as the most likely genetic group allocation for the data.....	50
Figure A.4: A scatter-plot depicting the relationship between the cross-validation error output from the ADMIXTURE program and the K -value. The most likely K -value occurs at the lowest cross validation error (Alexander et al., 2015).....	51
Figure A.5: Principal components analysis of 37 495 autosomal SNP's from the Alberta feral horse population broken down by location (top) and by equine zone (bottom). Depicts that there is variation across and within equine zones, suggesting the presence of subpopulations.....	52
Figure A.6: Bayesian clustering output from ADMIXTURE depicting the Alberta horse population and the primary breeds (>5% contribution) contributing to the genetic make-up.....	53

LIST OF TABLES

Table 3.1. Number of samples (N), the excess of homozygosity inbreeding coefficient (F_{HOM}), the minimum and maximum inbreeding statistic within each population, and the observed and expected homozygosity values.....	26
Table A.1. The inbreeding coefficients fro the Alberta feral horse population subdivided into samples and by equine zones.....	54

1. INTRODUCTION

1.1 Background

High levels of genetic diversity are instrumental in the processes that enable species to respond and adapt to environmental change (Lacy et al., 1997). Many different evolutionary processes interact in determining what genetic diversity there is within and among populations, including mutations, gene flow, genetic drift and natural and artificial selection (Hartl and Clark, 1997). Populations with greater genetic variation have an increased likelihood of possessing beneficial alleles, which may allow for new adaptations to take hold through selection (Darwin, 1859; Lacy et al., 1997). Notably, the impact that these evolutionary forces have in shaping the level of genetic diversity within a population can differ dramatically between large and small populations (Hartl and Clark, 1997).

Small populations with reduced gene flow are highly susceptible to non-adaptive evolutionary change through genetic drift. Genetic drift is the random deviation in allele frequencies from the expected due to the chance inheritance of alleles (Masel, 2011). In small populations, fluctuation in gene frequency will eventually lead alleles to extinction or fixation, and hence reduction in genetic diversity (Lohr et al., 2014). Large population size and gene flow typically reduces the effects of genetic drift. Subsequently, small, isolated populations are also disproportionately affected by evolutionary phenomena influencing genetic drift, such as founder and bottleneck effects (Hartl and Clark, 1997; Masel, 2011). The founder effect occurs when a population is founded by a small group of individuals. Over time, as the population grows and interbreeds, the descendants that the population is composed of carry a high frequency of alleles that were present in the founding individuals, conferring reduced diversity at those loci (Hartl and Clark, 1997; Jarvis et al., 2011; Masel, 2011). A small founding population is associated with high rates of inbreeding and increased genetic differentiation from the source population (Winton et al., 2015). Similarly, the bottleneck effect is caused by a reduction in genetic diversity following a dramatic decline in population size which results in the loss of a large portion of the population, resulting in a net loss of genetic diversity (Hartl and Clark, 1997). Genetically isolated populations do not have new genetic variation introduced by immigrating

individuals, which maximizes the effect of evolutionary processes like genetic drift, and the founder and bottleneck effects.

Inbreeding is an additional factor exerting influence over the genetic variability of small and isolated populations (Dunn et al., 2011; Hartl and Clark, 1997; Keller and Waller, 2002; Walling et al., 2011). The breeding of close relatives increases homozygosity within a population, which can lead to a decrease in fitness traits, known as inbreeding depression (Pemberton et al., 2017; Roff, 2002). Loss of viability, fertility, disease resistance and decreased birth weight, neonatal survival, litter size and body size are commonly documented effects of inbreeding (Bilski et al., 2013; Coltman et al., 1998; Dunn et al., 2011; Keller and Waller, 2002; Walling et al., 2011). However, not all increases in homozygosity reduce fitness as inbreeding depression is only expected to occur when the frequency of fitness-related recessive mutations is increased (Allendorf and Luikart, 2007). The partial dominance hypothesis suggests inbreeding causes an increase in deleterious alleles resulting in a decrease in fitness, although in the absence of inbreeding, deleterious alleles may still become fixed in a population due to founder effects or genetic drift (Keller and Waller, 2002; Lacy et al., 1997). The severity of inbreeding depression can be heavily influenced by a process known as genetic purging: the removal of deleterious alleles in a population through natural selection; though it is more effective when population numbers are high (Lacy and Ballou, 1998). Inbreeding depression is of particular concern for large mammals due to small population sizes and high levels of human influence on population structure; thus conservation efforts of large mammals often consider the genetic constituents of populations (Paetkau et al., 1998). Ultimately, a decrease in fitness may lead to the extinction of a small population or rare species. Genetic diversity within a population is therefore essential in mitigating the negative effects of inbreeding and reduced genetic heterozygosity on fitness.

Livestock are an important component of food security worldwide and hence the conservation of farm animal genetic resources is an international priority (FAO Commission on Genetic Resources for Food and Agriculture Assessments, 2015; Lauvie et al., 2014; Tamminen, 2015). The Food and Agriculture Organization of the United Nations (FAO) defines the term livestock broadly and considers it inclusive to all domesticated animal species (FAO Commission on Genetic Resources for Food and Agriculture Assessments, 2011). Domestication and selective breeding to increase the prevalence of desirable aesthetic and production traits has created breeds with specific phenotypes. Genetically diverse breeds allow animal production

systems to be efficient under a range of environmental stresses (Tamminen, 2015). Diversity loss within breeds limits future improvement and adaptive potential, and increases the susceptibility of breeds to inbreeding depression, such as reduced disease resistance and fertility, and an increase in the prevalence of genetic diseases (Hartl and Clark, 1997; Utsunomiya et al., 2015). Within domestic species 50% of all genetic variation is between breeds (FAO, 2011); and due to selective breeding practices and the pressures to increase production and efficiency, 8% of all livestock breeds have become extinct, 21% are at risk of extinction, and an additional 35% are of unknown status (Agri-food Canada, 2008; FAO, 2011). As a result, global conservation efforts are focusing on the conservation of livestock breeds as animal genetic resources (FAO Commission on Genetic Resources for Food and Agriculture Assessments, 2015).

To conserve the genetic diversity of livestock species, the available resources must first be inventoried, characterized and prioritized (FAO, 2011). Given limited resources available for conservation, primary consideration should then be given to breeds which: 1) are the most genetically distinct, 2) are at the greatest risk of extinction, 3) have the largest measure of breed ‘merit’, and 4) have the most limited within breed variation (Barker, 2001). Feral populations often harbour unique genetic variants that may be of commercial, scientific, aesthetic or cultural value, making them a vital target for conservation efforts. Hence, characterizing genetic variation of feral populations, including within population variability, divergence and relationships with domestic breeds can provide insight into their conservation value (Barker, 2001; Hall and Bradley, 1995; Onogi et al., 2017).

1.2 Horse Demographic History

1.2.1 Domestication, Expansion and Differentiation

The modern domestic horse, *Equus ferus caballus*, poses a particularly meaningful study subject, not only because of its role in modern agriculture, but its evolutionary history and connection to human civilization. The modern horse is a member of the Family Equidae, which originated in North America approximately 55 million years ago (MacFadden, 2005). Approximately 10 000 years ago *Equus* disappeared from the fossil records in North America due to a combination of climate and habitat-related pressures, though it has also been speculated that human overkill may have been a contributing factor (Librado et al., 2016; Smith et al.). Despite their disappearance from the Americas, horses became an important source of food to humans during the Upper

Paleolithic period in the Eurasian steppe (Achilli et al., 2012; Librado et al., 2017; Warmuth et al., 2012). Archaeological evidence suggests that they were extensively hunted, nearly to extinction, during the period prior to domestication.

Fossilized teeth with bit wear (Lippold et al., 2011) and mares milk in pottery suggest the horse was first domesticated in the Western Eurasian steppe 6000 years ago by the Boati culture (Outram et al., 2009). Additionally, low Y-chromosome diversity and high mitochondrial diversity of the modern horse indicates that high levels of introgression from wild mares occurred as the domestic horse began to expand east-to-west across the Eurasian steppe (Ling et al., 2010; Lippold et al., 2011; Warmuth et al., 2012). The domestication of the horse allowed for a level of familiarity with the species that encouraged alternative use for the animals such as riding and the pulling of chariots and plows (Carnegie Museum of Natural History, 2003). By 3000 B.C. horseback riding had become widespread across the Eurasian Steppe (Carnegie Museum of Natural History, 2003). Human exploitation of the horse in transportation and trade mitigated the range expansion of the horse across Europe (Carnegie Museum of Natural History, 2003). Since the horse's domestication, *Equus ferus caballus* has been subjected to artificial and natural selection, population bottlenecks and periodic admixture. Presently, the Food and Agriculture Organization of the United Nations (FAO) reports 905 genetically distinct horse breeds across the globe, with the horse range once again including North America (FAO Commission on Genetic Resources for Food and Agriculture Assessments, 2015).

1.2.2 Formation of Specialized Breeds and Further Horse Adaptation

The horse's domestication revolutionized warfare, transportation, agriculture, hunting, and industry (Carnegie Museum of Natural History, 2003; Lippold et al., 2011). With specialized purpose came selective breeding for traits of economic and functional significance, thus the concept of a breed developed. The FAO broadly defines a breed as a sub-specific group of animals with definable phenotypic characteristics that enable it to be separated from other similarly defined groups within the same species, or a group for which geographical and/or cultural separation from phenotypically similar groups has led to acceptance of its separate identity (FAO Commission on Genetic Resources for Food and Agriculture Assessments, 2007). Standardized breeds are based on morphology and performance traits such as endurance, size, strength, speed, and behaviour. Presently, the FAO reports 905 horse breeds globally, of which 10% are extinct, 22% are at risk of extinction, and an additional 53% are of an unknown status

(FAO Commission on Genetic Resources for Food and Agriculture Assessments, 2015). Of the thirty horse breeds native to North America, eight are extinct; this is likely due to the decline in effective population size associated with breed differentiation, as well as the favouring of certain breeds and traits (FAO Commission on Genetic Resources for Food and Agriculture Assessments, 2015). Throughout this document, I use the term ‘breed’ broadly, to encompass those populations that have not been defined or recognized as their own breed too date. Native to Canada are two pony breeds (Newfoundland and Lac la Croix), one horse breed (the Canadian horse), and four feral horse populations, including those on Sable Island, Nova Scotia, and the Alberta Foothills; all are classified as critical, endangered or at risk (Rare Breeds Canada, 2014). Feral populations may be beneficial to present or future breed improvement programs and thus an understanding of the genetic diversity and uniqueness within these populations is required.

1.3 Feral Horse Population of Sable Island

1.3.1 Background and Ancestry

Sable Island is a crescent-shaped sandbar, 49 km in length and 1.3 km wide at its widest point, located 275 km southeast of Halifax, Nova Scotia (43°55' N; 60°00' W; Debeffe et al., 2016). The island is characterized by sand dunes up to 30 m in height, freshwater ponds, vegetation, and its only terrestrial mammal, a population of feral horses. Historical documents indicate that a first shipment of horses was brought to the island from a Boston port in the mid-18th century by Andrew LeMercier (Frasier et al., 2016). A Boston merchant, Thomas Hancock also claimed to have brought horses to the island during this time frame, it has been suggested that the horses may have been confiscated from French settlers and placed on the island in 1755 after the Acadian expulsion (Christie, 1995; Frasier et al., 2016). Clydesdales, Morgan, and Thoroughbred breeding mares and stallions were introduced to the island and selected offspring were rounded up and brought to the mainland to be sold at market from 1801–1940 (Christie, 1995). In 1960 the horses became legally protected from human interference under the Sable Island Regulations of the Canada Shipping Act (Government of Canada, 2016). The shipping act remained in effect until June 2013 when Sable Island became Canada’s 43rd National Park Reserve. Presently, Parks Canada controls access to the island and the horse population. The population has been free from introgression since at least 1935 (Christie, 1995; Welsh, 1975). As a consequence of

the protection, population bottlenecks and genetic isolation has likely led to the population becoming inbred (Lucas et al., 2009).

Sable Island horses are physically similar to sport or light draft horses, ranging in colour from black, to bay and chestnut, with no grey or paint horses. Microsatellite and mtDNA studies concluded that the population shares a common ancestor with the Newfoundland, Shetland, Icelandic, and Norwegian Fjord breeds (Plante et al., 2007; Prystupa et al., 2012a; Prystupa et al., 2012b; Uzans et al., 2015). The studies conducted using microsatellite markers used between ten and thirty-eight microsatellite loci, calculating a range of heterozygosity between 0.58 and 0.647 (Lucas et al., 2009; Plante et al., 2007; Prystupa et al., 2012b). A study examining the maternal lineages of the Sable Island horses assessed a 378-bp region of the mitochondrial displacement loop, identifying 54 informative loci, this determined a mean nucleotide diversity of 0.0007 (Prystupa et al., 2012a). The horse population was also found to have the greatest genetic divergence and the least genetic variation of any Canadian breed, making it an ideal candidate for genetic conservation efforts (Plante et al., 2007; Prystupa et al., 2012a). Additional genome wide studies are required to determine the Sable Island horse ancestry and diversity levels with greater accuracy.

1.4 The Alberta Foothills Feral Horse Population

1.4.1 Background and Ancestry

The feral horse population living in the Alberta foothills has for many years been the subject of debate (Derworiz, 2014; Offin, 2016; Rieger, 2018). The population exists in the southeastern foothills of the Rocky Mountains and is most densely concentrated in the Sundre region (Alberta Environment and Parks - Feral Horses, 2018). The feral horses exist as a largely unmanaged population within an ecosystem shared with large predators and browsing herbivores. Horse roundups began in the 1930s but official attempts to manage the population were not made until 1956 when a joint Federal-Provincial agency instituted a policy of removal of the feral horses on the Eastern slopes of the Rocky Mountains (Salter, 1978). The extermination failed despite over two thousand horses being removed between 1962 and 1972 (Salter and Hudson, 1980). Public perception has shaped the feral horses' history and distribution; by 1974 horse-hunting permits had been phased out due to public scrutiny (Bearcroft, 1974). Concerns from the public regarding the inhumane treatment and capture methods of the horses led the Alberta Government

to create *Horse Capture Regulations* under the *Stray Animals Act* in 1993 (Alberta Provincial Government, 2016). The Act, primarily developed to ensure the humane treatment of the animals, subsequently limited the number of animals that could be captured by preventing the use of weapons, vehicles, and snares in captures, as well as requiring the provision of food and water for the duration of the capture. The horses have since been subjected to other pressures such as habitat range restriction due to habitat loss to infrastructure, settlement and agriculture, as well as culls. In 2013 the government established The Feral Horse Advisory Committee to review and provide management recommendations, including the acceptable number of horse captures / culls each year (Alberta Provincial Government, 2016). The horse represents a significant part of Alberta's heritage; the settlement of the West was largely dependent on the horse, and the horse is culturally significant to the indigenous peoples referred to as "The Plains Indians". However, horse presence raises concerns about negative impacts on native ecosystems and interference with resource industries, particularly cattle grazing. The Provincial government and public organizations such as the Wild Horses of Alberta Society (WHOAS) may have different perspectives of the feral horse population and its place on the foothills landscape. However, both agree that more research needs to be done on the population's origins and impacts on the ecosystem. Research will provide the insight necessary to make appropriate and educated management decisions.

Since 2013, the Alberta Environment and Parks has conducted annual aerial counts of the feral horse population, surveying six equine zones (2015 excepted). The Sundre equine zone has the highest concentration of horses and following a drop in 2013, has been increasing in numbers with 1015 individuals being counted in January 2018. In 1978, the total population was estimated to be approximately one thousand individuals, while the most recent feral horse counts indicate a total of 1721 horses (Alberta Environment and Parks - Feral Horses, 2018; Salter, 1978).

In contrast with the Sable Island feral horse population, the Alberta horse population has a poorly documented history, making it difficult to conclude the origin of the population. Historical records indicate that horses were used among the Blackfoot Confederacy at least from 1730, and these horses may have originated from the northward expansion of the Spanish mustang. Though, the horses may also have originated from stock used in logging and mining operations (Alberta Provincial Government, 2016). Throughout the population's range, horses

exhibit a wide range of phenotypic variation in height, weight, colour, and facial shape, which is consistent with a mixed breed heritage. However, this has not been verified using genetic tools.

1.5 SNP Array

Single Nucleotide Polymorphisms (SNPs) are genome positions at which there are at least two distinct nucleotides present within a population (LaFramboise, 2009). Variations in SNPs predispose an individual to having a specific phenotype and can influence susceptibility to diseases. Differentiated breeds, ethnicities and populations have different frequencies of alleles, which enable the identification of breeds and populations that are closely related. The equine genome project utilized the genome sequence of a Thoroughbred mare, a Quarter Horse and the partial sequence from seven additional breeds to develop a SNP map of *Equus ferus caballus* with more than three million markers (Finno and Bannasch, 2014). The SNP map developed is an excellent resource for studies investigating equine genetics; the discovery also conferred the development of the whole genome SNP microarray. SNP microarrays are a type of DNA microarray based on similar principals. Fragmented, single stranded DNA is hybridized with a strand, which is complementary to the SNP being harboured. The complementary strand carries either a yellow fluorescence which when bound to DNA glows brighter or dimmer dependent on the degree of complementarity of the strand or a red or green fluorescence which corresponds to the allele at the SNP location (LaFramboise, 2009). These techniques facilitate the identification of the allele at a locus.

SNP arrays allow for the direct analysis of sequence variations and whole genome coverage, which increases the accuracy of within and among individual diversity, as well as breed diversity. Previous studies conducted have utilized microsatellite markers to determine the ancestry of the Sable Island horses. Individual microsatellite markers are more informative relative to individual SNP loci, and benefit from widespread use and a greater amount of allelic variation (Kalinowski, 2002; Liu et al., 2005). Microsatellites, however, do not have a standardized method of allele scoring, resulting in difficulty of data sharing and comparison between studies that cross international borders (de Valk et al., 2009). In population structure analyses SNPs consistently outperform microsatellites due to the genome wide distribution and standardized scoring methods (Helyar et al., 2011; Liu et al., 2005; Nielsen et al., 2009). The use of SNP arrays to determine the ancestry and genetic structure of the Sable Island and Alberta Foothills populations will increase the accuracy of the estimate and facilitate future analyses.

1.6 Objectives and Predictions

1.6.1 Objectives

I aimed to use equine SNP microarrays to assess broad patterns of genetic diversity, with increased accuracy relative to previous studies, in the Sable Island and Alberta Foothills feral horse populations with a particular focus on the amount of variation present within each population, extents of inbreeding, and the breed origins of each.

1.6.2 Sable Island Predictions

Based on previous studies, I predict that the horses of Sable Island are most closely related to breeds of Nordic origin such as the Shetland, Icelandic, and Fjord breeds (Plante et al., 2007; Prystupa et al., 2012a; Prystupa et al., 2012b). Similarly, due to the isolated nature of the Sable Island population, I hypothesize that the population will have a substantial degree of genetic divergence relative to other breeds, as well as little genetic variation within the population.

1.6.3 Alberta Foothills Predictions

I predict that the Alberta Foothills population may have originated from the Spanish horse, but gene flow and admixture from other stray and feral populations has diluted the bloodline. The closest relatives of the Foothills population are likely Spanish and draft horses, which were commonly used in the mining and forestry industries in Western Canada. Due to a higher population size compared to Sable Island and recent admixture, it is likely that the Alberta feral horse population will be less genetically diverged from the breed of origin(s) and have a higher degree of genetic variation.

2. METHODS

2.1 Sable Island Study Site

Sable Island is a crescent-shaped sandbar, 49 km in length and 1.3 km wide at its widest point, and is located 275km southeast of Halifax, Nova Scotia (43°55'N; -60°00'W; Debeffe et al., 2016). The island is characterized by sand dunes with heights up to 30 m above sea level and its vegetation, which has been extensively studied and is primarily composed of marram grass (*Ammophila breviligulata* Fernald; Contasti et al., 2012; Freedman et al., 2012; Stalter and Lamont, 2006; Tissier et al., 2013). The horses span the length of the island but are concentrated most densely in the West end, and the vegetated areas, though density is subject to change with the season (Contasti et al., 2012; van Beest et al., 2014; Welsh, 1975).

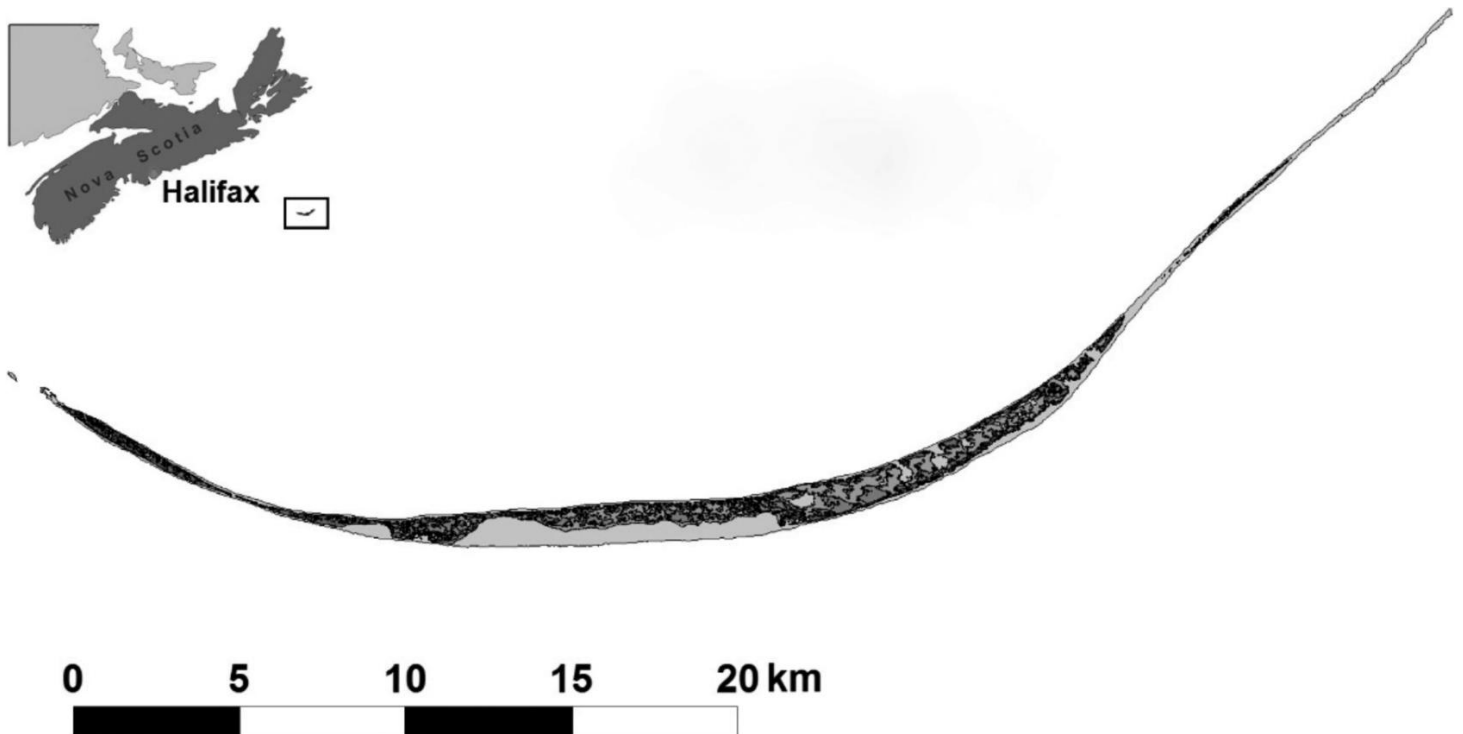


Figure 2.1: A map of Sable Island, Nova Scotia, Canada and its relative location to Halifax Nova Scotia (Cabrera et al., 2017).

2.2 Alberta Foothills Study Site

The horses range in the Eastern slopes of the Rocky Mountains; this range is divided into ‘equine zones’, assigned by the Alberta government, with the Sundre equine zone being the most densely populated region (Figure 2.2; Alberta Environment and Parks - Feral Horses, 2018). The foothills of the Rocky Mountains are predominantly forested with feral horses showing a preference for grasslands and shrublands (Girard et al., 2013a; Salter and Hudson, 1980). Where these are unavailable and during the winter season disturbed areas such as road side edges and cutblocks (harvested conifer forest) are preferentially selected by the horses (Girard et al., 2013a; Girard et al., 2013b). Accessibility and human presence also impacts horse habitat selection, with accessibility to habitat being improved by trails and roads but with increasing human use of trails and roads, they are decreasingly used by horses.

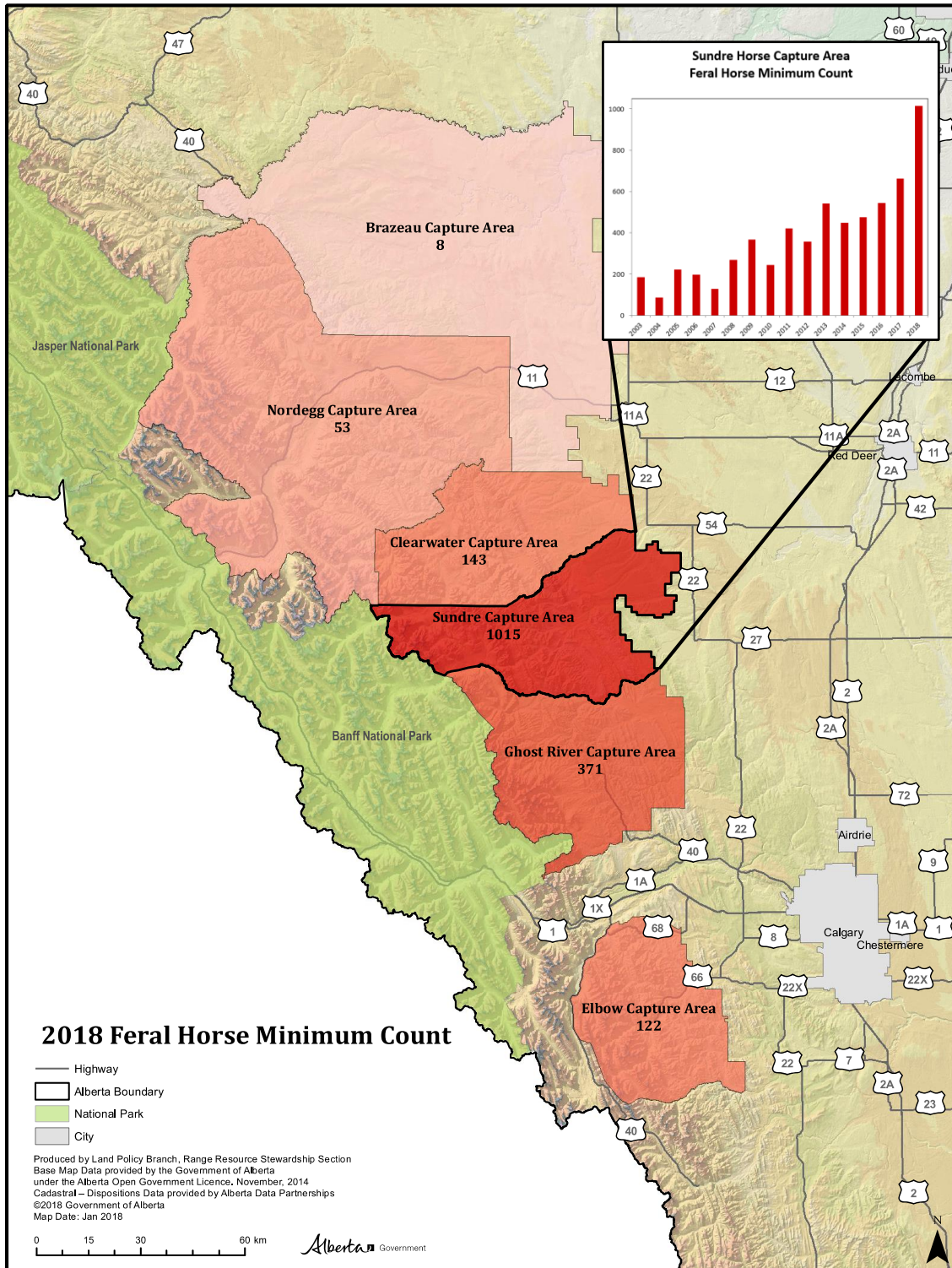


Figure 2.2: The 2018 feral horse survey count and equine zone divisions in Alberta, Canada (Alberta Environment and Parks, 2018).

2.3 DNA samples and genotyping

For the Sable Island population, DNA was obtained from twenty-seven hair samples collected from across the island in 2009 (Figure 2.3). For the Alberta population, hair and tissue samples were collected from 19 individuals who were captured for adoption or being sold at market. The Alberta samples were collected over a gradient from Bragg Creek to North of Sundre, with 12 samples coming from the Ghost/ Sundre equine zone, 3 from the Elbow zone, 2 from the Brazeau zone, and an additional 2 samples from an unknown location. Samples were genotyped using the Illumina (Illumina, Inc., San Diego, CA, USA) Equine GGP 65 array (outsourced to Geneseek, Lincoln, Nebraska). This data was then combined with Equine SNP50 genotypes from 795 individuals from 37 breeds from Petersen *et al.* (2013) and 34 individuals from 4 breeds from Hendrickson *et al.* (2013). Only SNPs present across the three data sets and located on autosomes, were retained.

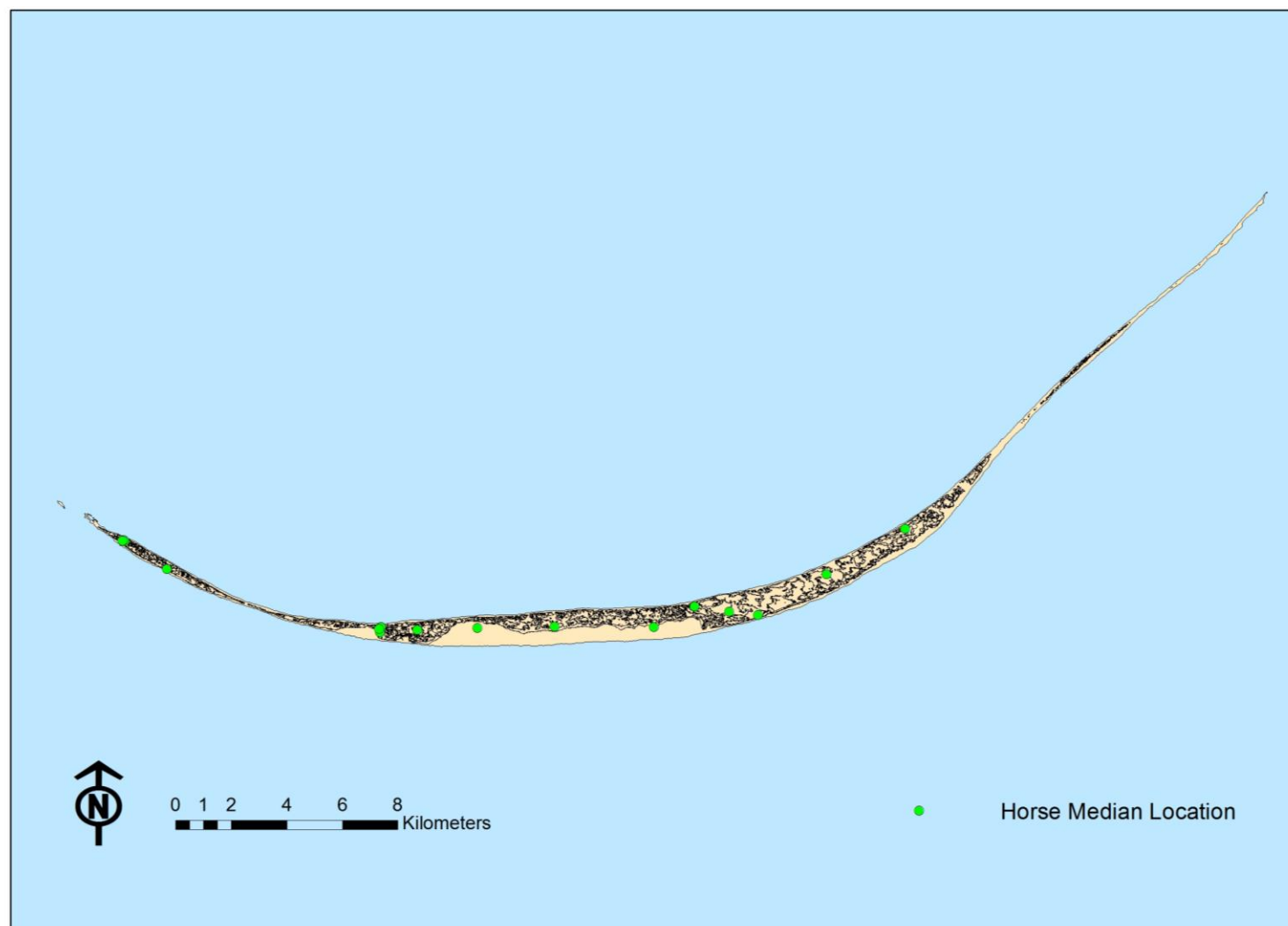


Figure 2.3: An ArcGIS map of Sable Island with the median locations plotted for the eighteen horses (of twenty-three successfully genotyped samples) from the Sable Island population, which have median locations, recorded.

2.5 Data Pruning

Firstly, the samples that failed to successfully genotype were removed from the populations, four samples were eliminated from each of the Sable Island and Alberta data sets at the suggestion of geneseek. SNPs with minor allele frequencies (MAF) less than 0.05 were removed. As in Petersen et al. (2013a), MAF were estimated while excluding the seven discovery breeds used to develop the Equine SNP50 array (Akhal Teke, Andalusian, Arabian, Icelandic, Quarter Horse, Standardbred, and Thoroughbred) to avoid introducing ascertainment bias. The data set was then filtered to exclude variants and individuals with a missing call-rate exceeding 10%. Finally, SNPs were pruned to exclude pairs of variants in high linkage disequilibrium (LD) using a window size of 50 SNPs and an R^2 cut-off value of 0.5 across all breeds. All data pruning was performed in PLINK (1.90 beta; Chang et al., 2014). After filtering 857 samples, and 28 204 SNPs remained with a total genotyping rate of 0.99577, including 23 Sable Island and 15 Alberta samples.

2.6 Within-Breed Diversity

2.6.1 Proportion of Polymorphic Sites

Small populations were excluded from the data set for the proportion of polymorphic sites analysis (the Colonial Spanish horse 3, the Florida cracker 7, the North American Peruvian 5, and the Spanish Barbs 5). The number of polymorphic sites per breed was determined in the Arlequin program (3.5.2; Excoffier and Lischer, 2010).

2.6.2 Homozygosity and Inbreeding

The frequency of expected (E_{HOM}) and observed homozygosity (O_{HOM}) were estimated using PLINK (-het) by the number of observed homozygous loci/number of non-missing loci. As well, the excess of homozygosity inbreeding coefficient (F_{HOM}) was estimated for each individual using PLINK and averaged within each of the 43 breeds using the option -het ($F_{\text{HOM}} = [\text{number of observed homozygous loci} - \text{number of expected homozygous loci}] / [\text{number of non-missing loci} - \text{number of expected homozygous loci}]$).

2.7 Among-Breed Diversity

2.7.1 Breed Divergence

Small populations were excluded from the data set to avoid introducing a bias, as small sample sizes will consequently lead to inflated F_{ST} estimates (Holsinger and Weir, 2009; Kalinowski, 2005; Morin et al., 2009; Wang, 2012). As a result four populations were removed for all among breed diversity analyses (the Colonial Spanish horse, the Florida cracker, the North American Peruvian, and the Spanish Barbs). Pairwise F_{ST} values across all pairs of breeds (populations) were estimated in Arlequin and significance tested by permuting individuals between populations 10 000 times.

2.7.2 AMOVA

An analysis of molecular variance (AMOVA) was conducted in Arlequin to estimate hierarchical fixation indices. The levels of genetic structure were defined as within individuals (all individuals), within breeds, within groups of populations (all other populations vs. Sable Island) and among groups of populations (all other populations vs. Sable Island). This was based on preliminary analyses, which suggested that Sable Island was highly diverged while Alberta was not.

2.8 Among-Breed Relationships

2.8.1 Principal Component Analysis

Principal component analyses (PCA) were conducted in PLINK using 37 495 autosomal SNP sites from the 43 horse breeds (not filtered for linkage disequilibrium as in Petersen et al., 2013). The top three principal components of the variance-standardized relationship were extracted and plotted.

2.8.2 Bayesian Cluster Analysis

The ADMIXTURE program was used to evaluate ancestry proportions for K ancestral populations; more stringent filtering for linkage disequilibrium ($R^2 = 0.1$) was used, as the program assumes no linkage of loci (1.3; Alexander and Novembre, 2009). As a result, 14 032 SNPs remained and were included in this analysis. I ran ADMIXTURE with K values ranging from 1 through to 44 ($n+1$), and the most likely K value was identified using ADMIXTURE's cross-validation procedures (Alexander et al., 2015).

3. RESULTS

3.1 Within-Breed Diversity

3.1.1 Proportion of Polymorphic Sites

The mean proportion of polymorphic sites per population was 91.82%, while the range spanned from 74.5 – 99.17%. The Sable Island population had the lowest proportion of polymorphic sites at 74.5%, whereas the Alberta feral horse population had a relatively high level of polymorphism (96.06%). The Sable Island population and the Clydesdale horse breed have the highest observed homozygosity at 116% of the expected. The Alberta population also has a high level of homozygosity at 110% of the expected, relative to the mean of 106% expected homozygosity. The Hanoverian horse breed, a German riding breed which allows outcrossing, has the lowest observed homozygosity and the only breed to be equal to the expected at 99.7%.

3.1.2 Homozygosity and Inbreeding

The excess of homozygosity inbreeding coefficient (F_{HOM}) averaged 0.119 across all breeds (Table 1). As expected, Sable Island horses exhibit a high degree of inbreeding, indicated by the highest degree of mean inbreeding ($F_{\text{HOM}} = 0.307$; range from 0.237 to 0.427, Table 1). The Alberta feral horse population has a mean inbreeding coefficient of 0.188, greater than the average (0.119).

Table 3.1. Number of samples (N), the excess of homozygosity inbreeding coefficient (F_{HOM}), the minimum and maximum inbreeding statistic within each population, and the observed and expected homozygosity values. The expected homozygosity value for all populations is 0.656.

	N	F_{HOM}	Min	Max	O(HOM)	% of Exp
Akhal Teke	18	0.091	0.020	0.255	0.687	104.696
Alberta	15	0.188	-0.023	0.406	0.721	109.894
Andalusian	18	0.130	0.051	0.323	0.701	106.828
Arabian	24	0.112	0.041	0.324	0.694	105.858
Belgian	30	0.152	0.092	0.202	0.708	107.979
Clydesdale	24	0.306	0.183	0.372	0.761	116.038

Colonial Spanish Horse	3	0.099	0.081	0.117	0.690	105.191
Caspian	18	0.060	-0.011	0.147	0.677	103.140
Exmoor	24	0.280	0.097	0.559	0.752	114.685
Fell Pony	21	0.150	0.117	0.224	0.708	107.876
Finn Horse	27	0.100	0.071	0.157	0.690	105.276
Florida Cracker	7	0.158	0.004	0.365	0.710	108.283
Franches-Montagnes	19	0.106	0.040	0.227	0.692	105.562
French Trotter	17	0.070	0.022	0.131	0.680	103.656
Hanoverian	15	-0.005	-0.043	0.029	0.654	99.739
Icelandic	25	0.155	0.120	0.280	0.709	108.135
Lusitano	24	0.098	0.015	0.195	0.690	105.159
Maremmano	24	0.006	-0.030	0.041	0.658	100.334
Miniature	21	0.127	0.093	0.204	0.699	106.861
Mangalarga Paulista	15	0.249	0.165	0.352	0.742	113.075
Mongolian	19	0.067	-0.075	0.096	0.679	103.511
Morgan	40	0.094	0.009	0.304	0.688	104.920
Mustang	11	0.072	0.019	0.237	0.681	103.768
North American Peruvian	5	0.130	0.041	0.251	0.701	106.817
New Forest Pony	15	0.060	0.020	0.108	0.676	103.124
Norwegian Fjord	21	0.179	0.124	0.253	0.718	109.414
North Swedish	19	0.182	0.119	0.263	0.719	109.574
Percheron	23	0.139	0.097	0.185	0.704	107.298
Peruvian Paso	21	0.084	0.046	0.144	0.685	104.393
PR Paso Fino	20	0.136	0.030	0.320	0.703	107.141
Paint	25	0.003	-0.037	0.063	0.657	100.153
Quarter Horse	40	0.008	-0.035	0.117	0.659	100.453
Sable Island	23	0.307	0.237	0.427	0.761	116.120
Saddlebred	25	0.080	0.030	0.127	0.683	104.209
Shetland	27	0.232	0.153	0.398	0.736	112.172
Shire	23	0.188	-0.041	0.290	0.721	109.876
Spanish Barbs	5	0.102	0.018	0.188	0.691	105.368
Standardbred Norway	25	0.112	0.050	0.165	0.694	105.871
Standardbred US	15	0.138	0.087	0.217	0.703	107.236
Swiss Warmblood	14	0.000	-0.036	0.068	0.656	99.993
Thoroughbred UK	19	0.058	0.014	0.116	0.676	103.048
Thoroughbred US	17	0.052	-0.011	0.094	0.674	102.712
Tuva	15	0.065	0.011	0.157	0.679	103.520

Mean	19.907	0.119	0.046	0.221	0.697	106.25
Total	856					
Min		-0.005	-0.075	0.029	0.654	99.739
Max		0.307	0.237	0.559	0.761	116.120

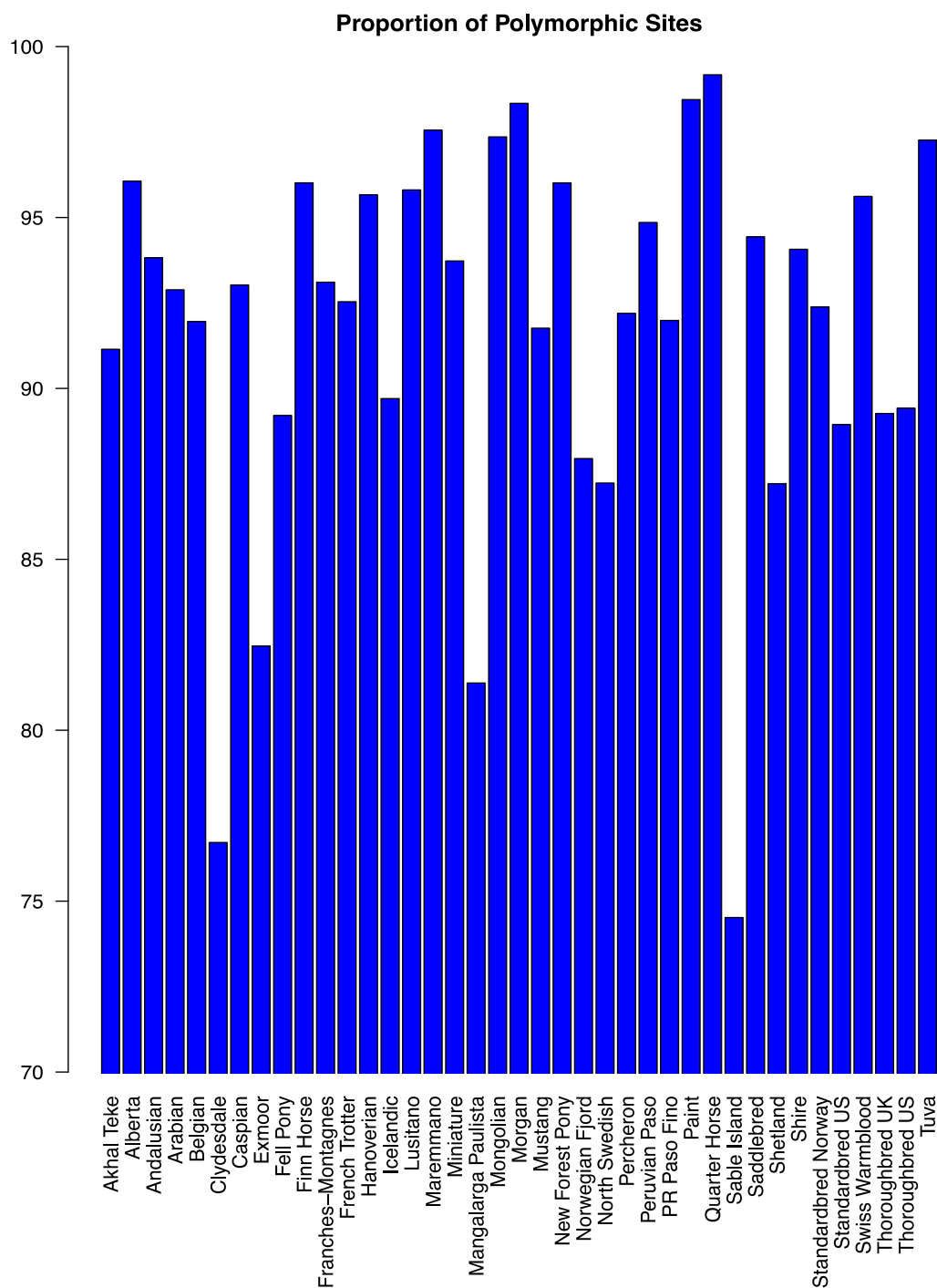


Figure 3.1: A bar plot representing the proportion of polymorphic sites found across each populations data set, the mean is 91.82%. Sable Island has the lowest proportion of polymorphic sites (74.5%), while the Quarter Horse has the largest proportion (99.17%).

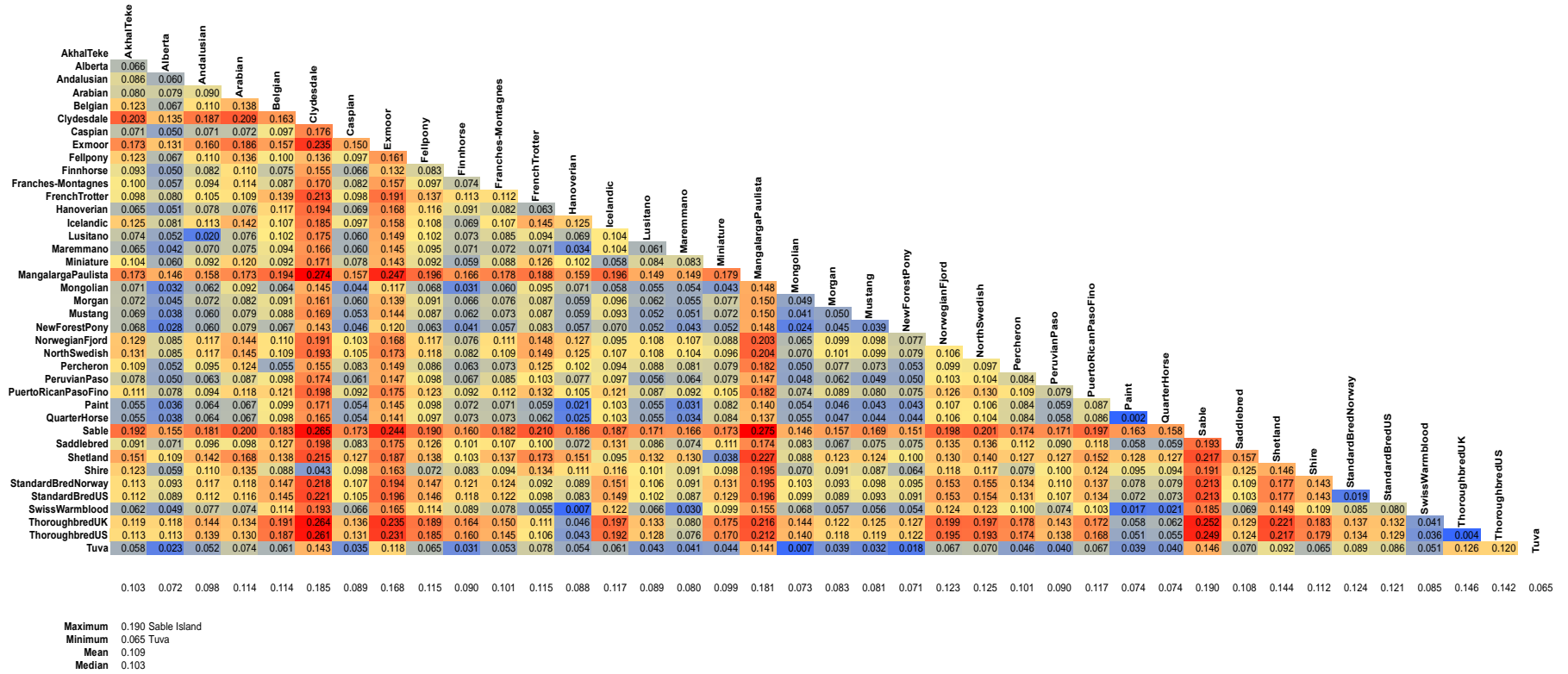
3.2 Among-Breed Diversity: Breed Divergence

3.2.1 Breed Divergence

Pairwise F_{ST} values were determined for 39 breeds and the means obtained for each breed (Figure 3.2). The mean F_{ST} value across all populations was 0.109; all F_{ST} values were highly significant ($p < 0.001$). The analysis identified the Sable Island feral horse population as the most distinct of all the horse breeds assessed with a mean F_{ST} of 0.190, with even its minimum F_{ST} value still indicating great genetic distinction (F_{ST} range is 0.146 Mongolian to 0.275 Mangalarga Paulista; Figure 3.1). Conversely, the Alberta feral horse population has a low mean genetic distinction ($F_{ST} = 0.072$) with values ranging from 0.023 – 0.155 (Tuva – Sable Island). The maximum F_{ST} observed across all pairings was 0.275 between the Sable Island feral horse population and the Mangalarga Paulista.

3.2.2 AMOVA

The AMOVA conducted on the set of 39 breeds indicated that 82.37% of the variation was within individuals ($p < 0.001$), 9.8% of genetic variance originated among populations ($p < 0.001$), and 0.61% of the variation was among individuals within populations ($p = 0.15$). An additional 7.2% of variance was found to be from among groups of populations, thus between Sable Island and the 38 remaining populations included in the assessment ($p = 0.017$).



3.3 Among-Breed Relationships

3.3.1 Principal component analysis

A principal component analysis was used to visualize variation across the 857 individuals. The first two components explained 59.6% of the variation present, while the third component accounted for an additional 12% of variation (Figure 3.3). The first two components had Sable Island horses grouping with the Mongolian, Finn and Exmoor horses (Figure A.2), but when including the third component Sable Island horses grouped predominantly on their own (Figure 3.3). Additionally, the Sable Island individuals grouped closely together, indicating minimal diversity levels.

In contrast, Alberta feral horses did not form a clear cluster, which is consistent with a history of admixture from various breeds (McVean, 2009). The primary cluster formed by the Alberta sample centered between draft breeds and the Standardbred horse, while other horses grouped more closely with the Morgan horse. This is highly indicative of mixed breed origins and the diversity present across the landscape.

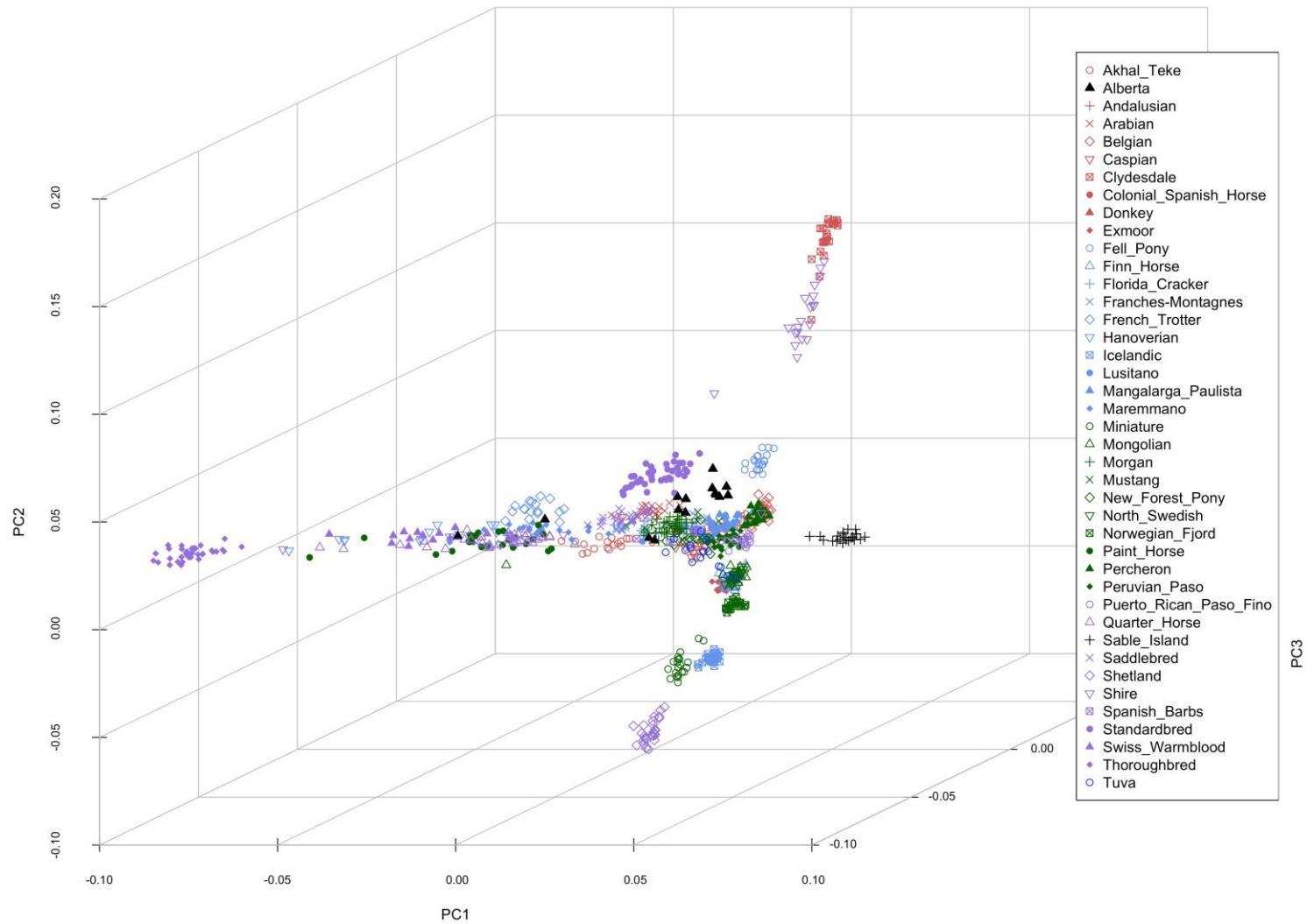


Figure 3.3: A three-axis principal component analysis of 43 horse breeds, based on 37 495 autosomal SNP sites. The leading three components explain approximately 72% of the variation (PC1 42.96%, PC2 16.6%, and PC3 12.76%).

3.3.2 Bayesian Cluster Analysis

In the Bayesian clustering analysis the marginal likelihood was maximized at $K = 25$ (Figure A.4). Results suggest that the Sable Island population is strongly differentiated from other horse breeds, having minimum genetic similarities to the breeds included in the analysis. Sable Island was separated out as a distinct sub-population at $K = 4$. At $K = 3$, Sable Island shared the majority of its group allocation with breeds of Nordic origins, including the Shetland, Icelandic, Miniature, Exmoor, Mongolian and Finn horse (Figure A.3). Conversely, the Alberta feral horse population appeared strongly admixed with great variation across individuals. On average however, excluding a suspected stray, the Alberta population consists of 15% Quarter and/or Paint horse (the paint horse is often distinguished by phenotype and is genetically very similar to the Quarter Horse), 15% Percheron (French draft breed), 9% Clydesdale, 7% Fell pony and 7% Belgian with the remainder being from various other breeds. I removed one outlier due to suspicion that the individual was a stray animal, as the individual was found to be 75% Quarter/paint horse and 25% Thoroughbred, leading me to believe that it was unlikely that the animal was a product of random breeding as she would be if she was apart of the feral population.

Alberta Admixture

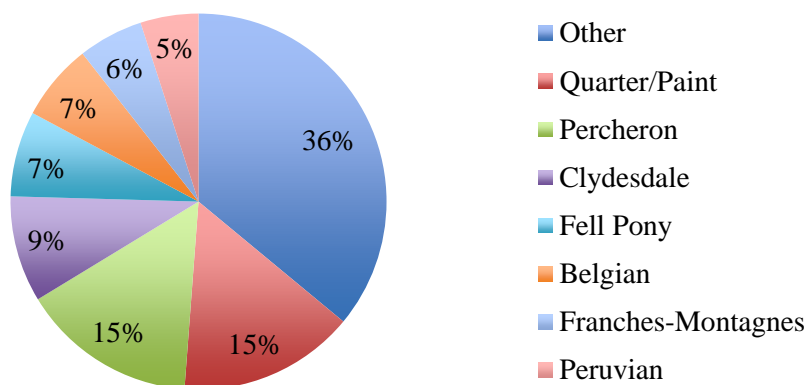


Figure 3.4: The Alberta percentage admixture obtained from the ADMIXTURE simulation results at $K = 25$, while excluding an individual identified as an outlier. The “Other” is a mixture of various breeds that all occurred at a mean of $< 5\%$.

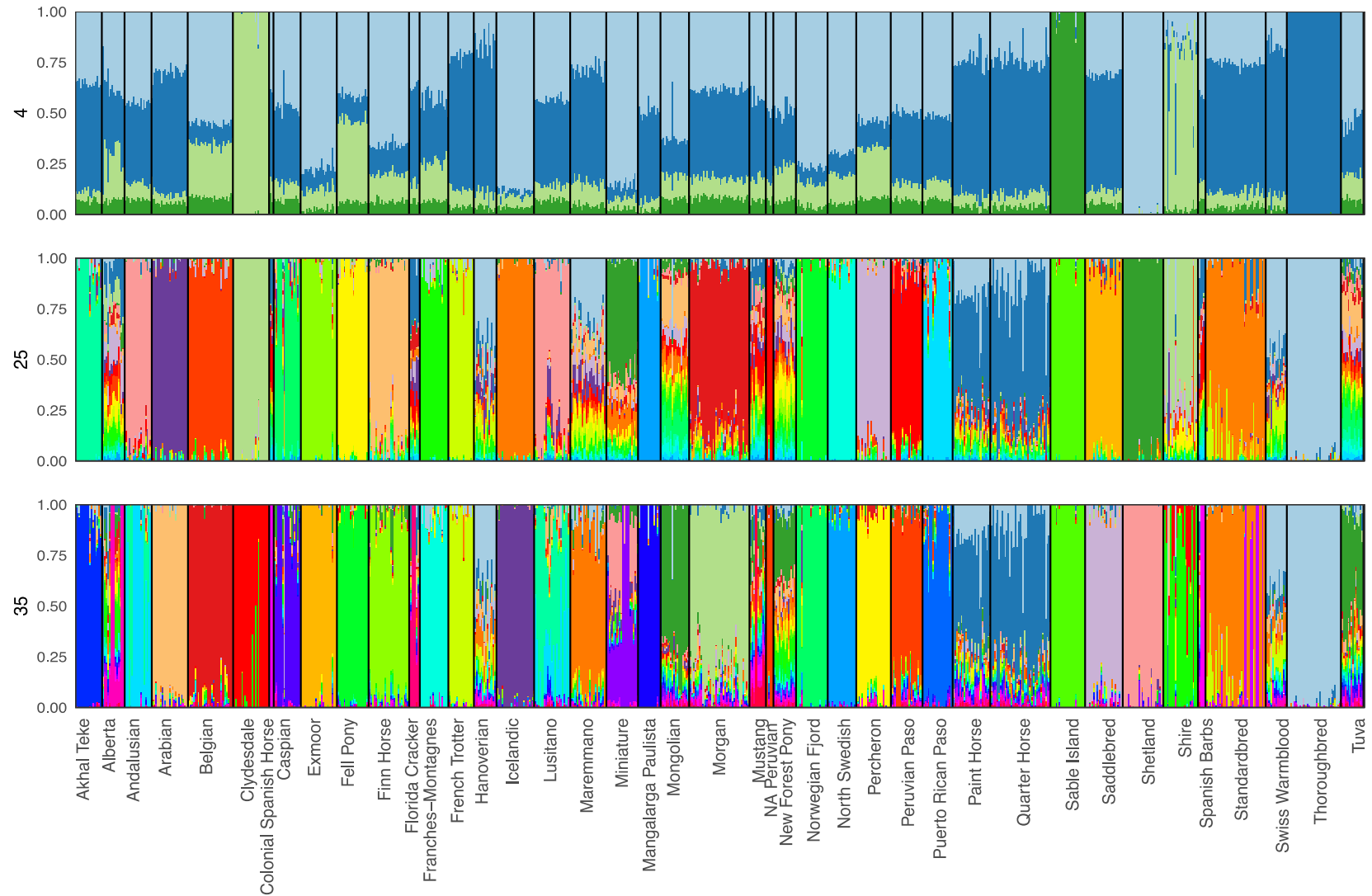


Figure 3.5: Bayesian clustering output from ADMIXTURE depicting the results for three values of K (4, 25, and 35) in 856 horses from 43 populations. ADMIXTURE identified $K = 25$ as the most likely genetic group allocation for the data.

4. DISCUSSION

Although the population genetic structure of domestic horses has been studied extensively (Petersen et al., 2013a; Plante et al., 2007; Prystupa et al., 2012b; Prystupa et al., 2012a), this study is the first to use SNP-based analyses to assess the genetic diversity levels, uniqueness, and ancestral relationships of the Sable Island and Alberta feral horse populations. My data indicate that Sable Island is the most inbred and the most genetically diverged population, related most closely to the Mongolian horse breed and Nordic breeds. In addition, the analyses indicate that the Alberta feral horse population is relatively inbred relative to other horse breeds and is of mixed-breed origins. These results have significant management implications for both populations.

4.1 Within-breed Diversity

I observed high levels of inbreeding in both the Sable Island and Alberta populations. This result was expected for Sable Island given its small population size, limited historical introgression, and previous analyses leading us to anticipate this result. However, the small population exceeds the inbreeding levels of even the most stringent and rare breeds. In addition to a small population size, this is perhaps a result of the intensity and frequency of bottleneck events on the population, including major fluctuations in population size (Ure et al., 2016). Moreover, the natural and polygynous social structure limits male mating success, potentially further decreasing the effective population size (Chesser, 1991; Storz, 1999). For example, in a feral equid population, the Asiatic wild ass, it was estimated that only 25% of adult males contributed to the next generation (Renan et al., 2015). This is predicted to have impacts on the amount of genetic variation and genetic drift, as over only four generations a substantial shift in allele frequencies was observed (Renan et al., 2015). The similar social structure of the horses on Sable Island suggests that comparable male contribution rates are expected, though analyses on the distribution of male mating success still need to be conducted. The implications of this in relation to inbreeding depression have, however, yet to be determined. In a similar feral livestock system (St-Kilda Soay Sheep) inbreeding depression was found to negatively influence first year survival in males, annual breeding

success in adults, and maternal inbreeding was found to cause depression in birth-weight and 4-month weight (Berenos et al., 2016; Pemberton et al., 2017).

Previous studies conducted on the population's inbreeding levels describe an observable gradient of the degree of inbreeding across the island, increasing from east to west (Lucas et al., 2009). The Sable Island samples included in my analyses are relatively well distributed across the island considering the increased population density in the west (Figure 2.3; Contasti et al., 2012). This suggests that our mean inbreeding levels are a good representation of the population as whole. In addition to a high degree of inbreeding, the Sable Island population exhibited the lowest proportion of polymorphic sites across the population and had among the lowest level of mean heterozygosity across individuals. Studies of a similar nature, evaluating polymorphism and heterozygosity in isolated feral livestock populations, like the St-Kilda Soay sheep population in Scotland, found similar levels of heterozygosity ($H_E = 0.223$) and polymorphism ($P_n = 0.667$) for 1397 SNPs examined (Kijas et al., 2009). The Soay sheep population has been isolated on the island of Hirta, free from introgression since the 1930s and has been the subject of broad studies since the 1950's (Walter, 2006). The population originating from domestic sheep shares many commonalities with the Sable Island population, including large over-winter mortalities (>50%; Walter, 2006) suggesting that similar levels of genetic drift and selection may be felt. This proposes that the two populations are expected to have similar levels of variation across the populations, though the lower number of SNPs assessed in the Soay population may account for the slightly lower level polymorphism observed. The correlation across these data sets is suggestive that the results obtained from this study are accurate.

While not as striking as in Sable Island horses, results also indicated levels of inbreeding were elevated in Alberta. The low level of diversity exhibited within individuals in the Alberta population, coupled with a high amount of population level diversity ($P_n = 0.96$; Figure 3.2), suggests that barriers to migration are present across the Alberta landscape, resulting in subpopulations. Failing to account for population substructure in an analysis such as this may lead to an overestimate of homozygosity and as a result an inflated inbreeding coefficient (Dharmarajan et al., 2013; Wahlund, 1928). This suggests our inbreeding estimates may or may not be inflated and asserts the need

for additional research into population genetic structure. Landscape composition can drive population distribution and gene flow; discontinuities in the landscape may inhibit movement and gene flow across populations, leading to the development of unique subpopulations by isolation (Sexton et al., 2014). These discontinuities may be in the form of anthropogenic fragmentation or natural landscape barriers that impede movement, functionally isolating populations. In Alberta, where the landscape is heterogeneous, it is likely that infrastructure, cities, major highways, or mountains and powerful river flows are responsible for limiting migration events between subpopulations. The effect of anthropogenic fragmentation on gene flow in an ungulate population in Alberta is well documented by Gubili et al. (2017). They found that even where anthropogenic fragmentation occurred at low levels, gene flow across caribou populations was limited to distances less than 100 km despite the high mobility of the species (Gubili et al., 2017). While caribou are migratory and are therefore likely to be highly impacted by anthropogenic barriers, these barriers may affect feral horse populations with equivalent mating structures similarly. Conversely, landscape barriers may not have led to differentiation but prevented already genetically different groups (ie. sub-populations founded by stray animals) from mixing. Assessing the equine zones individually demonstrates that there is potentially an inbreeding gradient present across the Alberta horses range (Table A.1). Using a principal component analysis (Figure A.4) I identified 2 potential subpopulations within the Ghost/ Sundre equine zone. The Brazeau equine zone also formed its own cluster, and has the highest level of mean inbreeding. Following the Brazeau equine zone, the subpopulation identified within the Ghost/ Sundre equine zone has the next highest level of inbreeding, followed by the Elbow equine zone and lastly, the remainder of the Ghost/ Sundre equine zone (Table A.1). The data set for each subpopulation is too small to draw conclusions with authority, though demonstrates the presence of multiple subpopulations across the Alberta landscape and illuminates an exciting avenue for future research efforts.

4.2 Among-breed Diversity

F_{ST} values identified the Sable Island population as the most genetically distinct horse population, reinforcing earlier results based on microsatellite data (Plante et al., 2007; Prystupa et al., 2012b). This is likely due to a founder event and subsequent bottlenecks,

small population size, isolation, and limited introgression. This finding was also apparent in the principal component analysis, as well as the Bayesian cluster analysis results, which highlights four populations, including the Sable Island population, as being the most distinct (Figure 3.3 and 3.5). This study adds support to the suggestion that the Sable Island population is a unique and important animal genetic resource, and that it should be prioritised for conservation efforts (Plante et al., 2007; Prystupa et al., 2012b; Prystupa et al., 2012a). The loss of a rare horse population, like Sable Island, would mean a major loss to among-breed diversity, which is important for the maintenance of biodiversity (Allendorf et al., 2010; Barker, 2001; Hoban et al., 2013).

In contrast to the Sable Island population, the degree of divergence in the Alberta feral horse population is low, though not negligible. Differentiation of the Alberta population is similar to that seen in landrace breeds, such as the Tuva, Mustang and New Forest Pony (Petersen et al., 2013a). Landrace breeds often originate from escaped animals and feral populations; they are locally adapted breeds that have developed over time through adaptation to their natural and cultural environment. These breeds are generally genetically uniform, though more diverse than members of a standardized breed (Petersen et al., 2013b). Landrace breeds have the potential to form quickly, developing in a couple of centuries. The processes acting to form landrace breeds are potentially similar to those influencing the development of the Alberta population. The horses exist with little to no management within an ecosystem with large predators and other grazing herbivores. While some of the sub-populations may be from recent admixture (i.e., outliers identified as suspected strays) from escaped or loose stock (Alberta Environment and Parks - Feral Horses, 2018), historical evidence maintains that feral horses have been present in Western Canada since the mid-18th century (Salter and Hudson, 2016). Because the Alberta population, or a subset there-of, is under similar selective pressures, this population may warrant future investigation or classification as a landrace breed.

4.3 Among-breed Relationships

Our data confirmed the close relationship between the Sable Island population and horses of Nordic origin (Plante et al., 2007; Prystupa et al., 2012a; Prystupa et al., 2012b), but further identified the Mongolian and Nordic breeds as the population's closest relatives. Strong historical and genetic evidence suggests that the Mongolian horse, which spread

across Europe with Attila the Hun in the 5th century and with Genghis Khan during the 13th century played an important role in the founding of Scandinavian horse breeds (Bjørnstad et al., 2003). Historical records indicate that horses were brought to the island from Boston in the United States during the mid-18th century by Andrew le Mercier, and Boston merchant Thomas Hancock also claimed to have brought horses to the island, though there are no official records of this latter introduction (Christie, 1995; Frasier et al., 2016). Since Thomas Hancock was regularly doing business in Halifax, Christie (1995) and others have suggested that Thomas Hancock perhaps introduced horses which originally belonged to the Acadian peoples who had been expelled from Nova Scotia. Results suggest that Sable Island horses may find their origins in horses brought to America by the Swedish and the Dutch in the 17th century. As such, the population of horses presently inhabiting the island are likely primarily a reflection of Nordic horse breeds 260 years ago as opposed to the modern Nordic horses included in the analyses. This relationship suggests that isolated feral livestock populations may resemble ancestral breeds and therefore be a good reservoir for recovering genetic variation that has been lost in more recent generations by intensive selection.

Results indicated that the Alberta population is of mixed breed origins. The PCA shows that the Alberta population does not group collectively, which is suggestive of high levels of genetic diversity and mixed ancestry (McVean, 2009). While the Bayesian cluster analysis ancestry results support the mixed breed heritage of the Alberta horses, it also identifies a strong relationship with draft breeds such as the Quarter/Paint horse, the Percheron and the Clydesdale. This relationship is consistent with historical evidence and predictions that the population has been introgressed with working horses. The US Mustang however, is primarily a mixture of Peruvian and Spanish breeds with the Peruvian horse (17%), Quarter / Paint horse (14%), and the Andalusian / Lusitano breeds (9%) being major contributors. Thus, speculation that the two are closely related is unlikely. The mixed ancestry observed in the Alberta population is also consistent with a population that has a low level of genetic divergence. While the population potentially originated from the northward migration of Spanish horses, the present stock appears to have been highly influenced by draft breeds.

While the results and the conclusions drawn were mostly as anticipated given the characteristics and the history of the populations, relationship analyses such as these are limited by the populations which are included in the data set. I obtained the reference data set from publicly available genomic data on domestic horse populations (Hendrickson, 2013; Petersen et al., 2013a); as such, the study was limited in the number of breeds included in the analysis to those that were publicly available and analyzed with the Illumina equine SNP chip. As a result, of the 905 FAO-recognized horse breeds, only 41 were used as a comparison to our feral populations (FAO Commission on Genetic Resources for Food and Agriculture Assessments, 2015). However, despite these restrictions impacting our data set, the breeds I predicted to be most closely related to the Sable Island and Alberta feral populations were included in the analyses.

4.4 Management Implications

Parks Canada incorporates into its mandate a goal to monitor the ecosystems for which they are responsible to ensure the maintenance of ecological integrity; this encompasses monitoring wildlife populations, like the Sable Island feral horse population, within its borders (Parks Canada, 2017). Due to the degree of isolation and previous analyses indicating inbreeding and uniqueness, Parks Canada took a particular interest in monitoring inbreeding levels on the island when it became a national park reserve in 2013 (Ure et al., 2016). Our study identifies Sable Island as highly inbred relative to even the most exclusive standardized breeds. Establishment of inbreeding depression levels will consequently present Parks Canada with the management implications of this knowledge.

The decision to act to reduce the degree of inbreeding in the population may entail the introduction of new breeding stock to the island. While successful genetic rescue of natural populations has seldom been documented, Hasselgren et al. (2018) documented positive progress after a genetic rescue of a small population of Scandinavian arctic foxes. The small population was highly inbred and suffering from inbreeding depression; the introduction of three males to the population increased the allelic richness in the population by 41%, decreased the mean inbreeding coefficient from 0.14 to 0.08 in five years, and increased juvenile survival and mating success (Hasselgren et al., 2018).

Additional studies suggest that initial benefits of migrant introductions may wane after two – three generations and populations may require continuous attention (Hedrick et al., 2014). The decision to introduce new stallions to Sable Island, however, could potentially have a detrimental impact, further reducing fitness, due to outcrossing depression (Sexton et al., 2014). The island population is highly differentiated from even its closest relations and is likely well adapted to its stressful environment; the introduction of non-adapted genotypes could disrupt local adaptations and introduce genetic incompatibilities (Edmands and Timmerman, 2003; Tallmon et al., 2004; Waller, 2015). Essentially, the benefits of heterosis introduced by gene flow could be masked by the inheritance of genes or behaviours that are poorly suited to the local conditions. A primary example of this was seen when an overall negative absolute fitness response was reported following the outcrossing of a highly genetically divergent marine copepod (*Tigriopus californicus*), generations F1 – F20 were all reported to have reduced survival and reproductive success relative to the parent colony (Hwang et al., 2012). Though, the study also reports that the outcome of genetic rescue is highly dependent on the compatibility of the hybridizing populations (Hwang et al., 2012). This suggests that a great deal of research would be required before action could occur. Alternatively to introducing new breeding stock to the island, Parks Canada could facilitate the movement of individuals across the island. Analyses indicating that individuals are more inbred in the West than the East and future analyses assessing landscape genomics of the horses will determine if this is a viable option to appreciate genomic variation (Lucas et al., 2009). To abstain from action, however, could lead the population to suffer the consequences of inbreeding depression, potentially leading to the loss of the population.

The inbreeding and PCA results highlighted genetically distinct subpopulations. The borders of the equine zones do not appear to reflect the genetic borders that are present across the population. Further investigation into landscape genomics will allow for re-evaluation of the equine zones/ management units based on genetic barriers. A primary example of using genetic structure in the identification of management units integrated the use of fine-scale spatial sampling and genetic cluster analysis of harbour seal (*Phoca vitulina*) populations (Olsen et al., 2014). Incorporating the use of genetic tools into population management unit assessments, in addition to identifying to populations of

greater conservation value, will enable selective culling from highly diverse populations as opposed to removing genetic variation from populations suffering from inbreeding.

One of the main objectives of this study was to use genomic tools and objective measurements to provide unbiased targets for conservation. A global focus to prevent additional loss of among breed diversity suggested that primary consideration should be given to breeds which: 1) are the most genetically distinct, 2) at the greatest risk of extinction, 3) have the largest measure of breed 'merit', and 4) have the most limited within-breed variation (Barker, 2001). Thus, our observations of within-breed variation, breed divergence, and among-breed relationships provide great insight into the uniqueness of the Sable Island and Alberta feral horse populations. The Sable Island horse population was identified as the horse population with the most limited within-breed variation and as the most genetically distinct population. In addition, the isolated Sable Island population demonstrated a potential link to what ancestral horse breeds looked like on a molecular level, this may contribute to the breed's value to society. The Alberta feral horse population was established to have a higher inbreeding level than initially anticipated, suggesting the presence of subpopulations, and was found to be poorly genetically diverged. While these data demonstrate that the Sable Island population is an ideal candidate for conservation efforts to preserve animal genetic resources, the Alberta population is not identified as a genetically distinct population. Despite this, the population's value to the people of Alberta and in Western culture may warrant a level of provincial protection.

4.5 Conclusions

This study aimed to assess broad patterns of genetic diversity in the Sable Island and Alberta Foothills feral horses characterizing the horse populations relative to well-defined breeds and discerning their origins. Investigating the genomic diversity of feral horse (*Equus ferus caballus*) populations in Canada can provide researchers with an important framework for genetic conservation and improve the understanding of the microevolution of natural populations. Using the Illumina EquineSNP74 BeadChip I identified the Sable Island feral horse population as a very genetically unique population, and possessing the least diversity. While the Alberta population was poorly genetically diverged, have low diversity levels within individuals and high diversity levels across the population,

indicative of the presence of subpopulations. These results suggest that the Sable Island population is an ideal candidate for animal genetic resource conservation efforts and further genomic level investigation. In addition, it indicates that the Alberta population should be explored using landscape genetics and subsequently management units should be re-evaluated based on genetic borders. In addition to our major findings, this study strongly illustrates the effectiveness of genetic isolation on genetic differentiation, by indirectly comparing populations with open and closed breeding systems. Genomic tools like SNP microarrays allow us to better understand and predict the consequences of reduced genetic variability. With habitat fragmentation being a common anthropogenic effect more populations are becoming genetically isolated and studies such as this will allow us to predict the evolutionary consequences of this isolation. This study will serve as the foundation for future genomic investigation into these feral populations, enabling exploration into queries that are ecologically and evolutionary relevant for natural populations that are being faced with rapidly changing environments, habitat fragmentation and isolation.

REFERENCES

- Achilli, A., Olivieri, A., Soares, P., Lancioni, H., Hooshier Kashani, B., Perego, U. a, Nergadze, S. G., Carossa, V., Santagostino, M., Capomaccio, S., et al. (2012). Mitochondrial genomes from modern horses reveal the major haplogroups that underwent domestication. *Proc. Natl. Acad. Sci. U. S. A.* 109, 2449–54.
- Agri-food Canada (2008). Canadian animal genetic resources program.
- Alberta Environment and Parks (2018). 2017 Feral horse count summary.
- Alberta Environment and Parks - Feral Horses (2018).
- Alberta Provincial Government (2016). Alberta Environment and Parks. *Feral Horses Alberta*.
- Alexander, D. H. and Novembre, J. (2009). Fast model-based estimation of ancestry in unrelated individuals. 1655–1664.
- Alexander, D. H., Shringarpure, S. S., Novembre, J. and Lange, K. (2015). Admixture 1.3 Software Manual.
- Allendorf, F. W. and Luikart, G. H. (2007). *Conservation and the Genetics of Populations*. First Edit. Malden, MA: Blackwell Publishing.
- Allendorf, F. W., Hohenlohe, P. A. and Luikart, G. (2010). Genomics and the future of conservation genetics. *Nat. Rev. Genet.* 11, 697–709.
- Barker, J. S. (2001). Conservation and management of genetic diversity: a domestic animal perspective. *Can. J. For. Res.* 31, 588–595.
- Bearcroft, N. (1974). *Wild horses of Canada*. J. A. Allen.
- Berenos, C., Ellis, P. A., Pilkington, J. G. and Pemberton, J. M. (2016). Genomic analysis reveals depression due to both individual and maternal inbreeding in a free-living

- mammal population. *Mol. Ecol.* 25, 3152–3168.
- Bilski, D. R., Pie, M. R. and Passos, F. C. (2013). Variable inbreeding effects across life-history stages in a captive carnivorous mammal population. *Anim. Conserv.* 16, 633–640.
- Bjørnstad, G., Nilsen, N. and Røed, K. H. (2003). Genetic relationship between Mongolian and Norwegian horses? *Anim. Genet.* 34, 55–58.
- Cabrera, D., Andres, D., McLoughlin, P. D., Debeffe, L., Medill, S. A., Wilson, A. J. and Poissant, J. (2017). Island tameness and the repeatability of flight initiation distance in a large herbivore. *Can. J. Zool.*
- Carnegie Museum of Natural History (2003). *Horses through time*. (ed. Olsen, S. L.) Lanham, MD: Roberts Rinehart.
- Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M. and Lee, J. J. (2014). Second-generation PLINK: rising to the challenge of larger and richer datasets.
- Chesser, R. K. (1991). Gene diversity and female philopatry. *Genetics* 127, 437–447.
- Christie, B. J. (1995). *The Horses of Sable Island*. Lawrencetown Beach: Pottersfield Press.
- Coltman, D. W., Bowen, W. D. and Wright, J. M. (1998). Birth weight and neonatal survival of harbour seal pups are positively correlated with genetic variation measured by microsatellites. *Proc. R. Soc. London, B* 265, 803–809.
- Contasti, A. L., Tissier, E. J., Johnstone, J. F. and McLoughlin, P. D. (2012). Explaining spatial heterogeneity in population dynamics and genetics from spatial variation in resources for a large herbivore. *PLoS One* 7,.
- Darwin, C. (1859). On the origins of species by means of natural selection. *London:*

Murray 247.

de Valk, H. A., Meis, J. F. G. M., Bretagne, S., Costa, J. M., Lasker, B. A., Balajee, S. A.,

Pasqualotto, A. C., Anderson, M. J., Alcázar-fuoli, L., Mellado, E., et al. (2009).

Interlaboratory reproducibility of a microsatellite-based typing assay for *Aspergillus fumigatus* through the use of allelic ladders: Proof of concept. *Clin. Microbiol.*

Infect. 15, 180–187.

Debeffe, L., Mcloughlin, P. D., Medill, S. A., Stewart, K., Andres, D., Shury, T.,

Wagner, B., Enkins, E., Gilleard, J. S. and Poissant, J. (2016). Negative covariance between parasite load and body condition in a population of feral horses.

Parasitology 1–15.

Derworiz, C. (2014). Plan to cull 200 wild horses triggers debate. *Calgary Her.*

Dharmarajan, G., Beatty, W. S. and Rhodes, O. E. (2013). Heterozygote deficiencies

caused by a Wahlund effect: Dispelling unfounded expectations. *J. Wildl. Manage.* 77, 226–234.

Dunn, S. J., Clancey, E., Waits, L. P. and Byers, J. A. (2011). Inbreeding depression in pronghorn (*Antilocapra americana*) fawns. *Mol. Ecol.* 20, 4889–4898.

Edmands, S. and Timmerman, C. C. (2003). Modeling factors affecting the severity of outbreeding depression. *Conserv. Biol.* 17, 883–892.

Excoffier, L. and Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of

programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* 10, 564–567.

FAO (2011). *Molecular genetic characterization of animal genetic resources.*

FAO Commission on Genetic Resources for Food and Agriculture Assessments (2007).

State of the art in the management of animal genetic resources.

FAO Commission on Genetic Resources for Food and Agriculture Assessments (2011).

Livestock statistics - Concepts, definitions and classifications.

FAO Commission on Genetic Resources for Food and Agriculture Assessments (2015).

The second report on the state of world's: animal genetic resources for food and agriculture. (ed. Scherf, B. D.) Rome, Italy.

Finno, C. J. and Bannasch, D. L. (2014). Applied equine genetics. *Equine Vet. J.* 46, 538–544.

Frasier, T. R., Lucas, Z., McLeod, B. A. and McLoughlin, P. D. (2016). The Horses of Sable Island. In *Sable Island, The Ecology and Biodiversity of Sable Island* (ed. Freedman, B.), pp. 271–298. Markham, ON: Fitzhenry & Whiteside.

Freedman, B., Catling, P. M. and Lucas, Z. (2012). Effects of feral horses on vegetation of Sable Island, Nova Scotia. *Can. Field-Naturalist* 125, 200–212.

Girard, T. L., Bork, E. W., Nielsen, S. E. and Alexander, M. J. (2013a). Seasonal variation in habitat selection by free-ranging feral horses within Alberta's forest reserve. *Rangel. Ecol. Manag.*

Girard, T. L., Bork, E. W., Nielsen, S. E. and Alexander, M. J. (2013b). Landscape-scale factors affecting feral horse habitat use during summer within the rocky mountain foothills. *Environ. Manage.*

Government of Canada (2016). Sable Island Regulations Règlement sur l'île de Sable.

Gubili, C., Mariani, S., Weckworth, B. V., Galpern, P., McDevitt, A. D., Hebblewhite, M., Nickel, B. and Musiani, M. (2017). Environmental and anthropogenic drivers of connectivity patterns: A basis for prioritizing conservation efforts for threatened populations. *Evol. Appl.* 10, 199–211.

Hall, S. J. G. and Bradley, D. G. (1995). Conserving livestock breed biodiversity. *Trends*

Ecol. Evol. 10, 267–70.

Hartl, D. L. and Clark, A. G. (1997). *Principles of Population Genetics*. Third Edit.

Sunderland, Massachusetts: Sinauer Associates, Inc.

Hasselgren, M., Angerbjörn, A., Eide, N. E., Erlandsson, R., Flagstad, Ø., Landa, A.,

Wallén, J. and Norén, K. (2018). Genetic rescue in an inbred Arctic fox (*Vulpes lagopus*) population. *Proc. R. Soc. B Biol. Sci.* 285, 20172814.

Hedrick, P. W., Peterson, R. O., Vucetich, L. M., Adams, J. R. and Vucetich, J. A.

(2014). Genetic rescue in Isle Royale wolves: genetic analysis and the collapse of the population. *Conserv. Genet.* 15, 1111–1121.

Helyar, S. J., Hemmer-Hansen, J., Bekkevold, D., Taylor, M. I., Ogden, R., Limborg, M.

T., Cariani, A., Maes, G. E., Diopere, E., Carvalho, G. R., et al. (2011). Application of SNPs for population genetics of nonmodel organisms: New opportunities and challenges. *Mol. Ecol. Resour.* 11, 123–136.

Hendrickson, S. L. (2013). A genome wide study of genetic adaptation to high altitude in feral Andean Horses of the páramo. *BMC Evol. Biol.* 13, 273.

Hoban, S. M., Hauffe, H. C., Pérez-Espona, S., Arntzen, J. W., Bertorelle, G., Bryja, J.,

Frith, K., Gaggiotti, O. E., Galbusera, P., Godoy, J. A., et al. (2013). Bringing genetic diversity to the forefront of conservation policy and management. *Conserv. Genet. Resour.* 5, 593–598.

Holsinger, K. E. and Weir, B. S. (2009). Genetics in geographically structured

populations: Defining, estimating and interpreting F_{ST} . *Nat. Rev. Genet.* 10, 639–650.

Hwang, A. S., Northrup, S. L., Peterson, D. L., Kim, Y. and Edmands, S. (2012). Long-term experimental hybrid swarms between nearly incompatible *Tigriopus*

californicus populations : persistent fitness problems and assimilation by the superior population. 567–579.

Jarvis, J. P., Cropp, S. N., Vaughn, T. T., Pletscher, L. S., King-Ellison, K., Adams-Hunt, E., Erickson, C. and Cheverud, J. M. (2011). The effect of a population bottleneck on the evolution of genetic variance/covariance structure. *J. Evol. Biol.* 24, 2139–2152.

Kalinowski, S. T. (2002). How many alleles per locus should be used to estimate genetic distances? *Heredity (Edinb)*. 88, 62–65.

Kalinowski, S. T. (2005). Do polymorphic loci require large sample sizes to estimate genetic distances? *Heredity (Edinb)*. 94, 33–36.

Keller, L. F. and Waller, D. M. (2002). Inbreeding effects in wild populations. *Trends Ecol. Evol.* 17, 230–241.

Kijas, J. W., Townley, D., Dalrymple, B. P., Heaton, M. P., Maddox, J. F., McGrath, A., Wilson, P., Ingersoll, R. G., McCulloch, R., McWilliam, S., et al. (2009). A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PLoS One* 4,.

Lacy, R. C. and Ballou, J. D. (1998). Effectiveness of selection in reducing the genetic load in populations of *Peromyscus polionotus* during generations of inbreeding. *Evolution (N. Y)*. 52, 900–909.

Lacy, R. C., Journal, S. and May, N. (1997). Importance of genetic variation to the viability of mammalian populations. *J. Mammal.* 78, 320–335.

LaFramboise, T. (2009). Single nucleotide polymorphism arrays: A decade of biological, computational and technological advances. *Nucleic Acids Res.* 37, 4181–4193.

Lauvie, A., Couix, N. and Verrier, E. (2014). No development, no conservation: elements

from the conservation of farm animal genetic resources. *Soc. Nat. Resour.* 27, 1331–1338.

Librado, P., Fages, A., Gaunitz, C., Leonardi, M., Wagner, S., Khan, N., Hanghoj, K., Alquraishi, S. A., Alfarhan, A. H., Al-Rasheid, K. A., et al. (2016). The evolutionary origin and genetic makeup of domestic horses. *Genetics* 204, 423–434.

Librado, P., Gamba, C., Gaunitz, C., Sarkissian, C. Der, Pruvost, M., Albrechtsen, A., Fages, A., Khan, N., Schubert, M., Jagannathan, V., et al. (2017). Ancient genomic changes associated with domestication of the horse. 445, 442–445.

Ling, Y., Ma, Y., Guan, W., Cheng, Y., Wang, Y., Han, J., Jin, D., Mang, L. and Mahmut, H. (2010). Identification of y chromosome genetic variations in Chinese Indigenous horse breeds. *J. Hered.* 101, 639–643.

Lippold, S., Matzke, N. J., Reissmann, M. and Hofreiter, M. (2011). Whole mitochondrial genome sequencing of domestic horses reveals incorporation of extensive wild horse diversity during domestication. *BMC Evol. Biol.* 11, 328.

Liu, N., Chen, L., Wang, S., Oh, C. and Zhao, H. (2005). Comparison of single-nucleotide polymorphisms and microsatellites in inference of population structure. *BMC Genet.* 6, 26.

Lohr, J. N., David, P. and Haag, C. R. (2014). Reduced lifespan and increased ageing driven by genetic drift in small populations. *Evolution (N. Y.)* 68, 2494–2508.

Lucas, Z. L., McLoughlin, P. D., Coltman, D. W. and Barber, C. (2009). Multiscale analysis reveals restricted gene flow and a linear gradient in heterozygosity for an island population of feral horses. *Can. J. Zool.* 87, 310–316.

MacFadden, B. J. (2005). Fossil Horses —. *Evolution (N. Y.)* 307, 1728–1730.

Masel, J. (2011). Genetic drift. *Curr. Biol.* 21, R837–R838.

- McVean, G. (2009). A genealogical interpretation of principal components analysis. *PLoS Genet.* 5,.
- Morin, P. A., Martien, K. K. and Taylor, B. L. (2009). Assessing statistical power of SNPs for population structure and conservation studies. *Mol. Ecol. Resour.* 9, 66–73.
- Nielsen, E. E., Hemmer-Hansen, J., Larsen, P. F. and Bekkevold, D. (2009). Population genomics of marine fishes: Identifying adaptive variation in space and time. *Mol. Ecol.* 18, 3128–3150.
- Offin, S. (2016). Alberta’s wild horses on birth control as part of pilot project. *Glob. News.*
- Olsen, M. T., Andersen, L. W., Dietz, R., Teilmann, J., Härkönen, T. and Siegismund, H. R. (2014). Integrating genetic data and population viability analyses for the identification of harbour seal (*Phoca vitulina*) populations and management units. *Mol. Ecol.* 23, 815–831.
- Onogi, A., Shirai, K. and Amano, T. (2017). Investigation of genetic diversity and inbreeding in a Japanese native horse breed for suggestions on its conservation. *Anim. Sci. J.* 88, 1902–1910.
- Outram, A. K., Stear, N. a, Bendrey, R., Olsen, S., Kasparov, A., Zaibert, V., Thorpe, N. and Evershed, R. P. (2009). The earliest horse harnessing and milking. *Science* (80-). 323, 1332–1335.
- Paetkau, D., Waits, L. P., Clarkson, P. L., Craighead, L., Vyse, E., Ward, R. and Strobeck, C. (1998). Variation in genetic diversity across the range of North American brown bears. *Conserv. Biol.* 12, 418–429.
- Parks Canada (2017). Science and Conservation, Ecological Monitoring.

- Pemberton, J. M., Ellis, P. E., Pilkington, J. G. and Bérénos, C. (2017). Inbreeding depression by environment interactions in a free-living mammal population. *Heredity (Edinb)*. 118, 64–77.
- Petersen, J. L., Mickelson, J. R., Cothran, E. G., Andersson, L. S., Axelsson, J., Bailey, E., Bannasch, D., Binns, M. M., Borges, A. S., Brama, P., et al. (2013a). Genetic diversity in the modern horse illustrated from genome-wide SNP data. *PLoS One* 8,.
- Petersen, J. L., Mickelson, J. R., Rendahl, A. K., Valberg, S. J., Andersson, L. S., Axelsson, J., Bailey, E., Bannasch, D., Binns, M. M., Borges, A. S., et al. (2013b). Genome-wide analysis reveals selection for important traits in domestic horse breeds. *PLoS Genet.* 9,.
- Pickrell, J. K. and Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* 8,.
- Plante, Y., Vega-Pla, J. L., Lucas, Z., Colling, D., De March, B. and Buchanan, F. (2007). Genetic diversity in a feral horse population from Sable Island, Canada. *J. Hered.* 98, 594–602.
- Prystupa, J. M., Hind, P., Cothran, E. G. and Plante, Y. (2012a). Maternal lineages in native canadian equine populations and their relationship to the nordic and mountain and moorland pony breeds. *J. Hered.* 103, 380–390.
- Prystupa, J. M., Juras, R., Cothran, E. G., Buchanan, F. C. and Plante, Y. (2012b). Genetic diversity and admixture among Canadian, Mountain and Moorland and Nordic pony populations. *Animal* 6, 19–30.
- Rare Breeds Canada (2014). Rare Breeds Canada.
- Renan, S., Greenbaum, G., Shahar, N., Templeton, A. R., Bouskila, A. and Bar-David, S. (2015). Stochastic modelling of shifts in allele frequencies reveals a strongly

- polygynous mating system in the re-introduced Asiatic wild ass. *Mol. Ecol.*
- Rieger, S. (2018). Experts challenge science behind Alberta's wild horse culls. *CBC News*.
- Roff, D. A. (2002). Inbreeding depression: tests of the overdominance and partial dominance hypotheses. *Evolution (N. Y.)*. 56, 768–75.
- Salter, R. E. (1978). Ecology of feral horses in Western Canada.
- Salter, R. E. and Hudson, R. J. (1980). Range relationships of feral horses with wild ungulates and cattle in Western Alberta. *J. Range Manag.* 33, 266–270.
- Salter, R. E. and Hudson, R. J. (2016). Distribution and management of feral horses in Western Canada. 5, 190–192.
- Sexton, J. P., Hangartner, S. B. and Hoffmann, A. A. (2014). Genetic isolation by environment or distance: Which pattern of gene flow is most common? *Evolution (N. Y.)*. 68, 1–15.
- Smith, F. A., Smith, R. E. E., Lyons, S. K. and Payne, J. L. *Body size downgrading of mammals over the late Quaternary*.
- Stalter, R. and Lamont, E. (2006). No TitleThe historical and axtant flora of Sable Island, Nova Scotia, Canada. *J. Torrey Bot. Soc.* 133, 362–374.
- Storz, J. F. (1999). Genetic Consequences of Mammalian Social Structure. *J. Mammal.*
- Tallmon, D. A., Luikart, G. and Waples, R. S. (2004). The alluring simplicity and complex reality of genetic rescue. *Trends Ecol. Evol.* 19, 489–496.
- Tamminen, S. (2015). Changing values of farm animal genomic resources. From historical breeds to the Nagoya Protocol. *Front. Genet.* 6, 1–9.
- Tissier, E. J., McLoughlin, P. D., Sheard, J. W. and Johnstone, J. F. (2013). Distribution of vegetation along environmental gradients on Sable Island, Nova Scotia. *Écoscience*

20, 361–372.

- Ure, D., McLoughlin, P., Frasier, T. and Lucas, Z. (2016). *Horse population status and trends, Sable Island National Park Reserve State of Park Assessment Technical Report*.
- Utsunomiya, Y. T., Pérez O’Brien, A. M. P., Sonstegard, T. S., Sölkner, J. and Garcia, J. F. (2015). Genomic data as the “hitchhiker’s guide” to cattle adaptation: Tracking the milestones of past selection in the bovine genome. *Front. Genet.* 5, 1–13.
- Uzans, A. J., Lucas, Z., McLeod, B. A. and Frasier, T. R. (2015). Small Ne of the isolated and unmanaged horse population on Sable Island. *J. Hered.* 106, 660–665.
- van Beest, F. M., Uzal, A., Vander Wal, E., Laforge, M. P., Contasti, A. L., Colville, D. and McLoughlin, P. D. (2014). Increasing density leads to generalization in both coarse-grained habitat selection and fine-grained resource selection in a large mammal. *J. Anim. Ecol.* 83, 147–156.
- Wahlund, S. (1928). Zusammensetzung von populationen und korrelationserscheinungen vom standpunkt der vererbungslehre aus betrachtet.
- Waller, D. M. (2015). Genetic rescue: a safe or risky bet? *Mol. Ecol.* 24, 2595–2597.
- Walling, C. a, Nussey, D. H., Morris, A., Clutton-Brock, T. H., Kruuk, L. E. and Pemberton, J. M. (2011). Inbreeding depression in red deer calves. *BMC Evol. Biol.* 11, 318.
- Walter, W. D. (2006). Soay sheep: Dynamics and selection in an island population. *J. Mammal.*
- Wang, J. (2012). On the measurements of genetic differentiation among populations. *Genet. Res. (Camb)*. 94, 275–289.
- Warmuth, V., Eriksson, A., Bower, M. A., Barker, G., Barrett, E., Hanks, B. K., Li, S.,

- Lomitashvili, D., Ochir-Goryaeva, M., Sizonov, G. V., et al. (2012). Reconstructing the origin and spread of horse domestication in the Eurasian steppe. *Proc. Natl. Acad. Sci.* 109, 8202–8206.
- Welsh, D. A. (1975). Population, behavioural and grazing ecology of the horses of Sable Island, Nova Scotia Thesis (Ph. D.)—Dalhousie University.
- Winton, C. L., Plante, Y., Hind, P., McMahon, R., Hegarty, M. J., Mcewan, N. R., Davies-Morel, M. C. G., Morgan, C. M., Powell, W. and Nash, D. M. (2015). Comparative genetic diversity in a sample of pony breeds from the U.K. and North America: A case study in the conservation of global genetic resources. *Ecol. Evol.* 5, 3507–3522.

APPENDIX A: PHYLOGENTIC RELATIONSHIP

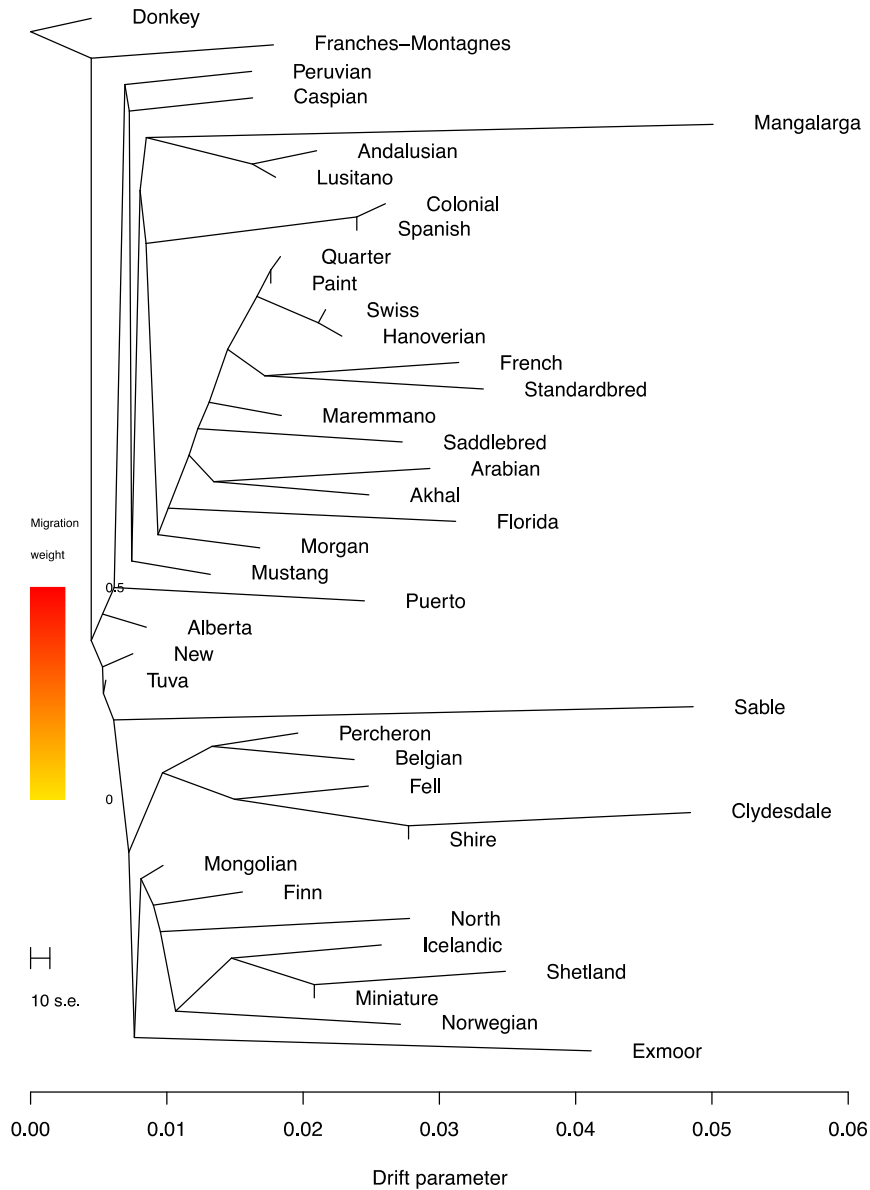


Figure A.1: TreeMix was used to create a maximum likelihood phylogeny of the 43 horse breeds included in our data set in addition to a Donkey sample to root the tree (1.13; Pickrell and Pritchard, 2012).

APPENDIX B: ALTERNATE PRINCIPAL COMPONENT ANALYSIS

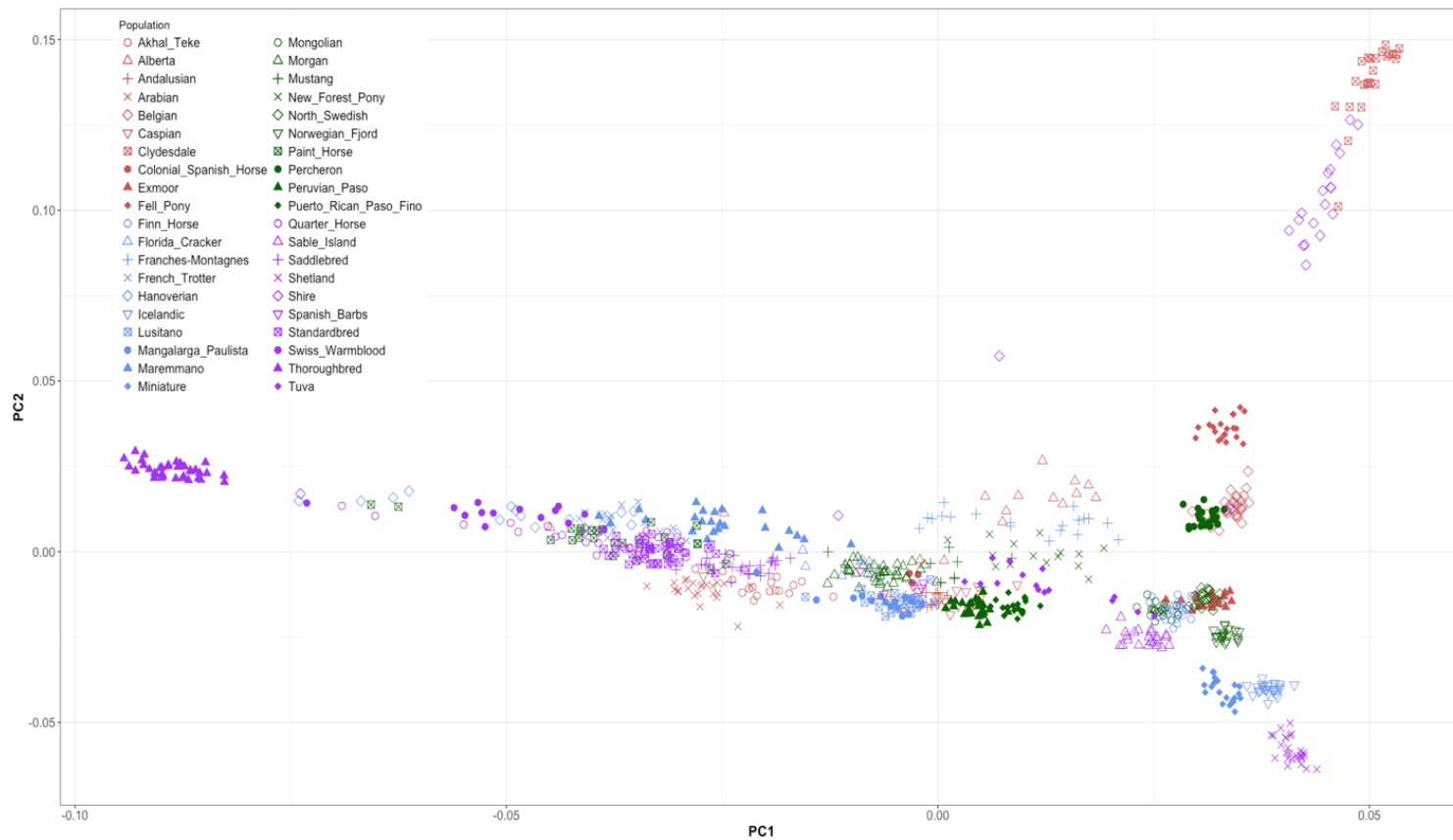


Figure A.2: A two-axis principal component analysis of 43 horse breeds, based on 37 495 autosomal SNP sites. The leading two components explain 59.6% of the variation (PC1 42.96% and PC2 16.6%).

APPENDIX C: BAYESIAN CLUSTER ANALYSIS AT $K = 3$ BEFORE SABLE ISLAND WAS SEPARATED AS A SUB-POPULATION

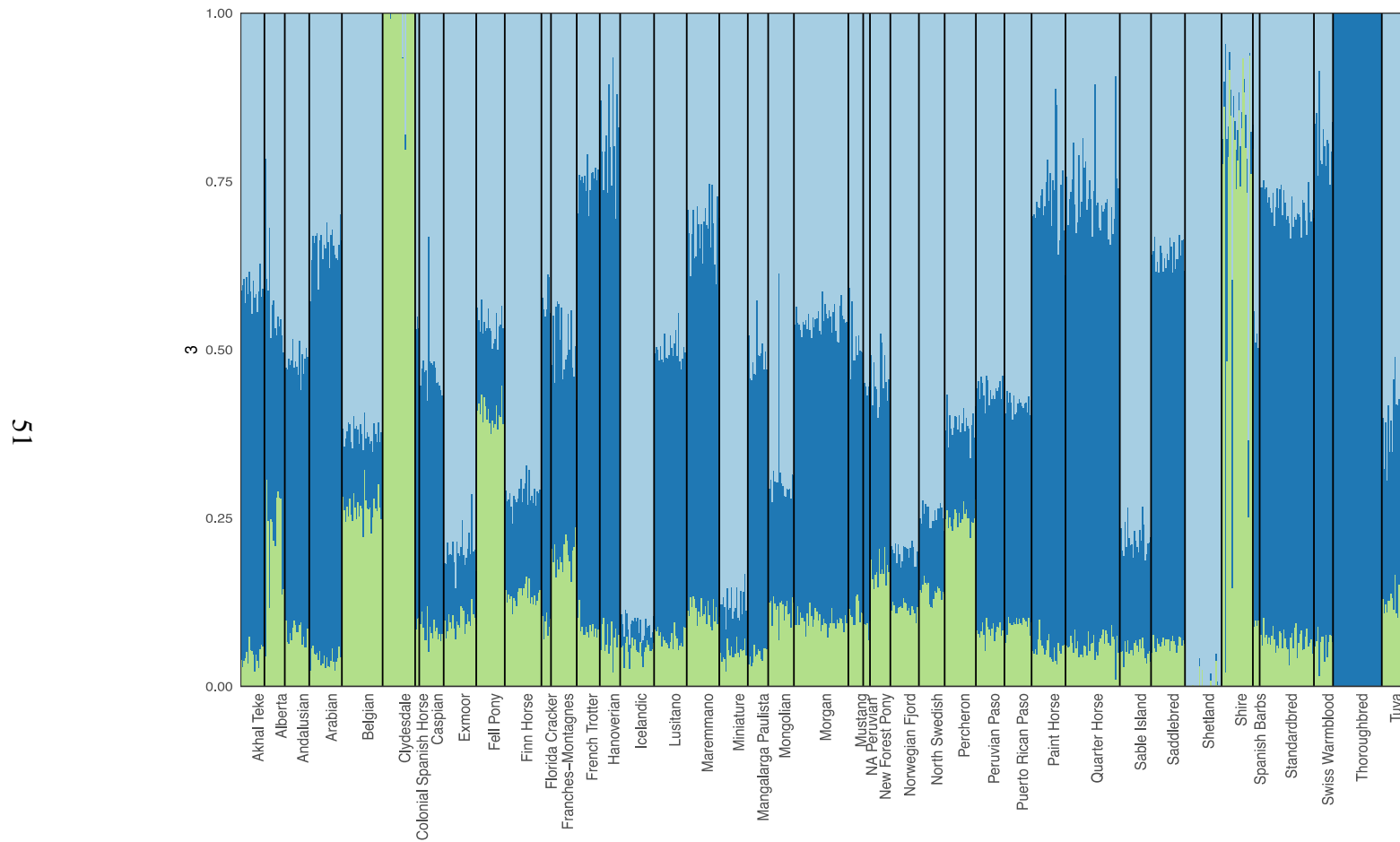


Figure A.3: Bayesian clustering output from ADMIXTURE depicting the results for a K -value of 3 in 856 horses from 43 populations. ADMIXTURE identified $K = 25$ as the most likely genetic group allocation for the data.

APPENDIX D: CROSS-VALIDATION ERROR AS A FUNCTION OF THE K -VALUE

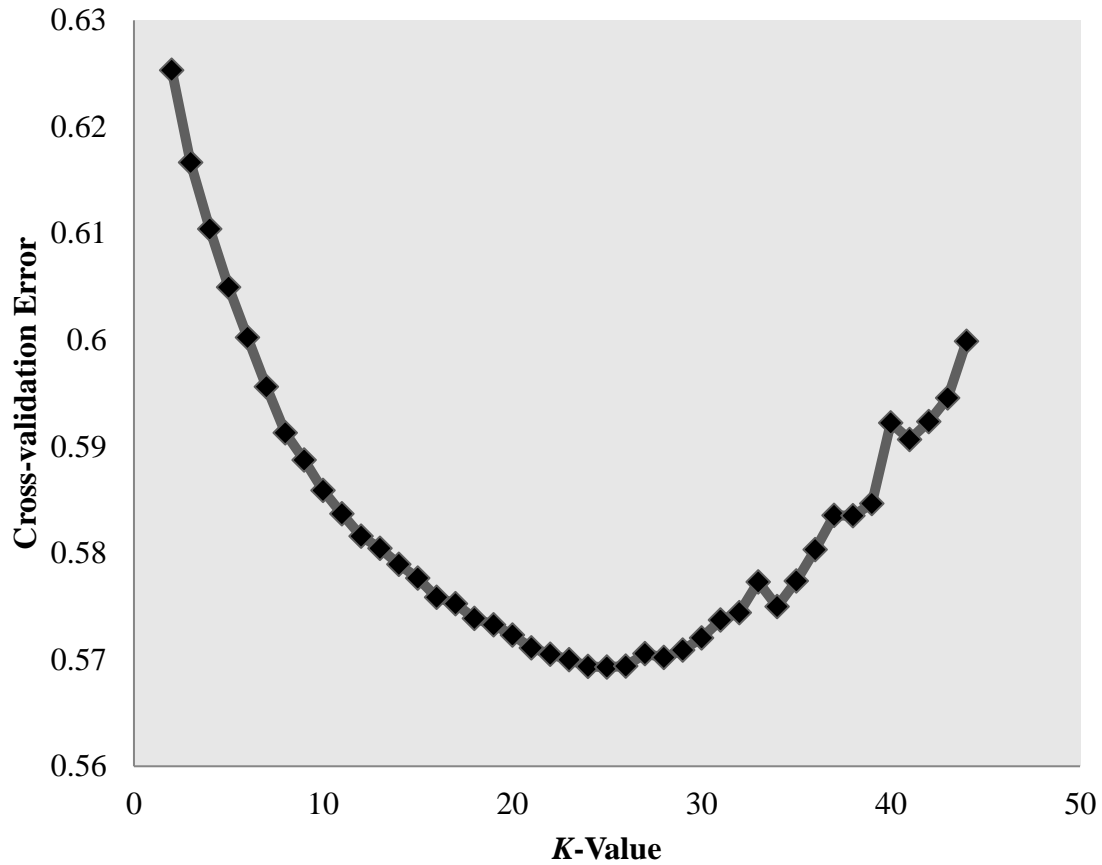


Figure A.4: A scatter-plot depicting the relationship between the cross-validation error output from the ADMIXTURE program and the K -value. The most likely K -value occurs at the lowest cross validation error (Alexander et al., 2015).

APPENDIX E: PRINCIPAL COMPONENT ANALYSES

EMPHASIZING THE ALBERTA POPULATION

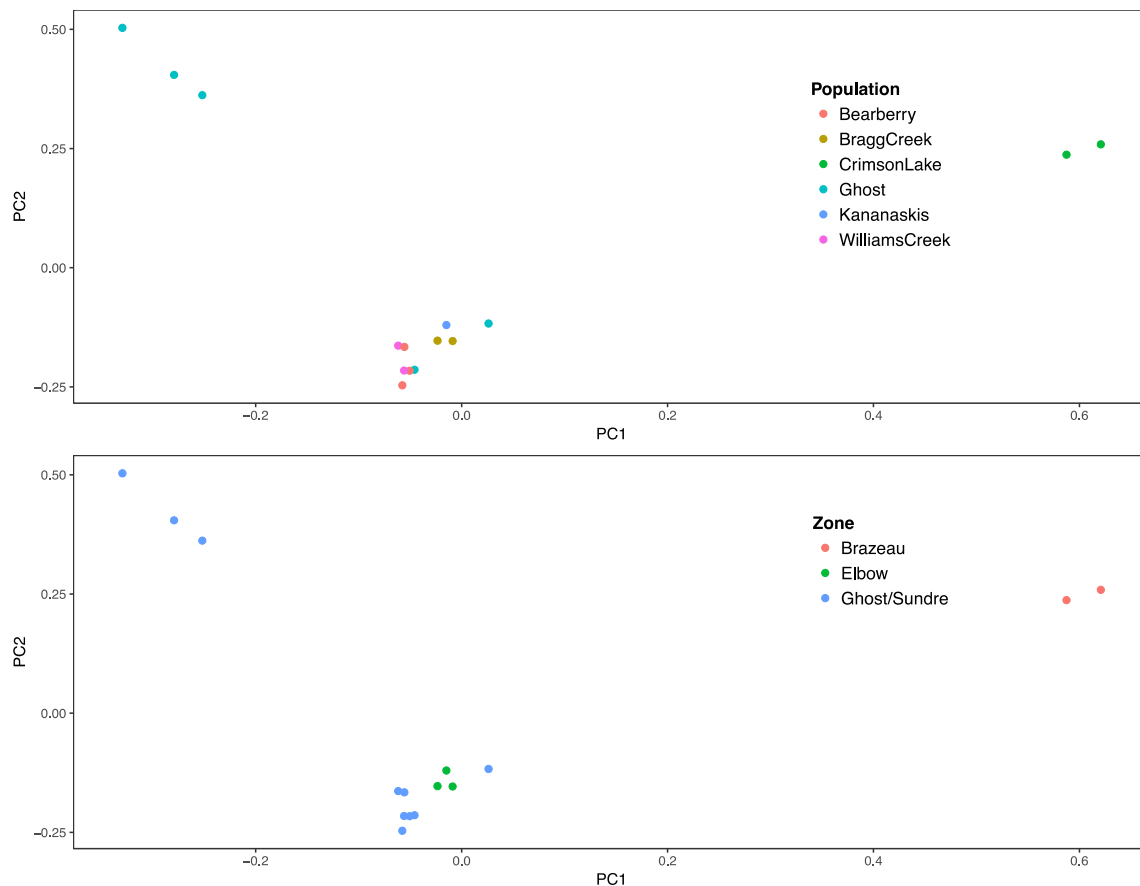


Figure A.5: Principal components analysis of 37 495 autosomal SNPs from the Alberta feral horse population broken down by location (top) and by equine zone (bottom) depicts that there is variation across and within equine zones, suggesting the presence of subpopulations.

APPENDIX F: BAYESIAN CLUSTER ANALYSIS OF THE ALBERTA POPULATION AND ITS CONTRIBUTORS

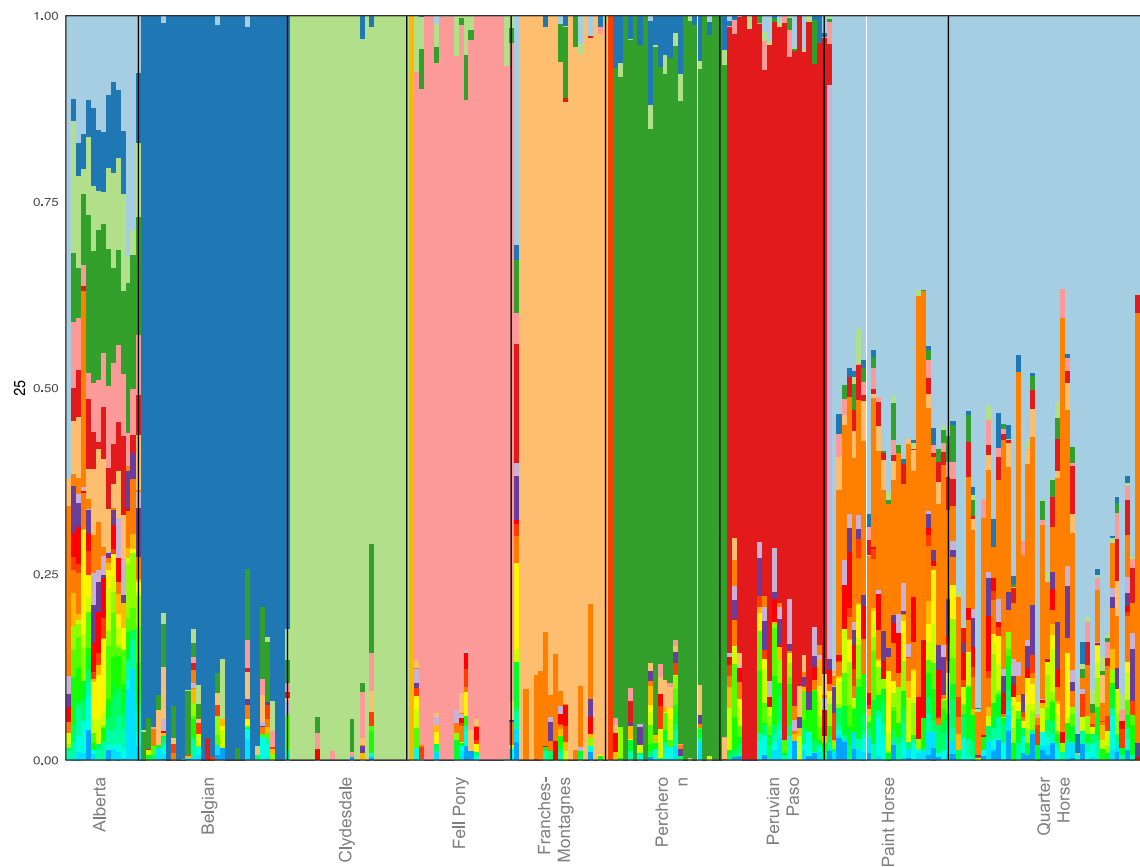


Figure A.6: Bayesian clustering output from ADMIXTURE depicting the Alberta horse population and the primary breeds (>5% contribution) contributing to the genetic make-up.

APPENDIX G: INDIVIDUAL INBREEDING COEFFICIENTS BY EQUINE ZONE

Table A.1: The inbreeding coefficients for the Alberta feral horse population subdivided into samples and by equine zones.

	F	Population	Zone	Mean F
AWH64	0.2819	Crimson Lake	Brazeau	0.3367
AWH65	0.3915	Crimson Lake	Brazeau	
AWH5	0.2119	Bragg Creek (Maclean creek)	Elbow	0.217733333
AWH6	0.3342	Bragg Creek (Maclean creek)	Elbow	
AWH7	0.1071	Rainy Summit/ Kananaskis	Elbow	
AWH63	0.07564	Bearberry	Ghost/Sundre	0.106852857
AWH66	0.04835	Bearberry	Ghost/Sundre	
AWH61	0.1664	Bearberry	Ghost/Sundre	
AWH10	-0.02309	Ghost area	Ghost/Sundre	
AWH11	0.2394	Ghost area	Ghost/Sundre	
AWH55	0.08497	Williams Creek	Ghost/Sundre	
AWH58	0.1563	Williams Creek	Ghost/Sundre	
AWH15*	0.1441	Ghost area 2	Ghost/Sundre2	0.250533333
AWH17*	0.406	Ghost area 2	Ghost/Sundre2	
AWH19*	0.2015	Ghost area 2	Ghost/Sundre2	

* Ghost/ Sundre outlier identified by PCA