January 2022

Genetic Breed Association And Contraceptive Response GWAS Of The Feral Horses (Equus Caballus) Of Theodore Roosevelt National Park

Melissa Amy Thompson

GENETIC BREED ASSOCIATION AND CONTRACEPTIVE RESPONSE GWAS OF THE FERAL HORSES (*Equus caballus*) OF THEODORE ROOSEVELT NATIONAL PARK

by

Melissa Amy Thompson
Bachelor of Science, Colorado State University, 2012

A Thesis

Submitted to the Graduate Faculty

of the

University of North Dakota

in partial fulfillment of the requirements

for the degree of

Master of Science

Grand Forks, North Dakota

May
2022

Name: Melissa A. Thompson

Degree: Master of Science

This document, submitted in partial fulfillment of the requirements for the degree from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.

DocuSigned by:
Turk Rhen
32ECCCC21FF5488...
Turk Rhen

DocuSigned by:
A6F957D6A106451...
Rebecca Simmons

DocuSigned by:
Blake McCann
183E89FBAE054E1...
Blake McCann

DocuSigned by:
Robert Newman
56046B66394D472...
Robert Newman

This document is being submitted by the appointed advisory committee as having met all the requirements of the School of Graduate Studies at the University of North Dakota and is hereby approved.

DocuSigned by:
Chris Nelson
2E0AF088C733403...
Chris Nelson
Dean of the School of Graduate Studies

4/20/2022
Date

PERMISSION

Title Genetic Breed Association and Contraceptive Response GWAS of the Feral Horses (*Equus caballus*) of Theodore Roosevelt National Park

Department Biology

Degree Master of Science

In presenting this thesis in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professors who supervised my thesis work or, in their absence, by the Chairperson of the department or the dean of the School of Graduate Studies. It is understood that any copying or publication or other use of this thesis or part 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 North Dakota in any scholarly use which may be made of any material in my thesis.

Melissa A. Thompson
April 15th, 2022

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank the National Park Service for funding this project. I also wish to thank my advisors, Turk Rhen and Rebecca Simmons, as well as my committee members for their advice and support. Special thanks to Blake McCann, who has done so much to provide me with guidance and opportunities for career and personal growth. In addition, I would like to thank Dan Baker for his continued support and for sharing data. I am grateful to Rytis Juras at Texas A&M University for his assistance with sample preparation. Finally, I would like to thank Lincoln Eddy for being my daily human interaction during a global pandemic, and Johanna Hodge for hours upon hours spent discussing horse ecology, genetics, behavior, management, and everything in between.

ABSTRACT

Theodore Roosevelt National Park (TRNP) is home to a herd of feral horses (*Equus caballus*) which were present on the landscape prior to the establishment of the park and are now maintained as a living history demonstration herd. I used genomic analyses to investigate the TRNP horses' relatedness to other breeds and the genetic basis for variation in response to a contraceptive vaccine. DNA from 118 horses was genotyped for 70k SNPs spread evenly across the genome. To clarify the relationship of this herd with other horses, I used population genomic analyses to compare the TRNP genotypes to a dataset of horses from 35 established breeds. These analyses indicate that the TRNP herd has experienced inbreeding and differentiation from other breeds, likely due to bottleneck events and isolation. The TRNP herd is an admixed population with no clear ancestral relationship to any one breed, but with greatest influence from draft breeds such as the Shire or Percheron. The genetic data do not support the oral history of Spanish origin. To identify genetic factors that influence the effectiveness of GonaCon contraception I conducted a genome-wide association study. While GonaCon-Equine has proven effective in reducing fertility among TRNP mares, there is individual variation in the duration of infertility. I found an association with SNPs on ECA18 for which the most likely candidate genes are STAT1 and STAT4, both involved in immune system function. Variation in STAT function could affect the immune response to the vaccine, leading to the variation observed in contraceptive efficacy. These findings will aid the TRNP management team in making informed decisions to improve management practices for the herd.

CHAPTER I

INTRODUCTION

History of Horses

The ancestors of modern-day horses (genus *Equus*) evolved in North America as one of many genera in the family Equidae (Librado & Orlando, 2021). They spread across the Bering Land Bridge into Eurasia around 0.95 million years ago, and analysis of ancient genomes indicates low levels of continued gene flow with two subsequent bidirectional dispersal events (Vershinina et al., 2021). The Bering Land Bridge was flooded and all horse species went extinct in the Americas around the time of the climatic transition from the Pleistocene into the Holocene some 12,500 years ago (Guthrie, 2003; Mann et al., 2015; Stewart et al., 2021). Horses persisted in Eurasia and were likely first domesticated in the Eurasian Steppes ~4,000-5,000 years ago, though the exact timeframe and location are still unknown (Anthony & Brown, 2011; Fages et al., 2019; Taylor & Barrón-Ortiz, 2021). High mitochondrial diversity in modern horses suggests that many mares contributed to the first domesticated stock (Librado & Orlando, 2021). Over time, types were developed in different regions based on the demands of work performance, environment, and reproductive characteristics. Intentional breeding practices further developed those types and pedigrees began to be kept for some valuable horses, with emphasis on a few influential stallions. Around 2,000 years ago horse Y-chromosome genetic diversity began to decline, with extremely low Y-chromosome diversity remaining today (Orlando, 2020). However, the concept of distinct breeds and breed registries only took hold in the last few centuries with an increased awareness of heritability (Hendricks, 1995). The domestication

process resulted in an overall loss of genetic diversity in the horse, with substantial decrease in the past 250 years (Orlando, 2020).

Domestic horses were first reintroduced to South and Central America by Spanish explorers in the 15th and 16th century, and as their use was adopted by many native peoples horses quickly spread throughout the Americas (Dobie, 1952; Thornhill, 2021). Later, English, French, and Dutch colonists brought their own horses to eastern North America (De Steiguer, 2021; Dobie, 1952). Horses of multiple origins were interbred in the development of new American breeds. For example the Quarter Horse, a prominent American ranch horse breed, arose from interbreeding of the English-influenced work horses of the US colonies with the Spanish-influenced horses of the Native Americans, and more recent interbreeding with the English Thoroughbred (Petersen et al., 2014).

Across the continent horses which had escaped or been released established free roaming populations. These populations experienced varying levels of gene flow. Before widespread settlement and construction of fences in the west, free-roaming horses experienced high population growth (Dobie, 1952). In the late 1700s and early 1800s, free-roaming populations were primarily found south of the Arkansas River in the region of Texas and Mexico. In the northern Great Plains few if any free-roaming populations were reported. The later 1800s saw the introduction of more horses to the Great Basin and the establishment of free-roaming populations which continue to exist today (De Steiguer, 2021). These populations often served as sources for new riding stock, and captured horses would sometimes be transported great distances for sale (Dobie, 1952). Settlers and ranchers commonly bred working horses by turning their stock out onto the open range and gathering the offspring as needed. Breeding practices became more deliberate in the late 1800s as ranchers selected purebred or pedigreed stallions to

turn out with local mares (De Steiguer, 2021). In southwestern North Dakota the largest horse ranch of the entire region imported purebred Percherons and a Thoroughbred stallion for use in this manner, producing both purebred and crossbred horses on the range (Huidekoper, 1947). When the economic value of horses dropped, these semi-maintained populations sometimes became unmanaged feral populations (De Steiguer, 2021). Though populations have since been fragmented and contained, there are still many feral horse populations in the United States today.

Study Area

A herd of feral horses has existed in Theodore Roosevelt National Park (TRNP) since before its establishment in 1947. TRNP is located within the northern Great Plains in southwest North Dakota, primarily featuring the badlands surrounding the Little Missouri river. The badlands are characterized by complex buttes, plateaus, and ridges eroded by wind and water. Year-round and seasonal streams carve out ravines and valleys feeding into the river. Vegetation is primarily mixed-grass prairie with wooded stands in valley bottoms and north-facing slopes (Hansen et al., 1984). The horses are found in the south unit of the park, which covers ~186.8 km^2. A boundary fence constructed around the perimeter of the unit contains reintroduced bison along with the horses. Elk, pronghorn, and bighorn sheep have also been reintroduced, and mule deer and white-tailed deer have had a continuous presence. The feral horses are maintained as a "living history demonstration", representing the free-roaming livestock that Theodore Roosevelt documented during his residency (Harmon, 1986). Of all these species, feral horses require the most time and energy to manage, due to their rapid population growth, politically charged atmosphere around their welfare, and ecological impact as a non-native species (B. McCann, personal communication). The origin and genetic health of the herd is a contentious topic when considering management strategies.

Because public sentiment is critical of lethal methods of population control, other solutions must be found. Commonly used methods of population management for feral horses include periodic capture and removal to reduce herd size and administration of contraceptive agents to reduce the growth rate of the population. The high growth rate of feral herds may be connected to positive selection for reproductive traits during domestication (Grange et al., 2009; Metzger et al., 2015). Both methods have been used on the herd at TRNP. Capture and removal of some proportion of the herd has been the primary tool for maintaining the population size in the past, but with varied concern for the continued genetic diversity of the herd (Cothran, 2000). Recently, an immunocontraceptive agent called GonaCon-Equine has proven to be effective in reducing foaling rates within the TRNP herd (Baker et al., 2018). Park managers are continuously seeking ways to improve management practices and are in the process of drafting a formal Horse Management Plan. It is important to have a scientific basis to make informed decisions on best management practices. In this project I used genome-wide analysis to help provide a thorough understanding of the genetics of the TRNP horses. To investigate the herd origins, I examined their relationship to other horse breeds using population genomic analyses. To identify genetic factors that influence the effectiveness of GonaCon contraception I conducted a genome-wide association study.

CHAPTER I REFERENCES

Anthony, D. W., & Brown, D. R. (2011). The secondary products revolution, horse-riding, and mounted warfare. *Journal of World Prehistory*, *24*, 131. https://doi.org/10.1007/s10963-011-9051-9

Baker, D. L., Powers, J. G., Ransom, J. I., McCann, B. E., Oehler, M. W., Bruemmer, J. E., Galloway, N. L., Eckery, D. C., & Nett, T. M. (2018). Reimmunization increases contraceptive effectiveness of gonadotropin-releasing hormone vaccine (GonaCon-Equine) in free-ranging horses (Equus caballus): Limitations and side effects. *PLoS ONE*, *13*(7), e0201570. https://doi.org/10.1371/journal.pone.0201570

Cothran, E. G. (2000). *Analysis of genetic variation in the feral horse herd of the Theodore Roosevelt National Park in 2000*.

De Steiguer, J. E. (2021). *Wild horses of the west: history and politics of america's mustangs*. University of Arizona Press.

Dobie, J. F. (1952). *The Mustangs*. Little, Brown and Company.

Fages, A., Hanghøj, K., Khan, N., Gaunitz, C., Seguin-Orlando, A., Leonardi, M., McCrory Constantz, C., Gamba, C., Al-Rasheid, K. A. S., Albizuri, S., Alfarhan, A. H., Allentoft, M., Alquraishi, S., Anthony, D., Baimukhanov, N., Barrett, J. H., Bayarsaikhan, J., Benecke, N., Bernáldez-Sánchez, E., … Orlando, L. (2019). Tracking five millennia of horse management with extensive ancient genome time series. *Cell*, *177*(6), 1419-1435.e31. https://doi.org/10.1016/j.cell.2019.03.049

Grange, S., Duncan, P., & Gaillard, J.-M. (2009). Poor horse traders: large mammals trade survival for reproduction during the process of feralization. *Proceedings of the Royal Society B*, *276*, 1911–1919. https://doi.org/10.1098/rspb.2008.1828

Guthrie, R. D. (2003). Rapid body size decline in Alaskan Pleistocene horses before extinction. *Nature*, *426*, 169–171. https://doi.org/10.1038/nature02070

Hansen, P. L., Hoffman, G. R., & Bjugstad, A. J. (1984). *The vegetation of Theodore Roosevelt National Park, North Dakota: a habitat type classification*.

Harmon, D. (1986). *At the open margin: the NPS's administration of Theodore Roosevelt National Park*. Theodore Roosevelt Nature and History Association.

Hendricks, B. Lou. (1995). *International encyclopedia of horse breeds*. University of Oklahoma press.

Huidekoper, A. C. (1947). *My experience and investment in the Bad Lands of Dakota and some of the men I met there*. Wirth Brothers.

Librado, P., & Orlando, L. (2021). Genomics and the evolutionary history of equids. *Annual Review of Animal Biosciences*, *9*, 81–101. https://doi.org/10.1146/annurev-animal-061220

Mann, D. H., Groves, P., Reanier, R. E., Gaglioti, B. V, Kunz, M. L., & Shapiro, B. (2015). Life and extinction of megafauna in the ice-age Arctic. *PNAS*, *112*(46), 14301–14306. https://doi.org/10.1073/pnas.1516573112

Metzger, J., Karwath, M., Tonda, R., Beltran, S., Águeda, L., Gut, M., Gut, I. G., & Distl, O. (2015). Runs of homozygosity reveal signatures of positive selection for reproduction traits in breed and non-breed horses. *BMC Genomics*, *16*(1), 1–14. https://doi.org/10.1186/S12864-015-1977-3/FIGURES/5

Orlando, L. (2020). Ancient genomes reveal unexpected horse domestication and management dynamics. *BioEssays*, *42*, 1900164. https://doi.org/10.1002/bies.201900164

Petersen, J. L., Mickelson, J. R., Cleary, K. D., & McCue, M. E. (2014). The american quarter horse: Population structure and relationship to the thoroughbred. *Journal of Heredity*, *105*(2), 148–162. https://doi.org/10.1093/jhered/est079

Stewart, M., Christopher Carleton, W., & Groucutt, H. S. (2021). Climate change, not human population growth, correlates with Late Quaternary megafauna declines in North America. *Nature Communications*, *12*(965). https://doi.org/10.1038/s41467-021-21201-8

Taylor, W. T. T., & Barrón-Ortiz, C. I. (2021). Rethinking the evidence for early horse domestication at Botai. *Scientific Reports*, *11*(7440), 7440. https://doi.org/10.1038/s41598-021-86832-9

Thornhill, C. A. (2021). Reanalysis of equid faunal remains from the Blacks Fork River site (48SW8319): A unique look at a protohistoric horse in Wyoming. *Plains Anthropologist*, *66*(257), 58–73. https://doi.org/10.1080/00320447.2020.1819180

Vershinina, A. O., Heintzman, P. D., Froese, D. G., Zazula, G., Cassatt-Johnstone, M., Dalén, L., Der Sarkissian, C., Dunn, S. G., Ermini, L., Gamba, C., Groves, P., Kapp, J. D., Mann, D. H., Seguin-Orlando, A., Southon, J., Stiller, M., Wooller, M. J., Baryshnikov, G., Gimranov, D., … Shapiro, B. (2021). Ancient horse genomes reveal the timing and extent of dispersals across the Bering Land Bridge. *Molecular Ecology*. *30*(23), 6144-6161. https://doi.org/10.1111/mec.15977

CHAPTER II

POPULATION GENOMICS PROVIDE INSIGHT INTO ANCESTRAL RELATIONSHIP OF THE FERAL HORSES OF THEODORE ROOSEVELT NATIONAL PARK

INTRODUCTION

Genetic data is a useful tool in furthering our understanding of organisms such as horses. Heritability of various traits has played a large role in historical breeding programs and the selective development of phenotypically diverse breeds, such as draft horses and miniature ponies (Makvandi-Nejad et al., 2012). As the field of genetics has rapidly advanced in recent years, the potential to use new genomic methods to understand phenotypes and solve various problems has increased. There have been studies to identify markers associated with diseases and coat color traits (Imsland et al., 2016; Raudsepp et al., 2019; Santschi et al., 1998) and studies to evaluate the genetic health of horse populations (Cosgrove et al., 2020; Shahsavarani & Rahimi-Mianji, 2010; Winton et al., 2020). Genomic approaches can be used to gain insights into population structure, relatedness, and genetic diversity. Many studies have evaluated genetic variation within and between horse breeds, often using mitochondrial DNA (mtDNA) or microsatellite markers to compare allelic diversity and frequencies (Dorji et al., 2018; Glowatzki-Mullis et al., 2006; Khanshour et al., 2015; Yang et al., 2018). While microsatellites are highly informative their use is usually limited to a small number of markers, often only 15-50 loci. The development of a horse reference genome allows for a genome-wide analysis using thousands of single nucleotide polymorphisms (SNPs) (Wade et al., 2009). Commercial SNP genotyping assays make gathering large datasets easy and affordable, and provide a standard set

of markers, making it easier to build upon and compare published datasets. Genome-wide SNPs have been used in many horse breeds to measure genetic diversity, identify regions of diversifying selection, and make inferences about the origins of breeds (Cosgrove et al., 2020; Gurgul et al., 2019; Petersen et al., 2013).

Population genetics analyses have also been done for a few populations of feral horses. A study done on a feral horse population in southern Spain using microsatellites found that those horses were distinct from domestic breeds and had a lower inbreeding coefficient (Vega-Pla et al., 2006). The feral horses of Sable Island in Canada have been studied using both microsatellites and SNPs. That population was found to have lower genetic diversity than domestic breeds and most similarity to Nordic breeds (Plante et al., 2007; Tollett, 2018). These prior studies of domestic breeds and feral herds suggest that population genetics can be a useful tool for investigating breed associations and the potential origins of feral herds such as the horses of Theodore Roosevelt National Park.

TRNP herd History

At the time Theodore Roosevelt National Park (TRNP) was established as a National Memorial Park in 1947, many of the horses roaming within the boundary of the park were owned and branded by local ranchers. The horses had either escaped or been turned loose to forage and reproduce on their own, so that they and their offspring could be recaptured for later use (McLaughlin, 1989). In 1954 as the park began work on the task of erecting a perimeter fence for the reintroduction of bison, an effort was made to round up the horses, which at that time were considered trespass livestock, and return them to their owners. Approximately 125 horses and mules were captured of an estimated 200 head present within the park boundary, 99% of which bore brands as evidence of ownership (McLaughlin, 1989). After unsuccessful attempts over the

next decade to remove all remaining horses, park administration decided in the 1970s to maintain the horses as a "historic livestock display" or "living history demonstration", reminiscent of the free-roaming livestock that Theodore Roosevelt documented during his residency (Harmon, 1986).

Reports vary as to the number of horses left in TRNP after these removal attempts, with some suggesting that the remaining founder individuals were one gray stallion and two mares. The consensus among reports, however, is that there were only about 16 individuals present in 1965 (McLaughlin, 1989; Harmon, 1986; TRNP records). Every few years thereafter the park conducted roundups to control population size by removing a portion of the herd. Ideal population size was initially set at 35-60 individuals (National Park Service, 1978). A habitat use and forage analysis recommended a population maximum of 90 individuals to prevent overgrazing of some forage species (Marlow et al., 1992). More recently, policy was changed to a new population objective of 70-140 animals, following a genetic analysis which found low effective population size (Cothran, 1992). Ten roundups were conducted from 1978 to 2013, where each time the population was reduced by an average of 52% (TRNP records); thus, the TRNP horse population has undergone eleven genetic bottleneck events as a result of historical herd management. The herd has mostly been a closed population since the perimeter was fenced, with the exception of six stallions which were introduced to the park in 1981 and 1982, at the same time as the removal of several established TRNP stallions from the herd. These included an Arabian, a Shire-Paint cross, a Quarter horse, and three feral stallions from a Wyoming herd, although each had varying levels of reproductive success within the population (McLaughlin, 1989). The Shire-Paint cross stallion was highly successful: he maintained a large band of mares for almost a decade and was considered the most dominant stallion in the park (TRNP records).

An estimated 15% of the population could be traced to this single stallion in 1991 (Cothran, 1992). The decision to augment the herd by introduction of "well-bred" domestic stallions was later reversed in favor of maintaining the historic type; in 1991 and 1997 attempts were made to remove the introduced stallions and their known offspring (TRNP records).

McLaughlin (1989) interviewed TRNP employees and local ranchers in the southwestern North Dakota area to collect an oral history of the herd. She suggests that the horses which eluded capture to remain in the park were the wildest and were descended from "Indian ponies". The horses that were commonly used by Native Americans of the 1800s were primarily derived from horses brought to the Americas by Spanish explorers (Sponenberg, 1992). The horses used for ranch work in late nineteenth century North Dakota were often "Indian type" horses crossbred with other European breeds from the eastern US (McLaughlin, 1989). A phenotypic evaluation conducted on the TRNP herd in 1994 based on physical conformation and coat colors found evidence of crossbreeding in all TRNP horses and some presence of Spanish-type features, but recommended genetic evaluation for further resolution (Sponenberg, 1994). Some locals have suggested that the TRNP herd is a unique population due to this association with the historical Spanish-type horse. This assertion, while popular among individuals interested in TRNP horses, has not been substantiated with genetic data.

Previous genetic work has been done on the TRNP horses to estimate genetic diversity. Seven red blood cell antigen loci were tested in the 1990s to calculate genetic variability measures for the herd such as expected heterozygosity (H_e), observed heterozygosity (H_o), effective number of alleles (A_e), and estimated inbreeding level (F_{IS}). These values were compared over a nine-year period from 1991-2000, as well as compared to average values for domestic horse breeds and other feral horse populations. There was a statistically significant

decrease in H_o over the sampling period, with the final value falling below the average for both domestic and feral horses (H_o=0.305, compared to 0.370 and 0.359 for domestic and feral populations, respectively). The report also found a lower value of A_e than average (Cothran, 2000). All of the allele variants found were previously described in domestic breeds; no unique alleles were found in the TRNP population (Cothran, 1992). A later analysis of mitochondrial DNA (mtDNA) and twelve short tandem repeat (STR) loci from the present-day TRNP herd was conducted by Ovchinnikov et al. (2018). H_o and A_e values for the TRNP herd were again lower than the average for other feral herds and most domestic breeds. Three mtDNA haplotypes were found, two of which had sequences similar to published American Paint Horse mtDNA genotypes and one of which had no close match to published sequences. This suggests that at least two different populations or maternal sources contributed to the TRNP herd. The STR analysis was inconclusive in determining the ancestry of the park horses and showed TRNP horses as being distinct from other breeds. The authors suggest that new genetic technologies could provide further information (Ovchinnikov et al., 2018).

Here, I use genome-wide SNP genotypes to examine the relationship of the TRNP horses to established horse breeds. Based on the known history of the herd, I hypothesized that TRNP horses would be most genetically similar to Quarter Horses and American Paint Horses (the most common ranch horses in the USA today), Shires, or Spanish-type breeds.

METHODS

Sample Collection

Hair samples were collected during a regularly scheduled roundup in 2013. From these, I selected samples of 87 horses which had been re-released into the park, including almost all adult mares and reproductive band stallions. In 2017, additional tissue samples were collected via

biopsy dart from 12 individuals that had evaded the 2013 capture. With the addition of these individuals my sample set represented an approximate census of the herd as it was at the end of 2013 (99/107=92.5%). In 2020 I collected an additional 18 tissue samples from young mares born post-roundup by biopsy dart under research permit number THRO-2020-SCI-0013 and NPS IACUC approval (ND_THRO_McCann_HorseBiopsyDarting_2020.A1). These 18 mares were of interest due to their inclusion in a study of the contraceptive vaccine GonaCon™ (Baker et al., 2018), and their use in a GWAS study of GonaCon effectiveness (Chapter III). With the addition of a hair sample from a 2017 foal with a coat color phenotype of interest, my full dataset had 118 samples. This includes 85% (94/111) of the adult individuals in the herd as of spring 2022 and represents 98% (177/181) of the herd when including offspring of sampled individuals.

DNA Extraction, Genotyping, and Sample Selection

Genomic DNA was extracted from hair follicle and tissue samples by the Animal Genetics Laboratory at Texas A&M University using Gentra Puregene Tissue kits (Qiagen) following manufacturer's protocols. Individuals were then genotyped at Neogen Genomics Laboratory (Lincoln, NE) for over 70k SNPs located evenly across the horse genome using the Illumina Equine GGP 70k BeadChip. I combined the resulting genotypes with a dataset of 793 horses from 35 different breeds which had been genotyped using the Illumina Equine GGP 50k BeadChip (Petersen et al., 2013). The mean sample size for breeds in that dataset was 22.63; to prevent a comparatively large number of TRNP samples from skewing the analyses, I used a representative selection of 23 TRNP individuals for my final dataset. Using pedigree data from TRNP records going back to the 1980s, I identified first degree relatives and excluded the younger individual (i.e., offspring) to reduce bias due to relatedness, which left 32 individuals. I then performed a random selection to reduce the dataset to the final 23 individuals.

Data Pruning

I performed quality control filtering using SNP & Variation Suite v8.9.0 (SVS) (Golden Helix, Inc., Bozeman, MT, www.goldenhelix.com), with the methods used by Petersen et al. (2013). I first removed markers with call rate ≤0.95, and then samples with call rate ≤0.95. This eliminated any markers which were not included in both the 70k and 50k genotyping arrays. I next removed SNPs with a minor allele frequency (MAF) of 0.05 or less. I mapped the remaining SNPs to EquCab3.0 (www.ncbi.nlm.nih.gov/assembly/GCF_002863925.1/) and filtered to include only autosomal loci. This resulted in a final set of 815 samples and 38,786 SNPs. I further filtered the dataset for linkage disequilibrium (LD) using a window size of 50 and an increment of 5, with an LD threshold of r^2=0.5. After LD filtering, this second version of the dataset retained 815 samples and 28,505 SNPs.

Among-breed Relationships

I conducted a Principal Components Analysis (PCA) on the dataset pruned for MAF and call rate, as per Petersen et al. (2013). PCA constructs reduced-dimension vectors to explain the variation in genotypes (Patterson et al., 2006). I computed the principal components in SVS using an additive model with the option selected to normalize each marker's data by its standard deviation. I then plotted the first three principal components against each other to visualize relationships among individuals of different breeds and the TRNP horses. I also calculated pairwise values of Wright's fixation index (F_{ST}) between all breeds and TRNP horses in SVS using the LD pruned dataset.

Phylogenetic Analysis

I converted the LD pruned dataset to phylip format using vcf2phylip (Ortiz, 2019). To reduce computing power and computation time, I randomly selected a maximum of 10 samples

from each breed for inclusion along with the same 23 TRNP samples used for the PCA and F_{ST} analyses. All phylogenetic analyses were performed via the CIPRES Science Gateway v3.3 (Miller et al., 2010). I converted individual SNP data to FASTA formats using NCLconverter v2.1 (Lewis, 2003). I aligned the resulting data using ClustalW v2.1 with standard parameters (Thompson et al., 1994), and used RAxML v8.2.12 (Stamatakis, 2014) to construct the phylogeny for these samples, along with 1000 bootstrap replicates. Bootstrap values were calculated by hand using a majority rule consensus tree with Consense (Felsenstein 1986-2008). Only bootstrap values of ≥70% are reported. I visualized the resulting phylogeny and edited the appearance of the tree using FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

Bayesian Cluster Analysis

I used the program ADMIXTURE on the LD pruned dataset to estimate ancestral clusters (Alexander et al., 2009). ADMIXTURE employs model-based ancestry estimation to assign individuals to clusters with similar ancestry based on genotypes. For each individual sample ADMIXTURE returns the proportion of their genome that can be assigned to each ancestral cluster based on allele frequencies. Since there were 36 breeds including the TRNP samples in the dataset, I ran the program for K = 1 through K = 36, K being the assumed number of ancestral populations. To determine the value of K that was the best fit in describing the dataset I used ADMIXTURE's values for cross-validation (CV) error at each value of K, using the default setting of 5-fold CV. Lower error values are considered better.

Estimates of Inbreeding

I calculated the average inbreeding coefficient (F_{IS}) for each breed in the LD pruned dataset using SVS. Another approach to evaluating inbreeding or relatedness by descent is to identify runs of homozygosity (ROH) within the genome. These homozygous-by-descent (HBD)

segments are created when an individual inherits two copies of the same stretch of chromosome from a common ancestor (Ceballos et al., 2018; Peripolli et al., 2016). I used the R package RZooRoH to identify ROHs and model the generational age of common ancestors based on segment length (Bertrand et al., 2019). To evaluate the state of the TRNP herd I included all 118 TRNP samples, did not prune for LD, and used the RZooRoH default model. I calculated the generation time of the TRNP horses in Vortex10 (Lacy & Pollak, 2021) using records of demographic rates. The resulting generation time of 10.48 years closely matched the 10 years previously reported for feral horses (National Research Council, 2013). I also selected several other breeds with known population history for comparison.

Since inbreeding can increase the occurrence of deleterious recessive traits, I selected nine individuals from my sample set to test for known genetic diseases. These individuals were chosen as prominent contributors to the present-day herd, based on number of their descendants currently residing within the park. These samples were tested for six genetic diseases using the Quarter Horse & Related Breeds Disease Panel (5-panel plus) at UC Davis Veterinary Genetics Laboratory (Davis, CA).

RESULTS

Among-breed Relationships

The first three principal components from my PCA account for 62.19% of total variation in the dataset (Table 1). The first principal component (PC1) explains 36.90% of the variance in the dataset. The second principal component (PC2) explains 14.06% of the variance, with a more

Table 1. First ten principal components from PCA.

Principal Component	Percentage of Variance Explained/Eigenvalues
1	36.90
2	14.06
3	11.23
4	10.16
5	8.98
6	8.36
7	7.42
8	7.18
9	6.07
10	5.63

gradual decline in explanatory value from PC2 to PC10 (Table 1). On a plot of PC1 by PC2, the breeds follow the pattern described by Petersen et al. (2013), with Thoroughbreds separated at the left end of PC1 and, at the right end, draft breeds and pony breeds separated on PC2 (Figure 1). The five Spanish-type breeds, including two Iberian breeds and three South American breeds, are tightly clustered together.

TRNP horses fall near the center of the plot. The TRNP cluster overlaps on PC1 with such breeds as Morgan, Lusitano, Mangalarga Paulista, Andalusian, Franches-Montagnes, New Forest Pony, Peruvian Paso, Tuva, Caspian, and Puerto Rican Paso Fino. On PC2, the TRNP horses are separated from these breeds in the direction of the draft horses such as Shire, Clydesdale, and Fell Pony. When looking at PC3 on the plot of PC1 by PC3 there is more overlap of the TRNP horses with the Tuva, New Forest Pony, and Caspian, as well as with the Akhal-Teke and French Trotter (Figure 2).

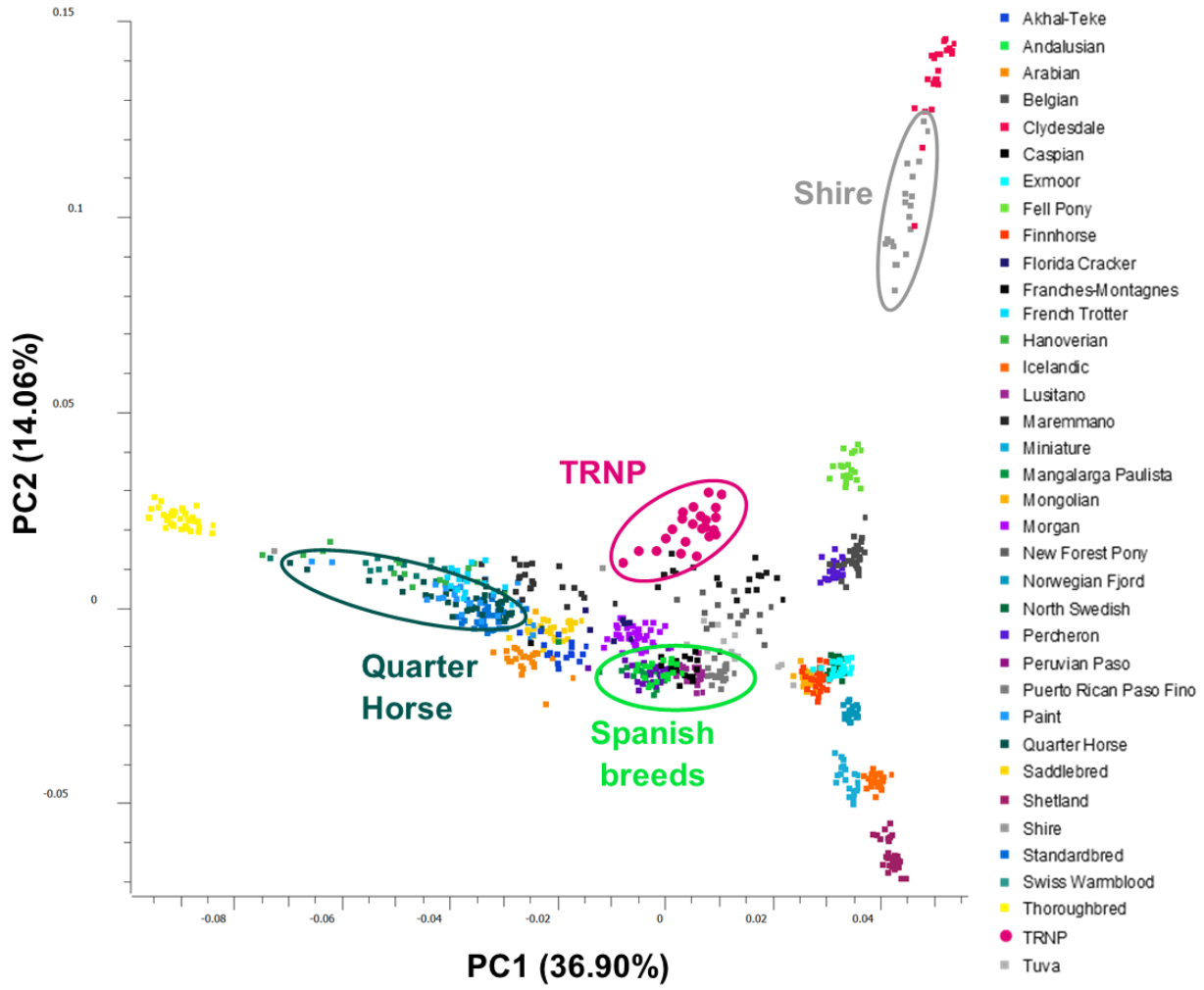

Figure 1. Principal components analysis of genetic variation among horse breeds and feral horses from TRNP, showing one plane of the cloud of points (PC1 by PC2). Points represent individuals. PC1 captures 36.90% of the variation in the dataset. PC2 captures 14.06% of the variation. TRNP horses fall into a cluster in the center of the plot, indicated by pink circles. Other breeds of interest are circled.

A preliminary run of the PCA which included all 118 TRNP samples resulted in a strong sample bias completely separating the TRNP horses from all other breeds on PC2. This disappeared when using the sample size corrected dataset of unrelated horses. Uneven sampling of populations is known to distort PCA results (McVean, 2009).

Mean F_{ST} value among breeds was 0.108, with a minimum of 0.002 between Paint and Quarter Horse and a maximum of 0.273 between Clydesdale and Mangalarga Paulista (Table 2).

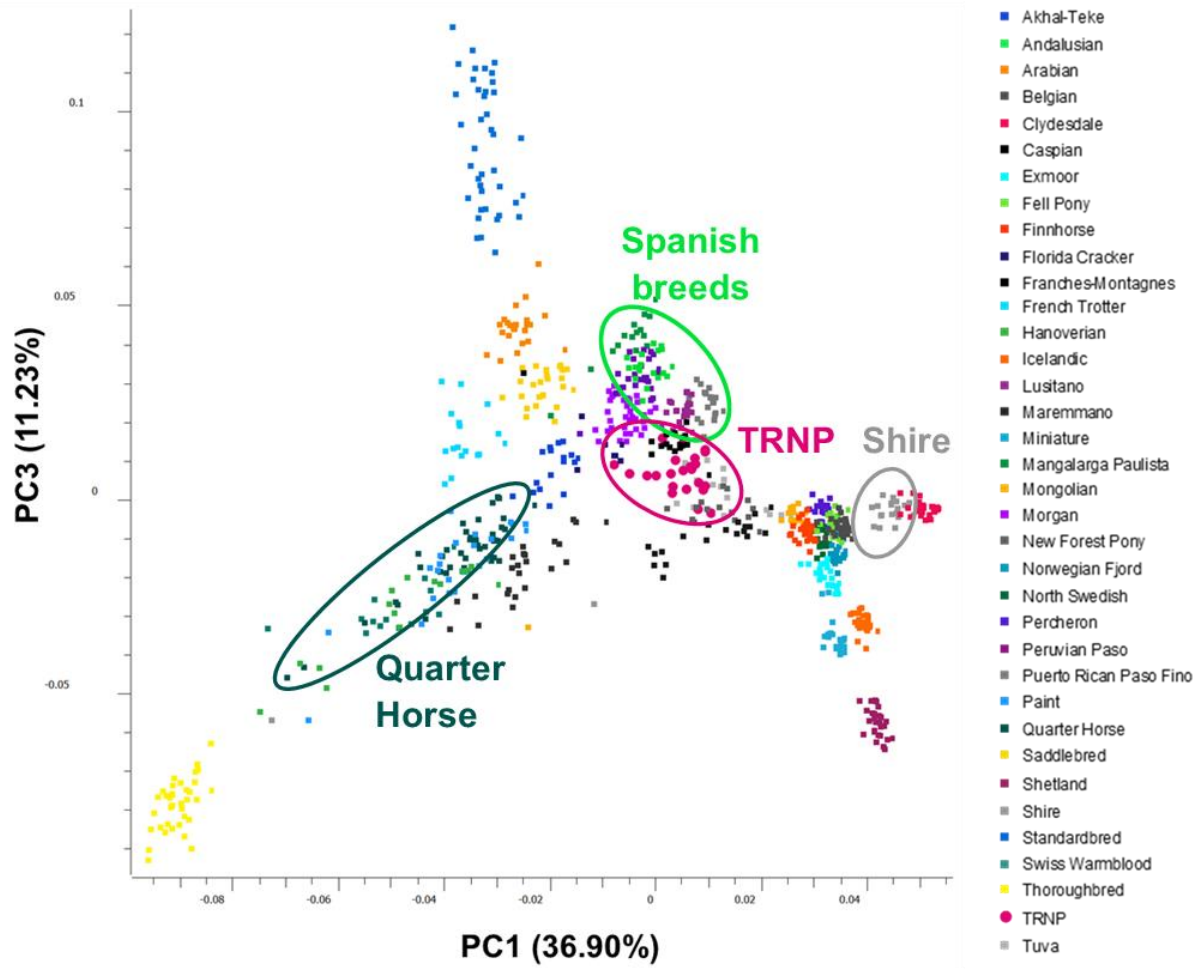

Figure 2. Principal components analysis of genetic variation among horse breeds and feral horses from TRNP, showing the same cloud of points from another plane (PC1 by PC3). Points represent individuals. PC3 captures 11.23% of the variation in the dataset. TRNP horses fall into a cluster in the center of the plot, indicated by pink circles. Other breeds of interest are circled.

F_{ST} values between the TRNP horses and other breeds ranged from 0.104 to 0.217. Breeds with the lowest level of genetic differentiation from the TRNP horses were the Tuva (0.104), New Forest Pony (0.107), Quarter Horse (0.108), Paint Horse (0.108), Mongolian horse (0.111), Maremmano (0.115), and the Morgan (0.115). The most dissimilar breeds were the Mangalarga Paulista (0.217), Clydesdale (0.213), Exmoor (0.203), Shetland (0.184), and Thoroughbred (0.180).

Table 2. Pairwise F_{ST} values calculated for 36 populations using SVS. Rows are sorted by value in comparison to TRNP.

	TRNP	Tuva	New Forest Pony	Quarter Horse	Paint	Mongolian	Maremmano	Morgan	Swiss Warmblood	Caspian	Finnhorse	Peruvian Paso	Percheron	Hanoverian	Lusitano	Franches-Montagnes	Andalusian	Miniature	Shire	Belgian	Akhal-Teke	Saddlebred	Fell Pony	Arabian	Puerto Rican Paso Fino	Standardbred	French Trotter	Icelandic	North Swedish	Norwegian Fjord	Florida Cracker	Thoroughbred	Shetland	Exmoor	Clydesdale
Tuva	.104																																		
New Forest Pony	.107	.018																																	
Quarter Horse	.108	.040	.044																																
Paint	.108	.039	.042	.002																															
Mongolian	.111	.007	.024	.055	.053																														
Maremmano	.115	.041	.044	.034	.031	.054																													
Morgan	.115	.038	.045	.046	.045	.049	.054																												
Swiss Warmblood	.125	.051	.054	.020	.016	.068	.031	.056																											
Caspian	.126	.035	.046	.054	.054	.044	.060	.059	.066																										
Finnhorse	.127	.031	.041	.073	.072	.031	.071	.066	.089	.066																									
Peruvian Paso	.127	.040	.050	.059	.058	.049	.064	.062	.074	.060	.067																								
Percheron	.127	.046	.052	.085	.084	.050	.082	.077	.100	.082	.064	.084																							
Hanoverian	.127	.054	.057	.024	.021	.071	.034	.059	.007	.069	.091	.077	.103																						
Lusitano	.127	.042	.051	.055	.054	.054	.061	.062	.065	.059	.072	.055	.088	.068																					
Franches-Montagnes	.135	.053	.057	.072	.070	.060	.072	.075	.077	.082	.074	.084	.074	.082	.085																				
Andalusian	.136	.051	.060	.064	.063	.061	.070	.072	.075	.069	.081	.063	.095	.078	.019	.093																			
Miniature	.138	.044	.052	.083	.082	.043	.084	.077	.099	.078	.058	.079	.079	.102	.083	.088	.092																		
Shire	.140	.064	.063	.094	.095	.070	.091	.091	.108	.097	.083	.100	.078	.110	.101	.093	.109	.098																	
Belgian	.140	.062	.067	.098	.099	.065	.095	.091	.114	.098	.075	.098	.055	.117	.101	.087	.110	.092	.088																
Akhal-Teke	.142	.057	.067	.056	.055	.071	.066	.072	.061	.070	.093	.078	.108	.066	.074	.099	.085	.103	.120	.123															
Saddlebred	.143	.071	.075	.059	.058	.083	.074	.066	.069	.083	.101	.090	.112	.073	.085	.107	.095	.112	.125	.128	.092														
Fell Pony	.146	.065	.064	.097	.097	.068	.096	.090	.114	.097	.083	.098	.086	.116	.102	.097	.109	.092	.072	.100	.123	.126													
Arabian	.149	.073	.078	.066	.066	.092	.075	.081	.073	.071	.110	.087	.124	.075	.075	.113	.088	.120	.134	.139	.080	.097	.136												
Puerto Rican Paso Fino	.153	.068	.076	.087	.086	.074	.093	.089	.103	.091	.092	.079	.108	.105	.086	.112	.094	.104	.123	.122	.110	.119	.123	.118											
Standardbred	.155	.084	.090	.074	.072	.098	.087	.087	.080	.101	.115	.104	.127	.084	.100	.118	.111	.125	.136	.140	.108	.102	.140	.113	.130										
French Trotter	.155	.078	.083	.061	.058	.094	.071	.086	.054	.097	.112	.103	.124	.063	.093	.111	.104	.125	.133	.139	.097	.100	.137	.107	.131	.090									
Icelandic	.156	.061	.070	.103	.102	.058	.104	.095	.122	.097	.068	.097	.094	.125	.103	.107	.112	.059	.115	.107	.124	.131	.109	.142	.121	.144	.145								
North Swedish	.161	.069	.079	.105	.105	.070	.105	.101	.123	.104	.082	.104	.097	.125	.108	.109	.116	.096	.116	.110	.129	.136	.118	.145	.130	.148	.148	.108							
Norwegian Fjord	.162	.067	.077	.105	.106	.065	.106	.098	.124	.102	.076	.102	.099	.127	.108	.111	.117	.087	.117	.110	.128	.135	.116	.143	.126	.146	.148	.095	.105						
Florida Cracker	.162	.077	.085	.075	.074	.091	.086	.086	.088	.097	.110	.097	.123	.089	.097	.118	.107	.122	.136	.139	.108	.111	.139	.111	.131	.124	.123	.142	.151	.149					
Thoroughbred	.180	.123	.124	.059	.055	.141	.078	.120	.039	.132	.160	.140	.173	.045	.130	.145	.140	.170	.176	.186	.115	.126	.183	.131	.167	.129	.108	.191	.190	.192	.146				
Shetland	.184	.091	.099	.127	.127	.088	.130	.121	.149	.126	.102	.126	.125	.151	.131	.136	.141	.038	.144	.137	.150	.157	.138	.168	.151	.169	.171	.095	.139	.128	.174	.213			
Exmoor	.203	.117	.118	.141	.144	.116	.145	.139	.165	.148	.131	.147	.148	.168	.149	.157	.159	.142	.161	.157	.171	.175	.161	.185	.174	.184	.190	.157	.172	.168	.196	.224	.185		
Clydesdale	.213	.142	.142	.164	.170	.145	.165	.160	.191	.174	.155	.174	.153	.192	.174	.170	.186	.170	.043	.163	.200	.198	.136	.208	.197	.205	.211	.185	.192	.190	.223	.249	.213	.233	
Mangalarga Paulista	.217	.141	.147	.137	.139	.149	.150	.149	.155	.155	.166	.148	.181	.160	.148	.178	.157	.180	.194	.195	.170	.175	.197	.172	.181	.185	.188	.196	.204	.204	.204	.206	.226	.246	.273

Phylogenetic Tree

The results of the maximum likelihood analysis revealed a star phylogeny pattern with short internal branches and long external branches, typical of a lineage which has undergone rapid differentiation (Figure 3). Samples with the same breed assignment were placed mostly into their own respective clades. External nodes had some high confidence bootstrap values, but most of the internal branches joining breeds were not well supported (less than 70% bootstrap values).

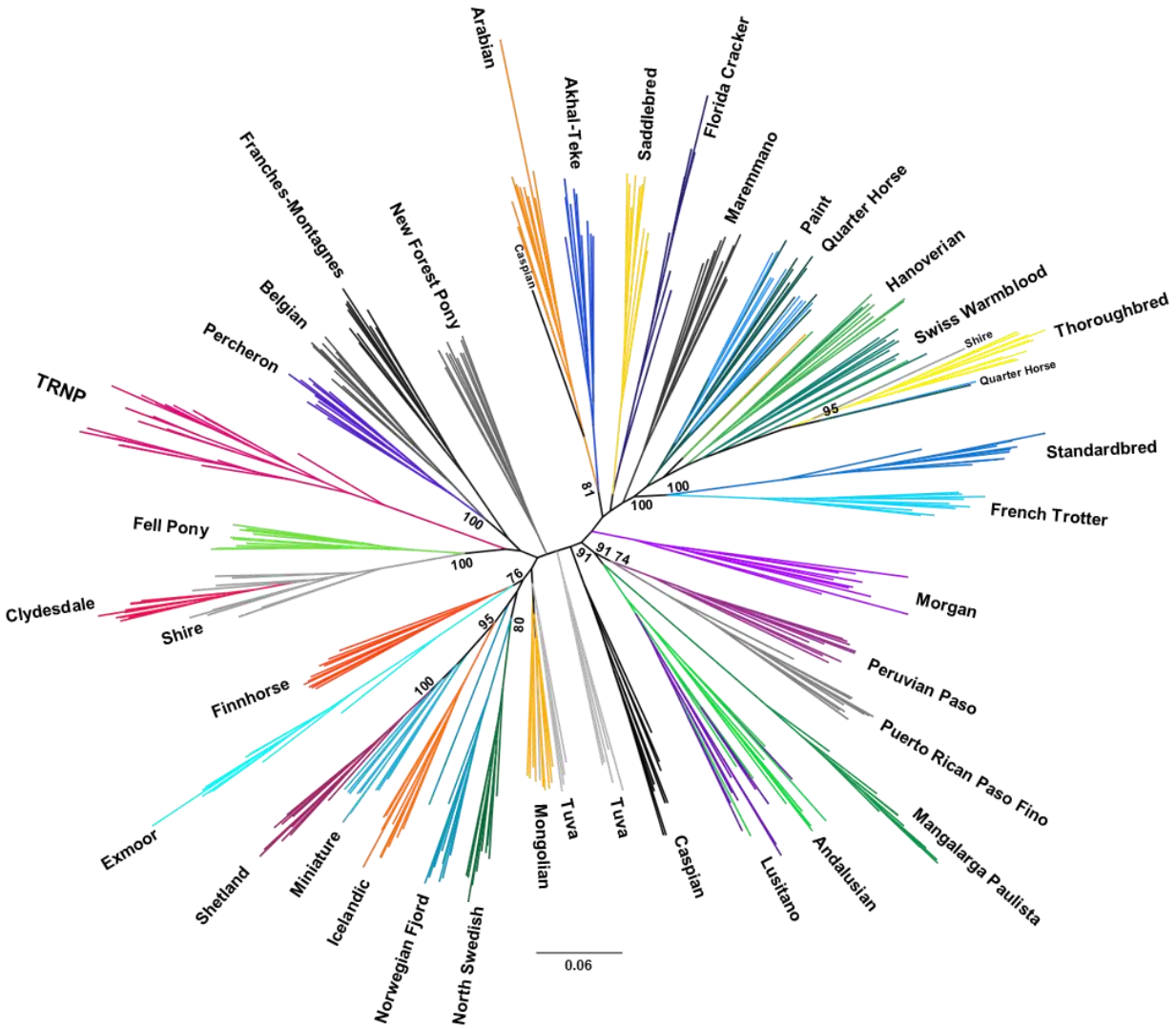

Figure 3. Maximum likelihood tree with bootstrap values for horse breeds, including TRNP horses. Only bootstrap values with confidence of 70% and higher are given. TRNP horses were placed without confidence amongst draft breeds. Spanish-type breeds diverge from the rest with high confidence. The scale bar reflects the uncorrected p-distance for these individuals.

TRNP horses were placed without confidence on a single branch amongst the draft-type breeds such as the Shire, Percheron, and Franches-Montagnes. The next nearest branches included the pony-type breeds. All Spanish-type breeds diverged together from the rest of the tree with high confidence (91%). Branch lengths for TRNP horses are slightly exaggerated due to overrepresentation in sample size, but placement of the branches is not affected by uneven sample size.

Bayesian Clustering Analysis

The lowest CV error returned by ADMIXTURE was observed at K=25 distinct populations, though CV error values of K in the 20-28 range were of similarly low values (Figure 4). At K=25 the TRNP horses were grouped into their own cluster with minimal assignment to other clusters (Figure 5). No individuals from other breeds had notable assignment to the TRNP horse cluster. Four TRNP individuals had a proportion of 0.039-0.075 of their genomes assigned to another cluster which included Quarter Horse, Paint, and Florida Cracker samples, and a 0.021-0.033 proportion assigned to a cluster including Thoroughbred, Hanoverian, Shire, Quarter Horse,

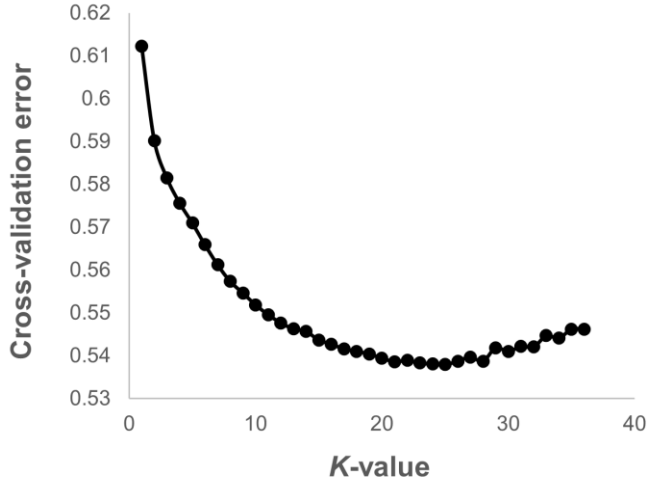

Figure 4. Cross-validation error for each value of K, calculated by ADMIXTURE.
The most likely value of K was chosen by the lowest error value.

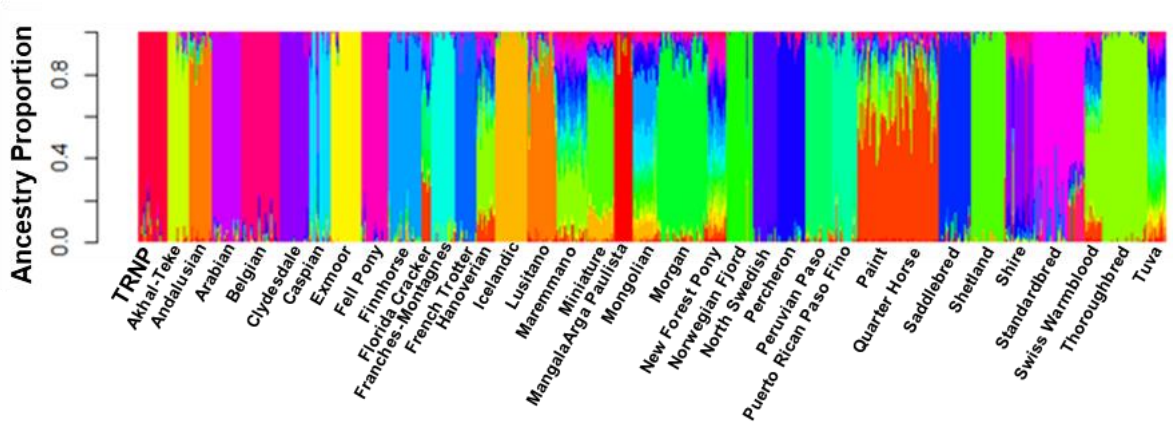

Figure 5. Ancestry estimation using ADMIXTURE modeling. The number of ancestral populations (clusters) K=25 was chosen based on ADMIXTURE's CV error calculation. Vertical lines represent individuals, with colors representing the proportion of their genome attributed to each ancestral cluster. The TRNP horses make up their own red cluster (leftmost) with minimal shared ancestry from other clusters.

Paint, Swiss Warmblood, and Maremmano samples. Two TRNP individuals had 0.010-0.031 of their genome assigned to each of three clusters including Andalusian, Lusitano, Percheron, Shire, and Clydesdale samples. The individuals from other breeds that had the highest proportion of assignment with the TRNP cluster (range 0.032-0.054) were a Swiss Warmblood, Hanoverian, Morgan, Saddlebred, two Paints, Maremmano, Quarter Horse, Puerto Rican Paso Fino, and a New Forest Pony.

Some of the relationships among breeds seen in the PCA and phylogenetic analysis were also reflected in the ADMIXTURE results. Several of the clusters included multiple draft breeds in one cluster, such as the Shire and Clydesdale, and the Miniature and Shetland ponies were placed into their own cluster. Another cluster contained warmblood type breeds such as the Thoroughbred, Hanoverian, Swiss Warmblood, Paint and Quarter Horse, and Maremmano. In addition, the Spanish breeds Andalusian and Lusitano were assigned to their own cluster.

In the K=20 to K=28 range of low CV errors the TRNP horses clustered similarly to the K=25 results. From K=20 through K=8 the TRNP horses were still assigned to their own cluster. At K=7, the TRNP horses were 62.1-85.0% assigned to a cluster that included many different breeds of draft type, such as Belgian, Percheron, Franches-Montagnes, and North Swedish Horse. At such a low number of ancestral populations most of the clusters separated into groups of draft breeds, pony breeds, Spanish and Arabian breeds, or warmblood breeds.

Estimates of Inbreeding

TRNP horses had relatively high values for inbreeding coefficients compared to other breeds. The F_{IS} of TRNP horses was 0.180, while the average F_{IS} for all other breeds in the dataset was 0.116 (standard deviation of 0.079). Only seven breeds had an F_{IS} higher than the TRNP horses (Table 3).

Table 3. F_{IS} values for all breeds.

Breed	F_{IS}	**Breed cont.**	F_{IS}
Hanoverian	-0.007	Arabian	0.109
Swiss Warmblood	-0.003	Standardbred	0.117
Paint	0.001	Andalusian	0.129
Quarter Horse	0.005	Miniature	0.132
Maremmano	0.005	Puerto Rican Paso Fino	0.141
Thoroughbred	0.046	Percheron	0.144
Caspian	0.062	Fell Pony	0.155
New Forest Pony	0.066	Icelandic	0.157
French Trotter	0.068	Belgian	0.157
Mongolian	0.070	Florida Cracker	0.159
Tuva	0.071	TRNP	0.180
Saddlebred	0.082	Norwegian Fjord	0.184
Peruvian Paso	0.088	North Swedish	0.187
Morgan	0.096	Shire	0.194
Lusitano	0.099	Shetland	0.237
Akhal-Teke	0.105	Mangalarga Paulista	0.252
Finnhorse	0.106	Exmoor	0.285
Franches-Montagnes	0.109	Clydesdale	0.310

RZooRoH identified 9195 ROHs among the 118 TRNP samples, on all autosomes and in all individuals. The average proportion of the genome covered by ROHs for TRNP individuals was 0.2172 (standard deviation 0.0656; min 0.0947; max 0.4728) (Figure 6). The highest proportions were in generation classes 4 and 8 (Figure 7). With a generation time of 10 years this corresponds to common ancestors approximately 40-80 years ago, suggesting bottlenecking or founder events around that timeframe. For comparison, Figure 7 also shows the Clydesdale and Florida Cracker, which have both undergone recent genetic bottlenecks, the Puerto Rican Paso Fino, which experienced a more distant bottleneck during the importation of Spanish horses to the Americas, and the Quarter Horse, which has multiple sources of recent admixture.

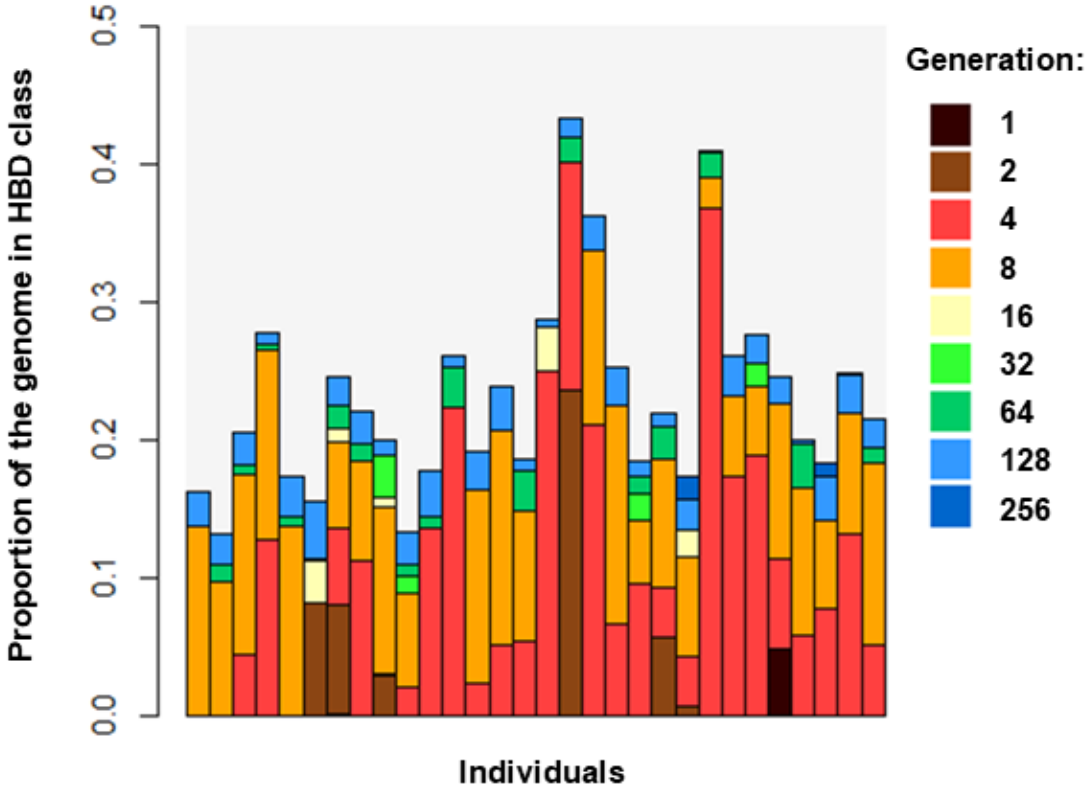

Figure 6. Example proportion of the genome in each homozygosity by descent class for TRNP individuals. Thirty individuals were randomly selected for ease of viewing. HBD classes represent inbreeding level based on number of generations removed to common ancestor, with lower generation numbers corresponding to more recent inbreeding.

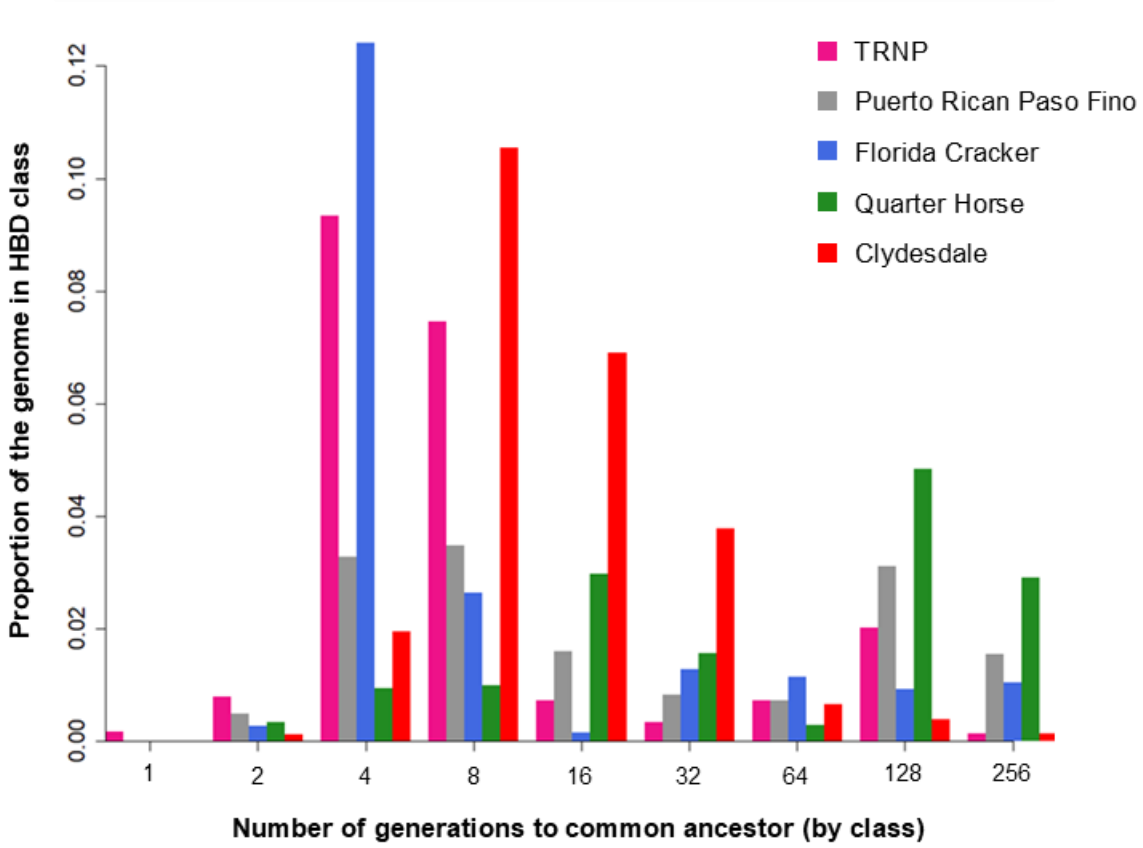

Figure 7. Average proportion of the genome in each HBD generation class for TRNP horses in comparison to four other breeds of known population history. Clydesdale and Florida Cracker have undergone recent bottlenecks, Puerto Rican Paso Fino an older bottleneck, and Quarter Horses have been recently admixed. HBD classes represent inbreeding level based on number of generations removed to common ancestor, with lower generation numbers corresponding to more recent inbreeding.

Of the nine prolific individuals tested for known genetic diseases, none carried any copies of disease-causing alleles. Considering this result and the lack of observational evidence for any of these diseases in other individuals, it is likely that none of these diseases are present within the TRNP herd. This alleviates some concerns over inbreeding increasing the rate of occurrence of deleterious recessive traits in the herd.

DISCUSSION

By using multiple approaches to observe the population genetics of TRNP horses, I identified overall patterns that reflect the history of this herd. These analyses place the TRNP horses well within the diversity seen in the modern domestic horse but do not show a strong signal of relatedness to any one breed, consistent with previous work and the evolutionarily recent development of most horse breeds. There are some emergent patterns, however, which I discuss here.

The PCA plot reflects variation in genotypes and identifies unique populations and associations with phenotypic traits. Highly specialized breeds can be found as distinctly separated from other breeds due to strong selection pressures and inbreeding (Petersen et al., 2013; Figure 1). Draft horses such as the Clydesdale and ponies such as the Shetland have been under strong selection for their size, while the Thoroughbred and Standardbred (Figure 2) have been selected for their racing ability. The PCA plot also shows admixture between populations, as seen in the locations of Quarter Horse and Paint individuals. The Quarter Horse breed was developed by crossing English-influenced work horses of the US colonies with the Spanish-influenced ponies of the Native Americans, and more recent interbreeding with the English Thoroughbred (Petersen et al., 2014). Horses with white-spotting coat phenotypes ("paints") were excluded from the Quarter Horse registry and were instead registered as American Paint horses (Brooks et al., 2020). Quarter Horses and Paints are located between the Thoroughbred cluster and the Morgan cluster on the PCA plot. Morgan horses originated in the Northeastern US during the late 1700s and contributed founding individuals to many other North American breeds (Battell, 1894). Placement of these breeds near the Morgan cluster on the PCA plot reflects their shared history.

The TRNP horses fall together in the center of the plot, potentially pointing to the lack of deliberate artificial selection and specialization, as expected with a freely breeding population (Figures 1 & 2). The TRNP points are not as tightly clustered together as some of the other breeds, indicating admixture between multiple sources in their recent history (McVean, 2009). Some TRNP points are separated from the center of the cluster, falling toward the Shires and other drafts. This separation may reflect the influence of the Shire-Paint stallion introduced to the park in the 1980s. Percherons, another draft breed, may also have contributed to the population genetics of the herd, given that several southwestern North Dakota ranchers in the late 1800s imported purebred Percherons and crossed them with local horses (Crawford, 1931; Huidekoper, 1947). On the PCA plot the five Spanish-type breeds are all tightly clustered together (Figure 1). The TRNP points align with Spanish-type breeds on PC1 but are separated from them on PC2, potentially indicating some affinity between the herd and these breeds.

The general distribution of domestic breeds in PCA is consistent across multiple reports (Funk et al., 2020; Ovchinnikov et al., 2018). Due to this consistency some inferences may be made in comparing separate analyses. A Canadian population of feral horses appeared in a comparable position to the TRNP horses in a PCA plot using the same Petersen et al. (2013) dataset. This population ("Alberta Foothills") has ranged in size from 1000-1700 individuals and likely experienced continual gene inflow from multiple draft breeds as well as Quarter Horses/Paints (Tollett, 2018). A second, isolated feral population of about 500 individuals ("Sable Island") was more tightly clustered on PC1 and PC2 with coldblood pony breeds such as Mongolian, Finnhorse, and Exmoor, and was noticeably separated from the main cloud of points on PC3 (Tollett, 2018). The TRNP herd, though considerably smaller than the Sable Island herd, does not show such divergence on PC3.

The maximum likelihood tree reflects the pattern seen in the PCA plot, with Thoroughbreds located in one portion of the tree while drafts and ponies are found on the other side of the tree. As in the PCA plot, the TRNP horses are found to be more similar to draft breeds, though the relationship in Figure 3 is not strongly supported. However, there is strong support for separation of the Spanish breeds from the TRNP horses, reflecting low genetic similarity between these two lineages.

Artificial selection for differing characteristics during the domestication process along with transport and husbandry of horses created various breeds or types within a short evolutionary timeframe. Recombination, or gene flow between branches, can also affect the shape of phylogenetic trees, lengthening the terminal branches (Li et al., 2019; Schierup & Hein, 2000). My analyses indicate that horse breed relationships are reconstructed in a star-like phylogeny, with short internal branches and long external branches. This pattern is commonly seen within domestic species, such as dogs, cattle and buffalo, goats, and chickens (Mannen et al., 2020; Quan et al., 2020; Rout et al., 2008; Sun et al., 2020; Vonholdt et al., 2010). Other trees built with data from different sources show a similar pattern for horse breeds (Felkel et al., 2018; Khanshour et al., 2013; Vilà et al., 2001). Because horse breeds can hybridize it is difficult to reconstruct that genetic history with a simple bifurcating phylogeny. This is reflected in the low support for the placement of TRNP horses and some breeds.

A common thread across these analyses is that TRNP horses make up a distinct population among these domestic horse breeds. They are separated into their own ancestral cluster by ADMIXTURE, form their own cluster in the PCA plot of PC1 versus PC2, and phylogenetic analysis places them on their own clade of the tree. These analyses indicate that TRNP horses display detectable genetic differentiation from other breeds. Genetic differentiation can be driven

by selection, mutation, reduced gene flow, genetic drift, and nonrandom mating. Although the oldest formal breed registries have only existed for approximately two hundred years, horses have been under artificial selection during their domestication for at least 4,000 years (Orlando, 2020). The TRNP herd has experienced reduced artificial selection but has had limited gene flow and a small population size (80-200 individuals) since the 1950s. A few individuals have been introduced, intentionally and unintentionally, but the last introduction occurred in the 1980s. Genetic differentiation could have resulted from the isolation and repeated bottleneck events experienced by the TRNP herd, which exacerbate the effects of genetic drift. Essentially, the TRNP horses are more similar to each other in allelic combinations than they are to any other horse breeds.

Ovchinnikov et al. (2018) report low values of genetic variability (observed heterozygosity and allelic diversity) in the TRNP herd compared to both domestic breeds and other feral herds. This is reflected in inbreeding coefficients from genome-wide analysis of SNPs, with the TRNP horses having higher values of F_{IS} than most other breeds. Low diversity in comparison with other breeds can also help explain the high F_{ST} values, since a population with a limited allelic pool is less likely to share the alleles of other populations. The F_{ST} values between TRNP and other breeds were all near or higher than the average among all breeds.

The ROH analysis also shows that the TRNP horses have experienced relatively recent inbreeding. The high proportion of ROHs in the genome due to inheritance from common ancestors 4 and 8 generations ago coincides with the herd's isolation and known bottleneck events (Figures 6 & 7). It was 60-70 years ago (6-7 generations) in the 1950s and 60s when the initial bottleneck occurred following the establishment of the park. The majority of the horses on the land were rounded up and returned to their owners at the same time as the park perimeter was

fenced and the remaining population was isolated. A reduction of the population to 16 individuals, followed by low gene flow into the population, explains the current presence of large chromosome segments inherited from common ancestors. The presence of these ROHs and a relatively high inbreeding coefficient confirm that the bottlenecks experienced by the TRNP herd have had an effect on the population.

This can also be observed in the inbreeding coefficients of other breeds which have experienced varying populations bottlenecks. Clydesdales experienced a bottleneck following agricultural mechanization and their use in WWI and WWII during the 1920s to 40s (Hendricks, 1995). This inbreeding is consistent with a higher proportion of their genome exhibiting ROHs assigned to 8 and 16 generations ago (Figure 7). A more recent and severe reduction in population size occurred with the Florida Cracker. When the Florida Cracker Horse Association was founded in 1989 there were only 31 individuals (Conant et al., 2012; *Florida Cracker Horse Association*), which is reflected by the highest HBD level occurring at the 4 generation class. These breeds have a clear signature of inbreeding in comparison to the Quarter Horse, a breed with recent admixture from multiple sources. In comparison, the TRNP results align with expectations based on their recent management history.

Inclusion of samples from other feral horse populations across North America may yield further relationships in these analyses. TRNP horses were in approximately the same position of the PCA plot as feral horses from the Alberta foothills, but both of these populations were distinct from feral horses on Sable Island (Tollett, 2018). Feral herds across the continent have diverse origins, some known from historical documentation and some substantiated by genetic analyses. Microsatellite analysis of another Canadian population of feral horses placed them in a group with draft breeds (Cothran & McCrory, 2014). Previous genetic work on feral horses

reported similar sources for some but not all eastern US herds (Conant et al., 2012; Goodloe et al., 1991). Two herds in Nevada were shown to be strongly differentiated from each other, suggesting variation in origin (Ashley, 2004). Genetic analysis of additional feral horses may also provide insight into the potential for re-adaptation of horses to the wild (Naundrup and Svenning, 2015). Signals of selection for high altitude adaptations were documented in an Andean feral horse population (Hendrickson, 2013). Further work could look for signals of selection in the TRNP herd.

Conclusions

These genetic analyses do not reveal the exact ancestry of the TRNP horses but do allow for assessment of their overall similarity to defined horse breeds. Admixture between breeds and the relatively recent development of many breeds prevents us from distinguishing similarities strictly due to ancestry. As these breeds have experienced continued artificial selection for certain characteristics, they differ from horse populations that existed in the late 1800s. In addition, the TRNP herd is an admixed population with influence from multiple sources. Admixture can create new allele combinations and frequencies, contributing to the apparent differentiation of the population. The reduced genetic diversity seen in the herd due to repeated bottlenecks and genetic drift likely contributed to their genetic differentiation.

Genetic similarity to some breeds was indicated across multiple analyses. The strongest signal is from the draft breeds. In particular, the Shire connection suggests that the park was not entirely successful at removing the offspring of the Shire-Paint stallion introduced in the 1980s. Records indicate that this stallion was the most dominant stallion in the park until his removal and that he had the most reproductive success of all the introduced stallions (McLaughlin, 1989). Due to the local popularity of Percheron horses in the late 1800s the draft influence may have

also been present before his introduction (Crawford, 1931; Huidekoper, 1947; McLaughlin, 1989). However, recent draft admixture can be attributed to this stallion. Similarity to Paint and Quarter Horse was not as prominent as expected, although the F_{ST} analysis places them as among the most similar breeds to the TRNP horses. Another breed mentioned across multiple analyses is the Morgan, the oldest remaining North American horse breed originating in the late 1700s and early 1800s as a working breed for use on farms (Battell, 1894). Perhaps the ranch horses of early western North Dakota were more similar to the early work horses of the eastern US and to the early Quarter Horses of Texas, before their differentiation into strictly kept breeds.

Based on McLaughlin's oral history of the herd, we would expect to see some connection of the TRNP horses with one or all of the Spanish breeds. In the PCA, the breeds of Spanish origin (Andalusian, Lusitano, Mangalarga Paulista, Peruvian Paso, Puerto Rican Paso Fino) cluster together tightly (Figure 1). The TRNP horses fall within that range on PC1 but diverge from those breeds on PC2. The F_{ST} values between the TRNP horses and the Spanish breeds are higher than average, however, and nine other non-Spanish breeds have a lower F_{ST}. Finally, the maximum likelihood tree separates the Spanish breeds from the TRNP horses. Based on this evidence it is unlikely that there is a strong ancestral connection to the Spanish breeds. Considering the history of horses in the Americas we cannot rule out all previous Spanish influence, but the genetic data do not support the oral history of predominantly Spanish origin for TRNP horses.

CHAPTER II REFERENCES

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research, 19*(9), 1655–1664. https://doi.org/10.1101/GR.094052.109

Ashley, M. C. (2004). Population genetics of feral horses: implications of behavioral isolation. *Journal of Mammalogy, 85*(4), 611–617. https://academic.oup.com/jmammal/article/85/4/611/858926

Baker, D. L., Powers, J. G., Ransom, J. I., McCann, B. E., Oehler, M. W., Bruemmer, J. E., Galloway, N. L., Eckery, D. C., & Nett, T. M. (2018). Reimmunization increases contraceptive effectiveness of gonadotropin-releasing hormone vaccine (GonaCon-Equine) in free-ranging horses (Equus caballus): Limitations and side effects. *PLoS ONE, 13*(7), e0201570. https://doi.org/10.1371/journal.pone.0201570

Battell, J. (1894). The morgan horse and register. In *Register Printing Company* (Vol. 1). Register Printing Company.

Bertrand, A. R., Kadri, N. K., Flori, L., Gautier, M., & Druet, T. (2019). RZooRoH: An R package to characterize individual genomic autozygosity and identify homozygous-by-descent segments. *Methods in Ecology and Evolution, 10*, 860–866.

Brooks, S. A., Palermo, K. M., Kahn, A., Hein, J. (2020). Impact of white-spotting alleles, including *W20*, on phenotype in the American Paint Horse. *Animal Genetics, 51*(5), 707-715. https://doi.org/10.1111/age.12960

Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M., & Wilson, J. F. (2018). Runs of homozygosity: Windows into population history and trait architecture. *Nature Reviews Genetics, 19*(4), 220–234. Nature Publishing Group. https://doi.org/10.1038/nrg.2017.109

Conant, E. K., Juras, R., & Cothran, E. G. (2012). A microsatellite analysis of five Colonial Spanish horse populations of the southeastern United States. *Animal Genetics, 43*(1), 53–62. https://doi.org/10.1111/J.1365-2052.2011.02210.X

Cosgrove, E. J., Sadeghi, R., Schlamp, F., Holl, H. M., Moradi-Shahrbabak, M., Miraei-Ashtiani, S. R., Abdalla, S., Shykind, B., Troedsson, M., Stefaniuk-Szmukier, M., Prabhu, A., Bucca, S., Bugno-Poniewierska, M., Wallner, B., Malek, J., Miller, D. C., Clark, A. G., Antczak, D. F., & Brooks, S. A. (2020). Genome diversity and the origin of the arabian horse. *Scientific Reports, 10*(1), 9702. https://doi.org/10.1038/s41598-020-66232-1

Cothran, E. G. (1992). *Genetic marker analysis of the Theodore Roosevelt National Park feral horse herd.*

Cothran, E. G. (2000). *Analysis of genetic variation in the feral horse herd of the Theodore Roosevelt National Park in 2000*.

Cothran, E. G., & McCrory, W. P. (2014). *A preliminary genetic study of the wild horse (Equus caballus) in the Brittany Triangle (Tachelach'ed) region of the ?Elegesi Qayus (Nemiah) Wild Horse Preserve of British Columbia*. http://www.lrgaf.org/articles/Wild Horse DNA Report 2015.pdf

Crawford, L. F. (1931). *History of North Dakota: Vol. I*. American Historical Society.

Dorji, J., Tamang, S., Tshewang, T., Dorji, T., & Dorji, T. Y. (2018). Genetic diversity and population structure of three traditional horse breeds of Bhutan based on 29 DNA microsatellite markers. *PLOS ONE*, *13*(6), e0199376. https://doi.org/10.1371/JOURNAL.PONE.0199376

Felkel, S., Vogl, C., Rigler, D., Jagannathan, V., Leeb, T., Fries, R., Neuditschko, M., Rieder, S., Velie, B., Lindgren, G., Rubin, C.-J., Schlotterer, C., Rattei, T., Brem, G., & Wallner, B. (2018). Asian horses deepen the MSY phylogeny. *Animal Genetics. 49*(1). 90-93. https://doi.org/10.1111/age.12635

Florida Cracker Horse Association. (n.d.). Retrieved March 5, 2022, from https://floridacrackerhorseassociation.com/about-us/

Funk, S. M., Guedaoura, S., Juras, R., Raziq, A., Landolsi, F., Luís, C., Martínez, A. M., Musa Mayaki, A., Mujica, F., Oom, M. do M., Ouragh, L., Stranger, Y. M., Vega-Pla, J. L., & Cothran, E. G. (2020). Major inconsistencies of inferred population genetic structure estimated in a large set of domestic horse breeds using microsatellites. *Ecology and Evolution*, *10*(10), 4261–4279. https://doi.org/10.1002/ECE3.6195

Glowatzki-Mullis, M. L., Muntwyler, J., Pfister, W., Marti, E., Rieder, S., Poncet, P. A., & Gaillard, C. (2006). Genetic diversity among horse populations with a special focus on the Franches-Montagnes breed. *Animal Genetics*, *37*(1), 33–39. https://doi.org/https://doi.org/10.1111/j.1365-2052.2005.01376.x

Goodloe, R. B., Warren, R. J., Cothran, E. G., Bratton, S. P., & Trembicki, K. A. (1991). Genetic variation and its management applications in Eastern U.S. feral horses. *The Journal of Wildlife Management*, *55*(3), 412–421. https://about.jstor.org/terms

Gurgul, A., Jasielczuk, I., Semik-Gurgul, E., Pawlina-Tyszko, K., Stefaniuk-Szmukier, M., Szmatoła, T., Polak, G., Tomczyk-Wrona, I., & Bugno-Poniewierska, M. (2019). A genome-wide scan for diversifying selection signatures in selected horse breeds. *PLOS ONE*, *14*(1), e0210751. https://doi.org/10.1371/JOURNAL.PONE.0210751

Harmon, D. (1986). *At the open margin: The NPS's administration of Theodore Roosevelt National Park*. Theodore Roosevelt Nature and History Association.

Hendricks, B. Lou. (1995). *International encyclopedia of horse breeds*. University of Oklahoma press.

Hendrickson, S. L. (2013). A genome wide study of genetic adaptation to high altitude in feral Andean Horses of the páramo. *BMC Evolutionary Biology*, *13*(237). https://doi.org/10.1186/1471-2148-13-273

Huidekoper, A. C. (1947). *My experience and investment in the Bad Lands of Dakota and some of the men I met there*. Wirth Brothers.

Imsland, F., McGowan, K., Rubin, C. J., Henegar, C., Sundström, E., Berglund, J., Schwochow, D., Gustafson, U., Imsland, P., Lindblad-Toh, K., Lindgren, G., Mikko, S., Millon, L., Wade, C., Schubert, M., Orlando, L., Penedo, M. C. T., Barsh, G. S., & Andersson, L. (2016). Regulatory mutations in TBX3 disrupt asymmetric hair pigmentation that underlies Dun camouflage color in horses. *Nature Genetics*, *48*(2), 152–158. https://doi.org/10.1038/ng.3475

Khanshour, A., Conant, E., Juras, R., & Cothran, E. G. (2013). Microsatellite analysis of genetic diversity and population structure of arabian horse populations. *Journal of Heredity*, *104*(3), 386–398. https://doi.org/10.1093/JHERED/EST003

Khanshour, A., Juras, R., Blackburn, R., & Cothran, E. G. (2015). The legend of the Canadian horse: genetic diversity and breed origin. *Journal of Heredity*, *106*(1), 37–44. https://doi.org/10.1093/JHERED/ESU074

Lacy, R. C., & Pollak, J. P. (2021). *Vortex: a stochastic simulation of the extinction process* (10.5.5). Chicago Zoological Society.

Lewis, P. O. (2003). NCL: a C++ class library for interpreting data files in NEXUS format. *Bioinformatics*, *19*(17), 2330–2331. https://doi.org/10.1093/BIOINFORMATICS/BTG319

Li, G., Figueiro, H. V, Eizirik, E., & Murphy, W. J. (2019). Recombination-aware phylogenomics reveals the structured genomic landscape of hybridizing cat species. *Molecular Biology and Evolution*, *36*(10), 2111–2126. https://doi.org/10.1093/molbev/msz139

Makvandi-Nejad, S., Hoffman, G. E., Allen, J. J., Chu, E., Gu, E., Chandler, A. M., Loredo, A. I., Bellone, R. R., Mezey, J. G., Brooks, S. A., & Sutter, N. B. (2012). Four loci explain 83% of size variation in the horse. *PLOS ONE*, *7*(7), e39929. https://doi.org/10.1371/JOURNAL.PONE.0039929

Mannen, H., Yonezawa, T., Murata, K., Noda, A., Kawaguchi, F., Sasazaki, S., Olivieri, A., Achilli, A., & Torroni, A. (2020). Cattle mitogenome variation reveals a post-glacial expansion of haplogroup P and an early incorporation into northeast Asian domestic herds. *Scientific Reports 2020 10:1*, *10*(1), 1–7. https://doi.org/10.1038/s41598-020-78040-8

Marlow, C. B., Gagnon, L. C., Irby, L. R., & Raven, M. R. (1992). *Feral horse distribution, habitat use, and population dynamics in Theodore Roosevelt National Park*.

McLaughlin, C. (1989). *The history and status of the wild horses of Theodore Roosevelt National Park*.

McVean, G. (2009). A genealogical interpretation of principal components analysis. *PLoS Genetics*, *5*(10), e1000686. https://doi.org/10.1371/JOURNAL.PGEN.1000686

Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES science gateway for inference of large phylogenetic trees. *Gateway Computing Environments Workshop (GCE)*, 1–8.

National Park Service. (1978). *Environmental assessment: proposed feral horse reduction, Theodore Roosevelt National Memorial Park*.

National Research Council. (2013). *Using science to improve the BLM Wild Horse and Burro Program : a way forward*.

Naundrup, P. J., & Svenning, J.-C. (2015). A geographic assessment of the global scope for rewilding with wild-living horses (Equus ferus). *PLoS ONE*, *10*(7), e0132359. https://doi.org/10.1371/journal.pone.0132359

Orlando, L. (2020). Ancient genomes reveal unexpected horse domestication and management dynamics. *BioEssays*, *42*. https://doi.org/10.1002/bies.201900164

Ortiz, E. M. (2019). *vcf2phylip v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis.* (v2.0). Zenodo. https://doi.org/10.5281/ZENODO.2540861

Ovchinnikov, I. V., Dahms, T., Herauf, B., McCann, B., Juras, R., Castaneda, C., & Cothran, E. G. (2018). Genetic diversity and origin of the feral horses in Theodore Roosevelt National Park. *PLoS ONE*, *13*(8). https://doi.org/10.1371/journal.pone.0200795

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLOS Genetics*, *2*(12), e190. https://doi.org/10.1371/JOURNAL.PGEN.0020190

Peripolli, E., Munari, D. P., Silva, M. V. G. B., Lima, A. L. F., Irgang, R., & Baldi, F. (2016). Runs of homozygosity: current knowledge and applications in livestock. *Animal Genetics*, *48*, 255–271. https://doi.org/10.1111/age.12526

Petersen, J. L., Mickelson, J. R., Cleary, K. D., & McCue, M. E. (2014). The american quarter horse: Population structure and relationship to the thoroughbred. *Journal of Heredity*, *105*(2), 148–162. https://doi.org/10.1093/jhered/est079

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., da Câmara Machado, A., Distl, O., Felicetti, M., Fox-Clipsham, L., Graves, K. T., Guérin, G., Haase, B., Hasegawa, T., Hemmann, K., … McCue, M. E. (2013). Genetic diversity in the modern horse illustrated from genome-wide SNP data. *PLoS ONE*, *8*(1). https://doi.org/10.1371/journal.pone.0054997

Plante, Y., Vega-Pla, J. L., Lucas, Z., Colling, D., de March, B., & Buchanan, F. (2007). Genetic diversity in a feral horse population from Sable Island, Canada. *Journal of Heredity*, *98*(6), 594–602. https://doi.org/10.1093/JHERED/ESM064

Quan, J., Cai, Y., Yang, T., Ge, Q., Jiao, T., & Zhao, S. (2020). Phylogeny and conservation priority assessment of Asian domestic chicken genetic resources. *Global Ecology and Conservation*, *22*, e00944. https://doi.org/10.1016/J.GECCO.2020.E00944

Raudsepp, T., Finno, C. J., Bellone, R. R., & Petersen, J. L. (2019). Ten years of the horse reference genome: insights into equine biology, domestication and population dynamics in the post-genome era. In *Animal Genetics* (Vol. 50, Issue 6, pp. 569–597). Blackwell Publishing Ltd. https://doi.org/10.1111/age.12857

Rout, P. K., Joshi, M. B., Mandal, A., Laloe, D., Singh, L., & Thangaraj, K. (2008). Microsatellite-based phylogeny of Indian domestic goats. *BMC Genetics*, *9*(1), 1–11. https://doi.org/10.1186/1471-2156-9-11/FIGURES/5

Santschi, E. M., Purdy, A. K., Valberg, S. J., Vrotsos, P. D., Kaese, H., & Mickelson, J. R. (1998). Mammalian genome endothelin receptor B polymorphism associated with lethal white foal syndrome in horses. *Mammalian Genome*, *9*, 306–309.

Schierup, M. H., & Hein, J. (2000). Consequences of recombination on traditional phylogenetic analysis. *Genetics*, *156*(2), 879-891. https://academic.oup.com/genetics/article/156/2/879/6051420

Shahsavarani, H., & Rahimi-Mianji, G. (2010). Analysis of genetic diversity and estimation of inbreeding coefficient within Caspian horse population using microsatellite markers. *African Journal of Biotechnology*, *9*(3), 293–299. http://www.academicjournals.org/AJB

Sponenberg, D. P. (1992). The colonial Spanish horse in the USA: history and current status. *Archivos de Zootecnia*, *41*(extra), 335–348. https://www.researchgate.net/publication/28057629

Sponenberg, P. (1994). *Sponenberg evaluation of Roosevelt National Park horses, October 1994*.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/BIOINFORMATICS/BTU033

Sun, T., Huang, G., Sun, J., Wang, Z., Teng, S., Cao, Y., Hanif, Q., Chen, N., Lei, C., & Liao, Y. (2020). Mitogenome diversity and maternal origins of Guangxi buffalo breeds. *Animals 2020, Vol. 10, Page 547*, *10*(4), 547. https://doi.org/10.3390/ANI10040547

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, *22*(22), 4673–4680.

Tollett, C. M. (2018). *Genomic diversity and origins of the feral horses (Equus ferus caballus) of Sable Island and the Alberta Foothills*. University of Saskatchewan.

Vega-Pla, J. L., Calderó, J., Rodríguez-Gallardo, P. P., Martinez, A. M., & Rico, C. (2006). Saving feral horse populations: does it really matter? A case study of wild horses from Doñana National Park in southern Spain. *Animal Genetics*, *37*, 571–578. https://doi.org/10.1111/j.1365-2052.2006.01533.x

Vilà, C., Leonard, J. A., Götherström, A., Marklund, S., Sandberg, K., Lidén, K., Wayne, R. K., & Ellegren, H. (2001). Widespread origins of domestic horse lineages. *Science*, *291*(5503), 474–477. https://doi.org/10.1126/SCIENCE.291.5503.474

Vonholdt, B. M., Pollinger, J. P., Lohmueller, K. E., Han, E., Parker, H. G., Quignon, P., Degenhardt, J. D., Boyko, A. R., Earl, D. A., Auton, A., Reynolds, A., Bryc, K., Brisbin, A., Knowles, J. C., Mosher, D. S., Spady, T. C., Elkahloun, A., Geffen, E., Pilot, M., … Wayne, R. K. (2010). Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature*, *464*, 898–902. https://doi.org/10.1038/nature08837

Wade, C. M., Giulotto, E., Sigurdsson, S., Zoli, M., Gnerre, S., Imsland, F., Lear, T. L., Adelson, D. L., Bailey, E., Bellone, R. R., Blocker, H., Distl, O., Edgar, R. C., Garber, M., Leeb, T., Mauceli, E., MacLeod, J. N., Penedo, M. C. T., Raison, J. M., … Lindblad-Toh, K. (2009). Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science*, *326*(5954). https://doi.org/10.1126/science.1178158

Winton, C. L., McMahon, R., Hegarty, M. J., McEwan, N. R., Davies-Morel, M. C. G., Morgan, C., & Nash, D. M. (2020). Genetic diversity within and between British and Irish breeds: The maternal and paternal history of native ponies. *Ecology and Evolution*, *10*(3), 1352–1367. https://doi.org/10.1002/ece3.5989

Yang, L., Kong, X., Yang, S., Dong, X., Yang, J., Gou, X., & Zhang, H. (2018). Haplotype diversity in mitochondrial DNA reveals the multiple origins of Tibetan horse. *PLOS ONE*, *13*(7), e0201564. https://doi.org/10.1371/JOURNAL.PONE.0201564

CHAPTER III

GWAS REVEALS ASSOCIATION BETWEEN SNPS ON ECA18 AND LONG-TERM EFFICACY OF GONACON IMMUNOCONTRACEPTION IN FERAL HORSES

INTRODUCTION

Preventing overpopulation is a primary concern for wildlife managers, especially for populations with high growth rates, a lack of natural predators, and for species that generate human-wildlife conflict (Messmer, 2000). One standard approach is to decrease population size by lethal or live-capture culling, effectively increasing the mortality rate. A common alternative, especially for larger mammalian species, is to decrease the birth rate through contraceptive vaccination. Because female reproductive success is the limiting factor in population growth, contraceptive efforts are often focused on females (Garrott, 1995). Immunocontraception has been used effectively in many different species, wild and domestic, with contraceptive vaccines targeting one of two proteins that are required for reproduction (Kirkpatrick et al., 2011). Some vaccines use porcine zona pellucida protein (PZP) as an antigen (Kirkpatrick et al., 2009). Antibodies to PZP bind to the zona pellucida and block fertilization in immunized females. The second target protein is gonadotropin-releasing hormone (GnRH), a key hypothalamic hormone that regulates synthesis and secretion of follicle-stimulating hormone (FSH) and luteinizing hormone (LH) from the pituitary gland (Marques et al., 2022).

Immunization against GnRH has been increasingly used to manage wild populations. The National Wildlife Research Center developed an immunocontraceptive named GonaCon™,

which consists of a GnRH peptide combined with an adjuvant to trigger an immune response (Miller et al., 2004). This generates neutralizing antibodies against GnRH, thereby inhibiting FSH and LH synthesis and secretion. In females this vaccine blocks ovulation and estrus, resulting in infertility (Miller et al., 2004). In male white-tailed deer, GonaCon vaccination leads to reduced testosterone levels, reduction of testes size, and inhibition of reproductive behavior (Killian et al., 2005). However, infertility is temporary in both males and females, lasting only as long as GnRH-antibody titers remain high enough to block GnRH activity (Miller et al., 2000; Pinkham et al., 2022). The efficacy of GonaCon in reducing reproduction has been evaluated in a wide variety of mammalian species, such as prairie dogs and fox squirrels, wild boar, feral cats, white-tailed deer, and elk (Krause et al., 2014; Massei et al., 2012; Miller et al., 2008; Powers et al., 2014; Vansandt et al., 2017; Yoder & Miller, 2010).

GonaCon has also been considered for use in feral horse herds. Feral horses on public lands exhibit a high population growth rate with about 20% increase per year (National Research Council, 2013) and cause ecological damage at high density (Eldridge et al., 2020), necessitating some form of population control. Public sentiment favors contraception as a horse management tool over capture and removal, yet the public often responds poorly to permanent sterilization of horses. Reversibility of the GonaCon vaccine provides an appealing option for management of feral horse populations.

Baker et al. (2018) began research on the long-term effectiveness of GonaCon-Equine in free-roaming feral horses in Theodore Roosevelt National Park (TRNP) in 2009, with 25 mares receiving a primary vaccination that fall and a secondary vaccination four years later, by hand-injection at a roundup. Due to the success of the treatment, the study was expanded. To determine optimal reimmunization interval, 38 additional mares in three treatment groups were

treated by remote delivery via dart, with the last group receiving their secondary vaccination in 2016 (Baker et al., manuscript in prep). Collectively, the reimmunization treatments were divided into groups based on the interval between primary and secondary injections: 4-year intervals, 2-year intervals, 1-year intervals, and 6-month intervals. A control group was treated with a saline injection on the same schedule as the 4-year group. Foaling rates of these mares were monitored from 2009 through 2021, with records of observational pedigree and foaling accounts going back to the 1980s. While the treatment was shown to be effective in reducing foaling rates, the duration of infertility is variable; some mares had one or two years of infertility while others remained infertile for the duration of the study. Factors influencing individual mares' responses to treatment are as yet unknown.

Individual variation in reproductive rates following immunocontraception could be due to many factors, both environmental and/or genetic. Some genetic components of fertility are known, but many aspects of reproduction are complex and not fully understood. In humans, genes have been identified for some female reproductive disorders/diseases as well as markers contributing to fertility traits such as reproductive lifespan and ovarian function (Gajbhiye et al., 2018). The genetics of fertility traits in dairy cattle has also been a subject of interest, with several candidate markers identified (Ma et al., 2019). However, there are currently no known genes for fertility in horses (Laseca et al., 2021; Raudsepp, 2020).

Variation in immunocontraceptive efficacy could also be due to variation in genes underlying immune responses. As with fertility, immune response is complex and difficult to attribute to any single gene. Research into human disease-resistance has demonstrated that epigenomic changes in gene expression play a role beyond purely genetic factors (Yamamoto et al., 2021). Of known genes, many are found within the Major Histocompatibility Complex

(MHC), a highly polymorphic region of the genome found across many species that contains genes associated with immune function and infectious disease (Trowsdale & Knight, 2013). In horses, the MHC is located on chromosome 20 and has been associated with several equine specific diseases (Viļuma et al., 2017).

Many genes affecting a disease or phenotypic trait have been identified through genetic association studies. Genome-wide association studies (GWAS) can simultaneously identify multiple candidate markers for complex traits, especially when using thousands or hundreds of thousands of single nucleotide polymorphisms (SNPs) distributed across the genome (Montgomery et al., 2014). Once a quantitative trait locus (QTL) has been identified, further experiments can be conducted to determine the functional effects of the polymorphism and the biochemical pathways driving phenotypic variation. GWAS have been used to identify causal mutations for horse coat colors (Brooks et al., 2010; Imsland et al., 2016), genetic diseases (Finno et al., 2018), and genes involved in body size variation (Makvandi-Nejad et al., 2012). Two loci influencing risk of equine recurrent uveitis, an autoimmune disease, were identified by GWAS in 144 German warmblood horses (Kulbrock et al., 2013). Some QTL in cattle have even been associated with immune response to a vaccination (Leach et al., 2012).

Here, I use SNP genotypes to conduct a GWAS in search of QTL associated with long-term effectiveness of GonaCon-Equine. I hypothesize that genes involved with fertility and/or immune response will impact the duration of infertility following vaccination. This research will improve our understanding of the mechanism by which the vaccine works and inform future use of GonaCon-Equine in other feral herds. More broadly, findings from this study may also provide insight into why vaccine induced immunity persists or wanes over time, which is an outstanding question in immunology (Kennedy et al., 2020).

METHODS

Sample and Data Collection

Hair samples were collected from all study mares during a regularly scheduled roundup in 2013. In 2017, tissue samples were collected via biopsy dart from four mares that had evaded the 2013 capture and were included in the later GonaCon treatment groups. I collected an additional 18 tissue samples from young study mares born in 2014 by biopsy dart under research permit number THRO-2020-SCI-0013 and NPS IACUC approval (ND_THRO_McCann_HorseBiopsyDarting_2020.A1). This resulted in a total of 88 samples, which included all treated and control mares that had received both injections (Table 4). I used foaling data as collected by Baker et al. (2018 and manuscript in prep) for all mares. I also generated pedigrees from TRNP records for the herd going back to the 1980s. Many pedigree relationships of the past two decades have been confirmed through genetic testing, and older relationships were recorded from behavioral observations.

Table 4. GonaCon treatment groups, named for time interval between primary and booster GonaCon injection. Control group received saline injections at the same schedule as 4-year group.

Treatment Groups	Number of Individuals
Control	25
4-year	25
2-year	11
1-year	14
0.5-year	13

DNA Extraction and Genotyping

Genomic DNA was extracted from hair follicle and tissue samples by the Animal Genetics Laboratory at Texas A&M University using Gentra Puregene Tissue kits (Qiagen) following manufacturer's protocols. Individuals were then genotyped at Neogen Genomics Laboratory (Lincoln, NE) for over 70k SNPs located evenly across the horse genome using the Illumina Equine GGP 70k BeadChip (Illumina Inc., San Diego, CA).

Data Pruning

I performed quality control filtering of the genotypes using SNP & Variation Suite v8.9.0 (SVS) (Golden Helix, Inc., Bozeman, MT, www.goldenhelix.com). I removed markers with call rate ≤0.95 and with a minor allele frequency (MAF) of 0.05 or less. I removed markers with Hardy-Weinberg equilibrium (HWE) P-values $<1x10^{-6}$ as recommended by Marees et al. (2018). All samples had a call rate ≥0.90 and all but three had a call rate ≥0.98, so I retained all individuals due to the already limited sample size. I mapped the remaining SNPs to the horse reference genome EquCab3.0 (https://www.ncbi.nlm.nih.gov/assembly/GCF_002863925.1/). Since my dataset only included females, I retained the SNPs on the X-chromosome. The dataset contained a total of 47,495 SNPs after pruning.

Data Analysis

I conducted a single-locus mixed linear model (MLM) GWAS using SVS's implementation of EMMAX (Kang et al., 2010; Vilhjalmsson, 2012). Infertility following GonaCon booster was treated as a quantitative trait. Because treatment groups had variable durations since the booster, I calculated a normalized measure of infertility as the proportion of years infertile: (# years infertile)/(total # of years since booster). Infertility values ranged from 1 for those mares that

exhibited complete infertility, to 0 for mares that exhibited complete fertility since the booster (Figure 8). I included Treatment Group as a fixed effect interaction with genotype in the model. Since many of the mares were related to each other, I computed a relationship matrix in SVS using pedigree data, which was included as a random effect kinship matrix to account for relatedness. The final model can be represented as Infertility = Genotype x Treatment Group + Kinship. The MLM analysis used an additive genetic model and included Bonferroni multiple testing corrections and false discovery rate (FDR) calculations. I used $P=5x10^{-8}$ for a genome wide significance threshold but also considered SNPs under a less stringent threshold of $P=5x10^{-7}$ as suggested by Chen et al. (2021). I referenced EquCab3.0 to identify genes that contained significant SNPs and genes within broader regions of multiple significant SNPs. I included genes found within 200 kb of significant SNPs (Brodie et al., 2016). To investigate biological function of these candidate genes I used The Human Gene Database (www.genecards.org).

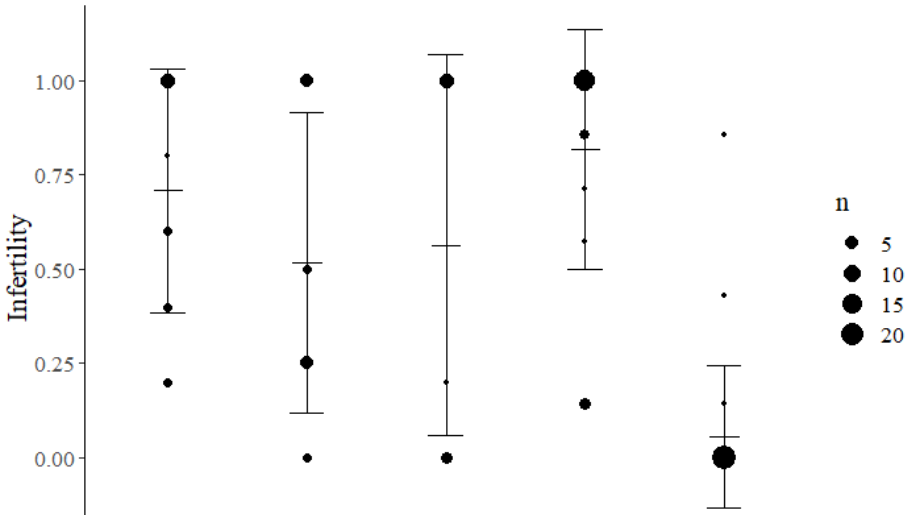

Figure 8. Infertility after treatment with GonaCon-Equine for mares in each reimmunization group, with error bars indicating mean and standard deviation. Infertility was calculated as the proportion of years infertile following booster injection.

I also evaluated several other variations of the MLM to test for potentially confounding variables. To verify that significant associations were due to the treatment group by genotype interaction, I dropped the interaction from the model and only included main effects, summarized as Infertility = Genotype + Treatment Group + Kinship. To determine whether there were any differences in basal infertility among mares prior to assignment to treatment groups, I tested a MLM with infertility prior to GonaCon immunization as the dependent variable. Infertility prior to injection was calculated in the same manner as infertility following the booster: (# years infertile)/(total # of years between sexual maturity and initial injection). The model used included the same interaction term and kinship matrix as above, summarized as Infertility Before Injection = Genotype x Treatment Group + Kinship. I also analyzed several models with additional variables such as age at first injection and injection delivery method (hand injection or dart delivery) as covariates. Those models had no impact on the results of the simpler model reported below.

RESULTS

The mixed linear model GWAS for infertility after GonaCon treatment using the treatment group by genotype interaction yielded 10 significant SNPs at the $p=5x10^{-7}$ threshold, with one SNP also significant at $p=5x10^{-8}$ (Table 5). These 10 SNPs were all located within a single 2.10 Mb region on chromosome 18 of the horse reference genome (Figure 9). All 10 SNPs were also significant by Bonferroni multiple testing correction, with P-values <0.05. The FDR for each was <0.005. A Q-Q plot of observed versus expected P-values shows deviation of significant SNPs from values expected when assuming no association (Figure 10). This tail strongly suggests that genetic variants truly influence GonaCon efficacy.

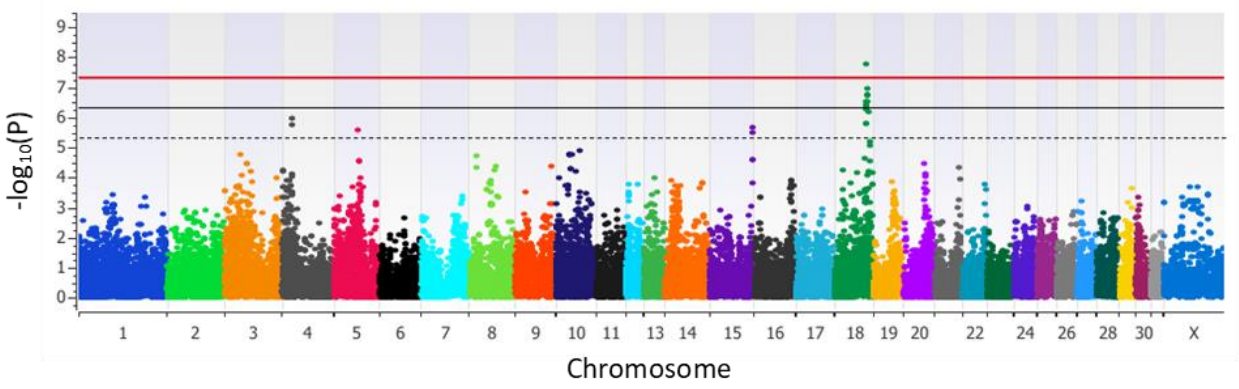

Figure 9. Manhattan plot of SNP association with infertility in TRNP mares following treatment with GonaCon-Equine. Red line denotes the genome wide significance threshold of $P=5x10^{-8}$ where $-\log_{10}(P)=7.3$. Black line at 6.3 denotes the genome wide significance threshold of $P=5x10^{-7}$. Dashed line at 5.3 denotes the relaxed threshold of $P=5x10^{-6}$ used in evaluating suggestive loci on other chromosomes.

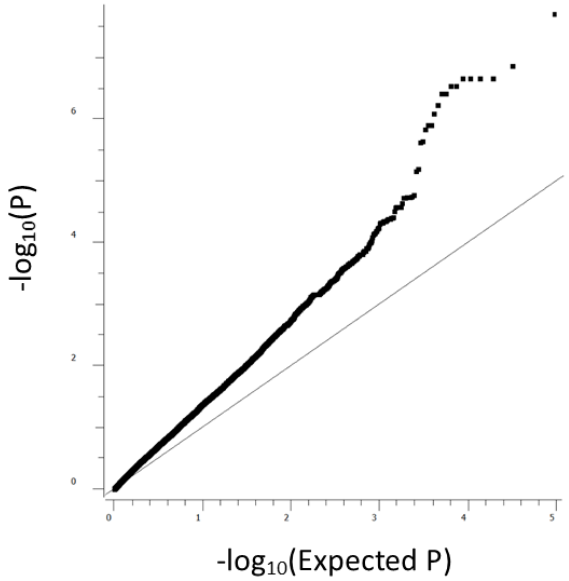

Figure 10. Q-Q plot of observed versus expected P-values from GWAS for infertility in TRNP mares following treatment with GonaCon-Equine.

None of the significant SNPs were located within the coding region of a gene, however one was located within an intron of *NGFI-A binding protein 1* (NAB1) (Table 5). The region of ECA18 with significant SNPs contained ten annotated genes and two uncharacterized protein coding regions (Table 6). Two of those genes, *signal transducer and activator of transcription 1*

Table 5. SNPs that exhibit significant genome-wide associations with GonaCon response in TRNP mares using a mixed linear model with threshold $P< 5x10^{-7}$.

Marker ID	Chr.	Position	P-Value	-log10(P)	Bonferroni P	FDR	Gene
BIEC2_417453	18	67111422	1.79225E-08	7.74660	0.00084	0.00084	NAB1
BIEC2_417645	18	68859296	1.16564E-07	6.93344	0.00549	0.00274	-
Affx-101317912	18	68836692	1.95723E-07	6.70836	0.00922	0.00154	-
Affx-103083998	18	68836582	1.95723E-07	6.70836	0.00922	0.00184	-
Affx-101279033	18	68760139	1.95723E-07	6.70836	0.00922	0.00230	-
BIEC2_439199	18	68737439	1.95723E-07	6.70836	0.00922	0.00307	-
BIEC2_417676	18	69211986	3.05797E-07	6.51457	0.01440	0.00180	-
Affx-102073543	18	68647709	3.05797E-07	6.51457	0.01440	0.00206	-
BIEC2_417588	18	68141026	4.02225E-07	6.39553	0.01894	0.00189	-
BIEC2_417587	18	68125695	4.02225E-07	6.39553	0.01894	0.00210	-

Table 7. SNPs on other chromosomes which are above the threshold of $P=5x10^{-6}$. Asterisks mark genes with known immune system function and parentheses indicate a gene located ~34 kb away from that SNP.

Marker ID	Chr.	Position	P-Value	-log10(P)	Bonferroni P	FDR	Gene
BIEC2_852319	4	20272692	1.14076E-06	5.94280	0.05389	0.00415	DDC
BIEC2_852290	4	20206184	1.83229E-06	5.73701	0.08655	0.00541	IKZF1*
Affx-101571882	15	92073632	2.30546E-06	5.63724	0.10890	0.00641	-
BIEC2_909642	5	51399392	2.64538E-06	5.57751	0.12496	0.00694	(PTPN22)*
Affx-102026979	15	92050892	3.25996E-06	5.48679	0.15399	0.00810	-

(STAT1) and *signal transducer and activator of transcription 4* (STAT4), have known roles in the immune system. A flanking gene, *major facilitator superfamily domain containing 6* (MFSD6) encodes a transmembrane protein which is predicted to enable protein binding and receptor activity of MHC class 1 molecules.

Table 6. Genes found within the 2.10 Mb ± 200 kb region on chromosome 18 of EquCab3.0 that contains SNPs with $P < 5x10^{-7}$. Asterisks mark genes with known immune system function.

Chr.	Start	Stop	Gene Name
18	66908950	66981047	MFSD6
18	66984476	67008007	NEMP2
18	67096736	67106335	LOC102148954
18	67099791	67138617	NAB1
18	67177759	67210309	LOC102149070
18	67330485	67410571	GLS
18	67414096	67452831	STAT1*
18	67464261	67559483	STAT4*
18	67654558	67834776	MYO1B
18	68053863	68062353	NABP1
18	68189795	68203091	CAVIN2
18	68298154	68523485	TMEFF2

Inspection of the Manhattan plot in Figure 9 suggests potential regions of interest on chromosomes 4, 5, and 15, despite SNPs falling below the most stringent thresholds for genome-wide significance. Five SNPs within those regions had P-value $< 5x10^{-6}$ (Table 7) and fourteen genes were located within a 200 kb window around those loci (Table 8). One of the SNPs on ECA4 is located within an intron of the *IKAROS family zinc finger 1* (IKZF1) gene which regulates lymphocyte differentiation. The nearest annotated gene to the SNP on ECA5 is *protein tyrosine phosphatase non-receptor type 22* (PTPN22), which regulates T-cell receptor signaling. There were no genes located within the 200 kb window surrounding the SNPs on ECA15.

Table 8. Genes found within 200 kb of SNPs which are above the threshold $P=5x10^{-6}$. Asterisks mark genes with known immune system function.

Chr.	Start	Stop	Gene Name
4	20122309	20219653	IKZF1*
4	20258323	20265158	FIGNL1
4	20269062	20359213	DDC
4	20372622	20577702	GRB10
5	51253066	51256536	OLFML3
5	51257521	51306709	HIPK1
5	51319364	51360409	DCLRE1B
5	51327263	51336444	AP4B1
5	51363965	51411556	LOC100064949
5	51431066	51437642	BCL2L15
5	51423450	51492288	PTPN22*
5	51493896	51531777	RSBN1
5	51534055	51581007	PHTF1
5	51587886	51811736	MAGI3
15	-	-	-

None of the other models yielded significant SNPs for infertility. When the treatment group by genotype interaction was dropped from the model, SNPs on ECA4, ECA5, ECA15, and ECA18 were no longer significant (Figure 11a). In addition, these SNPs were not associated with infertility in mares prior to their treatment with GonaCon (Figure 11b). The difference between models including vs. excluding the genotype by environment interaction and models before vs. after GonaCon immunization strongly indicate that genetic variants specifically influence GonaCon efficacy. Addition of other covariates (i.e., age at immunization, vaccine delivery method) to the original model did not appreciably change the GWAS results, and so were not included in the final model described above.

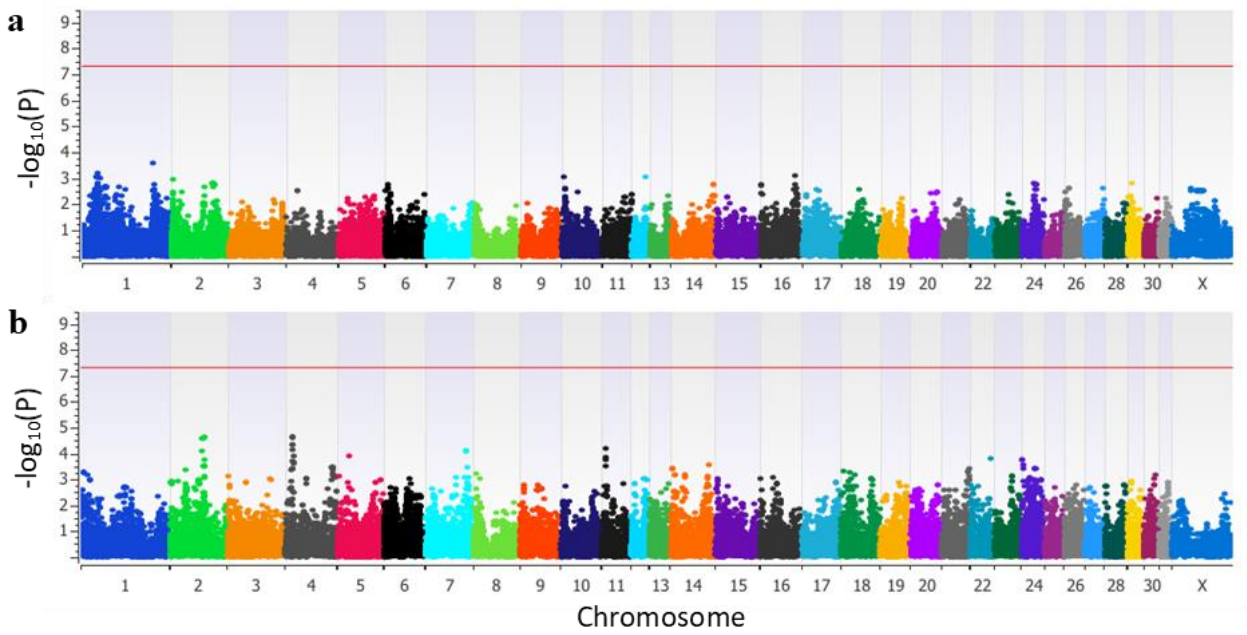

Figure 11. Manhattan plots from other MLMs to test for SNP association with infertility in TRNP mares following treatment with GonaCon-Equine. Red line denotes the genome wide significance threshold of $P=5x10^{-8}$. a) SNP significance in infertility after GonaCon treatment when treatment group is included as a covariate rather than an interaction with genotype. b) SNP significance when infertility before first injection is used as dependent variable.

DISCUSSION

This study demonstrates that genetic variants influenced the response of TRNP mares to GonaCon vaccination. The lack of significant SNPs when the GonaCon treatment by genotype interaction was dropped from the model indicates that these associations cannot simply be explained by genotype alone or by GonaCon vaccination alone. In addition, there were no associations between genotype and infertility prior to GonaCon vaccination. Taken together, these findings indicate that individual differences in infertility are the results of genetic variation

in the immune response to the GonaCon vaccine. Indeed, genetic variants are a major factor influencing immune responses to pathogens and vaccines in humans (Scepanovic et al., 2018).

The most likely candidate genes to influence GonaCon efficacy in TRNP mares are STAT1 and STAT4 on ECA18. The STAT family of genes are transcription activators with a wide range of functions, some of which include innate and adaptive immune response to a variety of pathogens (Horvath, 2000). Together with Janus kinases (JAK) they form the JAK-STAT signaling pathway by which cell surface receptors trigger transcription in the nucleus (Stark & Darnell, 2012). STAT1 and STAT4 variants have been shown to influence disease risk and immune responses in humans (Hedl et al., 2020). STAT1 is involved in the signaling pathway that activates antiviral responses, and reduced activity of STAT1 is connected to increased infection susceptibility (Katze et al., 2002; Najjar & Fagard, 2010). Mutations in STAT1 cause different clinical disorders depending on whether the mutations are recessive or dominant and whether they result in STAT1 deficiency or STAT1 gain-of-function (Mizoguchi & Okada, 2021). Loss of STAT1 function leads to increased susceptibility to mycobacterial and viral infection, and gain-of-function is associated with autoimmune disease in addition to increased infection susceptibility (Boisson-Dupuis et al., 2012). STAT1 plays a key role in inhibition of the IL-17 pathway, which can result in impaired immunity (Liu et al., 2011). The cytokine IL-17 induces tissue inflammation and is produced by Th17 memory cells (Korn et al., 2009), which are themselves a major cause of autoimmune disease in mice and humans when dysregulated (McGeachy, 2013). In horses, STAT1 was upregulated following infection with African horse sickness virus (Wall et al., 2021). STAT4, which is adjacent to STAT1 on ECA18, is also an important component in immune response to many diseases and is linked to inflammation and T-cell differentiation (Yang et al., 2020).

Genetic variation in STAT1 and/or STAT4 function within the TRNP herd could affect the mares' ability to mount a strong and long-lasting immune response to the GnRH peptide. However, the significant SNPs on ECA18 are not found within any of the candidate genes. In human response to an influenza vaccine STAT1 expression levels were linked to antibody response, with increased antibody titers strongly correlated with higher STAT1 expression in the first day after vaccination (Bucasas et al., 2011). Thus, we hypothesize that the causal SNP or SNPs in the TRNP mares are likely to influence expression of STAT1 and/or STAT4 either in basal conditions or in response to GonaCon vaccination. Future studies could test this hypothesis by drawing blood from mares and measuring STAT1 and STAT4 expression and activity in white blood cells under basal condition and in response to signaling molecules that activate STATs.

The observation that two of three suggestive loci on other chromosomes also contain genes related to immune function (Table 7) supports my initial hypothesis. IKZF1 and PTPN22 have both been implicated in association with a range of immunodeficiencies (Kuehn et al., 2021; Tizaoui et al., 2021). In humans, variants in IKZF1, PTPN22, and STAT4 are associated with risk of developing the autoimmune disorder systemic lupus erythematosus (Ebrahimiyan et al., 2018; Hu et al., 2013; Kyogoku et al., 2004; Mohan & Putterman, 2015). There is some evidence of epistasis between IKZF1, STAT4, and other loci (Dang et al., 2014), and IKZF1 may be a regulator of STAT4 in human T cells (Hu et al., 2013). PTPN22 has roles in both the adaptive and innate immune system as a signal inhibitor of T-cell activation and a signal enhancer promoting production of type I interferon in myeloid cells, respectively (Stanford & Bottini, 2014).

While we think that genetic variation in immune response to GonaCon vaccination is the most plausible hypothesis, STAT family genes have also been associated with reproductive traits in cattle. STAT1 and STAT4 are located together on chromosome 2 in the cattle genome and were within a genome window associated with age at first calving in cattle (Dubon et al., 2021). STAT1 and the paralog STAT3 were linked to embryonic survival in cattle (Khatib et al., 2009). Several members of the STAT family as part of the JAK-STAT pathway have also been shown to play a role in milk production in dairy cattle (Khan et al., 2020). In horses, STAT genes were found to be differentially expressed in adult versus fetal ovaries (Hall et al., 2017), but not much is known about the association of STAT gene variation with horse reproductive traits. In mice, the JAK-STAT pathway has been shown to play a critical role in normal reproductive development and function. Cytokine receptors activate JAK-STAT signaling in GnRH neurons, which regulate the secretion of GnRH peptide. Knockout of the JAK2 gene resulted in impaired fertility (Wu & Wolfe, 2012). Thus, altered STAT expression/activity in GnRH neurons could also potentially affect reproductive function in the mares. This hypothesis could be tested through experiments measuring STAT expression and activity in GnRH neurons in a rodent model of GonaCon immunocontraception.

Further work needs to be done to identify the specific locus responsible for the genetic association as well as the mechanism causing phenotypic variation in the long-term efficacy of GonaCon. Given that this vaccine induces an immune response to the body's own reproductive hormone, it is reasonable to assume that genes in immune function pathways are involved in either the initial immune response to GonaCon injection or the maintenance of immunity over time (i.e., Th17 memory cells). Although other loci did not cross the more stringent genome-wide thresholds of significance, the presence of other immune related genes within noticeable

peaks suggests that genetic variation across multiple loci affects the immune response to GonaCon vaccination.

Finally, there are additional factors not explored here which may also be influencing efficacy of GonaCon injections. Baker et al. (manuscript in prep) evaluated differences between injection method of treatment groups and found that the two-year dart injection treatment group had a more variable response than the 4-year hand injection treatment group, though the difference was not statistically significant. Neither of the other dart injected groups (0.5-year and 1-year) were significantly different from the hand injected group. The use of darts does have the potential for greater tissue trauma at the injection site due to the velocity of both the dart impact and the expulsion of the dart's contents (Hampton et al., 2021). Differences in dart placement, ballistics, injection force, the animal's behavioral response, and individual physiological variation can impact the response to darting. In contrast, the placement and force of the injection can be more carefully regulated when administered by hand. Baker er al. hypothesize that variation in contraceptive success between treatment groups is due to booster interval rather than injection delivery method, however it is possible that tissue trauma caused by darting may influence the strength of the immune response to vaccination.

Awareness of a genetic influence on the effectiveness of GonaCon, and other immunocontraceptive agents, in reducing foaling rates in feral mares is important when planning for population management. Over multiple generations the continued use of an immunocontraceptive within a population could result in unintended selection for alleles that decrease responsiveness to the vaccine, reducing its effectiveness over time and potentially weakening the immune response to other pathogens (Cooper & Larsen, 2006). Those managing feral horse herds should take into consideration their long-term goals for population

management. While GonaCon immunocontraception is currently effective, selection for a reduced immune response to GonaCon may necessitate new approaches in the future. The genetic variants associated with GonaCon response may or may not also affect response to the immunocontraceptive PZP, which is prepared using a different adjuvant to trigger immune response (Naz & Saver, 2016). To identify whether the same QTL are associated with PZP response a similar GWAS could be performed on mares treated with PZP. If PZP response is associated with different genetic variants than those identified in this study, selection pressure could be reduced by alternating GonaCon and PZP use in the population. Immunocontraception will likely need to be used in combination with other management techniques, balancing the benefits of slowing population growth with the potential for reduced efficacy of the vaccine over time.

CHAPTER III REFERENCES

Baker, D. L., Powers, J. G., Ransom, J. I., McCann, B. E., Oehler, M. W., Bruemmer, J. E., Galloway, N. L., Eckery, D. C., & Nett, T. M. (2018). Reimmunization increases contraceptive effectiveness of gonadotropin-releasing hormone vaccine (GonaCon-Equine) in free-ranging horses (Equus caballus): Limitations and side effects. *PLoS ONE*, *13*(7), e0201570. https://doi.org/10.1371/journal.pone.0201570

Boisson-Dupuis, S., Kong, X. F., Okada, S., Cypowyj, S., Puel, A., Abel, L., & Casanova, J. L. (2012). Inborn errors of human STAT1: Allelic heterogeneity governs the diversity of immunological and infectious phenotypes. *Current Opinion in Immunology*, *24*(4), 364–378. https://doi.org/10.1016/j.coi.2012.04.011

Brodie, A., Azaria, J. R., & Ofran, Y. (2016). How far from the SNP may the causative genes be? *Nucleic Acids Research*, *44*(13), 6046–6054. https://doi.org/10.1093/nar/gkw500

Brooks, S. A., Gabreski, N., Miller, D., Brisbin, A., & Brown, H. E. (2010). Whole-genome SNP association in the horse: identification of a deletion in myosin Va responsible for Lavender Foal Syndrome. *PLoS Genet*, *6*(4), 1000909. https://doi.org/10.1371/journal.pgen.1000909

Bucasas, K. L., Franco, L. M., Shaw, C. A., Bray, M. S., Wells, J. M., Niño, D., Arden, N., Quarles, J. M., Couch, R. B., & Belmont, J. W. (2011). Early patterns of gene expression correlate with the humoral immune response to influenza vaccination in humans. *Journal of Infectious Diseases*, *203*(7), 921–929. https://doi.org/10.1093/infdis/jiq156

Chen, Z., Boehnke, M., Wen, X., & Mukherjee, B. (2021). Revisiting the genome-wide significance threshold for common variant GWAS. *G3: Genes, Genomes, Genetics*, *11*(2). https://doi.org/10.1093/g3journal/jkaa056

Cooper, D. W., & Larsen, E. (2006). Immunocontraception of mammalian wildlife: ecological and immunogenetic issues. *Reproduction*, *132*(6), 821–828. https://doi.org/10.1530/REP-06-0037

Dang, J., Shan, S., Li, J., Zhao, H., Xin, Q., Liu, Y., Bian, X., & Liu, Q. (2014). Gene-gene interactions of IRF5, STAT4, IKZF1 and ETS1 in systemic lupus erythematosus. *Tissue Antigens*, *83*(6), 401–408. https://doi.org/10.1111/TAN.12349

Dubon, M. A. C., Pedrosa, V. B., Feitosa, F. L. B., Costa, R. B., de Camargo, G. M. F., Silva, M. R., & Pinto, L. F. B. (2021). Identification of novel candidate genes for age at first calving in Nellore cows using a SNP chip specifically developed for Bos taurus indicus cattle. *Theriogenology*, *173*, 156–162. https://doi.org/10.1016/J.THERIOGENOLOGY.2021.08.011

Ebrahimiyan, H., Rezaei, R., Mostafaei, S., Aslani, S., Goulielmos, G. N., Jamshidi, A., & Mahmoudi, M. (2018). Association study between STAT4 polymorphisms and susceptibility to systemic lupus erythematosus disease: A systematic review and meta-analysis. *Meta Gene*, *16*, 241–247. https://doi.org/10.1016/J.MGENE.2018.03.010

Eldridge, D. J., Ding, J., & Travers, S. K. (2020). Feral horse activity reduces environmental quality in ecosystems globally. *Biological Conservation*, *241*. https://doi.org/10.1016/J.BIOCON.2019.108367

Finno, C. J., Gianino, G., Perumbakkam, S., Williams, Z. J., Bordbari, M. H., Gardner, K. L., Burns, E., Peng, S., Durward-Akhurst, S. A., & Valberg, S. J. (2018). A missense mutation in MYH1 is associated with susceptibility to immune-mediated myositis in Quarter Horses. *Skeletal Muscle*, *8*(1), 1–13. https://doi.org/10.1186/S13395-018-0155-0/FIGURES/5

Gajbhiye, R., Fung, J. N., & Montgomery, G. W. (2018). Complex genetics of female fertility. *Npj Genomic Medicine*, *3*(29). https://doi.org/10.1038/s41525-018-0068-1

Garrott, R. A. (1995). Effective management of free-ranging ungulate populations using contraception. *Wildlife Society Bulletin*, *23*(3), 445–452. https://www.jstor.org/stable/3782953?seq=1&cid=pdf-

Hall, S. E., Upton, R., Mclaughlin, E. A., & Sutherland, J. (2017). Phosphoinositide 3-kinase/protein kinase B (PI3K/AKT) and Janus kinase/signal transducer and activator of transcription (JAK/STAT) follicular signalling is conserved in the mare ovary. *Reproduction, Fertility and Development*, *30*(4), 624–633. https://doi.org/10.1071/RD17024

Hampton, J. O., Arnemo, J. M., Barnsley, R., Cattet, M., Daoust, P.-Y., DeNicola, A. J., Eccles, G., Fletcher, D., Hinds, L. A., Hunt, R., Portas, T., Stokke, S., Warburton, B., Wimpenny, C., Hampton, J. O., Arnemo, J. M., Barnsley, R., Cattet, M., Daoust, P.-Y., … Wimpenny, C. (2021). Animal welfare testing for shooting and darting free-ranging wildlife: a review and recommendations. *Wildlife Research*, *48*(7), 577–589. https://doi.org/10.1071/WR20107

Hedl, M., Sun, R., & Abraham, C. (2020). Disease risk–associated genetic variants in STAT1 and STAT4 function in a complementary manner to increase pattern-recognition receptor–induced outcomes in human macrophages . *The Journal of Immunology*, *205*(5), 1406–1418. https://doi.org/10.4049/JIMMUNOL.1901112/-/DCSUPPLEMENTAL

Horvath, C. M. (2000). STAT proteins and transcriptional responses to extracellular signals. *Trends in Biochemical Sciences*, *25*(10), 496–502. https://doi.org/10.1016/S0968-0004(00)01624-8

Hu, S.-J., Wen, L.-L., Hu, X., Xian-Yong, Y., Cui, Y., Yang, S., & Zhang, X.-J. (2013). IKZF1: a critical role in the pathogenesis of systemic lupus erythematosus? *Modern Rheumatology*, *23*, 205–209. https://doi.org/10.1007/s10165-012-0706-x

Imsland, F., McGowan, K., Rubin, C. J., Henegar, C., Sundström, E., Berglund, J., Schwochow, D., Gustafson, U., Imsland, P., Lindblad-Toh, K., Lindgren, G., Mikko, S., Millon, L., Wade, C., Schubert, M., Orlando, L., Penedo, M. C. T., Barsh, G. S., & Andersson, L. (2016). Regulatory mutations in TBX3 disrupt asymmetric hair pigmentation that underlies Dun camouflage color in horses. *Nature Genetics*, *48*(2), 152–158. https://doi.org/10.1038/ng.3475

Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S.-Y., Freimer, N. B., Sabatti, C., & Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nature Publishing Group*, *42*(4). https://doi.org/10.1038/ng.548

Katze, M. G., He, Y., & Gale, M. (2002). Viruses and interferon: a fight for supremacy. *Nature Reviews Immunology*, *2*(9), 675-687. https://doi.org/10.1038/nri888

Kennedy, R. B., Ovsyannikova, I. G., Palese, P., & Poland, G. A. (2020). Current challenges in vaccinology. *Frontiers in Immunology*, *11*, 1181. https://doi.org/10.3389/fimmu.2020.01181

Khan, M. Z., Khan, A., Xiao, J., Ma, Y., Ma, J., Gao, J., & Cao, Z. (2020). Role of the JAK-STAT pathway in bovine mastitis and milk production. *Animals*, *10*(11), 2107. https://doi.org/10.3390/ANI10112107

Khatib, H., Huang, W., Mikheil, D., Schutzkus, V., & Monson, R. L. (2009). Effects of signal transducer and activator of transcription (STAT) genes STAT1 and STAT3 genotypic combinations on fertilization and embryonic survival rates in Holstein cattle. *Journal of Dairy Science*, *92*(12), 6186–6191. https://doi.org/10.3168/JDS.2009-2439/ATTACHMENT/55DFD172-75C2-45BC-95C5-988EA57DDF63/MMC1.PDF

Killian, G., Wagner, D., & Miller, L. (2005). Observations on the use of the GnRH vaccine GonaCon^{TM} in male white-tailed deer (Odocoileus virginianus). *Wildlife Damage Management Conferences -- Proceedings*, 133. https://digitalcommons.unl.edu/icwdm_wdmconfproc

Kirkpatrick, J. F., Lyda, R. O., & Frank, K. M. (2011). Contraceptive vaccines for wildlife: a review. *American Journal of Reproductive Immunology*, *66*, 40–50. https://doi.org/10.1111/j.1600-0897.2011.01003.x

Kirkpatrick, J. F., Rowan, A., Lamberski, N., Wallace, R., Frank, K., & Lyda, R. (2009). The practical side of immunocontraception: zona proteins and wildlife. *Journal of Reproductive Immunology*, *83*, 151–157. https://doi.org/10.1016/j.jri.2009.06.257

Korn, T., Bettelli, E., Oukka, M., & Kuchroo, V. K. (2009). IL-17 and Th17 Cells. *Annual Review of Immunology*, *27*, 485–517. https://doi.org/10.1146/annurev.immunol.021908.132710

Krause, S. K., Kelt, D. A., Gionfriddo, J. P., & Vuren, D. H. Van. (2014). Efficacy and health effects of a wildlife immunocontraceptive vaccine on fox squirrels. *The Journal of Wildlife Management*, *78*(1), 12–23. https://doi.org/10.1002/JWMG.635

Kuehn, H. S., Nunes-Santos, C. J., & Rosenzweig, S. D. (2021). Germline IKZF1 mutations and their impact on immunity: IKAROS-associated diseases and pathophysiology. *Expert Review of Clinical Immunology*, *17*(4), 407–416. https://doi.org/10.1080/1744666X.2021.1901582

Kulbrock, M., Lehner, S., Metzger, J., Ohnesorge, B., & Distl, O. (2013). A genome-wide association study identifies risk loci to equine recurrent uveitis in German Warmblood horses. *PLOS ONE*, *8*(8), e71619. https://doi.org/10.1371/JOURNAL.PONE.0071619

Kyogoku, C., Langefeld, C. D., Ortmann, W. A., Lee, A., Selby, S., Carlton, V. E. H., Chang, M., Ramos, P., Baechler, E. C., Batliwalla, F. M., Novitzke, J., Williams, A. H., Gillett, C., Rodine, P., Graham, R. R., Ardlie, K. G., Gaffney, P. M., Moser, K. L., Petri, M., … Behrens, T. W. (2004). Genetic association of the R620W polymorphism of protein tyrosine phosphatase PTPN22 with human SLE. *American Journal of Human Genettics, 75*(3), 504-507.

Laseca, N., Anaya, G., Peña, Z., Pirosanto, Y., Molina, A., & Demyda Peyrás, S. (2021). Impaired reproductive function in equines: from genetics to genomics. *Animals*, *11*(2), 393. https://doi.org/10.3390/ani11020393

Leach, R. J., O'neill, R. G., Fitzpatrick, J. L., Williams, J. L., Glass, E. J., & Varga, S. M. (2012). Quantitative trait loci associated with the immune response to a bovine respiratory syncytial virus vaccine. *PLoS ONE, 7*(3), e33526. https://doi.org/10.1371/journal.pone.0033526

Liu, L., Okada, S., Kong, X. F., Kreins, A. Y., Cypowyj, S., Abhyankar, A., Toubiana, J., Itan, Y., Audry, M., Nitschke, P., Masson, C., Toth, B., Flatot, J., Migaud, M., Chrabieh, M., Kochetkov, T., Bolze, A., Borghesi, A., Toulon, A., … Casanova, J. L. (2011). Gain-of-function human STAT1 mutations impair IL-17 immunity and underlie chronic mucocutaneous candidiasis. *Journal of Experimental Medicine*, *208*(18), 1635–1648. https://doi.org/10.1084/jem.20110958

Ma, L., Cole, J. B., Da, Y., & VanRaden, P. M. (2019). Symposium review: Genetics, genome-wide association study, and genetic improvement of dairy fertility traits. *Journal of Dairy Science*, *102*(4), 3735–3743. https://doi.org/10.3168/JDS.2018-15269

Makvandi-Nejad, S., Hoffman, G. E., Allen, J. J., Chu, E., Gu, E., Chandler, A. M., Loredo, A. I., Bellone, R. R., Mezey, J. G., Brooks, S. A., & Sutter, N. B. (2012). Four loci explain 83% of size variation in the horse. *PLOS ONE*, *7*(7), e39929. https://doi.org/10.1371/JOURNAL.PONE.0039929

Marees, A. T., de Kluiver, H., Stringer, S., Vorspan, F., Curis, E., Marie-Claire, C., & Derks, E. M. (2018). A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *International Journal of Methods in Psychiatric Research*, *27*(2). https://doi.org/10.1002/mpr.1608

Marques, P., Skorupskaite, K., Rozario, K. S., Anderson, R. A., & George, J. T. (2022). Physiology of GnRH and gonadotropin secretion. In K. Feingold, B. Anawalt, & A. Boyce (Eds.), *Endotext* (Updated 2022 Jan 5). MDText.com, Inc. https://www.ncbi.nlm.nih.gov/books/NBK279070/

Massei, G., Cowan, D. P., Coats, J., Bellamy, F., Quy, R., Pietravalle, S., Brash, M., & Miller, L. A. (2012). Long-term effects of immunocontraception on wild boar fertility, physiology and behaviour. *Wildlife Research*, *39*(5), 378–385. https://doi.org/10.1071/WR11196

McGeachy, M. J. (2013). Th17 memory cells: live long and proliferate. *Journal of Leukocyte Biology*, *94*(5), 921–926. https://doi.org/10.1189/JLB.0313113

Messmer, T. A. (2000). The emergence of human-wildlife conflict management: turning challenges into opportunities. *International Biodeterioration & Degradation*, *45*(3–4), 97–102. https://doi.org/10.1016/S0964-8305(00)00045-7

Miller, L. A., Gionfriddo, J. P., Fagerstone, K. A., Rhyan, J. C., & Killian, G. J. (2008). The single-shot GnRH immunocontraceptive vaccine (GonaCon^{TM}) in white-tailed deer: comparison of several GnRH preparations. *American Journal of Reproductive Immunology*, *60*(3), 214–223. https://doi.org/10.1111/J.1600-0897.2008.00616.X

Miller, L. A., Johns, B. E., & Killian, G. J. (2000). Immunocontraception of white-tailed deer with GnRH vaccine. *American Journal of Reproductive Immunology*, *44*(5), 266–274. https://doi.org/10.1111/J.8755-8920.2000.440503.X

Miller, L. A., Rhyan, J., & Killian, G. (2004). GonaConTM, a versatile GnRH contraceptive for a large variety of pest animal problems. *Proceedings of the 21st Vertebrate Pest Conference*, 269–273. https://digitalcommons.unl.edu/icwdm_usdanwrc/371

Mizoguchi, Y., & Okada, S. (2021). Inborn errors of STAT1 immunity. *Current Opinion in Immunology*, *72*, 59–64. https://doi.org/10.1016/j.coi.2021.02.009

Mohan, C., & Putterman, C. (2015). Genetics and pathogenesis of systemic lupus erythematosus and lupus nephritis. *Nature Reviews Nephrology 2015 11:6*, *11*(6), 329–341. https://doi.org/10.1038/nrneph.2015.33

Montgomery, G. W., Zondervan, K. T., & Nyholt, D. R. (2014). The future for genetic studies in reproduction. *Molecular Human Reproduction*, *20*(1), 1–14. https://doi.org/10.1093/MOLEHR/GAT058

Najjar, I., & Fagard, R. (2010). STAT1 and pathogens, not a friendly relationship. *Biochimie*, *92*(5), 425–444. https://doi.org/10.1016/J.BIOCHI.2010.02.009

National Research Council. (2013). *Using science to improve the BLM Wild Horse and Burro Program : a way forward*.

Naz, R. K., & Saver, A. E. (2016). Immunocontraception for animals: current status and future perspective. *American Journal of Reproductive Immunology*, *75*, 426–439. https://doi.org/10.1111/AJI.12431

Powers, J. G., Monello, R. J., Wild, M. A., Spraker, T. R., Gionfriddo, J. P., Nett, T. M., & Baker, D. L. (2014). Effects of GonaCon immunocontraceptive vaccine in free-ranging female Rocky Mountain elk (Cervus elaphus nelsoni). *Wildlife Society Bulletin*, *38*(3), 650–656. https://doi.org/https://doi.org/10.1002/wsb.434

Raudsepp, Terje. (2020). Genetics of equine reproductive diseases. *Veterinary Clinics of North America: Equine Practice*, *36*(2), 395–409.

Scepanovic, P., Alanio, C., Hammer, C., Hodel, F., Bergstedt, J., Patin, E., Thorball, C. W., Chaturvedi, N., Charbit, B., Abel, L., Quintana-Murci, L., Duffy, D., Albert, M. L., Fellay, J., Alcover, A., Aschard, H., Bousso, P., Bruhns, P., Cumano, A., … Toubert, A. (2018). Human genetic variants and age are the strongest predictors of humoral immune responses

to common pathogens and vaccines. *Genome Medicine*, *10*(1), 1–13. https://doi.org/10.1186/S13073-018-0568-8/TABLES/3

Stanford, S. M., & Bottini, N. (2014). PTPN22: the archetypal non-HLA autoimmunity gene. *Nature Reviews Rheumatology*, *10*(10), 602-611. https://doi.org/10.1038/nrrheum.2014.109

Stark, G. R., & Darnell, J. E. (2012). The JAK-STAT pathway at twenty. *Immunity*, *36*(4), 503-514. https://doi.org/10.1016/j.immuni.2012.03.013

Tizaoui, K., Terrazzino, S., Cargnin, S., Lee, H., Gauckler, P., Li, H., Shin, J. Il, & Kronbichler, A. (2021). The role of PTPN22 in the pathogenesis of autoimmune diseases: A comprehensive review. *Seminars in Arthritis and Rheumatism*, *51*(3), 513-522. https://doi.org/10.1016/j.semarthrit.2021.03.004

Trowsdale, J., & Knight, J. C. (2013). Major histocompatibility complex genomics and human disease. *Annual Review of Genomics and Human Genetics*, *14*, 301–323. https://doi.org/10.1146/annurev-genom-091212-153455

Vansandt, L. M., Kutzler, M. A., Fischer, A. E., Morris, K. N., & Swanson, W. F. (2017). Safety and effectiveness of a single and repeat intramuscular injection of a GnRH vaccine (GonaCon^{TM}) in adult female domestic cats. *Reproduction in Domestic Animals*, *52*, 348–353. https://doi.org/10.1111/rda.12853

Vilhjalmsson, B. J. (2012). *mixmogam*. https://github.com/bvilhjal/mixmogam

Viļuma, A., Mikko, S., Hahn, D., Skow, L., Andersson, G., & Bergström, T. F. (2017). Genomic structure of the horse major histocompatibility complex class II region resolved using PacBio long-read sequencing technology OPEN. *Scientific Reports*, *7*, 45518. https://doi.org/10.1038/srep45518

Wall, G. V., Wright, I. M., Barnardo, C., Erasmus, B. J., van Staden, V., & Potgieter, A. C. (2021). African horse sickness virus NS4 protein is an important virulence factor and interferes with JAK-STAT signaling during viral infection. *Virus Research*, *298*, 198407. https://doi.org/10.1016/J.VIRUSRES.2021.198407

Wu, S., & Wolfe, A. (2012). Signaling of cytokines is important in regulation of GnRH neurons. *Molucular Neurobiology*, *45*, 119–125. https://doi.org/10.1007/s12035-011-8224-y

Yamamoto, K., Suzuki, A., & Guerrini, M. M. (2021). Functional genetics for studying the human immune system. *International Immunology*, *33*(12), 647–651. https://doi.org/10.1093/intimm/dxab046

Yang, C., Mai, H., Peng, J., Zhou, B., Hou, J., & Jiang, D. (2020). STAT4: an immunoregulator contributing to diverse human diseases. *International Journal of Biological Sciences*, *16*(9), 1575. https://doi.org/10.7150/IJBS.41852

Yoder, C. A., & Miller, L. A. (2010). Effect of GonaCon™ vaccine on black-tailed prairie dogs: Immune response and health effects. *Vaccine*, *29*(2), 233–239. https://doi.org/10.1016/j.vaccine.2010.10.055

CHAPTER IV

IMPLICATIONS

Feral horse management is controversial with emotionally invested stakeholders on all sides of the argument, and has no one correct approach (Notzke, 2013; Scasta, 2019). The best strategy is to aim for an impartial approach informed by scientific evidence, with a willingness to try new policy. This work is meant to provide scientific data which can be referenced when making management decisions about the TRNP horse herd.

My SNP analyses show that the TRNP herd has experienced inbreeding and differentiation from other breeds, likely due to genetic drift, bottleneck events, and limited gene flow. Management should consider the effect of continued isolation on the herd's genetic diversity. Introduction of new individuals into the gene pool can increase diversity. Depending on long term management objectives, individuals for introduction could be chosen from a more genetically distant population to maximize variation, or from a more similar population to maintain the "type". My population genomic analyses show that the TRNP herd is an admixed population with draft breed influence, but no clear ancestral association with any of the breeds included in the dataset. The claim that the TRNP herd originated from Spanish type horses was not substantiated. Including samples from other feral herds across the country may identify more relationships, potentially suggesting a similar source from which to select animals for introduction.

To prevent further loss of genetic diversity, Cothran (1992) recommended removal of young individuals during capture operations, rather than adults, to retain the present genetic variation. This effectively increases the generation time, and can be similarly produced by the use of contraception to reduce population growth rate (Gross, 2000).

While immunocontraceptives such as GonaCon-Equine have proven effective in reducing fertility, we see individual variation in the duration of infertility. My genome-wide association study found a genetic association with the response of the TRNP mares to GonaCon vaccination. The variation observed in contraceptive efficacy is likely due to immune system function in response to the vaccine. If there is a genetic component to GonaCon treatment response, selection can act upon it. If only those individuals with less responsive immune systems reproduce, over time the population may develop genetically based non-response to immunocontraception, rendering that means of population control ineffective. Reduced immune system function also brings up concerns over the ability to respond to pathogens and infectious diseases. Immunocontraception should be used judiciously, and likely in combination with other management practices, considering both the benefits of slowing the population growth rate to retain genetic diversity and concerns over unintended selection altering allele frequencies.

CHAPTER IV REFERENCES

Cothran, E. G. (1992). *Genetic marker analysis of the Theodore Roosevelt National Park feral horse herd*.

Gross, J. E. (2000). A dynamic simulation model for evaluating effects of removal and contraception on genetic variation and demography of Pryor Mountain wild horses. *Biological Conservation*, *96*(3), 319–330. https://doi.org/10.1016/S0006-3207(00)00078-1

Notzke, C. (2013). An exploration into political ecology and nonhuman agency: The case of the wild horse in western Canada. *The Canadian Geographer / Le Géographe Canadien*, *57*(4), 389–412. https://doi.org/10.1111/CAG.12028

Scasta, J. D. (2019). Why are humans so emotional about feral horses? A spatiotemporal review of the psycho-ecological evidence with global implications. *Geoforum*, *103*, 171–175.