Norwegian University
of Life Sciences

Master's Thesis 2023 60 ECTS

Faculty of Biosciences,
Department of Animal and Aquacultural Sciences

Unraveling the Genetic Mysteries of the Norwegian Fjord-horse:

Identifying Harmful Haplotypes for Improved Breeding Strategies

Nathalie Almaas Smogeli
Genome Science

Supervisors and Evaluation Committee

Main Supervisor:	Professor Peer Berg, NMBU
Evaluation Committee:	Researcher Arne Gjuvsland, Norsvin/Geno

Acknowledgments

This is my thesis marking the end of a five-year study at the Norwegian University of Life Sciences. I would like to express my appreciation for the time as a student, for the exciting courses, and for engaging professors with a challenging syllabus. My years here have sparked my interest in further engagement in molecular genetics, bioinformatics, and animal science.

I want to express my gratitude to Professor Peer Berg for the guidance you have shown me. I humbly thank you for the experience of writing this thesis, the opportunity to learn by doing, and the questions that challenged me to think critically and deeply about my research. Being able to write about horses which holds a special place in my heart, as well as my newfound passion for research, has been a unique experience for which I am grateful to the project FeNomen. Further, I am thankful to Cathrine Brekke for the knowledge and expertise shared with me in the use of software. I also want to thank you for your advice to me in our first meeting telling me to enjoy the writing, it has motivated me throughout the work with the thesis.

I also want to give a special thanks to Dorota Monika Jaskuła for advice and suggestions regarding my bioinformatic struggles. I also want to thank you for the endless moral support, answering my frustration with encouragement.

I would be remiss in not mentioning Mother and Father for your belief in me and not at least the great support you show me. I would like to recognize my family and friends for the support you have shown me and the help you have offered. Lastly, I would like to mention Minnie, who has been the most patient emotional support during long days with reading and writing.

"Head like a brisling, neck like a spinning wheel, body like a turnip, and limbs as steel springs "
- Description of the Fjord-horse (Fjord Horse International, 2014)

“The eyes should be like the mountain lakes on a midsummer evening, big and bright.

A bold bearing of the neck like a lad from the mountains on his way to his beloved.

Well-defined withers like the contours of the mountains set against an evening sky.

The temperament as lively as a waterfall in spring and still good-natured.”

– Norwegian description of the Fjord-horse (Statement from Norwegian Fjord-horse Association on an international conference, 1996, as cited by Huggett, 2009).

Table of Contents

Abbreviations and Definitions

A1A1	A homozygote genotype of the minor allele (A1).
A1A2	A heterozygote genotype of one minor and one major allele (A1 and A2).
A2A2	A homozygote genotype of two major alleles (A2).
AI	Artificial Insemination
BLOCK	A continuous set of genetic variants that tend to be inherited together.
BP	Base pair; the unit of length in DNA.
BP1	A location on a chromosome, defined by a specific base pair position. In this study, it is used as a start location of a haploblock.
BP2	Another location on a chromosome, defined by a specific base pair position. In this study, it is used as the end location of a haploblock.
DNA	Deoxyribonucleic acid; the molecule that carries genetic information in most living organisms.
E. HET	Expected heterozygosity.
E. HOM	Expected homozygosity.
FR	Foaling rate; the percentage of mares that produce one foal after one breeding season. See the equation eq. (1), in chapter 2.2.1.1.
GWAS	Genome-wide association study; a study in which the genomes of many individuals are compared to identify genetic variants associated with a particular trait or disease.
HWE	Hardy-Weinberg equilibrium; a principle of population genetics stating that, in the absence of evolutionary forces, the frequencies of alleles and genotypes in a population will remain constant over time.
IBD	Identical by descent; a term used to describe genetic variants that are identical in sequence because they were inherited from a common ancestor.
Locus	(plural loci) the position that a gene occupies in a chromosome or within a segment of Genomic DNA (King et al., 2013)
LOD	Logarithm-of-Odds
Mucosa	A mucous tissue lining various tubular structures consisting of epithelium, lamina propria, and in the digestive tract, a layer of smooth muscle (*Mucosa*, 2012).
NCBI	National Center for Biotechnology Information; a public repository of genetic and genomic data.
NENT	National Committee for Research Ethics in Science and Technology (‘Den nasjonale forskningsetiske komité for naturvitenskap og teknologi’)

NESH	The National Research Ethics Committee (‘De Nasjonale Forskningsetiske komiteene’)
NHS	Norwegian Equine Centre (‘Norsk Hestesenter’)
O.HOM	Observed homozygosity.
OHET	Observed heterozygosity.
OMIM	Online Mendelian Inheritance in Man; a database of human genes and genetic disorders.
POS	Position; a location on a chromosome defined by a specific base pair position.
QC	Quality control; a process used to ensure that data is accurate and reliable.
QTL	Quantitative trait locus; a genomic region that is associated with a particular quantitative trait.
RNA	Ribonucleic acid; a molecule that plays a central role in the synthesis of proteins from genetic information.
ROH	Runs of homozygosity; continuous stretches of the genome in which an individual is homozygous for all variants.
RPKM	Reads per kilobase of transcript per million mapped reads; a measure of gene expression in RNA sequencing experiments.
SNP	Single nucleotide polymorphism; small variations in DNA sequence in which, at any given position, a single nucleotide is replaced by one of the other three nucleotides (King et al., 2013).
VEP	Variant Effect Predictor; a software tool used to predict the functional effects of genetic variants.

Abstract

The fertility in breeding-animals is especially critical in small populations in which there is a high degree of inbreeding. The Norwegian Fjord-horse population consists of few founder animals and is observed to underperform in foaling rate compared to several other horse breeds. The project is based on HD SNP genotypes from approximately 330 Fjord-horses. In order to identify potential recessive lethal alleles, the haplotype homozygote deficiencies were identified. Significant haplotypes with one or fewer homozygotes observed were characterized as the candidate loci. Further, the candidate loci were aligned with functional knowledge of equine, human, and mouse genomes. Candidate genes identified by this study were DEFA5L, DEFA22, KIA1109, and DPF2.

Norsk sammendrag

Fertiliteten til avlsdyrene er særdeles kritisk i små populasjoner med høy innavl. Fjordhest- populasjonen i Norge består av få stamfedre, og er blitt registrert til å ha en lavere føllings-rate enn flere andre hesteraser. Denne analysen er basert på høy-tetthets SNP genotyper fra ca. 330 Fjordhester. For å identifisere mulige resessive dødelige alleler undersøkes genomet for homozygote avvik i haplotyper. Signifikante avvik i haplotyper med én eller mindre observerte homozygoter ble karakterisert som kandidat-loci. Deretter, vil kandidat-loci analyseres i forhold til funksjonell kunnskap om genomet til hest, menneske og mus. Kandidatgenene som dette studiet kommer frem til er DEAF5L, DEFA22, KIA1109, og DPF2.

1 Introduction

1.1 The Structure of the Thesis

The thesis is divided into an introductory chapter introducing the reader, the problem and relevance, the possible implications of the results, and a review of related literature. Next is the theoretical chapter giving an overview of the Fjord-horse as a breed before describing the reproductive process in horses and the detection of genes. The thesis continues with a chapter describing the methods and considerations, followed by results from the study. Further, there is a discussion of the results and a final comment on the study. The thesis ends by listing the references, followed by the appendices. All tables are numerated after the chapter, then the order they reside in, e.g., the first table of chapter one would be Table 1.1, while the fifth table of chapter four would be Table 4.5.

The theory chapter includes theoretical background and knowledge about detecting candidate genes based on family data. However, the dataset does not include family data, so this analysis cannot be done on the current dataset. Within the time-scope of this thesis, only a limited amount of data was available. The data used in this study was approximately 330 unrelated animals, while the planned study population was 1200 parent-offspring trios and parent duos without offspring. A further study with a larger dataset will make the theoretical chapter more relevant, based on this study as part of the gene detection method.

1.2 Background

The Food and Agriculture Organization of the United Nations (FAO) has classified three horse breeds native to Norway as "critically at risk" and included them on their endangered species list (FAO, 2023a, 2023b, 2023c). The Global Plan of Action for Animal Genetic Resources (GPA) is an internationally recognized framework that aims to conserve the genetic materials of the animal breeds that member countries possess to prevent them from going extinct (Hoffmann & Scherf, 2010). Since Norway is the country of origin for the Dole, Fjord, Nordland/Lyngen-horse, and Norwegian trotter, it is responsible for preserving these breeds and maintaining sustainable genetic diversity.

According to Norsk Hestesenter (2021b), there are approximately 2100 Dole-horses, 2500 Fjord-horses, and 1500 Nordland/Lyngen-horses in the general breeding populations of these three Norwegian national breeds. However, only 12% of the mares are covered yearly (Norsk Hestesenter, 2021b), and the number of coverings is decreasing. In addition, the yearly number of foals born in the Dole, Fjord, and Nordland/Lyngen populations is only 120-130 (Furre & Norsk Hestesenter, 2017), which is far below the recommended number of more than 200 foals born annually (Olsen & Klemetsdal, 2010).

Reproductive performance in horses can be measured using a foaling or cyclic foaling rate. According to Norsk Hestesenter (2020), the foaling rates for Dole, Fjord, and Nordland/Lyngen-horses have varied between 50 to 62%, 50 to 65%, and 58 to 76%, respectively, in the last two decades. In Norway, it is customary to mate these horses using natural covering. However, for other breeds that use natural covering, the average foaling rates are much higher, such as 68% in Icelandic horses, 68% in Shetland ponies, and 82% in Thoroughbreds (Hanlon et al., 2012; Morel & Gunnarsson, 2000; van Buiten et al., 1998). These studies and several others have highlighted the impact of management factors, mare age, reproductive status, and inbreeding on the reproductive performance of other horse breeds.

The management of breeds through selection can result in the accumulation of homozygosity, which is undesirable (Maltecca et al., 2020). The Dole, Fjord, and Nordland/Lyngen-horses, the three national breeds of Norway, are facing inbreeding challenges. The Nordland/Lyngen population has maintained a

relatively stable average inbreeding degree of around 12% in the last four decades. At the same time, the Fjord- and Dole-horse breeds have experienced an increase in the inbreeding degree of 2% and 6%, respectively (Norsk Hestesenter, 2020). Studies have shown that increasing inbreeding coefficients can decrease foaling rates in trotter mares, Przewalski-horses, Standardbred Trotters, and Finn-horse (Bouman, 1977; Fomin et al., 1982; Klemetsdal & Johnson, 1989; Sairanen et al., 2009).

Inbreeding occurs when closely related animals mate, leading to inbreeding depression, which reduces the survival and fertility of potential offspring. This depression is primarily caused by recessive deleterious mutations in the population (Charlesworth & Willis, 2009). Mutations in genes essential for reproduction can negatively impact the fertilization, embryonic development, and survival of potential foals. While lethal mutations are generally eliminated by natural selection, recessive lethal mutations may persist in small populations due to genetic drift, leading to a higher frequency in related animals. This accumulation of detrimental alleles is typical in populations with intense selection management, population bottlenecks, or breed formation events. Todd, Thomson, et al. (2020) have identified a candidate recessive lethal haplotype in the Thoroughbred horse population. Recessive lethal mutations due to inbreeding may cause reproductive failures in horses, adversely affecting population health.

Identifying recessive lethal mutations can be challenging, particularly in small populations. One method to detect them is by homozygosity mapping. This method involves searching for the depletion or absence of homozygotes. This search can be done by examining every Single Nucleotide Polymorphism (SNP) or haplotype. The latter is preferred due to its expected linkage to causative variants and a lower rate of false positive discoveries (VanRaden et al., 2011). Haplotype homozygosity mapping has been successful in identifying recessive lethal alleles and loci in various livestock species (Casellas et al., 2017; 2012; Derks et al., 2017, 2019; Derks, Megens et al., 2018; Fritz et al., 2013; Hoff et al., 2017; VanRaden et al., 2011;).

There is a limited understanding of genetic inbreeding and recessive lethal loci/haplotypes in Norwegian breeds. Therefore, this study aims to enhance knowledge about reproductive health in the Norwegian Fjord-horse by identifying and validating genetic inbreeding and lethal recessive mutations leading to embryonic death. Identifying such mutations is crucial for making informed breeding decisions, improving health-related traits, and preventing them from reaching high frequencies. This study aims to increase fertility in small populations, such as the Norwegian national horse breeds, by identifying mutations impacting population health. Ultimately, the implementation of mutation-focused breeding management for Norwegian horse populations will lead to the development of sustainable populations.

1.3 Problem Statement

Fertility is affected by many factors and is generally accepted to be affected by genetics. Recently, genes related to fertility and embryonic death have been discovered in several livestock species, including horses. A concern for the Norwegian Fjord breeding community is the low fertility of the breed, the small population, with a downward trend in breeding activity. The Norwegian Fjord-horse population requires help and an effort to preserve the current diversity if the breed is to survive.
The aim is to identify potentially lethal recessive mutations causing embryonic death and subfertility in animals of the population.

The genomic investigation will be on SNP genotypes of the population, processed in statistical software programs to detect depletion or a complete lack of homozygote SNPs or haplotypes. Lastly, the candidate loci will be characterized against publicly available genome sequences or variant repositories

to identify candidate genes in the loci. The thesis will give a genetic foundation for breeding recommendations as a part of the project FeNomen. FeNomen will further communicate the project results to horse owners, breeders, and breed organizations contributing to sustainable populations of Norwegian national horses.

1.4 The Relevance of the Study

This study contributes to the discovery of genetic causes for fertility-related problems in horses. The social and practical relevance of the study is for the horse sector that may get the opportunity to correct breeding to avoid matings of carriers of candidate genes to improve fertility and, therefore, the breed's economic gain. Low foaling rates in some national breeds have been problematic for several years. Additionally, the low number of mares presented for breeding each year is considered a substantial threat to all Norwegian national breeds. Currently, the demand for adult horses is greater than the availability. This low number of foals born does not only affect the horse sector. It is also a national concern, as maintaining these breeds is vital to Norwegian national heritage. Norway has accepted the responsibility for these breeds by ratifying the Convention on Biological Diversity (United Nations Conference on Environment and Development, 1992). It is further committed to the Global plan of action for animal genetic resources (FAO, 2011). The increase in fertility and, therefore, the economic gain of keeping the Norwegian Fjord-horse may also increase national interest in keeping and breeding the Norwegian fjord.

1.5 Impact and Implications

1.5.1 Research Potential and Competence Building

Maintaining the local population is a national responsibility, and low fertility and inbreeding are dangerous to the population size if not managed. This study has the potential to help increase the population census size by more foals being born. The increase in population size is required for sustainable maintenance of the population, as stated by the strategy for the national horse breeds (Norsk Hestesenter, 2021a).

Infertility or subfertility in horse breeding leads to problems with profitability. Horses are expensive to keep and must regularly produce foals to maintain profitability. The unfortunate effects of inbreeding are avoided by selecting breeding-animals. The heritability of fertility and prediction of breeding values are essential to increase fertility in the population. Further, it is known that managemental and environmental factors also play a critical role in fertility. Although the Norwegian national breeds have similar management systems, there are differences in the foaling rate between the Nordland/Lyngen and the Dole and Fjord. This analysis will identify regions in the genome harboring potential recessive lethal variants.

Further, the identified candidate genes could be used to revise Fjord breeding schemes. Also, this study may guide studies in other breeds with fertility issues and transfer knowledge to other breeding organizations, breeders, and breeding advisors.

1.5.2 The Direct Benefit

The reproductive performance of horses is a complex trait, with many factors that may be decisive. The confirmation or elimination of genetics as a major cause of the low fertility in the Fjord may be essential

to explain infertility or subfertility in horses. Indications of genetic factors having significant effects on fertility would need consideration when selecting horses for breeding in the future. For novel recessive variants identified, the potential would be to incorporate them into breeding schemes with genotyping or screening of the horses selected for breeding. Incorporating into breeding schemes would avoid disadvantageous combinations of recessive lethal variants. The increase in the foaling rate in response to the identification of recessive lethal variants would depend heavily on the variants' prevalence in the population. However, if the result shows that genetic causes are less likely as a major cause of low fertility in the Fjord-horse, other aspects of fertility need investigation. In this case, the next step would be to investigate fertility managemental and environmental factors. The study will also increase the attention to the breed's status and may contribute to more mares being covered.

1.5.3 Risk assessment

Risks connected to this study is obtaining sufficient samples for DNA extraction. However, owners of endangered breeds are generally committed to the breed's conservation. Further, all the breeding associations are involved in the FeNomen project and motivate initiatives of the owners and breeders of the breed. Owner confidentiality of data used is ensured through the de-identification of records.

1.6 Literature Review

1.6.1 Quality Control and Methodology

The detection of recessive lethal traits depends on the genomic data of a population and cannot be done with only pedigree and phenotypic data. More recently, a method for detecting this without the phenotypic data has been developed, although it requires a large set of genotyped animals (VanRaden et al., 2011). This method identifies defects by looking at common haplotypes rarely occurring in a homozygous state. The idea is that the non-homozygous occurrence does not happen by chance in a large population.

VanRaden et al. (2011) obtained haplotypes with Fortran (v.2) findhap.f90, with haplotype sizes from 600-75 markers in length. When detecting recessive haplotypes concerning fertility, the genotype data was ~50K SNPs chips used on ~65’000 cattle of different breeds. The most frequent haplotypes within the breeds, not observed in homozygous states, were examined before looking at the general population. The haplotypes were examined to see if they had an inheritance similar to a genetic model. Further, the expected homozygous observations were calculated with two different methods. The phenotype was examined for conception and stillbirth rates. Lastly, the candidate haplotypes were validated by comparing carrier versus carrier and non-carrier matings.

In a study by Sahana et al. (2013), the quality control filtered out minor allele frequencies (<0.05) and low call rates (>0.65). Further, the default parameters in the software Beagle (v. 3.3) of shifting and scaling were used to impute missing markers and determine the phase. Afterward, the pedigree or haplotypes of potential carriers are utilized to verify the recessive lethal by observing the delivery of diseased animals or reduced fertility.

Pausch et al. (2015) detected four haplotypes affecting reproductive and rearing success, performed on ~25.500 genotyped Fleckvieh cattle. The quality control was on SNPs and animals with call rates threshold values at 95%, minor allele frequencies threshold below 2%, and significant deviation (P<10-6) from Hardy-Weinberg equilibrium. Animals deviating from normal when comparing pedigree and

genomic relationships and SNPs with more than 500 mendelian errors were excluded. Haplotypes were obtained with a window size of 0.75-10 Mb shifted in half window size across the genome. Only the haplotypes with a frequency above 2% were further characterized.

One haplotype and two alleles were discovered as potentially lethal in a study by Todd, Thomson, et al. (2020). The genotype data used was 156 Australian Thoroughbred horses genotyped with a 70 K Chip or 670 K Chip. They used Plink (v.1.9) to scan for deviations to the Hardy-Weinberg equilibrium with an absence of homozygotes for one allele, adjusted the p-values with false discovery rate correction in R ("qvalue" package), and searched for adjacent SNPs fulfilling these criteria. The candidate SNPs were also investigated in data from Japanese Thoroughbreds, which was added to the sample. The frequencies of the candidate SNPs were further characterized from genotype data in Swedish Warmblood and Norwegian-Swedish Coldblooded Trotters and several other breeds to obtain SNP frequencies. Detected calling errors of the candidate SNPs were recalled in R with an expectation-maximization algorithm.

1.6.2 Defining Expected Homozygosity

VanRaden et al. (2011) calculated the expected number of homozygotes with an actual mating pattern or assuming random mating. For random mating, the expected homozygous animals were calculated as the number of genotyped animals, divided by four, and multiplied by the square of the carrier frequency. However, this method may wrongly estimate if the inbreeding and allele frequency changes are ignored over an extended period. The actual mating pattern was calculated as the number of carrier sire and carrier maternal grandsire coverings divided by four, assuming allele frequencies of maternal grandsire equal to maternal grand dams. The expectance is commonly calculated as the number of homozygotes under the Hardy-Weinberg equilibrium given an allele frequency in the population (Howard et al., 2017; Todd, Thomson et al., 2020).

There are several reasons for observing homozygote deficit. Reducing the genotype fitness causes a deficit in a genotype relative to the population allele frequency. The reduction of fitness means that animals with a specific genotype are less likely to be viable than the other genotypes. Because of reduced fitness in homozygote animals, we will observe fewer or no homozygotes for the variant. The degree of the deficit may indicate the severity of the variant. This degree of severity means that a variant in a specific position in a gene may be lethal before blastocyst development. Also, another variant in that same position in the gene may be harmful but not lethal. In this last case, we will observe homozygotes of the variant, but there will be less than expected by the Hardy-Weinberg equilibrium. The severeness and at what development stage it affects will also determine the variant's name, i.e., embryonic lethal, detrimental allele.

1.6.3 Known Equine Recessive Lethal Disorders

This chapter presents a small summary of recessive disorders based on Griffin (2014a), focusing on the disorders observed in foals and young horses. In Appendices, a table shows a few more known recessive lethal diseases (Table 8.10), their location in the genome, what gene they are associated with, and the causative mutation for the disease.

Hereditary Equine Regional Dermal Asthenia (HERDA) is generally present only in the American Quarter horse. The disease can be observed when the horse is around two to three years old. The symptom of the disease is the hyper-extensible skin which will start falling off where there is pressure, e.g., the

saddle area.
Lavender Foal Disease is generally only present in the Arabian horse. The newborn foals have a lavender tint to their coat, are weak, and die shortly after birth.
Severe Combined Immunodeficiency (SCID) is generally only found in the Arabian breed. The affected foal is born seemingly ordinary but will soon present symptoms such as elevated temperature, respiratory stress, and diarrhea. Heterozygous (Carrier) horses have increased incidents of sarcoids. Homozygous foals rarely live longer than six months due to poor immune systems.
Cerebellar Abiotrophy (CA) is a neurological condition in the Arabian horse. The symptoms are a lack of balance in foals and often a danger to themselves and others as they fall frequently.
Glycogen Branching Enzyme Deficiency (GBED) is generally found only in the American Quarter and Paint-horse. The disease prevents the foal from storing sugars, resulting in the foal's death.
Junctional Epidermolysis Bullosa (JEB) is generally only found in Saddlebred, Belgians, and other draft horse breeds. The disease affects the skin making it too painful for foals to survive humanely. Heterozygote animals are unaffected by disease symptoms.
Overo Lethal White Syndrome (OLWS) is caused by a mutation causing a particular white coat pattern (frame overo) in horses, most frequent in the Paint horse breed. The disease presents itself in a foal with an underdeveloped digestive system when homozygous. A lack of nerves in the intestines causes the foal not to defecate, getting colic and eventually dying within five to six days.
Multiple Congenital Ocular Anomaly Syndrome (MCOAS) is generally only in the Rocky Mountain horse breed. The mutation is in the same gene producing a popular coat color for the breed. The heterozygous animals suffer less severe issues, but the homozygous animals suffer a wide range of ocular problems, e.g., megaloglobus, iridociliary cysts, and cataracts.
Fragile Foal Syndrome is found in all horse breeds. The disease affects the connective tissue, giving the foal slack joints and thin skin barely connected to the body. Heterozygote carrier animals are unaffected, but the homozygotes suffer greatly. The skin is easily damaged, giving lacerations, hematomas, and seromas. Affected foals are humanely put down to reduce their suffering.

1.6.4 Candidate Genes Related to Fertility in Horses

There are several candidate genes related to equine fertility. The study by Laseca et al. (2021) reviews the literature and summarizes the candidate genes in equines. The genes are in this study divided into genes related to gonad and sexual development, oocyte development, gametic interaction and embryo development, sperm quality, and fertility traits. Shortly after, Laseca et al. (2022) proposed new candidate genes. The tables from Laseca et al. (2021) and the candidate genes from Laseca et al. (2022) are included in the Appendices as Table 8.5-8.9.

1.6.5 Concerns Regarding the National Norwegian Fjord-horse

Three National Norwegian horse breeds are listed as endangered breeds by FAO. The Fjord-horse was listed as critically endangered by FAO (2023a, 2023b, 2023c), and with decreasing breeding activity, the risk of rapid accumulation of inbreeding is a concern. The general breeding population of the Norwegian Fjord in 2020 was ~2500 horses, and ~12% of the breeding population was covered last year (Norsk Hestesenter, 2021b). There has been a decline in the number of active stallions and mares and foals registered since 2000 (Rochmann, 2016).

There have been a handful of papers voicing concern about the breeding of the Norwegian Fjord-horse. The studies focus on the inbreeding coefficient in the population, the effective population size, and the

number of foals born per year. The inbreeding level of the Norwegian Fjord was estimated in a retrospective study in 1967 (Fjord Horse International, 2014). The estimation started in 1910 when the Fjord had an inbreeding coefficient of 0.010, then in 1930, it was at 0.47, and in 1957 revealed an inbreeding coefficient of 0.77 (Fjord Horse International, 2014). Later, the increase in the inbreeding rate of the Fjord was found acceptable according to international guidelines of 0.5-1% (Woolliams et al., 2015) and is the result of the introduction of foreign stallions in the 1980s and the use of young stallions (Johnsen & Seilen, 2009 as cited by Høiseth, 2017).

When looking at the effective population size of the Fjord-horse population, the breed is just above the minimum recommended number (Ne = 100) of unrelated breeding stock with Ne = 107 (Fjord Horse International, 2014). In a study by Olsen et al. (2020), the estimated effective population of the Norwegian Fjord-horse was Ne=~63-71, calculated from the pairwise coancestry and shared genomic segments. Further, with recommendations to the number of foals born per year, the Fjord-horse is below the minimum of 200 foals per year. The Fjord-horse foaling numbers range between 120-200 born foals over the last few years, with ~125 (2017), 170 (2019), 154 (2020), and 189 (2021) (Fjord Horse International, 2014). Olsen & Klemetsdal (2010) conducted a simulation study on small horse populations, concluding the need for using a considerable fraction of young stallions (3 years) in breeding for the Fjord-horse to maintain genetically sound with 200 foals registered per year. This number of foals registered annually would further maintain an effective population size of 100 animals (Rochmann, 2016).

2 Theoretical Foundations

2.1 The Fjord-horse

2.1.1 History

In the words of Fjord-Horse International (FHI) (2014), the Norwegian Fjord-horse is said to be one of the world's oldest and purest breeds. From horses of long travels and raids, the Fjord-horse became a farm horse for Norwegian farmers in the mountains west of Norway (Huggett, 2009). From 1840 to 1850, organized breeding in Fjords began formally (Fjord Horse International, 2014). In 1843 a government-funded stud was founded in Dovre to improve Norwegian breeding. Further, Dahle (2006, as cited by Olsen et al., 2020) points out that the stud failed due to poor management. Rochman (2016) adds that there was little knowledge about the genetic background of breeding. Rochman (2016) also states that the government purchased stallions to be used by the breeders but that no pure Norwegian Fjord was appropriate to improve the breed. The Fjord was nationally crossbred with the Norwegian Dole horse to make the Fjord bigger and sturdier for farming purposes. A few generations later, the crossbreeds show poor temperaments and unattractive coloring (Huggett, 2009).

Further, it was decided to make the breed purebred again in 1907 after the counties were dissatisfied with the crossbred Fjord-horses. A purebred stallion, Njål, was brought back into the breeding program. He is considered the father of modern Fjord-horses since all of the Fjord-horses today will have Njål in their pedigrees (Huggett, 2009) if you dig deep enough into their pedigrees. Nestaas (2010, as cited by Olsen et al., 2020) states that the mated Fjord-horses in 1921 had a 64% decline from the previous three years. In 1941 a law was passed restricting the use of stallions needing to be licensed if covering other breeders' mares' remaining practice until 2000 (Rochmann, 2016).

Mechanization has made the Fjord-horse redundant as a workforce and faced intense competition from other more specialized horse breeds. In the late 1960s, the number of horses in Norway was at its lowest (Dahle, 2010, as cited by Olsen, 2011). In 1992 the Rio Convention on Biodiversity was held, and Norway claimed the geographical origin of the Fjord-horse. The Fjord has continued to impress with a capability to adapt to different disciplines. It is not a specialized breed, and its versatility is an important trait. According to the Norwegian Equine Centre (Norsk Hestesenter, n.d.), the modern Fjord-horse can be seen in almost any discipline, including dressage, trail riding, jumping, eventing, light draft work, games, and therapeutic riding. They can also be seen in high-level competitions in dressage, driving, and reining (Norsk Hestesenter, n.d.).

2.1.2 Breeding

FHI (Fjord Horse International, 2014) states that before a breeding plan is actualized, the Norwegian Equine Centre (NHS) board must approve it. The breeding goals are managed towards a genetic improvement of traits. This means considering the market, economy, rules and regulations, genetic diversity, inbreeding, breed type, performance, health, temperament, fertility, species, and breed limitations. FHI continues that consideration is also given to the breeding population size within the breed when developing a breeding plan. In addition to these considerations, breeding and locations must be economical and practically feasible for horse owners. Regarding this, international breeds follow international breeding plans and generally do not have the same issue of inbreeding. The Norwegian horse breed populations are in a unique situation, with small populations being more or less

isolated from the outside world, depending on the rules within the breed (Fjord Horse International, 2014).

Breeding of the Fjord-horse in Norway is based upon Norwegian heritage with the breeding in-pasture system or natural covering in hand. A pasture system is a form of natural covering where one approved stallion runs with a pack of mares in a grazing mountain pasture in summer. The pasture system has one official release of horses in Sikkilsdalen every year in May. Also, FHI (Fjord Horse International, 2014) points out that government support and NHS subsidies are essential for the continuity of breeding and industry. FHI further states that NHS formulates the breeding for all Norwegian horse breeds in cooperation with the respective breed societies, guiding the selection of horses for future generations. NHS formulates the breeding plans for the Norwegian Fjord-horse with Odd Vangen, Tore Kvam, advisors, and the Norwegian Fjord-horse Association (NFL) (Fjord Horse International, 2014).

The first official breeding plan for the Fjord aimed to keep a purebred population with suitable breed types and versatility, healthy and functional horses. The breed has adapted to the market's requirements and is now found worldwide. However, as FHI (Fjord Horse International, 2014) highlights in their breeding plan, the Norwegian breeding population keeps dropping, with fewer mares and stallions active in breeding and evaluations. Matching suitable horses for breeding becomes increasingly more important with decreasing breeding activity. Artificial insemination has become available for some Fjord stallions through the work of the NFL. Artificial insemination is carefully considered for every stallion, and this is done not to threaten traditional heritage breeding methods (Fjord Horse International, 2014).

2.2 Equine Reproduction

The reproductive performance of an animal is affected by many factors, such as genetics, management, and environment. Reproductive performance can be a measure of a population or stripped down to an individual's performance, divided into male and female reproductive performance, or seen as a whole. Fertility measures reproductive performance and is defined as the ability to conceive and deliver offspring (Perdomo-González et al., 2021). This can also be seen as contributing to the next generation and is crucial for the survival of a population.

Essential traits in livestock commonly have economic aspects. Such traits may include meat color, quality, quantity, milk yield and composition, egg production, and fertility. Livestock fertility relations to financial success are significant. Although generally accepted, there are studies demonstrating genetic factors affecting reproductive health and reproduction-related traits, such as the maternal lineage or the inbreeding affecting the gestation length in mares (Gómez et al., 2020; Mantovani et al., 2020; Todd, Hamilton et al., 2020; Kuhl et al., 2015; Rodrigues et al., 2020; Valera et al., 2006). Further, the line, breed, and inbreeding affect the sperm traits as well (Dini et al., 2020; Gottschalk et al., 2016a, 2016b; Greiser et al., 2019).

Livestock species have taken into use technological tools to identify genes associated with traits, such as fertility in cattle and sheep, and included in breeding goals of, e.g., cattle (Demars et al., 2013; Ma et al., 2019; Dezetter et al., 2015; Laseca et al., 2021; Liu et al., 2008; Meyer et al., 1990; Phocas et al., 1998; Robinson, 2008). Palmer & Chavatte-Palmer (2020) points out that in equine breeding, the historical focus was on something different than fertility. The breeding interest has been on traits of conformation, color, athletic performance, and pedigree (Sairanen et al., 2009; Stock et al., 2015; Wolc & Olori, 2009). Insufficient information on fertility traits is accompanied by no inclusion of the feature in

breeding programs (Gómez et al., 2020; Mantovani et al., 2020; Mucha et al., 2012; Sabeva, 2011; Sairanen et al., 2009; Wolc & Olori, 2009).

Heritability for fertility has shown low values in various species, including horses (Cammack et al., 2009; Gómez et al., 2020; Mantovani et al., 2020; Sairanen et al., 2009). The lack of significant heritability values and scientific efforts is part of why fertility is not implemented as a selection objective in breeding programs. In other livestock species, the genetics of reproduction is better characterized and reported to improve fertility in cattle despite the low estimated heritability of the trait (Berglund, 2008). Similar progress should be feasible in equines if fertility is considered in breeding programs.

2.2.1 Measures of Reproductive Performance

The reproductive performance of livestock is measured in the number of progenies produced or the pregnancy rate. Equine Reproduction efficiency is poor compared to other livestock (Perdomo-González et al., 2021; Engelken, 1999; Menzies, 1999). Most horse breeds also lack organized data collection on reproductive traits. Bristol (Bristol, 1982, 1987, as cited by Morris & Allen, 2002) found that more than 85% of mares who ran with a stallion became pregnant. Also, mares had fertilization rates of 81-90%hen the oviducts were flushed (Ball et al., 1986, 1989). The cyclic pregnancy rate was as high as 85% to 95% in studies by Morris & Allen (2002) and Hemberg et al. (2004), with a foaling rate of 75% to 83%. However, more studies reveal lower fertility, such as in some retrospective studies with cyclic pregnancy rates of only 43-64% in Thoroughbred, Standardbred, and Quarter horses (Allen et al., 2007; Bruck et al., 1993; Sullivan et al., 1975; Woods et al., 1987, as cited by Morris & Allen, 2002). In different age groups, Allen et al. (2007) recorded a ~44-63% cyclic pregnancy rate, end-of-season pregnancy rates of ~74-91%, and a foaling rate of ~63-82%. As demonstrated in this section, there are many ways to determine equine fertility.

This chapter will only detail one of these measures, which is the most relevant to the thesis. The other fertility measures not mentioned in this chapter are, e.g., pregnancy rate, cyclic foaling rate, cyclic pregnancy rate, reproductive efficiency, and cyclic pregnancy loss.

2.2.1.1 Foaling Rate

The foaling rate (FR) is the percentage of mares that produce one foal after one breeding season. The foaling rate estimates the reproduction efficiency in a population and is used as a proxy for fertility. The foaling rate in the past two decades is varying in the Dole (50-62%), the Fjord (50-65%), and the Nordland/Lyngen-horse (58-76%) (Norsk Hestesenter, 2020).
In other breeds using similar methods for covering average foaling rates are reports for Icelandic horses (68%), Shetland ponies (68%), and Thoroughbreds (82%) (Hanlon et al., 2012; Morel & Gunnarsson, 2000; van Buiten et al., 1998). Sairanen et al. (2009) found average foaling rates of Standardbred ~73% and 66% in Finn horses. Müller-Unterberg et al. (2017) point out a difference in foaling rates between warmblood and coldblooded breeds, where the Black Forest Draught horse foaling rates are like other coldblooded breeds in Finland and France (Katila, Nivola, et al., 2010; Katila, Reilas, et al., 2010; Langlois & Blouin, 2004). Todd, Hamilton, et al. (2020) found no measurable effect of inbreeding on the foaling rate of Australian thoroughbred horses. The foaling rate is measured as follows;

$$\boldsymbol{FR}(\%) = \left(\frac{\boldsymbol{F}}{\boldsymbol{C}}\right) \times \mathbf{100}, \qquad \textbf{(eq.1)}$$

Where F is the born foals, and C is the covers of the horse.

2.2.2 Reproductive Physiology of Horses

Challenges in equine fertility can be understood through reproductive physiology. Horses are long-day seasonal breeders (Osbourne, 1966, as cited by Morris & Allen, 2002), mating in the spring and summer and giving birth in the summer with an 11-month gestation period (Spencer, 2013). Mares with twins are likely to experience abortion, stillborn or weak when born, so mares generally only produce one foal per year, and long generation intervals are standard. These restrictions in equine reproduction physiology make maintaining fertility rates important for the equine society economy and increase the size of endangered populations.

There is a significant difference between the equine physiological mating season and that humans enforce for age-related racing (Morris & Allen, 2002). According to Müller-Unterberg et al. (2017), conception rates are highest in April-June, as in German Thoroughbred mares (Ewert et al., 2013, as cited by Müller-Unterberg et al., 2017). Also, seasonally late foalings result in fewer cycles to becoming pregnant again (Langlois & Blouin, 2004). Further, accumulating sub-fertile mares repeatedly covered without success harms fertility measures (Hamann et al., 2005; Merkt & Klug, 2000, as cited by Müller-Unterberg et al., 2017).

2.2.3 Management

Fertility in equines is a complex trait with many factors, environmental and managemental, making it difficult to estimate the genetic contribution to fertility issues (Perdomo-González et al., 2021). Equine fertility is expected to differ between populations depending on the managemental factors and the genetic factors in the pool. Fertility has become a topic with increasing interest over the last two decades. Determining causal variants is difficult with the low heritability and linkage disequilibrium between genomic variants (Mahon & Cunningham, 1982). The ability of mares to reproduce has a low heritability (Mahon, 1982), greatly influenced by factors such as the environment and how they are managed (Laseca et al., 2021).

The overall reproductive efficiency of a group of horses is significantly influenced by several factors related to the mare, as well as the management of the stallion (Morris & Allen, 2002). The primary factors influencing fertility in horses are the timing and method of breeding, the age of the mare and stallion, the reproductive health of the mare, and the level of inbreeding (Morris & Allen, 2002; Müller-Unterberg et al., 2017; Sairanen et al., 2009). Other factors affecting fertility may be the horse's training level before breeding season, endometrium health, and veterinary tools such as hormone therapy.

2.2.3.1 Methods of Covering

The covering of a mare is limited to a handful of methods. Some breeds are restricted to using a particular method, and others use all available forms. The types of covers include in-hand covering, in-pasture, and artificial insemination. Comparing these measures is not always straightforward since the in-hand and artificial insemination are controlled, knowing the number of times a stallion has covered a mare. However, with in-pasture coverings, the horses are to breed by natural terms, and the number of covers done by a stallion on each mare is not easily obtainable.

Mating the horses in hand is the second form of breeding the Norwegian national horse breeds. When breeding horses in hand, the mare and stallion are accompanied by someone holding them, whereas

they otherwise perform the copulation without any other human help. When the copulation is over, the mare and stallion return to everyday life; later, the mare will be examined to determine whether the mare is pregnant. The gestation examination is done using veterinary tools or by reintroducing the mare and stallion in the next cycle to check if she is in heat. She will likely be pregnant if the mare does not respond in a specific way to the stallion.

Artificial insemination (AI) is named after the storage method of the semen or the location where the mare is inseminated in studies (on-site AI or frozen AI). With AI in horses, there are a few steps to go through. The first step is a collection of stallion semen with a breeding mount. Secondly, the semen is tested, diluted, and nutrients and antibiotics are added for the semen to survive before being transported, put in the refrigerator, or frozen. Lastly, the mare is inseminated by a veterinarian with a plastic pipette containing the liquid semen.
Artificial insemination on Norwegian national breeds is not allowed except for the Fjord, where mares can use AI from licensed stallions. Imported sperm of a stallion is not permitted to be used on Norwegian Fjord-horse mares, except if the Judges' panel for Fjords accepts the stallion (Fjord Horse International, 2014).

Morel & Gunnarsson (2000) conducted a study investigating the factors affecting fertility in Icelandic horses. The methods of covering were in-hand or in-pasture. Seventeen stallions were used to cover in-hand and ten in-pasture, where the adjusted foaling rate was 11.6% higher in pasture coverings. The method of breeding was found to be a significant ($P<0.05$) effect on the foaling rate but was also found to only account for a small part of the variation (~1%) with the other factors. Müller-Unterberg et al. (2017) found a significant effect of the breeding method on fertility in the Black Forest Draught horses with a $P<0.0001$ significance. The study found natural mating to have the highest cyclic foaling rate, followed by transported semen and on-site insemination. However, the study had few observations in mating types other than natural mating.
Contrary to this, Sairanen et al. (2009) report the on-site AI to be most successful, followed by transported semen and natural mating with much lower foaling rates. Frozen semen was also found to be the least efficient way of mating.

When comparing all the mating types, transported semen is easy for a mare owner. However, some factors reduce its efficiency. With inexperienced handlers, the quality of the transported semen may decay. Frozen semen gives lower pregnancy rates than transported semen (Samper & Hankins, 2001). Still, it is practical if the chosen stallion lives abroad. For the Morel & Gunnarsson (2000) study, the number of coverings did not affect the stallion foaling rate for the in-hand method. Also, for the in-pasture breeding, the period time of the year, the length of the period, and group size within each period had no statistically significant effect on stallion fertility. Similarly, other studies have not found an effect on the number of coverings and the fertility measure (Morris & Allen, 2002; Sharma et al., 2010).

2.2.3.2 Nutrition

The food composition of a mare is also suggested as a possible factor affecting mare fertility. The nutrition has been investigated as covered mares given different diets (Van Niekerk & Van Niekerk, 1998) or the body score condition of mares such as in Morel & Gunnarsson (2000).
Morel & Gunnarsson (2000) included the mare body condition in their investigation of stallion fertility in Icelandic horses. No statistical difference was found, although there were few mares in the two categories outside the 'normal' body score category. However, the adjusted foaling rate of the mares was ~71% for thin mares, ~68% for average, and ~66% for fat mares.

Niekerk & Niekerk (1998) performed a study looking at the dietary effect on fertility. In this study, 35% of mares fed low-quality proteins suffered early embryonic loss compared to 7% of the three other groups receiving higher-quality protein diets. On average, the same low-quality protein diet group lost about 25 kg during lactation, while the other groups had no reduced weight.

2.2.3.3 Age

The age of horses has been recognized as a significant factor influencing fertility by several studies (Allen et al., 2007; Held & Rohrbach, 1991; Hutton & Meacham, 1968; Laing & Leech, 1975; Morel & Gunnarsson, 2000; Ricketts & Alonso, 1991; Wealchli, 1990, as cited by Morris & Allen, 2002). The age of stallions seems to influence their fertility less than the age of a mare (Langlois & Blouin, 2004). The results are similar to Müller-Unterberg et al. (2017), where the stallion age did not influence the cyclic foaling rate in Black Forest Draught horses. Similarly, no significant effect between stallion fertility and age (age 5-18) is observed in Thoroughbred and Icelandic stallions (McDowell et al., 1992; Morel & Gunnarsson, 2000). Ewert et al. (2013, as cited by Müller-Unterberg et al., 2017) observed decreased pregnancy rates for Thoroughbred stallions above 16 years. The age of mares was found to be significantly ($P<0.001$) affecting stallion fertility by Morel & Gunnarsson (2000), especially in young mares. In this study, the mare's age explained 1.7% of the total variance of stallion fertility. They also found no sign of stallion age affecting stallion fertility over the breeding season. The mare's age was statistically significant ($P<0.0001$), with a decrease in the cyclic foaling rate for an advancing age in the mare. However, this was not true for the stallions in the Black Forest Draught horses (Morel & Gunnarsson, 2000). There was also lower semen quality in very young and stallions above age 11 in a study by Dowsett and Knott (1996).

The mare age is often found to significantly affect fertility, such as in the young (age 3-9) Black Forest Draught mares having significantly better cyclic foaling rates than the older age groups (Müller-Unterberg et al., 2017). Further confirmed by others describing the general trend was a lower foaling rate with increasing mare age (Langlois & Blouin, 2004; Sairanen et al., 2009). Sharma et al. (2010) recorded breeding ~250 Thoroughbred mares over seven consecutive years, revealing a ~69% foaling rate. The foaling rate was significantly higher in young mares (age 3-7) than that in older mares (age 13-17 and >18). The younger mares also had a smaller percentage of pregnancy losses. The latter is supported by Malheiros de Souza et al. (2020), who found that the pregnancy rate was unaffected by the mare age or age group. However, age significantly impacted embryo loss, which was smaller in young mares. In a retrospective study of Thoroughbred mares by Morris & Allen (2002), the difference between age groups (3-8 and 14-18) significantly affected the mare's fertility. Further, the pregnancy loss between young mares (3-8) and the other age groups differed significantly.

2.2.3.4 Status of mare

The breeding status of the mare has been recognized as an influencing factor in fertility since the 20th century (Ricketts & Alonso, 1991; Woods et al., 1987, as cited in Morris & Allen, 2002). The breeding status can be divided into categories and may differ slightly between reports. Generally, the different statuses include lactating mare, barren and maiden, and variations of these categories. For example, intentionally barren or non-intentionally barren mare, where the difference is whether the mare was covered last year. Morel & Gunnarsson (2000) found the highest adjusted foaling rate difference between lactating mares (65.5%) and barren mares that were not covered the previous year (70%), although there was not a significant difference. However, the mare status for the individual stallion

showed a significant effect (P<0.001), accounting for a big part of the variance of stallion fertility. The same was true when they compared all lactating mares against non-lactating mares. Unlike this study, the mares with foal at foot had the highest cyclic foaling rate by Müller-Unterberg et al. (2017), which agrees with a few other reports (Langlois & Blouin, 2004; Sairanen et al., 2009).

Interestingly, 50% of mares with two or more consecutive aborted or barren years were again barren or aborted the following season (Sharma et al., 2010). Supporting the previous research by Morris & Allen (2002), suggesting that resting a mare is not enhancing the foaling rates of the mares covered the following season. Allen et al. (2007) found the 42 Day pregnancies per cycle to significantly differ between the barren mares, the foaling, and the maiden mares. The maiden mares had the highest foaling rate, followed by the foaling and barren mares. The findings of Bruck et al. (1993) contradict the previously stated and found no effect on the reproductive status in Thoroughbred mares.

2.2.3.5 Inbreeding

The selection of breeding animals is usually based on their performance in a sport and desired morphological traits (Belloy & Bathe, 1996; Perdomo-González et al., 2020; Todd et al., 2018). An animal excelling in a discipline is more likely used for breeding purposes, creating an uneven contribution of genetic material to the next generation (Perdomo-González et al., 2021). A variable genetic contribution to the next generation increases a population's inbreeding, thus increasing the risk of effects from recessive detrimental mutations (Charlesworth & Willis, 2009).

In addition to selective breeding, population bottlenecks (Schubert et al., 2014) may cause detrimental variants to accumulate in the equine genome. Further increase in the frequency may be due to selective breeding, inbreeding, and genetic drift (Moyers et al., 2018).

Although lethal mutations are under negative selection, this is not the case for recessive lethal mutations with the potential of reaching high frequencies in a small population. The statistical likelihood of identical recessive lethal mutations in unrelated animals is low; however, in the case of mating closely related animals, the likelihood increases with the level of inbreeding (Charlesworth & Willis, 2009). Gómez et al. (2009) also state that the lack of genetic control in planned coverings affects the inbreeding observed in horses.

It is generally accepted that inbreeding is associated with reduced fitness and performance and affects many traits, including reproduction (Kjöllerström et al., 2015; Klemetsdal & Johnson, 1989; Langlois & Blouin, 2004; Leroy, 2014). Although studies report different inbreeding effects, several reported decreased foaling rates with increasing inbreeding coefficients (Bouman, 1977; Fomin et al., 1982; Shemarykin & Vysokolova, 1986, as cited by Perdomo-González et al., 2021). There are reports of correlation (r=0.47) between inbreeding coefficients in mares and the number of matings before conception by Fonesca et al. (1977, as cited by Perdomo-González et al., 2021). They showed that the inbreeding coefficient of potential offspring affects reproductive efficiency. The reproductive efficiency was affected by an increasing number of unsuccessful matings as the level of inbreeding rose. Klemetsdal and Johnson (1989) report that inbreeding affects early abortion significantly in Norwegian Trotter mares. Also, Keller & Waller (2002) have pointed out inbreeding as a critical factor in the reduced fertility of wildlife animals.

Fertility traits are also shown to correlate with genomic variants in the X chromosome, where the homozygosity increase showed a drastic reduction in the fertility of mares (Laseca et al., 2020). Metzger

et al. (2015) detected inbreeding (ROH) in genomic loci with 139 genes, including several associated with spermatozoa health and embryonic development.

However, some studies do not provide an obvious, direct, positive correlation, as seen in the previous section, such as the study on Standardbred trotters and Finn horses by Sairanen et al. (2009). In this study, the Standardbred horses had higher inbreeding (~10%) and higher foaling rates (~73%) than the Finn horse (~4% and 66%, respectively). However, the inbreeding degree significantly negatively affected the foaling rates within the breeds. In Standardbreds, foaling rates significantly dropped when the predicted foals had an inbreeding coefficient above 15%. Similarly, Finn horses had lower foaling rates when the predicted foals' inbreeding coefficients ranged from approximately 4% to 10%. Cothran et al. (1984, as cited by Muller-Ünterberg et al., 2017) report reduced foaling and conception rates with increasing inbreeding in the Standardbred Pacers but the opposite in the Standardbred Trotter. The mares' productivity declined by 0.5-1% for a French horse population, with a 1% increase in the inbreeding coefficient (Langlois & Blouin, 2004). For the inbreeding coefficient of sires, the opposite results were observed.

Some studies find no relationship between inbreeding and reproduction traits. For example, Mahon & Cunningham (1982) found the inbreeding level effect on mare fertility not significant; however, the general inbreeding of the horses was low (<1%). Müller-Unterberg et al. (2017) investigated fertility in Black Forest Draught Horses and did not detect a significant effect of inbreeding coefficients on the mares' fertility. However, this report's population size is limited and had fewer matings between highly inbred animals. Further, non-significant results can also arise when reproductive efficiency is recorded as a binary trait and a deficient registration practice of breeding results. The consequences of this make for overestimating foaling rates and, therefore, underestimating the effect of inbreeding coefficients on reproductive efficiency (Müller-Unterberg et al., 2017). Comparing studies may also be challenging because of differences in experimental designs, animals examined, and methods of inbreeding estimation.

It is suggested that inbreeding depression is predominately caused by recessive detrimental variations in a population (Charlesworth & Willis, 2009). The last decades have confirmed variable inbreeding depression in several other livestock and pet animals such as the pig, cattle, and canine (González-Recio et al., 2007; Lazzari et al., 2011; Leroy et al., 2015; Makanjuola et al., 2021; Saura et al., 2015; Wall et al., 2005; Zhang et al., 2018). A possible explanation of the fluctuations in inbreeding effect on reproduction may be based on factors such as the rate of inbreeding increase, selective pressures, and the genetic variation differing in the populations studied (Todd, Hamilton, et al., 2020). Pedigree completeness is also known to affect the inbreeding levels of an animal, making the old studbooks also essential to include in studies. Foaling rates may also appear higher than what they are if breeders are to report on the outcome of a mating event, but mostly they only report live-born foals and have a lower reporting degree of matings with no live-born foal outcome.
For instance, there is documented deficiency of reported breeding outcomes in the Finn horse and French horse breeds (Katila et al., 2010; Langlois & Blouin, 2004). Inbreeding has been found to have negative consequences, such as inbreeding depression in studies. However, it also has a broad spectrum of effects - negative, neutral, and even positive - on livestock, according to research by Perdomo-González et al. (2021). Derks et al. (2019) report the results of recessive lethal haplotypes carrier matings, decreasing litter sizes of ~15-22% in pigs, and also found identified lethal recessives to contribute to observed heterosis effects for fertility.

The effects of inbreeding can be devastating in populations with small effective population sizes (Avdi & Banos, 2008; Gamboa et al., 2009), and the genetic details of reduced reproductive health must be

investigated in small populations (Klemetsdal & Johnson, 1989; Olsen & Klemetsdal, 2010). Three Norwegian national horse breeds are facing challenges with inbreeding; the Dole, Fjord-horse, and Nordland/Lyngen-horse. Recent population bottlenecks exist, such as the Nordland/ Lyngen-horse population today descending from 5 to 10 founders. Also, all the Fjord-horses today can be traced back to one founder stallion, Njål 166 (Grønntun, 2020), after an incentive to make the Fjord-horse a pure breed. Breeding incentives from four decades ago in the Nordland/Lyngen-horse have kept the inbreeding degree in the population at 12%. In contrast, the Dole and Fjord-horse have average inbreeding degrees increasing from 8% to 14% and 6% to 8%, respectively (Norsk Hestesenter, 2020).

2.3 Gene Detection

Searching for genes associated with a phenotype can be compared to finding a needle in a haystack. Despite the difficulty, identifying genes linked to specific traits is not a new concept, as noted by Sharmaa et al. (2015). In the 1990s, microsatellite markers were primarily utilized in QTL mapping, as Lipkin et al. (1998) outlined. In contrast, Sharmaa et al. (2015) highlight that in recent years, whole-genome sequencing and SNP panels have been used in association studies with phenotype and pedigree data. The use of genomic strategies in horses has been increasing in recent years, but there is still untapped potential for understanding the horse genome (Raudsepp et al., 2019; Laseca et al., 2021). The increase in genomic gene detection methods is due to an adaption of practices, obtained reference sequences, and increases in genotyped horses (Kalbfleisch et al., 2018; Nandolo et al., 2018). The benefits of genomic methods and information are to improve pedigree qualities, increase genetic gains, identify deleterious alleles, understand selection history, and conserve diversity in a population (Wooliams & Oldenbroek, 2017). Gómez et al. (2020) point out that advancing genomic methods makes obtaining better and more reliable information cheaper. However, it is challenging to get large-scale, accurate phenotypic datasets and to model environmental effects in complex traits (Gómez et al., 2020; Laseca et al., 2021).

Genomic evaluation techniques are primarily employed in human and model animals for the detection of mutations, copy number variants, homozygosity, and Genome-Wide Association (Dini et al., 2020; Laseca et al., 2021; McQuillan et al., 2008; Visscher et al., 2012). Scott et al. (2022) describe gene detection as two approaches. The classical approach follows a linear series of events: defining the phenotype, identifying multi-case families, gathering samples, genotyping, analyzing data for initial disease gene position, determining the minimum candidate locus, and then sequencing genes within this locus to find the causative mutation. Unlike the classical approach, more modern methods are non-linear and involve numerous evolving steps. These steps are continuously redefined, refined, or substituted by subsequent ones. Each step has vital factors that must be considered in a complex trait study. Also, the order and emphasis of these steps on the approach will vary between studies. The steps within a complex trait study consist of (1) phenotype, (2) study design, (3) analysis, and (4) follow-up (Scott et al., 2022).

2.3.1 Basic Theoretical Principles

2.3.1.1 Hardy-Weinberg Frequencies

Calculating the allele- and genotype frequencies in populations and observing the fluctuations of frequencies over generations is a fundamental part of population genetics. The Hardy-Weinberg mathematical model describes allele and genotype frequencies and their relationship. The accompanying assumptions to the Hardy-Weinberg model are for the population to be an 'ideal'

population without selection, mutation, drift, and migration (gene flow). Hardy-Weinberg expectations multiply two allele frequencies to give the likelihood of a genotype. For example, the expected homozygosity is the fraction of homozygote animals in a population calculated from the allele frequencies observed;

$$\mathbf{E(HOM)} = \boldsymbol{p^2}\text{ ,} \qquad \textbf{(eq. 2)}$$

Where EH is expected homozygosity, and p^2 is the squared allele frequency. The genotype frequency is then multiplied by the population size to obtain the number of expected homozygotes. These estimations of expectancies help look at populations deviating from an ideal population to find why. Theoretically, any break in the principles may lead to deviations in the Hardy-Weinberg Equilibrium (HWE). If one were to test the deviations and then reject the null hypothesis (that there is equilibrium), we would have no information on what forces act on the population. Also, one does not expect a perfect equilibrium in most populations we investigate. Instead, we search for proof that there is no equilibrium and seek to describe why there is not or what effect and consequences this may have.

The break of the equilibrium may be due to nonrandom mating. Nonrandom mating, also called selection, can be where an organism's fitness is connected to a genotype giving the organism an advantage or disadvantage in any critical moments in life (e.g., finding mates, conceiving, finding food, and avoiding predators). The equilibrium break may also be due to the population not being 'ideal,' which practically no population is. Several factors can influence genetic variation within and between populations, including selection, mutation, drift, and gene flow. Depending on the circumstances, these factors can increase or decrease genetic variation.

Within populations, selection can reduce and increase genetic variation depending on the type of selection present. If a balancing selection is present, it will maintain genetic variation. However, if a directional selection like a recessive detrimental or advantageous allele is present, the genetic variation will gradually be reduced depending on the selection coefficient. Mutation mainly increases the genetic variation within a population by introducing new variants, but this may be short-lived if the mutation is not distributed across the population. Drift mainly reduces genetic variation within a population, but in a large population, the effect of drift is lower and may increase the genetic variation of an allele. However, genetic drift generally reduces genetic variation in the long term. Gene flow mainly increases genetic variation within a population because new variants may be introduced. However, a mutant may also migrate from the population, leaving the population with less genetic variation.

Between populations, selection can reduce and increase genetic variation within a species depending on the type of selection. Mutation mainly increases the genetic variance between populations because two separate populations are unlikely to have the same mutation at the same position. Drift may increase or reduce genetic variance between populations depending on the population size, where the effect of the force is more considerable in smaller populations. Gene flow mainly reduces the genetic variance between populations by exchanging genetic material between populations, which means that the populations do not diverge completely.

To identify deviations from the HWE, one can calculate the expected frequencies and statistically compare them to the observed frequencies using the Exact Test or the Chi-Square Test. The Exact Test is best to use when the number of animals is low. The exact tests calculate the probability of each genotype combination occurring and assume that any scenario where this probability is lower than 0.05 is not consistent with HWE. The Chi-square test looks at the deviation between the observed number of genotypes and the expected number from HWE to see whether the deviation is significant. The difference between the genotypes is then squared, divided by the expected number, summed, and compared to a likelihood table (Chi-square distribution). If the Chi-square value we got is higher than the

table value, with degrees of freedom and a statistical power we want to have, then the null hypothesis can be thrown out. The null hypothesis is always that the population is in HWE.

2.3.1.2 Linkage Disequilibrium

Linkage disequilibrium is the non-random association of two loci, such as genes or a marker and a gene. Further, the two loci are rarely separated by recombination and, therefore, also closely located on the chromosome or in other means, such as being in a recombination cold spot or genetic hitchhiking. The linkage between two loci is described continuously between -1 and 1, where one means they are always observed together.

Drift and geneflow as evolutionary forces may generate linkage disequilibrium and inbreeding, and selection maintain it. The recombination in the genome is what breaks up the linkage between loci. From this comes the term 'rate of decay' describing the decline of association between the linked loci when recombination occurs between them.
The linkage disequilibrium can be estimated in numerous ways, for example, by square correlation coefficient. However, the standardization of a linkage disequilibrium measure is that it depends on allele frequencies and therefore is difficult to compare across populations and loci.

2.3.2 Detecting Candidate Genes

When starting the process of gene detection, the trait phenotype or characteristics need to be evaluated as to if it is affected by genetic factors. Methods to estimate the genetic aspects of a trait are usually first determining the familial aggregation of the feature.

2.3.2.1 Familial Aggregation

Koch & Rampersaud (2022) states that all genetic traits' common characteristic is that they aggregate within families. Therefore, if a familial aggregation is observed, one can statistically determine the significance of this observation. Further, they describe the start of the analysis as deciding whether the trait seems to aggregate within one or more families without having any specific genetic model in mind. One can do this by identifying clusters of cases and controls and comparing the probability of the trait appearing in either group (Sahebi et al., 2013). Koch & Rampersaud (2022) mentions other statistical methods to decide the aggregation, including the family history approach, correlation coefficients, twin and adoption studies, recurrence risk of relatives of the affected, heritability, and segregation analysis. Lastly, granting there is statistical evidence of genetic factors of a trait, the causative genes need identification to confirm the proof.

2.3.2.1.1 Heritability Estimation

Heritability is generally the genetic proportion of the trait variation and may be defined in a broad or narrow sense (Sahebi et al., 2013). The heritability will always be between 0 and 1, 0 being no genetic factors and 1 being complete genetic control. In the broad sense, heritability is written as:

$$h^2 = \frac{G}{V} = \frac{G}{(G+B+E)}, \quad \textbf{(eq. 3)}$$

Where V is the phenotypic variance of the trait, G is the genetic variance of the phenotype, B is the within-family variance of the phenotype, and E is the random environmental variance of the phenotype, according to Koch & Rampersaud (2022). They further state that the genetic variance can be dismembered into additive, dominance, and epistatic variance. In a narrow sense, heritability measures the additive genetic variation of the feature (Sahebi et al., 2013). The correlation coefficient one may calculate between two relatives' trait values may be used in calculating the heritability (obs., this is not true for full siblings):

$$\boldsymbol{h^2 = \frac{r^2}{(IBD)}}, \qquad \textbf{(eq. 4)}$$

Where r^2 is the square of the correlation coefficient, and IBD is the estimate between related animals (for example, 0.5 shared genes IBD between a parent and an offspring) (Koch & Rampersaud, 2022). Besides, one may also estimate heritability with the covariance between two relatives or with the correlation from multiple relative pairs (Hartl & Clark, 1997, as cited by Koch & Rampersaud, 2022; Rice et al., 1997).

2.3.2.1.2 Segregation Analysis

When the aggregation pattern seems reliable with an effect of genes, the investigation takes the next step into segregation analysis to determine if the evidence of familial aggregation is substantial enough for linkage analysis. The segregation analysis intends to decide an inheritance pattern for the trait in question and the compatibility with predefined genetic models.

The segregation analysis uses the pedigrees of affected animals as the unit of study (Thomas, 2004). The power needed depends on the genetic model the trait follows. If the study seeks to detect recessive lethal alleles, the test requires family-based studies or massive population datasets with pedigrees and accurate phenotypes (Scott et al., 2022). Family-based studies include large extended families, smaller multi-case families, and discordant sib-pair studies (Scott et al., 2022).
When the investigator is investigating a pedigree and deciding which animals to include in the analysis, several things must be considered. The starting point is generally the proband and first to second-degree relatives (Thomas, 2004).
For example, segregation analysis is feasible without first-degree relatives but less potent. Obtaining information about all reported cases is an essential part of the job. However, excluding phenotypes of not-mentioned relatives is unacceptable. At the same time, careful consideration is needed not to exclude unconfirmed cases since this may contribute to the bias of the analysis (Thomas, 2004). The inclusion of subjects should be systematically built to avoid bias without investigating uninformative families.

Canning and Thomson (1977, as cited by Thomas, 2004) established a strategy to fulfill the conflicting requirements of the segregation analysis. The principles are (1) the initial identification of the family must follow a systematic characterizable scheme, (2) at every stage of the pedigree building process - ascertainment of phenotype information must be unbiased, (3) when deciding to extend a pedigree, one can use information about the phenotype accumulated so far, and knowledge of the family structure but must not be influenced by informally obtained information. As one branch has been extended, all phenotype data obtained must be included in the analysis (Thomas, 2004).

The genetic models (characterizable scheme) may be any of the following: one major gene, several major genes, polygenes, or a shared environments model. The models for simple inheritance (Major

gene effect) have their own rules on how the trait presents itself: autosomal dominant and recessive, the other X-linked dominant and recessive (Nicholas, 1987).
An example of these rules of display for a sample recessive autosomal trait is (1) the trait may skip generations, (2) all affected parents' matings result in affected offspring, (3) there is no significant difference between the sexes of the population, and (4) if the trait is rare- heterozygous matings result in ¼ affected offspring in the general population (Nicholas, 1987). Depending on mated genotypes, the population frequencies in simple recessive autosomal inheritance may be between ½ - $^1/_8$.
The model's parameters are estimated if a genetic model fits the data. Specific matings between genotypes would ease the fitting of a model but are often not possible (Nicholas, 1987) because a trait is usually discovered by chance in an affected offspring. Also, specific matings are not relevant for animals producing only one offspring or with long gestation periods, such as humans, cattle, and horses. For major gene effects on a trait, the power of inheritance models can be calculated easily with computer simulation (Given a phenotype, population, and pedigree) (Scott et al., 2022). The genetic models are challenging to determine for complex traits, and a computer simulation is required to calculate the power of several genetic models and compare them (Scott et al., 2022). The ascertainment of families is critical since a bias affects the best-fitting model estimate and leads to an incorrect conclusion of a genetic model (Greenberg, 1986) as cited by (Koch & Rampersaud, 2022).

Segregation analysis may be performed with different analytical approaches: the Mixed Model, the Transmission Probability Model, the Unified Model, and the Regressive Model (Bonney et al., 1984; Elston & Stewart, 1971; Lalouel, 1984; Lalouel & Morton, 1981; MacLean et al., 1984; Morton & MacLean, 1974 as cited by Koch & Rampersaud, 2022). Although computationally demanding, all of the proposed methods are accessible through software packages such as POINTER, SAGE, and PAP (Hasstedt, 1993; Lalouel & Morton, 1981; SAGE, u.d., as cited by Koch & Rampersaud, 2022).

According to Thomas (2004), the maximum likelihood estimation in segregation analysis is commonly used to fit genetic models in complex traits, maximizing the probability of the data given in the model. The estimation components are genotype distribution, transmission probabilities, method of ascertainment, and penetrance. A form of ascertainment correction is necessary, correcting for the likely to-be over-represented affected gene frequency compared to the general population (Thomas, 2004)

According to Nicolas (1987), Davie (1979) developed a quick and easy way to estimate a segregation frequency of a trait, in which the genetic model fitted may be statistically tested, called the Singles Method. This approach requires the inclusion of all reported families in the data. The procedure is challenging if there are sporadic cases, a misclassification of phenotypes, mutations, and environmental factors. The approach starts with estimating the fit of simple Mendelian inheritance. If the simple model is rejected, but there is doubtless a genetic factor involved, one can perform a second variant of the approach: Complex analysis. The complex analysis is intricate to perform and interpret correctly.
To use the simple form of the Singles method, one must either include (1) all families that have affected offspring in the analysis or (2) a randomized subset of such families (Nicholas, 1987). If either assumption is satisfied, the simple segregation frequency is estimated as:

$$\boldsymbol{p} = \frac{\boldsymbol{A}-\boldsymbol{A_1}}{(\boldsymbol{T}-\boldsymbol{A_1})}, \qquad \textbf{(eq. 5)}$$

With the variance of:

$$var(\boldsymbol{p}) = \frac{(T-A)}{(T-A_1)^3}\left\{A + A_1 + A_2\frac{(T-A)}{(T-A_1)}\right\}, \qquad \textbf{(eq. 6)}$$

A is the total number of affected offspring present in the data set, T is the total number of offspring in the dataset, A1 represents the number of families with only one affected offspring, and A2 represents the total number of families with two affected offspring. For uncertainty about whether the assumptions are fulfilled, Nicholas (1987) shows that an upper and lower limit estimate may also be calculated. In this case, the upper limit would be calculated as shown in eq. 5. The lower limit is calculated with the same equation (eq. 5), replacing A1 with the number of families and setting A2 equal to 0. The next step is to test whether the estimated frequencies fit the expectations of a specific genetic model (Nicholas, 1987).
The expectations of the genetic models are: for an autosomal dominant inheritance, the segregation frequency (p_0) will be $p_0 = 0.5$ in the offspring. For an autosomal recessive inheritance, $p_0 = 0.25$ in the offspring. The X-linked dominant will have an affected-parent-dependent frequency. For example, with a heterozygote-affected mother, the segregation frequency would be $p_0 = 0.5$ in all offspring. Further, if the mother is homozygote affected, the segregation frequency becomes $p_0 = 1$ for all offspring. However, if the father is affected, the segregation frequency becomes $p_0 = 0$ for male offspring and $p_0 = 1$ for female offspring. The expectance of the specific genetic models is the null hypothesis (p_0) for the significance test (Nicholas, 1987). The frequency estimate test is calculated as follows:

$$\boldsymbol{z^2 = \frac{(p - p_0)^2}{\{var(p)\}},} \qquad \textbf{(eq. 7)}$$

Being approximately distributed as X^2 with one degree of freedom (Binomial distribution). If an upper and lower estimate is obtained, the most practical is to conclude that data is consistent with the null hypothesis if:

$$\boldsymbol{p_U + \sigma p_U > 0.25 > p_L - \sigma pL,} \qquad \textbf{(eq. 8)}$$

Where subscripts of U and L refer to upper and lower estimates, and sigma refers to the square root of the estimate variance (Steinberg, 1959, as cited by Nicholas, 1987).

2.3.2.2 Family-based Linkage analysis

If a genetic factor seems evident from previous tests, one would want to determine the location of the responsible gene or genes with linkage analysis. A gene can be localized through a phenomenon called recombination. The linkage between two loci is dependent on the recombination frequency between them, and two loci that do not segregate independently are said to be linked (Griffin, 2014b). Following Thomas (2004), linkage analysis is built on three basic principles: (1) chromosomes separate independently, so there is no linkage between them. (2) Recombination is measured as the recombination fraction (q), where q = 0 is no recombination and perfect linkage, and q = 0.5 is recombination as in independent markers (Crawford & Dumitrescu, 2022; Thomas, 2004). (3) Additionally, the recombination between markers depends on their physical distance. From these principles- a low recombination rate between a marker and a gene gives the likelihood of the marker being physically close to the causative gene.

The general idea is that affected animals of a disease are expected to have inherited the same segment from a common ancestor (Griffin, 2014b). The causative loci are expected to be identical between the diseased animals. The investigation continues by determining the markers' genotypes for members of multiple case families and counting the recombination fraction of haplotype in offspring from double heterozygous parents (Thomas, 2004). Every offspring is classified as recombinant or nonrecombinant between a specific marker and the trait phenotype (Crawford & Dumitrescu, 2022).

Crawford & Dumitrescu (2022) give insight into Linkage analysis and states that it can be a powerful approach to mapping genes with Mendelian inheritance and high penetrance. When collecting a population for a linkage study, the large extended families give the highest power. However, smaller multiple-case families may also give good results because the analysis scores (logarithm-of-odds (LOD)) are additive. Crawford & Dumitrescu (2022) explain that the ideal family structure for linkage analysis is three-generational families with many offspring to determine the phase of the haplotypes. Additionally, if the phase is known, choosing recombinant and nonrecombinants is simple for highly penetrant traits. However, if the phase is unknown, recombinants from the probands are scored with assumed and alternate phases. With unknown phases, more offspring are needed to detect linkage (Crawford & Dumitrescu, 2022).

There are two general approaches to linkage analysis or linkage testing, one parametric and the other nonparametric linkage analysis. If there is enough information on the parameters gained from obtaining an inheritance model from segregation analysis and sufficient data on members and families, then parametric linkage analysis would be the best choice (Blanton, 2022; Griffin, 2014b). However, this is often not the case. Therefore, Griffin (2014b) continues that if the inheritance model is complex, but the genetic component of the trait is undeniable, the nonparametric linkage analysis is used.

Using a specific inheritance model, the Parametric Linkage Analysis compares the segregation pattern between a trait and a set of markers (Griffin, 2014b). This test allows for specific continuous and discontinuous functions of penetrance, age, and gender. Assuming that rare alleles cause rare diseases, the causal allele frequencies are estimated based on the trait frequency in the population (Griffin, 2014b). Blanton (2022) highlights some advantages to parametric linkage analysis that must be addressed. The first is the statistical power, given that the inheritance model is correct. Secondly, it utilizes all members' phenotypic and genotypic data. Thirdly, it calculates the recombination frequency between the marker and trait locus, and lastly, it conducts a statistical test to assess linkage and genetic heterogeneity. (Blanton, 2022).

Crawford & Dumitrescu (2022) points out that several approaches were developed to address many of the typical limitations of the parametric approach. Nonparametric studies have proven more potent in investigating complex traits and estimating the IBD segments between affected pedigree members. Some methods used in this analysis are suited for late-onset disorders or pedigrees without several generations, such as affected relative pairs or sib-pairs and concordant-discordant sib-pairs (Crawford & Dumitrescu, 2022; Haseman & Eiston, 1972; Risch, 1990). According to Griffin (2014b), the most straightforward nonparametric approach is the affected sib-pair test, which requires genotype data on parents with two affected offspring.

Further, the probability of known genotyped markers segregating with the disease is calculated for parametric and nonparametric linkage analysis. The LOD score is a step further than the linkage ratio and is most common (Griffin, 2014b). According to Crawford & Dumitrescu (2022), the probability of a linkage existing between the marker and the studied gene is determined by estimating a likelihood ratio (LR). LR compares the data assuming linkage to the data occurring by chance. Then, the likelihood ratio is typically converted to a LOD score.

Commonly, linkage analysis starts with two-point linkage analysis to establish the linkage in the genome. A two-point LOD score is used to assess the level of evidence for linkage between a locus and a marker, while a multipoint LOD score is used to evaluate the linkage between a locus and multiple markers. The LOD score is dependent on the recombination fraction (q) and is calculated by comparing the likelihood of the observed pattern under conditions of linkage and non-linkage (Griffin, 2014b). The calculation of

LOD scores compares a range of different values for the recombination fraction. The LOD score gives the likelihood of the observed data. A LOD score of 3 or greater indicates a statistically significant linkage, but if the score is minus two or less, then linkage is unlikely (Sahebi et al., 2013).

Suppose a linkage between the trait phenotype and marker in the genome is established. In that case, determining the smallest region that could contain the causative gene is the next step of the analysis (Blanton, 2022). As described by Blanton (2022), two main methods are used for localizing disease genes: haplotype analysis and multipoint linkage analysis. This approach operates under the premise that alleles in closely-linked marker loci reside on the exact chromosome copy and are transmitted together. As such, when affected animals share overlapping haplotype segments compared to unaffected animals, this can assist in pinpointing the location of the gene of interest. Haplotype analysis provides visual confirmation of statistical testing and complements the results obtained from multipoint linkage analysis. With the physical order of the markers, specific determinations of alleles passing from parent to offspring and identified recombination events allow flanking the smallest region that could contain the causative gene. The haplotype analysis can be done by hand or with computer software such as SimWalk2 (Weeks et al., 1995, as cited by Blanton, 2022).

Blanton (2022) continues that the multipoint linkage analysis is a statistical tool used to localize a gene to its flanking markers utilizing genetic maps where the order of markers and positions are known. The analysis hypothetically positions the causative gene between different markers and calculates the likelihood of the pedigree for every new position of the gene. Further, the multipoint LOD scores are often graphed, demonstrating the LOD scores between each marker. The locus in the graph with the highest LOD score is the most likely location of the causative gene. Multipoint linkage analysis has a few significant advantages. First, it may recover linkage information through haplotype inference. This recovery makes the study less sensitive to uninformative or missing genotypes at markers. Second, it provides a locus of interest for fine mapping. Third, the loci with low LOD scores can be further used for exclusion mapping. Some researchers demonstrate the method as useless since the LOD curves often peak throughout extended loci (Blanton, 2022). Others, such as Greenberg & Abreu (2001), argued that the sensitivity of multipoint linkage analysis may be overestimated. However, they continue that the use of multipoint LOD scores in combination with the admixture test (multipoint HLOD) can still serve as a potent tool, even for complex traits. For multipoint linkage analysis, the complexity of the computation increases with the number of markers and pedigree members. Several algorithms are developed, such as the Elston-Stewart “peeling” algorithm, Lander-Green algorithm, and the Markov Chain Monte Carlo (MCMC) type, such as SimWalk2 (Elston & Stewart, 1971; Lander & Green, 1987; Sobel & Lange, 1996 as cited by Blanton, 2022).

2.3.2.2.1 Homozygosity mapping

Detecting recessive lethal mutations may be difficult, especially in small populations. Identifying carriers of detrimental recessives is critical to reducing the mutation frequency, but it may be challenging with pedigree and phenotype data (Wooliams & Oldenbroek, 2017). Carriers of the lethal allele mated with other carriers may go unreported. Genomic tools, however, can identify carriers with high reliability. When the recessive lethal expression is neonatal or post-natal, homozygosity mapping is frequently used to detect this. Homozygosity mapping searches for depletion or lack of homozygotes (Derks et al., 2017). This deficit search is done because recessive detrimental mutations do not affect the heterozygote.

On the contrary, they are lethal to the harboring allele homozygote, changing the Mendelian genotype ratio (2:1 instead of 1:2:1). Homozygosity mapping has been a powerful tool for detecting recessive disease genes in related families for several decades (Daetwyler et al., 2014; Lander & Botstein, 1987; Todd, Thomson, et al., 2020; VanRaden et al., 2011). This method does not require the DNA of any family members other than the affected offspring (Lander & Botstein, 1987). Ideally, the study also has genotypes of the parents of the cases (Wooliams & Oldenbroek, 2017). It can then come with a genome-wide linkage association analysis using markers such as SNPs (Gibbs & Singleton, 2006). Using a moderate or high-density SNP chip will give sufficient evidence. The SNPs on each homologous pair of chromosomes are arranged in order according to an available reference genome (Wooliams & Oldenbroek, 2017).

This method heavily relies on the population size of genotyped animals. By using tens of thousands of genotyped animals, even very rare deleterious haplotypes (frequency < 2%) can be detected, as demonstrated by Derks et al. (2017). On the other hand, Derks et al. (2017) also point out that using specialized SNP assays, known as 'SNP chips', to genotype a large number of domestic animals is unlikely to uncover the causal variants of severe syndromes. This is because these assays are designed to favor high minor allele frequency, resulting in a bias. Consequently, the haplotype-based approach is more efficient for identifying harmful variation than single SNPs. The causal variant is often in low linkage disequilibrium with the single SNPs, but haplotypes can capture it more effectively. A significant decrease in haplotype homozygosity indicates reduced viability, and these haplotypes are likely to carry detrimental mutations that cause embryonic lethality in a homozygous state. Studies in mice indicate that about 30% of mouse genes exhibit a lethal knockout phenotype in a homozygous state, suggesting many potential embryonic lethal genes in Mammalia (Ayadi et al., 2012).

Homozygosity mapping may be done with a “single-SNP” approach or by looking at haplotypes. Looking at single SNPs may identify many homozygote-depleted areas but also give false positives. The haplotype approach looks at a set of variants to identify the depletion of homozygosity (VanRaden et al., 2011) and is expected to be in linkage with a causative variant. In this technique, there is a haplotype scan manually or by software. Also, neighboring alleles on parental chromosomes are likely inherited together, meaning small loci on either side of a recessive lethal allele are likely homozygous and contain some SNP marker. When looking at SNP genotypes in chromosomal order, a candidate locus will be where several successive SNP markers are homozygous in all the cases and none of the controls (Wooliams & Oldenbroek, 2017). In early embryo lethality, the recessive lethal may be identified through the lack of homozygotes in a population. However, this requires a large population of healthy genotypes. This method may look at every SNP or haplotype (VanRaden et al., 2011). The latter is preferred by many due to the expected linkage to causative variants and fewer false positive discoveries.

Haplotypes are inspected and scanned for homozygous loci shared by affected animals and not with unaffected family members. Calculating haplotypes is a complex and time-consuming process that relies heavily on accurate allele frequency data (Kruglyak et al., 1995). However, haplotypes that are never seen in homozygous individuals can suggest the presence of recessive mutations that are likely embryonic lethal. These methods can be handy for studying rare phenotypes in small populations with limited available data (Häfliger et al., 2021). Without familial DNA in the analysis, the haplotype approach is prone to more type I errors and cannot correct them without the familial information (Kirk & Cardon, 2002). Haplotype homozygosity detecting of recessive lethal alleles has successfully identified recessive lethal alleles and loci in livestock species, such as cattle (Casellas et al., 2017; Fritz et al., 2013;

Hoff et al., 2017; VanRaden et al., 2011), mice (Casellas et al., 2012), pigs (Charlier et al., 2016; Derks et al., 2017, 2018, 2019), chickens (Derks, Megens, et al., 2018).

2.3.3 Characterization of Known Genes

Suppose an association between a locus of markers and the transmission of the phenotype is detected. In that case, the linked locus may harbor several genes with known functions relevant to the investigated phenotype. Testing the locus association with the phenotype can be done by comparing genotypes of cases and controls. Retrieving associations between a locus and genes may reveal already identified genes with known functions, plausibly related to the phenotype, called candidate genes (Thomas, 2004).

Further, Thomas (2004) states that typical studies of candidate gene associations use a population-based case-control design with unrelated controls, tested with a Chi-square test. The case-parent trio design for association studies uses the non-transmitted alleles from parents as a comparison group rather than using unrelated controls. The McNemar test is applied to assess whether a specific allele is transmitted more frequently than expected by chance, referred to as the transmission disequilibrium test (TDT). Nevertheless, it is crucial to remember that a significant association does not necessarily indicate a causal effect of the gene on the phenotype and could be due to linkage or an erroneous outcome from an unsuitable choice of controls.

3 Methods and Considerations

This chapter describes the design of the experiment. Test materials used in the experiment are described before explaining the statistical methods used to analyze the data material. Ethical research considerations are in the last chapter.

3.1 Data collection

The 328 Norwegian Fjord-horse genotype samples came from an inbreeding study by Olsen et al. (2020). The genotypes from this study were blood samples collected from December 2015 to March 2016. The samples were taken by veterinarians or a team from the project in eastern Norway. The animals selected for analysis have unknown relatedness, assumed to be random in this analysis. Genotyping was done with Affymetrix High-Density Axiom Equine Genotyping Array (Applied Biosystems).

The gathered genotype information from other animals in the population would have included a minimum of 500 new animals related to the ones in this dataset. This would have increased the population data by approximately 2.5 times, from ~330 animals to ~830, drastically increasing the results' reliability. The new data would also have facilitated a segregation analysis and the reverse genetic screening of the haplotypes as done in other literature. The analysis was planned for when this data was to arrive as a reverse genetic approach analyzing trios and duos in the data (Häfliger et al., 2021; VanRaden et al., 2011). The reverse genetic approach would reveal the haplotypes segregating from parents to offspring and the detrimental haplotypes in carrier matings with reduced fertility. With the many studies having successful results with this SNP-based homozygosity mapping and reverse genetic screening, there is no doubt that the methods are valuable for analyzing recessive lethal alleles. However, to be precise, there should be a large amount of data or family data, preferably both (Charlier et al., 2016; VanRaden et al., 2011). These analyzing methods would have thrown out haplotypes with no phenotypic effect.

3.2 Analysis

The code description chapter will describe the functions used to perform the analysis, while the remaining chapters provide an overview of the method used. The analysis used Rstudio (v.4.1.0) (Rstudio Team, 2020) with different packages.

3.2.1 Software

3.2.1.1 PLINK

Plink (v.1.9) is a powerful and flexible open-source software for analyzing genomic data. Shaun Purcell is developing PLINK with the support of others (Chang et al., 2015). The software was developed for genome-wide association studies (GWAS) to compare genetic variations across many animals to identify genomic loci associated with traits or diseases. PLINK can handle large datasets and perform various analyses, including quality control, inbreeding identification, and statistical calculations. PLINK also supports a wide range of input file formats, conversion of formats, and a command-line interface with options for customizing analyses. PLINK is widely used in the genetics research community and has been cited in numerous scientific publications.

3.2.1.2 SHAPEIT

SHAPEIT (v2.r904) (Delaneau et al., 2013, 2014) is a widely used open-source software tool that helps geneticists and researchers to determine the haplotype or combination of alleles inherited from each parent at each location in the genome, a process known as phasing. Accurate phasing is crucial in GWAS as it allows the identification of genetic variations associated with diseases or traits. SHAPEIT (v2) employs a statistical algorithm based on a hidden Markov model (HMM) that takes advantage of linkage disequilibrium (LD) between nearby genetic markers to infer haplotypes. It also provides users with tools for quality control and error detection, enabling them to identify and rectify any inconsistencies in the data. The software supports various input file formats, including PLINK binary PED and MAP files, VCF files, and Oxford Gen format (".gen") files. It is highly parallelized, running efficiently on high-performance computing clusters. With its speed, efficiency, and ability to handle large datasets, SHAPEIT (v2) is an essential tool for geneticists and researchers in genomics.

3.2.1.3 GHap

GHap (v.3.0.0) (Utsunomiya et al., 2016, 2020) is an R package that facilitates haplotype estimation and association analysis in GWAS. It is a user-friendly tool designed to help geneticists and researchers identify genetic variations that may be associated with a particular disease or trait by analyzing genotype data. GHap implements several statistical methods for haplotype estimation and association analysis, including the haplotype frequency estimation method based on the expectation-maximization (EM) algorithm, the haplotype-based association test (HBAT), and the permutation-based association test (PBAT). In addition to these methods, GHap includes functions for handling genetic data in various file formats, performing data quality control, and conducting several types of analyses, including single-marker and haplotype association testing. Furthermore, the package provides functions for generating publication-quality graphics and visualizing the results of the studies. GHap is a valuable tool for geneticists and researchers engaged in GWAS who require haplotype analysis and association testing, with its easy-to-use interface and compatibility with multiple data formats.

3.2.2 Code Description

This chapter will describe the code, while the actual code remains on the authors GitHub page under the 'Detecting Recessive Detrimental Alleles' repository (https://github.com/NattyandMinnie/Detecting-Recessive-Detrimental-Alleles). The code is described in sections following the GitHub scripts.

1. **Updating SNP Position (Part 1)**

This R script filters out the positions relevant to the dataset and writes it to a file necessary for the Plink updating of the positions in the Fjord-horse dataset to EquCab3, initially aligned to EquCab2. The first part of the code loads the essential package "tidyverse" (v.1.3.1) (Wickham et al., 2019), providing tools for manipulating and visualizing data. Throughout the scripting, the "%>%" is used; this is a pipe operator that allows the chaining of operations on data. Further, the file containing SNP positions from the complete file was read in with the "read.csv" function, skipping the first 20 lines of information comments and removing whitespace. The file containing SNP positions in the Fjord-horse dataset is read with the "read_delim" function, skipping three lines of information comments. This table also gets whitespace removed in the 'SNP' column, setting the 'BP' column as numeric before removing all white space in character columns using "mutate," "str_remove," and "str_trim." Desired columns from the two

datasets are subset into two new datasets, renaming the column containing SNP ids' to be "SNP" with the "names" function. Next, the two new datasets are joined with "left_join" and remove the genetic positions, not in the Fjord-horse dataset with "na.omit". Columns in the new, joined table are renamed and subset a second time only to include the SNP identification and the new position. The missing values rows ('---') in the 'New_pos' column were changed to be '0' using the assignment operator "< - "before writing this output to a text file with "write_delim". The next part does the same for chromosome number but includes one line where the finished data gets a value replaced using "mutate," "str_replace_all", and regex pattern. All the SNPs with a chromosome position starting with "NW_" was replaced with "0" not to complicate further analysis.

2. Updating SNP Positions (Part 2)

This Bash script updates the genomic position and chromosome number for the SNPs in the dataset. The script also includes a part that will list duplicated variants but only list the SNP identifications necessary for filtering them out before analysis. First, the scripts load Plink using the "module load" command, ensuring the module is available. Then the input files are specified after the "—file" option, used when one has a .ped and .map file. Further, the "—chr-set" option is used to specify the number of chromosomes in the dataset, as the default is human (22 autosomes). The "—update-chr" and "—update-map" options specify the file paths for the updated SNP positions on the chromosome and genomic position, respectively. "—make-bed" determines that the output files should be in plink binary format, as often needed by plink in combination with options. The "—list-duplicate-vars" option with the "ids-only" argument lists the duplicated SNPs without including the other information, such as position, making it directly usable in later exclusion of the duplicates. This option also leaves out the first variant of all duplicates so that only one SNP will remain on each duplicated position when they are excluded in later analysis. Finally, the "—out" option specifies the prefix and file path of the output plink files. This script provides a straightforward and efficient way to update SNP positions in a Plink dataset using Plink software.

3. Plink Quality Control

This Bash script provides quality control of the data and removes problematic SNPs to ease the accuracy and reliability of analysis done in later scripts. Plink offers many more options for filtering than the ones used. The data quality control (QC) is executed with the Plink software. First, the "plink" module is loaded. Then the input files are specified as "—bfile," pointing to a prefix of binary plink files. "—make-bed" specifies that the output should be in a plink binary format, and the "—chr-set" specifies the number of chromosomes in the dataset as the data is non-human. The QC steps performed by this script are set under various plink options. The "—geno" option specifies that SNPs with a call rate below a certain threshold (default is 0.1) should be removed from the dataset. The "--alleleACGT" option specifies that SNPs with ambiguous or non-standard alleles should be removed from the dataset. The "--exclude" option specifies a list of SNPs to be excluded based on the "rsID" of the SNP, which is listed in the "plinkduplicates.dupvar" file from the previous script. Finally, the "--out" option specifies the prefix of the output plink files. This script provides a straightforward and efficient way to perform QC on a Plink dataset using Plink software. By removing problematic SNPs, this QC step can improve the accuracy and reliability of downstream analyses.

4. Plink Hardy Weinberg Equilibrium

This Bash script uses the "plink" software to calculate the HWE deviation for each SNP in a Plink dataset. The script begins by loading the "plink" module using the "module load" command, ensuring that the plink software is available for use in the script. Next, the script runs the "plink" command to calculate the HWE deviation for each SNP in the input plink files. The input files for this command are specified using the "--bfile" option, which points to a prefix of the input binary plink files. The "--hardy" option specifies that the HWE deviation should be calculated. The "--nonfounders" option specifies that only non-founder animals should be used for the calculation, as pedigree was unavailable. The "--chr-set" option specifies the number of chromosomes in the dataset. Finally, the "--out" option defines the prefix of the output files. By calculating HWE deviation for each SNP, this script provides a critical quality control step to identify SNPs that deviate significantly from HWE. Such SNPs may have genotyping errors, population stratification, or other downstream analysis issues. Therefore, this script can ensure that the data used for analysis are high quality and reduce the risk of false-positive findings.

5. Separate Chromosomes

This Bash script separates the input plink dataset by chromosome for use with the SHAPEIT software. The script begins by loading the "plink" module, then uses a "for" loop to process each chromosome in the dataset. The "seq 1 35" command generates a sequence of numbers from 1 to 35, which are then used to specify the chromosome number in the Plink dataset. The script runs the command for each chromosome to extract the SNPs located on the specified chromosome. The input files for this command are determined using the "--bfile" option, which points to the input plink binary files, and the "--chr-set" option, which defines the total number of chromosomes in the dataset. The "--chr" option is used to specify the current chromosome being processed. The "--recode" option specifies that the output should be in the plink recode format (.ped and .map files). Finally, the "--out" option defines the prefix of the output files. The output of this script is a set of plink files, one for each chromosome, containing only the SNPs located on that chromosome. These files can be used as input to the SHAPEIT software, which requires chromosome-separated data to perform phasing.

6. Phasing with SHAPEIT

This Bash script is used for phasing genomic data with SHAPEIT. The script loads the required modules and then iterates through all 35 (31, X, Y, XY, and mitochondrial) chromosomes using a "for" loop. The script calls SHAPEIT for each chromosome and provides the necessary inputs: the genotype data in ped/map format, the output file names for the phased haplotypes, and the sample information. The output files are written to a file named "ChrX.phased.haps" and "ChrX.phased.sample", where X is the chromosome number.

7. GHap with SHAPEIT files

This Bash script runs R scripts and submits several batch jobs using the "sbatch" command. The first line loads the R module version 4.1.0, allowing the R scripts to be run. The following line runs an R script called ghap1.R using the Rscript command. The following six lines use the sbatch command to submit different batch jobs for scripts run_ghap.sh with arguments 2 through 7. These jobs are given names

with the --job-name option. Overall, this script runs a series of R and batch scripts needed for the haplotyping of the data.

8. GHap haploblocks

This simple Bash script loads the R module and runs a specific R script called "ghap$1.R", where "$1" is a command line argument passed to the script when it is run. The script expects to be called with a single argument (e.g., ‘./script.sh 2’), which specifies which "ghap" script to run. The "$1" variable is then used to replace the "$1" in the filename of the R script to be executed. The purpose of this script is to run a set of R scripts sequentially as separate jobs on a computing cluster or system, with each script potentially performing a different task.

8.1 Make executable

This command changes the permission of two files, "run_ghap.sh" and "ghap2-7.R," by adding execute permission. The "chmod" command stands for "change mode" and modifies access permissions for files and directories. The "+x" option adds the execute permission to the files, meaning the files can be run as scripts. After running this command, the two files (explained in 7. And 8.) can be executed as shell scripts by running "./run_ghap.ssh" or "./ghap2-7.R".

9. R scripts for GHap

This R script is the R code haplotyping with the package "GHap." Part of the script was written separately for every window size and sliding size wanted in the analysis. The separate R scripts made it possible to run each haplotyping separately with the scripts above (7-8.1). The scripts are the same for all but have varying window sizes and slide sizes and start with one script performing the following tasks:

Loads the required R packages "GHap," sets the working directory with "setwd," and creates ghap-files (.samples, .markers, and .phase) from the SHAPEIT output files (.haps and .samples) for each chromosome. The first line here extracts the sample IDs from the .sample file of chromosome 1 and saves them to a file called "Chr.samples". It is unnecessary to do the same for every chromosome, as the sample IDs are the same for all chromosome numbers. Next, the code loops over each chromosome, extracting the marker data (i.e., position, rsID) from the .haps file, and saves it to a file called "Chr.markers". Then, it loops over each chromosome, extracts the non-binary phased data (phasing information for each animal) from the .haps file, and saves it to a "Chr.phase" file. Finally, the "GHap" package is used to compress ("ghap.compress") the non-binary phased data (phase/markers/samples) for each chromosome into binary phaseb files, with the prefix "Chr."

The next section of the code is a part of the "GHap" haplotyping with different window sizes. This part is repeated for every desired size. The code performs haplotyping and statistical analysis on genetic data. The general structure of the scripts starts with loading the necessary packages ("GHap", and "tidyverse") and setting the working directory by "setwd." The second part of the code loads the phased genotype data in the ".phase" format into the R environment using the "ghap.loadphase()" function. The third block of code generates blocks and performs haplotyping on the phased data using the "ghap.blockgen()" and "ghap.haplotyping()" functions. It then loads the resulting haplotype data into the

R environment using the "ghap.loadhaplo()" function. Statistical calculations on the haplotype data are performed using the "ghap.hapstats()" function. It calculates the probability of observing fewer or greater homozygotes than expected ("Pless" and "Pexcess," respectively) and calculates the overall P-value based on these probabilities. The last block of code subsets and annotates the haplotype statistics data based on different P-value thresholds before saving them as separate files with the "write_delim" function. Finally, the code closes the graphics device.

10. Annotating SNPs (Part 1)

The haplotype statistics from the previous code chunk were written out to a file called "hapstatsX.txt," where X is the sliding size of the haplotype provided and unique to every run. This code was created to calculate adjusted p-values, filtering out significant haplotype blocks and finding the SNPs within these haplotypes. The code is in R and starts with loading the package "tidyverse" and reading in the haplotype statistics files with the "read_delim" function, assigning them to an object. The next part is the adjustment of p-values, correcting the values for multiple testing with the "Bonferroni" argument as the correction method before assigning them to a new column within the haplotype statistics table. The haplotype statistics tables are then filtered with the "filter" function only to include haploblocks that have less or equal to one observed homozygote ("O.HOM") and an adjusted p-value ("padj") less than 0.001. The haplotype statistics for all the different windows- and slide sizes were then bound together with the "rbind" function. Then, the ".map" and ".hwe" files are read into the environment with the functions "read_table" and "read_delim," respectively. The tables get columns renamed with the "names" function, and an empty column is removed with "subset" and "select" before a column joins the two tables with unique row content ("SNP," being the SNP identifications).

Further, the joined table has the positions column "BP" made numeric, renaming a column with the dplyr (v.1.0.8) "rename" function. The joined table was then filtered to only contain SNPs within the specific chromosome number "CHR" and base pair position "BP" of the significant haplotypes and assigned to separate objects before being bound together to one table. The data frame is sorted to be ascending in the chromosome number. Also, before creating the specific variant identification needed for Ensembl, the base pair positions were made into a separate column in the table, dropping variants missing the first allele "A1". This column was then written to a text file using "write."

11. Ensembl VEP Online Tool

This code runs a program called "VEP," which stands for Variant Effect Predictor. This program predicts the effects of genetic variants, such as single nucleotide polymorphisms (SNPs), on the functions of genes and their encoded proteins. The web tool was initially used, but Ensembl provides the command line for every web run, which this section uses to describe the VEP analysis. The documentation for the VEP script is used to gather information about the functions (*Variant Effect Predictor - Command Line VEP*, n.d.). The first option used in this code was "—appris" telling VEP to use the APRIS database to add isoform annotation for this transcript. Next is the "—biotype," telling VEP to include the biotype of the transcript in the output before "buffer_size" with the argument "5000". The buffer size option (default "5000") sets a given number of variants read into memory simultaneously. "—check_existing" tells VEP to check if there are known variants co-located with the input of data comparing alleles and variants on an allele-specific basis. The option "—distance" with the argument "5000" tells VEP to annotate variants within 5000 base pairs of a known feature such as gene or regulatory element. "—mane" tells VEP to

prioritize Matched Annotation from NCBI and EMBL-EBI (MANE) transcripts when predicting the effects of variants.

Further, the "—sift" option with the "b" argument tells VEP to predict the amino acid substitution effect on protein function based on sequence homology and physical properties of amino acids. The "b" means to include the prediction and score in the output. "--species equus_caballus" sets the species to be analyzed to Equus caballus, or the domestic horse, whereas "—symbol" tells VEP to include gene symbols in output. "—transcript_version," tells VEP to include version numbers, and "—tsl" adds the transcript support level to the output. "---cache" tells VEP to use a cache of stored data to speed up annotation. The "—input_file" option sets the input data with a file path and filename. However, this code is copied from a code-line version from the web tool; therefore, no file is included in the command line here. The same goes for the output file, which the "—output_file" option can specify. The VEP output in the web tool can be downloaded in different formats.

12. Annotating SNPs (Part 2)

This code manipulates data on a file containing genetic information from the Ensembl database. The code reads in the text file from Ensembl VEP using "read.delim", passing the arguments that the data frame has a header and is separated by tabs ("\t"). Further, the chromosome number is extracted from a column called 'Location' using the "mutate" and "str_extract" function into a new column called 'CHR.' "str_extract" need a regex pattern to extract, where ".+?:" specifies it should extract all an infinite number ("+") of symbols/characters/numbers ("."), until the last ("?") being a colon ":". The colon symbol is removed by the "mutate" and "str_remove" functions. The same is done for the base pair position in the 'Location' column. The basepair positions were the same for "basepair start" – "basepair end," and the latter was extracted using the regex pattern "-.+" extracting all ("+") symbols/characters/digits (".") after the "-" symbol into a new column called 'POS'. After that, the hyphen is removed from the 'POS' column. Then the dataset gets many of its columns removed, keeping only the most exciting columns using the "subset" function with the "select = -c()" argument to deselect named columns. The rows where the gene column has the value of "-" are filtered out and assigned an object before they are grouped ("group_by") by the 'Gene' value and again filtered only to keep unique rows within the dataset. Next, the 'BP' column is renamed to 'POS' ("mutate") and made numeric ("as.numeric") before joining this table with the table containing SNPs and their Hardy-Weinberg statistics (created in step 10). The resulting data has only the rows of the significant SNPs.

3.2.3 Quality Control

The dataset collected for previous studies (Olsen et al., 2020) had the SNP genotypes called with Axiom Analysis Suite (v. 2.0.0.35). The genotypes were filtered for a Dish QC with a threshold of 0.82. The Dish QC excludes low-quality DNA by the expected clear signals of non-polymorphic loci. Further, the QC call rate was applied, filtering out the sample SNPs with less than 0.97 calls. The genotype data obtained in this study have 506128 SNPs and 328 subjects.

Before filtering the genotype data, a file with updated (EquCab3) SNP positions and identifications from Thermo Fisher Scientific was used to identify and correct the SNPs' new positions and chromosome numbers in the dataset (initially EquCab2). The SNPs assigned unknown chromosome locations were excluded from further analysis. Several duplicated variant positions were also discovered, and all but one per position were removed from the dataset. Further, the Hardy-Weinberg Equilibrium (HWE)

deviation was calculated for each SNP in a Plink dataset. The dataset was split up into files based on their chromosome number, as needed by the SHAPEIT (v2.17) (Delaneau et al., 2014) phasing software.

3.2.4 Homozygosity Mapping

SHAPEIT (v2.17) (Delaneau et al., 2014) conducted phasing on the data further used in Rstudio to detect haplotypes. In Rstudio, the package "GHap" (Utsunomiya et al., 2016) was used to obtain haplotype blocks and calculate the statistics of the blocks. The windows were of different sizes (700, 500, 300, 100, 40, and 10kbp), sliding down the genome with different-sized jumps (150, 100, 60, 20, 5, and 1kbp, respectively). A two-sided Poisson test giving p-values was calculated separately to highlight the differences between the excess and depletion of homozygotes. Further, the p-values were calculated, correcting for multiple testing by the "Bonferroni" method. Only the haplotypes with less or equal to one homozygote observed and an adjusted p-value of 0.001 were kept for further analysis. The SNPs found within the significant haplotype blocks were used in Ensembl Variant Effect Predictor (VEP) (v.109) tool (McLaren et al., 2016).

The significant haplotype blocks obtained related to a known gene can be seen in Table 4.1. The table has fewer columns than the original table from the analysis because not all the columns are relevant or interesting for the thesis. Further, many estimations of P were obtained in the original table. The 'Padj' is the Bonferroni correction of the P-values and, therefore, the only one included in this table. The author manually added the length column to easily distinguish the block sizes in the different chromosomes. The estimated numbers of some columns (e.g., 'E.HOM', 'Padj') were limited to a reasonable number of digits, whereas the original table would include 6-15 digits. Further, the haploblock names were shortened to make the table easier to look at, interpretable, and comparable. The block column gives a unique identification for each haplotype search with a chromosome identificatory, followed by a block identification (i.e., chr_1_B1234, here; B1234). The 'CHR' column specifies which chromosome the block is and the positions of the block found as start position 'BP1' and end position 'BP2'. Further, SNP alleles within the window are included in the 'ALLELE' column, followed by the block frequency within the relative population. 'O.HOM.', 'O.HET', and 'E.HOM' are observed homozygosity, heterozygosity, and expected homozygosity, respectively. GHap calculated these values with the Hardy-Weinberg expectance. 'Padj' is the Bonferroni adjusted P-values of all the haploblocks found by GHap, and 'Size' is the respective length of the block in kilobase pair (kbp). The table is sorted ascending by chromosome, start- and end-position of the blocks.

The dataset's '.hwe' and '.map' files were merged using the SNP identifier before filtering only to include the SNPs within the significant haploblocks. The plink data on the SNPs can be seen in Table 8.3 (the result section version is Table 4.3). The first column is the chromosome number where the SNP resides, followed by the SNP identification and the physical position of the SNP (BP). The major and minor alleles ('A1' and 'A2' respectively), and the genotypes as; A1A1/A1A2/A2A2. The observed and expected heterozygosity is calculated, and the respective P value by Plink ('hwe' option) is calculated. The table is sorted ascending by chromosome, then position. The SNPs were further used with Ensembl to assess the locations' possible association with known genes, loci, and other features provided by Ensembl. The author adds new columns extracting each genotype into a separate column. The author calculates the expected homozygotes as follows;

$$\boldsymbol{E}(HOM) = q^2 N, \qquad \textbf{(eq.9)}$$

N is the number of animals in the population (Todd, Thomson, et al., 2020).

The author also calculates the survival rate of the significant loci assumed to be independent. The calculation of the survival rate was done as follows;

$$\prod_{i=1}^{n}(1-p_i^2), \quad \textbf{(eq.10)}$$

Where p_i is the haploblock frequency, and n is the number of haploblocks. This calculation will say something about the effect of the result from the analysis. For example, for the haploblock frequency of 0.3, the lethality is the frequency of the minor allele homozygote, which then is at 0.09. If there are five independent loci, the total survival rate would be one minus the lethality rise to the power of fifth, i.e., 0.624 survival rate. This means there is a ~62% chance of surviving in this population if all five candidate loci are true.

3.2.5 Characterization of Candidate Loci

Candidate haplotypes were characterized against publicly available ENSEMBL- VEP (v.109) (Cunningham et al., 2022). The ENSEMBL- Variant Effect Prediction (VEP) tool is web-based to search for loci in nearby loci of a given dataset. The detection settings were at default: 5000 bp upstream or downstream of a given variant. ENSEMBL -VEP reports back a candidate gene or loci in the nearby area of the variant and other information. The variants reported are further investigated on horses and orthologs in cows, dogs, mice, and humans to determine the effects and function of the genes.

In the output table from Ensembl, all impacts were classified as "Modifier" by Ensembl. Because of this, the “Impact” column was removed in Table 4.3. The table describes the SNP's chromosome, position, strand, and allele from the input. Further, it associated the SNP with a gene loci and the respective gene name if it is known. Ensembl also predicted the consequence of SNP alleles as to where it is to the respective gene. The 'downstream_gene' means the SNP is downstream of the gene, and the 'intron' is within an intron of a gene. More information about the meaning of the consequences can be found on the Ensembl page (https://www.ensembl.org/info/genome/variation/prediction/predicted_data.html). Duplicates of the results were deleted if all columns were identical.

A search in the NCBI database was done for the known genes found by Ensembl (Table 8.4). DEFA 22 and DEFA5L are predicted to be orthologous to several other DEFA genes in mice and humans. However, both DEFA was found to be orthologues with the human DEFA4 and mouse orthologue defa-ps8 by the “orthologs” section of NCBI linking to a website called ‘Orthodb’ (https://www.orthodb.org/). The many orthologues are therefore not listed, and only the ones found most orthologous by ‘Orthodb’. The same is for the U6, KIAA1109, and OR5AS1 genes without NCBI-linked orthologues.

3.2.6 Phasing Uncertainty

Phasing uncertainty was to be obtained on the phased chromosomes. This is an option to be done by SHAPEIT “—output-graph”. The likelihood of the obtained haplotypes would have been obtained and presented in the thesis; however, this was not completed due to limitations in the server used for analysis. The idea of this uncertainty calculation is to take a large dataset and divide it into smaller subsets relevant to an analysis or question. The algorithm then calculates the most likely haplotype pair based on the subset of variants for each sample. Additionally, it includes the probability or confidence level of that haplotype being correct.

3.3 Ethical Considerations of Research

Research ethics is a joint effort between research quality and ethics, with values, norms, and institutional agreements (Den nasjonale forskningsetiske komité for samfunnsvitenskap og humaniora (NESH), 2021). This means a researcher has a professional and ethical obligation to all the study objects. Norwegian researchers follow The National Research Ethics Committees (NESH) guidelines. Anonymity and confidentiality are essential when research involves people, as in a genetic analysis of privately owned animals. NESH puts a responsibility on the researcher to keep a fundamental respect for human dignity. According to NESH, the law about personal data secures that the data material is anonymized and that privacy is maintained so that the research cannot harm the participants. The following cautions were taken to preserve the integrity of the study:

- Genotype data are assigned a number instead of the names of the horses.
- There is no contact information about the owners in any of the datasets.
- Participation in the study was voluntary with informed consent.

The National Committee for Research Ethics in Science and Technology (NENT) guidelines for animals used in research specify that animals are moral objects that deserve respect (Den nasjonale forskningsetiske komité for naturvitenskap og teknologi (NENT), 2018). With this, the researcher must maintain and respect animal welfare when preparing and conducting the research. In this study, the animals have been the subjects of the investigation. Further, the genotype data used in previous studies comply with the laws for animal research.

"I acknowledge that I am a part of an international community of researchers. I will practice my activities per the recognized standards for good research practice. I shall conduct my research honestly and truthfully and show respect for humans, animals, and nature. I shall use my knowledge and skills to the best of my judgment for the good of humanity and sustainable development. I shall not allow interests based on ideology, religion, ethnicity, prejudice, or material advantages to overshadow my ethical responsibility as a researcher." - (The National Committee for Research Ethics in Science and Technology (NENT), 2016, p.20)

4 Results

Analysis of genotype data from Norwegian Fjord-horse (n = 328) identified 48 haploblocks with a homozygote deficiency associated with a known gene. The haploblocks (Table 4.1) included haploblocks of size 100-10kbp, distributed over six chromosomes (2, 12, 18, 27). The different window sizes are not found on all chromosomes; the medium-sized (100-40kbp) is found on chromosomes 12 and 27; the small haploblocks are found on all significant chromosomes. The table of all significant haploblocks detected is in the appendices (Table 8.1). It is crucial to keep in mind that the data represented in the result and discussed are only the haploblocks and SNPs with known genes associated with them.

The blocks within chromosome 2 are two blocks of 10kbp, which overlap and together span 11kbp. The number of SNPs within these blocks is seven and nine, whereas the last block contains the first block's seven SNPs. The frequency of both haplotypes was 30%, where 196 animals were observed as heterozygotes for the haploblocks. The expected homozygosity was ~29 animals, with zero observed homozygotes. The Bonferroni-correction P-value of these blocks is significant at ~8e-7.

In chromosome 12, there are two different loci with significant blocks. GHap detected both loci with 10, 40, and 100kbp window sizes. The first locus is at chromosome 12, with blocks overlapping from position 15720001-15900000 (total length 180kbp) by 100kbp window sizes. Here, the alleles are between five and six, overlapping +/- one SNP of the preceding block. The frequency observed of these 100kbp blocks was 27-29%, with 178-187 observed heterozygotes. The observed homozygotes of these blocks were at 0, while the expected varied between ~24-26 animals. The adjusted p-value was between ~4e-6 and ~5e-5. The 40kbp window size detects haploblocks from 15780001-15855000 (total length 75kbp) with four to five SNPs within each block. The SNPs have frequencies between 27-29%, with 178-187 heterozygotes. The observed homozygosity within these blocks was zero, and the expected was between 24-26 animals with an adjusted p-value of 1e-4 to 9e-6. Lastly, the overlaps of 10kbp are at 15806001-15819000 (total length 13kbp), with two to three SNPs. The frequency of the haploblocks is at 29%, where there are 188 heterozygous, zero observed, and ~27 expected homozygotes. The adjusted p-value for these blocks was at 9e-6.

The second locus is a smaller region, with two 100kbp blocks detecting in the same position (29220001-29320000) (total length of 100kbp). The first block has five SNPs with the alleles AAGTA, while the other has the alleles AAGCA. Both the blocks are observed as heterozygotes (189 and 198, respectively), while none are observed in homozygotes (expected from ~27-30 animals). The next is the 40kbp block detected in 29280001-29320000 (total length of 40kbp) with three SNPs (GCA) observed with a 44% frequency, 291 heterozygotes, ~65 expected homozygotes and a p-adjusted value of 3e-22. The two blocks of 10kbp are found in 29310001-29321000 (11kbp total size). The first block has the same SNPs as the 40kbp block and the same statistics besides the p-adjusted being 4e-22. The last 10kbp block contains an extra SNP. This block has 42% frequency in the study population, 272 observed and ~56 expected homozygotes, and a p-adjusted of 1e-18.

In chromosome 18, only three blocks of size 10kbp were detected as significant. The haploblocks overlap in position (74587001-74599000) (12kbp length). The first two blocks contain the same SNPs, and since they are the same size, the exact statistics. They contain three SNPs with a frequency of 33%, where 211 animals are heterozygotes. Additionally, there is one observed homozygote, while the expected homozygotes were at ~35 animals. The p-adjusted was at 1e-7. For the other haploblock, the three SNPs

and one additional SNP are included with a population frequency of 29%. The expected homozygotes were ~28, one observed homozygote, and 191 observed heterozygotes.

Lastly, chromosome 27 also contains several overlapping haploblocks of different sizes. The most extended haploblock size here was at 100kbp with positions 33720001-33900000 (total length of 180kbp). The first three blocks of this size seem to contain the same SNPs, while the last contains one additional SNP. The frequency of the four first haploblocks is at 35%, with 231 observed heterozygotes. The expected homozygotes were at ~41, and the p-adjusted value was at 3e-12. For the last haploblock of this size, the frequency became 31%, with 201 observed heterozygotes and ~31 expected homozygotes. The adjusted p-value was 7e-8, and neither of the blocks had any observed homozygotes. The 40kbp haploblocks span the area of 33780001-33850000 (total length of 70kbp). All the haploblocks of this size contain the same SNPs and, therefore, the same statistics. The frequency of the alleles was 35%, with 231 observed heterozygotes, ~41 expected homozygotes, and zero observed homozygotes. The p-adjusted value is at 9e-12. For the smallest window size (10kbp), the observations are at 33806001-33824000 (total length of 18kbp). All the observations contain the same two SNPs, zero observed and ~41 expected homozygotes. There are also 231 observed heterozygotes and a p-adjusted value of 9e-12.

22 SNPs are found within the significant haploblocks with gene association (Table 4.2). This number of SNPs may initially seem like few, but the overlapping windows of the haploblocks must be remembered. Many SNPs have a depletion of minor homozygotes and significant deviation from the HWE. The power of the deviation is seen by the p-value (“P”), and the deficit is not always prominent. The table containing all the significant SNPs is in appendices (Table 8.2).

Table 4.3 from Ensembl returns many loci associated with the SNPs but only a handful of genes with known functions. The consequences of the known genes are intron, downstream- and upstream of the respective genes. Several loci are associated with a specific position, but with different meanings, i.e., either as a downstream gene variant or an intron variant of a gene.

The tissue expression of the known genes in this analysis was not found in several orthologues. The expression of the defensins was restricted expression to the large intestine of adult mice, and the human orthologue found a restricted expression in the bone marrow. The NCBI summary description implicates the gene in the immune system defense in the mucosa. Further, DPF2 is expressed in the testis and ovary among many other tissues in humans and thymus and limbs among many other tissues in mice. The summary of the gene indicates that the gene has an essential role in the cell response to environmental factors. The KIAA1109 was not found in the human, mouse, or horse in the NCBI database. When looking into the zebrafish KIAA1109, the human and mouse orthologue- the BLTP1 and bltp1, respectively, appears. NCBI also list the human and mouse orthologue name "KIA1109" or "kia1109" as a 'previously known' name of the genes for human and mice. The expression of BLTP1 was found to be in various tissues but most abundant in the ovaries and thyroid. For the mice, the expression of the gene was found in the adult testis and "CNS," the central nervous system. The NCBI summary description of the gene indicates its role in cell differentiation, among other things, which may be interesting for this study. OR5AS1 has no tissue expression data but is described as a receptor in the nasal epithelium that interacts with molecules and triggers the perception of a smell. The gene is found in a range of animals, including humans. The U6 gene is not found as an NCBI-connected orthologue of humans or mice, but after a few web searches, NCBI provides a gene of humans and mice called RNU6-1 and rnu6 known as "U6". These genes are therefore included as the orthologues of the U6 equine loci. Unfortunately, there is little information on the gene orthologues, without proper tissue expression data or a description and prediction of the gene functions.

Table 4.1: A shortened version of Table 8.1 showing the haploblocks that were found significant. This table only keeps the haplotypes for loci associated with functional genes. The columns in the table consist of BLOCK (haplotype block identification), CHR (Chromosome), Size (the size of the haploblock in kbp), BP1, and BP2 (Start and end position of the haploblock). Further, the ALLELE (Alleles of SNPs observed), FREQ (frequency of haploblock in population), O.HOM (Observed homozygotes), O.HET (Observed heterozygotes), E.HOM (Expected homozygotes), and Padj (Bonferroni corrected P-values).

BLOCK	CHR	Size	BP1	BP2	ALLELE	FREQ	O.HOM	O.HET	E.HOM	Padj
B106260	2	10	106259001	106269000	TCAATAA	0.30	0	196	29.3	8.667e-7
B106261	2	10	106260001	106270000	TCAATAATT	0.30	0	196	29.3	8.667e-7
B787	12	100	15720001	15820000	ACCGA	0.29	0	187	26.7	4.872e-6
B788	12	100	15740001	15840000	CCGACC	0.27	0	178	24.2	5.958e-5
B789	12	100	15760001	15860000	CCGACC	0.27	0	178	24.2	5.958e-5
B3157	12	40	15780001	15820000	CCGA	0.29	0	187	26.7	9.728e-6
B790	12	100	15780001	15880000	CCGACC	0.27	0	178	24.2	5.958e-5
B3158	12	40	15785001	15825000	CCGAC	0.29	0	187	26.7	9.728e-6
B3159	12	40	15790001	15830000	CGACC	0.27	0	178	24.2	1.190e-4
B3160	12	40	15795001	15835000	CGACC	0.27	0	178	24.2	1.190e-4
B3161	12	40	15800001	15840000	CGACC	0.27	0	178	24.2	1.190e-4
B791	12	100	15800001	15900000	CGACC	0.27	0	178	24.2	5.958e-5
B3162	12	40	15805001	15845000	CGACC	0.27	0	178	24.2	1.190e-4
B15807	12	10	15806001	15816000	CG	0.29	0	188	26.9	9.010e-6
B15808	12	10	15807001	15817000	CGA	0.29	0	188	26.9	9.010e-6
B15809	12	10	15808001	15818000	CGA	0.29	0	188	26.9	9.010e-6
B15810	12	10	15809001	15819000	CGA	0.29	0	188	26.9	9.010e-6
B3163	12	40	15810001	15850000	GACC	0.27	0	179	24.4	9.062e-5
B3164	12	40	15815001	15855000	GACC	0.27	0	179	24.4	9.062e-5
B1462	12	100	29220001	29320000	AAGTA	0.29	0	189	27.2	2.746e-6
B1462	12	100	29220001	29320000	AAGCA	0.30	0	198	29.9	1.931e-7
B5857	12	40	29280001	29320000	GCA	0.44	0	291	64.5	3.408e-22
B29311	12	10	29310001	29320000	GCA	0.44	0	291	64.5	4.201e-22
B29312	12	10	29311001	29321000	GCAA	0.42	0	272	56.4	1.460e-18
B74588	18	10	74587001	74597000	CCA	0.33	1	211	34.6	1.540e-7
B74589	18	10	74588001	74598000	CCA	0.33	1	211	34.6	1.540e-7
B74590	18	10	74589001	74599000	CCAC	0.29	1	191	28.4	6.200e-5
B1687	27	100	33720001	33820000	GA	0.35	0	231	40.7	3.978e-12
B1688	27	100	33740001	33840000	GA	0.35	0	231	40.7	3.978e-12
B1689	27	100	33760001	33860000	GA	0.35	0	231	40.7	3.978e-12
B6757	27	40	33780001	33820000	GA	0.35	0	231	40.7	7.942e-12
B1690	27	100	33780001	33880000	GA	0.35	0	231	40.7	3.978e-12
B6758	27	40	33785001	33825000	GA	0.35	0	231	40.7	7.942e-12
B6759	27	40	33790001	33830000	GA	0.35	0	231	40.7	7.942e-12
B6760	27	40	33795001	33835000	GA	0.35	0	231	40.7	7.942e-12
B6761	27	40	33800001	33840000	GA	0.35	0	231	40.7	7.942e-12
B1691	27	100	33800001	33900000	GAA	0.31	0	201	30.8	7.755e-8
B6762	27	40	33805001	33845000	GA	0.35	0	231	40.7	7.942e-12
B33807	27	10	33806001	33816000	GA	0.35	0	231	40.7	9.791e-12
B33808	27	10	33807001	33817000	GA	0.35	0	231	40.7	9.791e-12
B33809	27	10	33808001	33818000	GA	0.35	0	231	40.7	9.791e-12
B33810	27	10	33809001	33819000	GA	0.35	0	231	40.7	9.791e-12
B33811	27	10	33810001	33820000	GA	0.35	0	231	40.7	9.791e-12
B6763	27	40	33810001	33850000	GA	0.35	0	231	40.7	7.942e-12
B33812	27	10	33811001	33821000	GA	0.35	0	231	40.7	9.791e-12

B33813	27	10	33812001	33822000	GA	0.35	0	231	40.7	9.791e-12
B33814	27	10	33813001	33823000	GA	0.35	0	231	40.7	9.791e-12
B33815	27	10	33814001	33824000	GA	0.35	0	231	40.7	9.791e-12

Table 4.2: Significant SNPs within the haplotypes connected with known genes, a shorter version of Table 8.2. Within the table, the columns are CHR (chromosome number), SNP (SNP identification), BP (bp position), A1, and A2 (The minor and major allele, respectively). Further, the GENO column is sorted with the minor homozygote (A1A1), followed by the heterozygote (A1A2), and the major allele homozygote (A2A2). PLINK calculated the P value as the deviation from Hardy-Weinberg expectations ('P') and the expected homozygote animals ('E.HOM').

CHR	SNP	BP	A1	A2	GENO	P	E.HOM
2	AX-103711548	106263332	C	T	0/1/327	1	0,5
2	AX-104314691	106265474	T	C	1/18/309	0.2581	10,0
2	AX-104340845	106266479	G	A	12/109/207	0.7331	66,5
2	AX-104003928	106267079	G	A	14/120/194	0.5264	74,0
2	AX-104521486	106267713	C	T	0/1/327	1	0,5
2	AX-103702712	106268127	A	G	15/259/49	3.3699e-31	149,5
2	AX-104191493	106268639	G	A	0/13/314	1	7,5
2	AX-104867635	106269154	C	T	1/18/309	0.2581	10,0
2	AX-104699852	106269311	C	T	18/121/189	0.8803	78,5
12	AX-104766198	15809758	T	C	0/22/306	1	11,0
12	AX-104429419	15815391	G	T	21/166/138	0.002106	107,0
12	AX-104580592	15816191	C	A	2/37/289	0.3611	20,5
12	AX-104160918	29220764	G	A	53/111/160	3.701e-5	112,5
12	AX-104495525	29230538	G	A	29/83/213	2.653e-5	73,5
12	AX-103064397	29317733	A	G	0/15/312	1	8,5
12	AX-103275076	29317857	T	C	26/227/71	4.736e-15	143,5
12	AX-104611438	29319554	C	A	0/75/252	0.01258	38,5
18	AX-104804861	74596153	G	A	24/68/228	2.365e-6	66,0
18	AX-102958495	74598729	T	C	0/51/276	0.2393	26,5
27	AX-104754829	33814845	G	A	22/209/96	1.783e-10	127,5
27	AX-104376096	33815537	G	A	80/115/131	5.431e-7	139,5
27	AX-104958326	33888828	G	A	89/118/113	3.733e-6	156,0

Table 4.3: A shortened version of Table 8.3, showing the genes associated with specific SNPs. In this table, the columns are Chr (chromosome number), Position (bp position), Allele (major allele), and Symbol (Gene symbol).

Chr	Position	Allele	Symbol
2	106263332	T	KIAA1109
2	106265474	C	KIAA1109
2	106266479	A	KIAA1109
2	106267079	A	KIAA1109
2	106267713	T	KIAA1109
2	106268127	G	KIAA1109
2	106268639	A	KIAA1109

2	106269154	T	KIAA1109
2	106269311	T	KIAA1109
12	15809758	C	OR5AS1
12	15815391	T	OR5AS1
12	15816191	A	OR5AS1
12	29220764	A	DPF2
12	29230538	A	DPF2
12	29317733	G	DPF2
12	29317857	C	DPF2
12	29319554	A	DPF2
18	74596153	A	U6
18	74598729	C	U6
27	33814845	A	DEFA22
27	33815537	A	DEFA22
27	33888828	A	DEFA5L

Table 4.4: A calculation of the survival rate in the Norwegian Fjord-horse population. The survival rate is calculated per assumed independent locus, where the P-adjusted was most significant, with the total survival rate in the bottom right corner.

BLOCK	CHR	BP1	FREQ	O.HOM	O.HET	E.HOM	Padj	Survival rate
B106260	2	106259001	0.30	0	196	29.3	8.667e-7	0.910
B15808	12	15807001	0.29	0	188	26.9	9.010e-6	0.916
B5857	12	29280001	0.44	0	291	64.5	3.408e-22	0.806
B74588	18	74587001	0.33	1	211	34.6	1.540e-7	0.891
B33815	27	33814001	0.35	0	231	40.7	9.791e-12	0.878
							Sum	0.526

The survival rate was calculated for the significant loci (Table 4.4.). The survival rate per locus was quite simar, between 0.806-0.916. The haplotype in chromosome 12:15807001 had the most prominent survival rate at 0-916, followed by chromosomes 2 and 18. The lowest survival rate was found in chromosome 12:29280001 at 0.806, followed by the haploblock on chromosome 27 at 0.878. In total, this becomes a survival rate of 0.526 in the population.

4.1 Haplotype Candidate Genes

4.1.1 Defensins (DEFA22 & DEFA5L)

Paneth cell-specific alpha-defensin 22 (DEFA22) is one of the orthologues to the mice gene defa-ps8 and Gm15313 and the human DEFA4 (*Equus Caballus (Domestic Horse; Equine) - DEFA22*, n.d.). Paneth cell-specific alpha-defensin 5L (DEFA5L) is one of the mouse orthologues of defa-ps8, and the human orthologue is DEFA4 (*Equus Caballus (Domestic Horse; Equine) - DEFA5L*, n.d.). Zhai et al. (2023) recently wrote an in-depth review of the defensins, their effect on female and male reproductive tracts, and their role as host defense and reproductive tract fertility defenders. The information on defensins in general, their role, and their effects have been studied and reviewed by many, but Zhais' (2023) review provides a comprehensive overview.

Defensins are small, positively charged molecules produced by immune cells and cells lining various body surfaces. They are essential in the body's defense against harmful bacteria, viruses, and other pathogens (Lehrer & Lu, 2012). Defensins have a characteristic structure, with three sulfur atoms forming bonds to stabilize the molecule. There are three types of defensins, alpha, beta, and theta, which differ in how their sulfur bonds are arranged. Alpha- and beta-defensins are found in animals and humans, while theta-defensins are only found in non-human primates (Cole et al., 2002; Lehrer & Lu, 2012).

The alpha-defensin loci in monkeys and horses have not been fully annotated, but it appears that they express at least 20 alpha-defensins currently in the digestive tract, called Paneth cells (Bruhn et al., 2009; Figueredo et al., 2009; Tanabe et al., 2004). Cows and dogs, however, seem to lack alpha-defensin genes, and the human Paneth cells only express two alpha-defensin. Figueredo et al. (2009) also suggest that the vast number of pseudogenes of alpha-defensins in mice, rats, and monkeys suggest that loss-of-function mutations are frequent in the genes. They also indicate that loss of function in an alpha-defensin does not influence enteric immunity much, especially in species with numerous functional genes. The researchers also point out numerous alpha-defensin genes in mice, monkeys, and horses, but only a handful of their peptides are found at measurable levels.

Zhai et al. (2023) refer to Wira et al. (2005), which says that the cells in the outer cell lining in the digestive, respiratory, and urogenital systems secrete antimicrobial peptides and cytokines. Numerous comprehensive reviews have covered the complex mechanisms behind the antimicrobial activity of defensins, as well as their significance in the mucosal immunity of both the respiratory and digestive systems (Klotman & Chang, 2006; Lehrer & Lu, 2012; Ouellette, 2011; Xu & Lu, 2020; Zhai et al., 2023). The outer cell lining of the reproductive tract is continually subjected to the risk of microbial invasion, much like the cell lining of the digestive and respiratory systems. Microbial invasion can arise from internal and external sources and may result in various reproductive tract infections, including sexually transmitted diseases, internal infections, and infections from medical care. Infections can adversely affect an individual's health and may result in unfavorable reproductive outcomes, including miscarriage, premature delivery, and infertility (Brunham & Paavonen, 2020; Sharma et al., 2022).

Defensins are present in many parts of the reproductive system, and their expression can be affected by hormones, age, and the presence of microorganisms (Cao et al., 2010; Fleming et al., 2003; Quayle et al., 1998; Yu et al., 2013). Defensins also play a role in several reproductive processes, such as maturation and transport of sperm, binding of sperm to the egg, implantation, fetal development, and parturition. (Burris et al., 2020; Figueredo et al., 2009; Tollner et al., 2011).

Alpha-defensins are grouped into myeloid (bone marrow) and enteric (intestines) peptides based on their expression and organization. The human neutrophil peptides 1-4 (HNP1-4) are produced in the bone marrow and are present in various immune system cells (Gabay et al., 1989; Selsted & Ouellette, 2005). On the other hand, the enteric alpha-defensins HNP5 and HNP6 are mainly expressed in the Paneth cells found at the base of intestinal glands, while HD5 is present in the outer cell-linings in the reproductive and urinary tracts (Com et al., 2003; Jones & Bevins, 1992, 1993; Ouellette, 2011; Quayle et al., 1998; Spencer et al., 2012).

Human defensins possess a range of properties such as antibacterial, antiviral, antifungal, and immune regulatory effects, which play a crucial role in various physiological processes, including inflammation, development, and cancer (Holly et al., 2017; Jin & Weinberg, 2019; Lai & Gallo, 2009; Ordonez et al., 2017; Semple & Dorin, 2012). The defensins' antibacterial function is the most extensively researched. The antimicrobial action of defensins involves a three-step process. Firstly, the defensins bind to the

bacterial membrane through electrostatic attraction. Secondly, defensins interact with the bacterial plasma membrane, increasing the membrane's permeability. Finally, the increased permeability of the bacterial membrane causes leakage of the inner cell contents, leading to cell death (de Leeuw et al., 2010; Lehrer & Lu, 2012; Selsted & Ouellette, 2005; Shafee et al., 2017; Zasloff, 2002). Additionally, they can decrease bacterial infection by neutralizing toxins secreted by bacteria. (Kim et al., 2005). The bacterial defense mechanisms in alpha- and beta-defensins differ because of their chemical properties (Lehrer & Lu, 2012). Defensins also have an immune defense against viruses. The defensin concentration and environmental factors such as time, temperature, pH, and ionic strength influence the level of viral suppression by defensins. The alpha-defensins also interact with enveloped viruses by inhibiting fusion to host cells or being absorbed, causing a leak in the envelope (Holly et al., 2017). Defensins have been shown to regulate innate and adaptive immune responses, among other things (Chertov et al., 1996).

Defensin expression can be regulated by various stimuli that bind to receptors on defensin-expressing cells. These stimuli include proteins, lipids, lipoproteins, and nucleic acids (Tang et al., 2012). The defensin expression during pregnancy differs from the upper female reproductive tract to the lower reproductive tract, as the body needs to expel pathogens and keep the growing fetus. The immune cell population in the lower reproductive tracts of females is more diverse than in the upper tract. More specifically, the HNP1-3 has been implicated in female reproductive tract viral and bacterial infection (Guthrie et al., 2015; Levinson et al., 2012; Shust et al., 2009). Defensins have been linked to the maternal defense against bacteria (Das et al., 2007; King, Kelly et al., 2007; King, Paltoo et al., 2007). HNP4 and HD5 are present in all parts of the female reproductive system. HNP4 is found in the vulva, vagina, cervix, uterus, and to a lesser extent, in the endometrium, myometrium, and ovary. Additionally, HD5 is found in the vagina, ectocervix, endocervix, endometrium, and fallopian tubes in a study by Quayle et al. (1998). Yarbrough et al. (2015) found that the female reproductive system's immune response is reduced during days 14-28 of the menstrual cycle to facilitate fertilization and pregnancy. This downregulation of the immune response means that also the defensins vary in expression throughout the menstrual cycle. In the male reproductive tract, the beta-defensins are the most investigated with a biased expression to the male reproductive system, especially in puberty (Patil et al., 2005; Sangeeta & Yenugu, 2019; Yamaguchi et al., 2002; Zaballos et al., 2004). There are few publications regarding alpha-defensins and the male reproductive tract. However, one study by Robertson et al. (2020) found a low level of transcripts of HNP1-3 in the testis and higher levels of HNP3 in the caput and cauda. Another study reported high levels of HNP1-3 expression in normal spermatozoa (Com et al., 2003).

Alpha-Defensins have been found in female reproductive tracts as a response to infection by Human Immunodeficiency Virus (HIV), Human Papilloma Virus (HPV), Herpes Simplex Virus (HSV), Neisseria gonorrhoeae, Chlamydia trachomatis, Gardnerella vaginalis, and Trichomonas vaginalis. The alpha-defensins have also been found as an immune response in male reproductive tracts by HIV, Neisseria gonorrhoeae, Chlamydia trachomatis, and Escherichia coli.

Defensin levels in fertilization, implantation, and pregnancy phases have been reported. During implantation, HNP1-3 is expressed by the endometrial stromal cells, embryonic fluid, outer cells of a blastocyst, and the birthing custard of newborn babies (Akinbi et al., 2004; Das et al., 2007; Espinoza et al., 2003; Svinarich et al., 1997). HD-5 is also present in the fetal membrane by Svinarich et al. (1997). Additionally, Espinoza and colleagues (2003) discovered that HNP1-3 levels increase during microbial invasions of the innermost layer of the placenta, premature delivery, and rupture of fetal membranes. Studies investigating defensin expression concerning male fertility are mainly done on beta-defensins, mostly in animals (Zhai et al., 2023).

4.1.2 DPF2

The Double Plant-Homeo-Domain Fingers 2 (DPF2) codes for a zinc finger protein transcription factor. Although the exact pathways and functions of the gene are not well mapped out, it has been shown to have a regulatory role in bone marrow differentiation. The gene is also recognized as an early-apoptosis-signaling gene and an epigenetic regulator by recognizing histone modification (Gabigs et al., 1994; Huber et al., 2017; Lim et al., 2010; Tando et al., 2010; Wong et al., 2006).

Mutations in DPF2 have been associated with variants of the Coffin-siris syndrome (CSS, OMIM #618027), a syndrome characterized by intellectual disability, dysmorphic features, and organ system anomalies (Vasileiou et al., 2018; Vasko et al., 2021). The gene has also been implicated in several types of cancer (Huber et al., 2017). In the findings of Vasileou et al. (2018), splice-sites, frameshift, and missense mutations were found in DPF2 of CSS individuals. The missense mutations were hypothesized to have affected the binding sites of the proteins. In humans, ten individuals have been recognized with CSS with a mutation in the DFP2 gene (Milone et al., 2020).

Zhang et al. (2019) show that Dpf2 maintains the pluripotency and embryonic stem cell differentiation potential. Deletion of the mouse Dpf2 gene led to changes in gene expression associated with stem cell maintenance, blastocyst formation, and pathways controlling pluripotency. Additionally, it caused an increase in genes involved in cell death and a decrease in genes regulating cell proliferation.

Reynolds et al. (2021) found a missense mutation in the DPF2 gene and five other genes in cows that explain the proportions of inbreeding depression. They also link DPF2 to hoof conformation abnormalities, although non-significant because of insufficient data for sufficient power. Further, the missense mutation g.44213160A>G was an amino acid substitution p.Lys216Arg, significant on bodyweight and stature phenotype association. However, the study did connect several genes to novel syndromes in cows, although known in other species.

4.1.3 KIAA1109

The Transmembrane Protein (KIA1109) gene was discovered in humans decades ago but remains uncharacterized. According to Kikuno (1999), the human reference genome (GrCh38) lists one transcript with experimental evidence and predicted transcripts with several non-protein coding transcripts of unknown significance. They also found that the gene is expressed in the human ovary and brain/central nervous system.

Mutations in the human ortholog (KIA1109) are generally considered lethal (Cabet et al., 2020), but a few mutations cause a disorder called Alkuraya-Kučinskas syndrome (AK) (OMIM #617822). AK is a genetic disorder characterized by intellectual disability and developmental delay. KIAA1109 loss of function mutations has been linked to perinatal deaths and premature termination of pregnancies. Surviving individuals with KIAA1109 missense mutations typically exhibit delayed development, mild to moderate learning difficulties, a lack of speech, reliance on support for standing and walking, muscle weakness and wasting, stereotypic movements, dysmorphic features, and early-onset epilepsy (Cabet et al., 2020; Gueneau et al., 2018). Individuals with KIAA1109 mutations have shown difficulties in concentration and have displayed self-injurious behavior, such as head-banging, to express frustration or anger. According to reports, three surviving individuals from two families have a missense variant in one of their two copies of the gene, while four individuals from two related families are homozygous for a new missense mutation (Gueneau et al., 2018; Kumar et al., 2020). Several autoimmune disorders are

associated with KIAA1109. These disorders include moderate to severe asthma, rheumatoid arthritis, type 1 diabetes, anterior uveitis, and allergic sensitization, endometrial- and prostate cancer (Balasopoulou et al., 2016; Bønnelykke et al., 2013; Bouzid et al., 2014; Kiani et al., 2015; Louahchi et al., 2016; Plaza-Izurieta et al., 2011; Qiao et al., 2019; Shrine et al., 2019; Tindall et al., 2010; Yang et al., 2011).

The gene has been linked to fertility through its suggested role in embryo development and the phagocytosis regulation process, where cells absorb microorganisms, foreign substances, and apoptotic cells (Jeng et al., 2019; Shamseldin et al., 2018). Mutations in the *Drosophila* ortholog of KIAA1109, *tweek*, usually results in lethality during metamorphosis, but a few flies survive to adulthood and exhibit severe locomotion deficits and seizures (Kane et al., 2019; Verstreken et al., 2009). A study on Chinese hamster ovary cells found FSA (KIAA1109) to play an essential role in mammalian epithelial cell growth and differentiation (Cao et al., 2006; Kuo et al., 2006).

4.1.4 Olfactory Receptors (OR5AS1)

Olfactory receptor five subfamily AS member 1 (OR5AS1) gene is related to the large gene group of olfactory receptors. Olfactory receptors (OR) are responsible for the perception of smell (NCBI). Mammalian genomes harbor approximately 1000 olfactory receptor genes and pseudogenes, with different mutational events in species giving the orthologs a complex relationship (Young et al., 2002). Humans were found to have 339 intact OR genes and 297 OR pseudogenes (Malnic et al., 2004), and they were unevenly distributed on 21 human chromosomes. In a study on the olfactory receptors' mice were found to have more olfactory receptors than humans, and the receptors were found to have multiple transcriptional variants and alternative isoforms (Malnic et al., 2004; Young et al., 2002; Zhang & Firestein, 2002).

4.1.5 U6

U6 spliceosome RNA (U6) is a subunit of the spliceosome (*RNA, U6 SMALL NUCLEAR, 1; RNU6-1*., n.d.). Its precise function is unknown but thought to have a catalyzing role in the spliceosome and splice site specificity (Fortner et al., 1994; Kandels-Lewis & Séraphin, 1993; Lesser & Guthrie, 1993; Valadkhan & Manley, 2001). The spliceosome is a catalysator of the splicing of precursor mRNA in eukaryotes (Butcher & Brow, 2005). The process of splicing is necessary to remove non-coding loci (introns) and to fuse the remaining coding loci (exons) before it is translated into proteins (Butcher & Brow, 2005). The U6 is similar in the yeast homolog, suggesting a conserved essential role in pre-mRNA processing (Brow & Guthrie, 1988, Abstract).

5 Discussion

5.1 Summary and Interpretation of Findings

For the first time, the genome of the Norwegian Fjord-horse has been analyzed for haplotype homozygosity depletion. The recessive detrimental alleles affect natural selection against homozygote animals explaining embryonic lethality, lower foaling rates, or exclusion from breeding programs. Unfortunately, the effects of detrimental alleles are only noticeable when the alleles reach a higher frequency (Häflinger et al., 2021), and carrier matings may go unnoticed.

Detection of recessive detrimental variants is difficult, particularly in small populations. In this study, a genome-wide haplotype homozygosity mapping was performed to detect recessive harmful alleles. This method was used with only genomic data, as done in other studies (Derks, Lopes, et al., 2018; VanRaden et al., 2011). Studies using this analyzing method include more animals (~4K-~75K). However, this study has a lot more SNPs (~670K) included than most of the previous studies (50K-80K SNPs) (Derks et al., 2017; Derks, Lopes, et al., 2018; Fritz et al., 2013; Hoff et al., 2017; Pausch et al., 2015; VanRaden et al., 2011). The number of SNPs in this study is a strength and gives a higher resolution of where the detrimental variants may reside. Daetwyler et al. (2014) also used a few animals (234 animals); however, that study also used whole genome sequencing of the animals providing them with an amount of data and resolution that is higher than in this study.

After the scan and correcting for multiple testing, the analysis reveals several significant overlapping haplotype blocks along different chromosomes. Having several overlapping blocks is consistent with other studies such as Häflinger et al. (2021) and Derks et al. (2017). Haplotypes significant after correction and missing a homozygote genotype (or one observed homozygote) were kept for further characterization. The strict criteria for keeping a haplotype were set to avoid false positives in this study and strengthen the reliability of the results. Only two loci with haplotypes included had one homozygote observed (chr12; 13260001-1362000 and chr18). The haplotype in chromosome 12:13260001-1362000 did not reside close to any known gene functions and was therefore not included as a causative haplotype. However, the haplotypes of chromosome 18 were later associated with the U6 gene. The overlapping blocks from the scan had the same observations of homozygotes; therefore, other summary statistics were the same. Such overlapping haploblocks appear when there are segments (in this case, not observed as homozygous by one allele) larger than the window size used in the analysis. It may also appear as such if no new SNP is included with the window-shift. In this case, the same SNPs are observed and have the same statistics. This study included the SNPs from all the overlapping blocks in the Ensembl characterization, resulting in more SNPs without a known gene nearby.

The variation in window size made a big difference in detecting haplotype blocks with a missing homozygote and significant HWE deviation. The largest window size was 700kbp, sliding down the genome by 300kbp at a time. The smallest window size was 10kbp, sliding down 1kbp each shift. The two largest window sizes (500-700kbp) did not detect significant haplotypes with missing homozygotes. Not detecting significant haplotypes of larger sizes can be true if the animals in the dataset are unrelated or very distantly related, as in this data. It may also be an old mutation with low LD around the mutation. The likelihood of unrelated animals sharing identical alleles is small, and when the population has a common ancestor, the chances increase. The ancestral background will affect the size of the shared fragments within the population data and is likely to keep the shared haplotype block segments small if there is a distant relationship between the animals and larger if there are recent common ancestors.

Further, the smallest window size at 10kbp detected far more significant haploblocks than any of the other window sizes collectively. As expected, the 10kbp window size detected significant blocks in areas where larger blocks were found. This window size also reveals haplotype blocks in chromosomes and loci that are insignificant by any different window size. This is consistent with the animals having a common ancestor at some point in their pedigree, which, as mentioned, all Norwegian Fjords do have a pedigree going back to one fjord-horse. Additionally, several blocks found by the 10kbp window size harbor many SNPs. Therefore, they are more likely to be actual homozygote stretches than longer blocks with few SNPs. That means a 10kbp stretch with ten homozygote SNPs is more likely to be a true homozygote compared to a 100kbp stretch with only two SNPs. This is because the amount of data points within the haplotypes may be too few to be sure that the entire segment is homozygote.

When looking at the frequency of the haploblocks detected, the effect would be expected to impact fertility significantly. This expected impact on fertility is expected to be larger than observed in the Fjord-horse, compared to other breeds. The survival rate was calculated for the population if the haploblocks in this study were true (Table 4.4), and the total survival rate was calculated to be 52.6%. Such a low survival rate would impact the foaling rate of the Fjord-horse population, presumably greater than what is observed.

Two chromosomes only had a few small haploblocks (10kbp) detecting homozygosity depletion. On chromosome two, only one SNP was significant as deviating from the HWE, while the other SNPs had a too low frequency of the minor allele to detect a deviation. In chromosome 18, only two SNPs were associated with a known gene. However, the first SNP shows a significant deviation and has a few observed homozygotes. The second SNP on this locus was insignificant, with a low frequency of the minor allele. The small size of haploblocks provides weak evidence of homozygote depletion because the fragment lengths are small. Small fragments are more likely to be observed as homozygotes due to older inbreeding. However, it is more probable that recessive lethal alleles reside within longer fragments of homozygosity.

Chromosomes 12 and 27 have several fragment sizes detected as significant, making these loci a stronger candidate than the ones on chromosomes 2 and 18. Although the first locus on chromosome 12 (at ~15 mill bp) is a stronger candidate with several window sizes detected, there is only one significant SNP in this locus. This one SNP also has observed homozygotes, although there is a depletion to what is to be expected (21 observed vs. 107 expected). The second locus (at ~29Mill bp) also has several window sizes detecting significant depletion of haplotypes. In this locus, a set of three SNPs has a significant depletion of 4e-22. Looking closer at the SNPs in this locus, three are significant (between 3e-5 and 4e-15), while the other two have low minor allele frequency. The significantly depleted SNPs in this locus also have observed homozygotes of the minor allele.

The haploblocks on chromosome 27 were also of different sizes (10-100kbp). All the overlapping haploblocks and SNPs within the haploblocks were found to be significant.

However, very few SNPs are observed in chromosome 27 and the first locus of chromosome 12 (at ~15 mill bp). Both chromosomes have only three SNPs observed within the haplotypes; for chromosome 18, only two SNPs were observed. While the second locus of chromosome 12 (at ~29 mill bp) have five observed SNPs, the haploblocks on chromosome 2 contain nine SNPs. The problem with this is that longer haploblocks with few SNPs may be a coincidence detected as having a depletion. This is particularly interesting as a coincidence when several flanking SNPs are observed with low minor allele frequency. Also, for short haplotypes with few SNPs, the possibility of identical alleles by chance and the chance for detecting false positives increases.

This, together with the low total survival rate of the assumed independent haploblock candidates, indicates that there are likely false positives candidates in this study. Therefore, the candidate haplotypes of this study are still expected to contain false positives.

Phasing uncertainty is a likelihood estimation of the obtained haplotypes by SHAPEIT. The expectance was that the uncertainty would have been poor certainty in parts of the genome since the population sample is small and assumed to not be in close relatives. The problem with using few animals in phasing haplotypes is an increased likelihood of specific haplotypes not being representative of the population. It could also be that the variation between unrelated animals causes SHAPEIT not to detect specific haplotypes. Preferably, one has family data when estimating haplotypes, as this allows for capturing rare haplotypes segregating within families and haplotypes common between families. The currently done analysis was on a small, unrelated population, and the estimated haplotypes may therefore be unrepresentative for the entire Norwegian Fjord-horse population.

The SNP effect prediction by Ensembl predicted all significant SNPs to be 'modifiers,' i.e., not introducing early stop codons or any immediately severe effects. This effect prediction means the mutation may affect the gene's sequence, affecting the gene's function. Therefore, the impact of the variant was not crucial for filtering out more likely variants in this analysis. There were expected to be several variants of lower effect since the significance and observed missing homozygote does not exclude these variants. It was odd that no lower-impact variants were found in this analysis. The expectance of the impact would have been many lower-impact predictions, a few moderate, and maybe one or two high-impact mutations. This expectation is due to the general expectations of most mutations being synonymous or harmless to the animals' fitness, where a few rare mutations are harmful, and a few have positive effects in eukaryotes. Observing no low-impact variant may be because of a selection of the animals. There are strict rules about the morphological traits of the Norwegian horse. Therefore, the animals with unwanted traits will not be eligible for breeding. This selection can be of traits such as white markings in the Fjord, being harmless considering the fitness of the animals. However, it is unwanted in the breeding population and may not be a part of the study population.
A reason for not detecting high-impact variants may be that the population size is too small to detect rare variants, the assumption about the population being unrelated may be wrong, there may be a low linkage between a detrimental variant and flanking SNPs, or the Ensembl variant detection method may not be the best for this population.

When investigating why the detected haplotypes were of higher frequencies (0.3-0.45), there were suspicions that the software favored the most frequent haplotypes. These frequencies were found rather odd and could be a possible side effect of using SHAPEIT which may favor the most abundant phases and disregard the less frequent phases. This favoring would further keep a high frequency when GHap generates haplotype blocks. Optimally, all the phases would have been written to the files, and from there, haplotype blocks could scan for deficiency in the uncommon phases. However, it may also be a side effect of the data used in the software since phases are better estimated with family data, i.e., mother, father, and offspring trio. The unknown relations of animals in the input data may make it impossible to estimate genome portions as a phase if no other animals in the population have identical genotypes. The phase could also be better assessed with a more extensive dataset than what was done in the current analysis.
However, it became evident that the haploblocks were done on all the estimated haplotypes. The number of haplotype blocks generated by GHap was, in total, over 13 Million, where the lowest frequency of a haplotype was 0.0015. It also became evident that the less frequent haplotype blocks naturally expect fewer homozygotes (i.e., less than one). These haplotype blocks would never be significant for homozygote depletion because no depletion is observed. There were several haplotype blocks with a frequency between 15-20 and an expected homozygote count of ~10-13 animals in the obtained haploblocks. However, these haplotype blocks were not deemed significant after the

Bonferroni adjustment. The high frequency of the haplotype blocks obtained is another side-effect of the study having few animals included.

The haplotype homozygosity depletion approach is based on haplotypes and causative variants being in perfect linkage disequilibrium, which is not always true. The genes detected by several window sizes of haplotype blocks are the most likely loci to harbor detrimental variants. Only a few significant SNPs are found within or nearby genes with known functions. Several significant loci without known gene functions may be a side effect of the population data not being closely related, breeding selection, detecting loci that are homozygote missing by chance, or without significant HWE depletion. These loci cannot be identified as plausible functional candidate genes and were not further characterized in this analysis. The genes with known gene functions found significant in this analysis include the DEFA22 and DEFA5L, DPF2, KIAA1109, OR5AS1, and U6 genes. The characterization of the genes included obtaining the tissue expression in human and mouse orthologues and reading the literature related to the gene suggested by NCBI and searches on other search engines. The actual harmful functions of the genes are not uncovered in this study. However, for some of the genes, there is literature investigating this. As far as the author knows, these genes have not been characterized as recessive deleterious candidate genes in horses or cows before.

This study's candidate Defensin alpha (DEFA) genes included the equine DEFA 5L and 22. Both were found to be partially orthologues of the human HNP4. The number of alpha-defensin genes varies between species, from the dogs and cows having none to the horse having 20 known gene annotations (Bruhn et al., 2009). The research on these peptides is most abundant in humans, where HNP1-3 is more abundantly expressed. The Alpha-defensins are peptides a part of the microbial immune response in mucus (Lehrer & Lu, 2012), in the respiratory systems, digestive system (Ouellette et al., 2011), and reproductive tracts (Zhai et al., 2023). The expression of alpha-defensins is found to be expressed in various parts of the menstrual cycle and during pregnancy (Zhai et al., 2023). There is also an expression difference within the female reproductive tracts. Here, the human HNP4 is found more abundantly in the lower part of the female reproductive tract (from the vulva with lowering expression towards the ovary) (Quayle et al., 1998). On days 15-28 of the menstrual cycle, the immune response is downregulating, allowing for sperm fertilization and implantation for pregnancy (Yarbrough et al., 2015). The role of defensins in general, and the HNP4 specifically, including this study detection of two different alpha-defensins, gives strong indications of this gene being a decent candidate gene.

Further, the Double Plant-Homeo-Domain Fingers 2 (DPF2) is associated with Coffin-siris syndrome (OMIM #618027) in humans and hoof conformation abnormality in cows (Reynold et al., 2021; Vasileou et al., 2018). The Coffin-siris syndrome only has ten individuals recognized with the mutation in the DPF2 gene. However, the mutation in the gene is predicted to have severe consequences for the affected individual (Vasko et al., 2021; Vasileou et al., 2018; Milestone et al., 2020). One recent article has also been published on DPF2s' role in stem cell differentiation, maintenance, and blastocyst formation (Zhang &Chronis et al., 2019). DPF2 is critical for normal development in human and mouse formation. The gene appears sensitive to different mutations, suggesting it is highly conserved with essential functions in developmental biology. The functions of the gene support the idea that DPF2 may be a candidate gene for recessive detrimental variants.

KIAA1109 is a gene whose functions remain unknown, although it was discovered long ago. According to Cabet et al. (2020), mutations in the gene are considered lethal. However, a few modifications have been detected in individuals with an autosomal recessive disorder called Alkuraya- Kučinskas syndrome (OMIM #617822) (Kumar et al., 2020). Some mutations have been associated with premature termination of pregnancies and perinatal death, where survivors are seriously affected by a range of

disabilities and behavioral problems (Gueneau et al., 2018; Cabet et al., 2020). The seriousness of mutations in this gene supports the discovery of a candidate gene in this study.

The Olfactory (OR5AS1) and the spliceosome RNA (U6) genes have little knowledge about them, making it difficult to conclude that they are strong candidate genes in this study. The functions of the Olfactory receptors are related to the perception of smell, and there are many homolog genes, with more in mice than in humans (Malnic et al., 2004; Young et al., 2002; Zhang et al., 2002). The relation to fertility in this gene is unknown. The U6 gene is known to code for a spliceosome subunit. However, the specific function of the subunit is not yet fully understood. As a part of the spliceosome, it does have a role in catalyzing splicing precursor mRNA, removing introns, and fusing the remaining exons before translation (Butcher & Brow, 2005).

5.2 Answering Thesis Question

These findings confirm the hypothesis that a genetic component may affect the fertility of the Norwegian Fjord-horse population. Despite many candidate loci, the proposed candidate genes are narrowed down to DEFA22, DEFA5L KIAA1109, and DPF2. These genes were selected to be the proposed candidate genes for fertility based on the mutation effects known to be potentially lethal and that the SNPs associated with these genes were significantly deviating from HWE.

5.3 Identified Gaps and Limitations of Study

The findings in the study must be seen in light of some limitations. The study is limited to a small population of the Norwegian Fjord-horse and does not include the genomic data of other breeds. Meaning the results from this study may not apply to other horse breeds.

The previous research on homozygote mapping to detect recessive detrimental variants is not few to all animals, but there are relatively few publications on horses. The haplotype homozygosity mapping publications are recent and different in their methodology for detecting candidate loci. Mainly the difference in the methods in studies makes it difficult to compare the results between studies.

The number of samples included in the study was fewer than initially planned. The study population consisted of 328 genotyped animals, where the selection of the animals is unknown. This analysis's planned number of animals was 1200 (i.e., ~800 more), where ~500 are already collected and processed. Unfortunately, the new data was not done processing in time for this study. The selection of this population provides a weakness since further analysis and verifications of the results are not possible. An analysis of these few animals is also expected to be error-prone, i.e., to show false positives and negatives. A similar study should be conducted on a more significant population size to verify the results, including a haplotype homozygosity mapping and segregation analysis. If the phenotype of the animals is also obtained, then an association analysis would be interesting to run on the data to see if there can be any significant detections with this method.

However, this study found continued data analysis with pedigree information and segregation analysis impossible as the animals in the data are unrelated and have no recorded phenotype. Charlier et al. (2016) suggest that the entire genome should be sequenced and mined for lethal variants. Following Charlier et al. (2016) suggestion was impossible in this study; however, this study used an HD SNP chip. The whole genome sequencing in this study is also inappropriate for all animals, as the frequency of the detrimental alleles is expected to be low and would not be cost-effective. However, after some animals

are confirmed as a carrier by the analysis, the whole genome sequencing of the animals would be a more precise way to obtain knowledge about the embryonic lethality in the Norwegian Fjord-horse population. The statistical evaluation of segregating haplotypes (VanRaden et al., 2011) was not possible in this study as the data of other horses related to the ones in the used dataset were delayed in processing.

While this thesis utilized an older version of SHAPEIT (v.2) (Delaneau et al., 2014), it is essential to acknowledge that newer versions may have additional features or improvements. However, the author deliberately chose to use this version based on recommendations and the availability of a helpful tutorial section. Additionally, the author chose to use PLINK 1.9 because of their familiarity with the software and its continued updates.

The conservation prediction of the genes was not conducted in this analysis. The conservation predictions help determine the likelihood of a variant being detrimental. If detected in other species, the prediction is made by looking at the area with a mutation to see if the gene has highly conserved domains. Genes with highly conserved domains are likely to be so because mutations in the loci prohibit the product from the gene from functioning as supposed to, which is critical for some developmental stages in several species.

This study did not compare software or packages with other software, leaving the approach vulnerable to internal errors. Neither did the study conduct a simulation run to estimate the software and packages' accuracy. The study is further limited by the researchers' limited access to data and inability to collect more samples. The author has no cultural or personal bias to declare for this study.

5.4 The Implications and Practical Applications of the Study

The data provide new evidence of the Norwegian Fjord-horse population genetics affecting the fertility in the population. This evidence can be important for improving the fertility of the population. However, further studies are necessary to confirm the results of this study, particularly studies with a more prominent test population. The results should then also be verified by the phenotype. Also, the validation in carrier matings would prove the genomic effects. Such validation is planned to be conducted by the author soon. Other studies should validate this study's findings before considering them with selection schemes.

6 Final Comment / Conclusion

In this study, the genotype data information extraction was done with a reverse genetic approach. With only genotype data, approximately 500K SNPs were examined in 328 animals from the Norwegian Fjord-horse population. The proposed candidate genes were KIA1109, DEFA22, DEFA5L, and DPF2, causing recessive lethality/ or unknown development disorders in equines. This study indicates that genetic factors diminish fertility in the Norwegian Fjord-horse population. It would be valuable for breeders, Fjord Horse International, Norwegian Fjord-horse Association, and the Norwegian Equine Centre to contemplate this matter. The opportunity to genetically improve the reproductive health of the Fjord-horse is essential to maintain a sustainable population.

7 References

Akinbi, H. T., Narendran, V., Pass, A. K., Markart, P., & Hoath, S. B. (2004). Host defense proteins in vernix caseosa and amniotic fluid. *American Journal of Obstetrics and Gynecology*, *191*(6), 2090–2096. https://doi.org/10.1016/j.ajog.2004.05.002

Allen, W. R., Brown, L., Wright, M., & Wilsher, S. (2007). Reproductive efficiency of Flatrace and National Hunt Thoroughbred mares and stallions in England. *Equine Veterinary Journal*, *39*(5), 438–445. https://doi.org/10.2746/042516407X1737581

Avdi, M., & Banos, G. (2008). Genetic diversity and inbreeding in the Greek Skyros horse. *Livestock Science*, *114*(2–3), 362–365. https://doi.org/10.1016/j.livsci.2007.11.007

Ayadi, A., Birling, M. C., Bottomley, J., Bussell, J., Fuchs, H., Fray, M., Gailus-Durner, V., Greenaway, S., Houghton, R., Karp, N., Leblanc, S., Lengger, C., Maier, H., Mallon, A. M., Marschall, S., Melvin, D., Morgan, H., Pavlovic, G., Ryder, E., … Herault, Y. (2012). Mouse large-scale phenotyping initiatives: Overview of the European Mouse Disease Clinic (EUMODIC) and of the Wellcome Trust Sanger Institute Mouse Genetics Project. *Mammalian Genome*, *23*(9–10), 600–610. https://doi.org/10.1007/s00335-012-9418-y

Azcona, F., Valera, M., Molina, A., Trigo, P., Peral-García, P., Solé, M., & Demyda-Peyrás, S. (2020). Impact of reproductive biotechnologies on genetic variability of Argentine Polo horses. *Livestock Science*, *231*. https://doi.org/10.1016/j.livsci.2019.103848

Balasopoulou, A., Stanković, B., Panagiotara, A., Nikčevic, G., Peters, B. A., John, A., Mendrinou, E., Stratopoulos, A., Legaki, A. I., Stathakopoulou, V., Tsolia, A., Govaris, N., Govari, S., Zagoriti, Z., Poulas, K., Kanariou, M., Constantinidou, N., Krini, M., Spanou, K., … Katsila, T. (2016). Novel genetic risk variants for pediatric celiac disease. *Human Genomics*, *10*(1), 34. https://doi.org/10.1186/s40246-016-0091-1

Ball, B. A., Little, T. V., Hillman, R. B., & Woods, G. L. (1986). Pregnancy rates at Days 2 and 14 and estimated embryonic loss rates prior to day 14 in normal and subfertile mares. *Theriogenology*, *26*(5), 611–619. https://doi.org/10.1016/0093-691X(86)90168-8

Ball, B. A., Little, T. V, Weber, J. A., & Woods, G. L. (1989). *Survival of Day-4 embryos from young, normal mares and aged, subfertile mares after transfer to normal recipient mares*.

Ben Braiek, M., Fabre, S., Hozé, C., Astruc, J. M., & Moreno-Romieux, C. (2021). Identification of homozygous haplotypes carrying putative recessive lethal mutations that compromise fertility traits in French Lacaune dairy sheep. *Genetics Selection Evolution*, *53*(1). https://doi.org/10.1186/s12711-021-00634-1

Berglund, B. (2008). Genetic Improvement of Dairy Cow Reproductive Performance. *Reproduction in Domestic Animals*, *43*(SUPPL.2), 89–95. https://doi.org/10.1111/j.1439-0531.2008.01147.x

Blanton, S. H. (2022). Linkage Analysis. In W. K. Scott & M. D. Ritchie (Eds.), *Genetic Analysis of Complex Disease* (3rd ed., pp. 105–160). John Wiley & Sons.

Bønnelykke, K., Matheson, M. C., Pers, T. H., Granell, R., Strachan, D. P., Alves, A. C., Linneberg, A., Curtin, J. A., Warrington, N. M., Standl, M., Kerkhof, M., Jonsdottir, I., Bukvic, B. K., Kaakinen, M., Sleimann, P.,

Thorleifsson, G., Thorsteinsdottir, U., Schramm, K., Baltic, S., … Bisgaard, H. (2013). Meta-analysis of genome-wide association studies identifies ten loci influencing allergic sensitization. *Nature Genetics*, *45*(8), 902–906. https://doi.org/10.1038/ng.2694

Bonney, G. E., Opitz, J. M., & Reynolds, J. F. (1984). On the statistical determination of major gene mechanisms in continuous human traits: Regressive models. *American Journal of Medical Genetics*, *18*(4), 731–749. https://doi.org/10.1002/ajmg.1320180420

Bouman, J. (1977). The future of Przewalski horsesEquus przewalskii [Plate 10] in captivity. *International Zoo Yearbook*, *17*(1), 62–68. https://doi.org/10.1111/j.1748-1090.1977.tb00867.x

Bouzid, D., Fourati, H., Amouri, A., Marques, I., Abida, O., Tahri, N., Penha-Gonçalves, C., & Masmoudi, H. (2014). Autoimmune diseases association study with the KIAA1109–IL2–IL21 region in a Tunisian population. *Molecular Biology Reports*, *41*(11), 7133–7139. https://doi.org/10.1007/s11033-014-3596-5

Bristol, F. (1982). Breeding behaviour of a stallion at pasture with 20 mares in synchronized oestrus. *Journal of Reproduction and Fertility. Supplement*, *32*, 71–77.

Bristol, F. (1987). Fertility of pasture bred mares in synchronized oestrus. *Journal of Reproduction and Fertility. Supplement*, *35*, 39–43.

Brow, D. A., & Guthrie, C. (1988). Spliceosomal RNA U6 is remarkably conserved from yeast to mammals. *Nature*, *334*(6179), 213–218. https://doi.org/10.1038/334213a0

Bruck, I., Anderson, G. Z., & Hyland', J. H. (1993). *Reproductive performance of Thoroughbred mares on six commercial stud farms*.

Bruhn, O., Paul, S., Tetens, J., & Thaller, G. (2009). The repertoire of equine intestinal alpha-defensins. *BMC Genomics*, *10*, 631. https://doi.org/10.1186/1471-2164-10-631

Brunham, R. C., & Paavonen, J. (2020). Reproductive system infections in women: lower genital tract syndromes. *Pathogens and Disease*, *78*(5). https://doi.org/10.1093/femspd/ftaa022

Burris, H. H., Riis, V. M., Schmidt, I., Gerson, K. D., Brown, A., & Elovitz, M. A. (2020). Maternal stress, low cervicovaginal β-defensin, and spontaneous preterm birth. *American Journal of Obstetrics & Gynecology MFM*, *2*(2), 100092. https://doi.org/10.1016/j.ajogmf.2020.100092

Butcher, S. E., & Brow, D. A. (2005). Towards understanding the catalytic core structure of the spliceosome. *Biochemical Society Transactions*, *33*(3), 447–449. https://doi.org/10.1042/BST0330447

Cabet, S., Putoux, A., Buenerd, A., Gueneau, L., Reymond, A., Thia, E. W. H., Lai, A. H. M., Schindewolf, E. M., Sanlaville, D., Lesca, G., & Guibaud, L. (2020). Prenatal cerebral imaging features of a new syndromic entity related to KIAA1109 pathogenic variants mimicking tubulinopathy. *Prenatal Diagnosis*, *40*(2), 276–281. https://doi.org/10.1002/pd.5589

Cammack, K. M., Thomas, M. G., & Enns, R. M. (2009). Reproductive Traits and Their Heritabilities in Beef Cattle. *The Professional Animal Scientist*, *25*(5), 517–528. https://doi.org/10.15232/S1080-7446(15)30753-1

Cao, D., Li, Y., Yang, R., Wang, Y., Zhou, Y., Diao, H., Zhao, Y., Zhang, Y., & Lu, J. (2010). Lipopolysaccharide-Induced Epididymitis Disrupts Epididymal Beta-Defensin Expression and Inhibits Sperm Motility in Rats1. *Biology of Reproduction*, *83*(6), 1064–1070. https://doi.org/10.1095/biolreprod.109.082180

Cao, Y.-W., Jiang, C.-L., & Jiang, T. (2006). Molecular cloning and preliminary analysis of a fragile site associated gene. *Biomedical and Environmental Sciences : BES*, *19*(5), 392–398.

Casellas, J., Cañas-Álvarez, J. J., González-Rodríguez, A., Puig-Oliveras, A., Fina, M., Piedrafita, J., Molina, A., Díaz, C., Baró, J. A., & Varona, L. (2017). Bayesian analysis of parent-specific transmission ratio distortion in seven Spanish beef cattle breeds. *Animal Genetics*, *48*(1), 93–96. https://doi.org/10.1111/age.12509

Casellas, J., Gularte, R. J., Farber, C. R., Varona, L., Mehrabian, M., Schadt, E. E., Lusis, A. J., Attie, A. D., Yandell, B. S., & Medrano, J. F. (2012). Genome scans for transmission ratio distortion regions in mice. *Genetics*, *191*(1), 247–259. https://doi.org/10.1534/genetics.111.135988

Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1). https://doi.org/10.1186/s13742-015-0047-8

Charlesworth, D., & Willis, J. H. (2009). The genetics of inbreeding depression. In *Nature Reviews Genetics* (Vol. 10, Issue 11, pp. 783–796). https://doi.org/10.1038/nrg2664

Charlier, C., Li, W., Harland, C., Littlejohn, M., Coppieters, W., Creagh, F., Davis, S., Druet, T., Faux, P., Guillaume, F., Karim, L., Keehan, M., Kadri, N. K., Tamma, N., Spelman, R., & Georges, M. (2016). NGS-based reverse genetic screen for common embryonic lethal mutations compromising fertility in livestock. *Genome Research*, *26*(10), 1333–1341. https://doi.org/10.1101/gr.207076.116

Chertov, O., Michiel, D. F., Xu, L., Wang, J. M., Tani, K., Murphy, W. J., Longo, D. L., Taub, D. D., & Oppenheim, J. J. (1996). Identification of Defensin-1, Defensin-2, and CAP37/Azurocidin as T-cell Chemoattractant Proteins Released from Interleukin-8-stimulated Neutrophils. *Journal of Biological Chemistry*, *271*(6), 2935–2940. https://doi.org/10.1074/jbc.271.6.2935

Cole, A. M., Hong, T., Ming Boo, L., Nguyen, T., Zhao, C., Bristol, G., Zack, J. A., Waring, A. J., Yang, O. O., Lehrer, R. I., & Babior, B. M. (2002). Retrocyclin: A primate peptide that protects cells from infection by T-and M-tropic strains of HIV-1. *PNAS*, *99*(4), 1813–1818. www.pnas.org/cgi/doi/10.1073/pnas.052706399

Com, E., Bourgeon, F., Evrard, B., Ganz, T., Colleu, D., Jégou, B., & Pineau, C. (2003). Expression of Antimicrobial Defensins in the Male Reproductive Tract of Rats, Mice, and Humans1. *Biology of Reproduction*, *68*(1), 95–104. https://doi.org/10.1095/biolreprod.102.005389

Crawford, D. C., & Dumitrescu, L. (2022). Study Design for Genetic Studies. In W. K. Scott & M. D. Ritchie (Eds.), *Genetic Analysis of Complex Disease* (3rd ed., pp. 58–74). John Wiley & Sons.

Cunningham, F., Allen, J. E., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., Austine-Orimoloye, O., Azov, A. G., Barnes, I., Bennett, R., Berry, A., Bhai, J., Bignell, A., Billis, K., Boddu, S., Brooks, L., Charkhchi, M., Cummins, C., Da Rin Fioretto, L., … Flicek, P. (2022). Ensembl 2022. *Nucleic Acids Research*, *50*(D1), D988–D995. https://doi.org/10.1093/nar/gkab1049

Das, S., Vince, G. S., Lewis-Jones, I., Bates, M. D., & Gazvani, R. (2007). The expression of human alpha and beta defensin in the endometrium and their effect on implantation. *Journal of Assisted Reproduction and Genetics*, *24*(11), 533–539. https://doi.org/10.1007/s10815-007-9173-2

de Leeuw, E., Li, C., Zeng, P., Li, C., Buin, M. D., Lu, W.-Y., Breukink, E., & Lu, W. (2010). Functional interaction of human neutrophil peptide-1 with the cell wall precursor lipid II. *FEBS Letters*, *584*(8), 1543–1548. https://doi.org/10.1016/j.febslet.2010.03.004

Delaneau, O., Howie, B., Cox, A. J., Zagury, J. F., & Marchini, J. (2013). Haplotype estimation using sequencing reads. *American Journal of Human Genetics*, *93*(4), 687–696. https://doi.org/10.1016/j.ajhg.2013.09.002

Delaneau, O., Marchini, J., McVeanh, G. A., Donnelly, P., Lunter, G., Marchini, J. L., Myers, S., Gupta-Hinch, A., Iqbal, Z., Mathieson, I., Rimmer, A., Xifara, D. K., Kerasidou, A., Churchhouse, C., Altshuler, D. M., Gabriel, S. B., Lander, E. S., Gupta, N., Daly, M. J., … Peltonen, L. (2014). Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. *Nature Communications*, *5*. https://doi.org/10.1038/ncomms4934

Den nasjonale forskningsetiske komité for naturvitenskap og teknologi (NENT). (2018). *Etiske retningslinjer for bruk av dyr i forskning* (1st ed.).

Den nasjonale forskningsetiske komité for samfunnsvitenskap og humaniora (NESH). (2021). *Forskningsetiske retningslinjer for samfunnsvitenskap og humaniora NESH* (5th ed.). www.forskningsetikk.no

Derks, M. F. L., Gjuvsland, A. B., Bosse, M., Lopes, M. S., Van Son, M., Harlizius, B., Tan, B. F., Hamland, H., Grindflek, E., Groenen, M. A. M., & Megens, H. J. (2019). Loss of function mutations in essential genes cause embryonic lethality in pigs. *PLoS Genetics*, *15*(3). https://doi.org/10.1371/journal.pgen.1008055

Derks, M. F. L., Lopes, M. S., Bosse, M., Madsen, O., Dibbits, B., Harlizius, B., Groenen, M. A. M., & Megens, H. J. (2018). Balancing selection on a recessive lethal deletion with pleiotropic effects on two neighboring genes in the porcine genome. *PLoS Genetics*, *14*(9). https://doi.org/10.1371/journal.pgen.1007661

Derks, M. F. L., Megens, H. J., Bosse, M., Lopes, M. S., Harlizius, B., & Groenen, M. A. M. (2017). A systematic survey to identify lethal recessive variation in highly managed pig populations. *BMC Genomics*, *18*(1). https://doi.org/10.1186/s12864-017-4278-1

Derks, M. F. L., Megens, H. J., Bosse, M., Visscher, J., Peeters, K., Bink, M. C. A. M., Vereijken, A., Gross, C., De Ridder, D., Reinders, M. J. T., & Groenen, M. A. M. (2018). A survey of functional genomic variation in domesticated chickens. *Genetics Selection Evolution*, *50*(1). https://doi.org/10.1186/s12711-018-0390-1

Dezetter, C., Leclerc, H., Mattalia, S., Barbat, A., Boichard, D., & Ducrocq, V. (2015). Inbreeding and crossbreeding parameters for production and fertility traits in Holstein, Montbéliarde, and Normande cows. *Journal of Dairy Science*, *98*(7), 4904–4913. https://doi.org/10.3168/jds.2014-8386

Dini, P., Bartels, T., Revah, I., Claes, A. N., Stout, T. A. E., & Daels, P. (2020). A retrospective study on semen quality parameters from four different Dutch horse breeds with different levels of inbreeding. *Theriogenology*, *157*, 18–23. https://doi.org/10.1016/j.theriogenology.2020.07.017

Dowsett, K. F., & Knott, L. M. (1996). THE INFLUENCE OF AGE AND BREED ON STALLION SEMEN. *Theriogenology*, *46*, 397–412.

Elston, R. C., & Stewart, J. (1971). A general model for the genetic analysis of pedigree data. *Human Heredity*, *21*(6), 523–542. https://doi.org/10.1159/000152448

Equus Caballus (Domestic horse; equine) - DEFA5L. (n.d.). OrthoDB V11. Retrieved March 11, 2023, from https://www.orthodb.org/?ncbi=100307027

Equus Caballus (Domestic horse; equine) - DEFA22. (n.d.). OrthoDB V11. Retrieved March 11, 2023, from https://www.orthodb.org/?ncbi=100306994

Espinoza, J., Chaiworapongsa, T., Romero, R., Edwin, S., Rathnasabapathy, C., Gomez, R., Bujold, E., Camacho, N., Kim, Y. M., Hassan, S., Blackwell, S., Whitty, J., Berman, S., Redman, M., Yoon, B. H., & Sorokin, Y. (2003). Antimicrobial peptides in amniotic fluid: defensins, calprotectin and bacterial/permeability-increasing protein in patients with microbial invasion of the amniotic cavity, intra-amniotic inflammation, preterm labor and premature rupture of membranes. *The Journal of Maternal-Fetal & Neonatal Medicine*, *13*(1), 2–21. https://doi.org/10.1080/jmf.13.1.2.21

FAO. (2011). MOLECULAR GENETIC CHARACTERIZATION OF ANIMAL GENETIC RESOURCES. In *FAO Animal Production and Health Guidelines* (Vol. 9).

FAO. (2023a). *Doelahest*. https://dadis-breed-datasheet-ext-ws.firebaseapp.com/?country=NOR&specie=Horse&breed=Doelahest&lang=en

FAO. (2023b). *Fjord Horse*. https://dadis-breed-datasheet-ext-ws.firebaseapp.com/?country=NOR&specie=Horse&breed=Fjordhest&lang=en

FAO. (2023c). *Nordland/Lyngen Horse*. https://dadis-breed-datasheet-ext-ws.firebaseapp.com/?country=NOR&specie=Horse&breed=Nordlandshest%20lyngshest&lang=en

Figueredo, S. M., Weeks, C. S., Young, S. K., & Ouellette, A. J. (2009). Anionic amino acids near the Pro-α-defensin N terminus mediate inhibition of bactericidal activity inmouse Pro-cryptdin-4. *Journal of Biological Chemistry*, *284*(11), 6826–6831. https://doi.org/10.1074/jbc.M807024200

Fjord Horse International. (2014). *Breeding Plan for the Fjord Horse*. http://www.fjordhorseinternational.org/FjHI/images/BreedingPlan/Breeding%20plan%20FINAL.pdf

Fleming, D. C., King, A. E., Williams, A. R. W., Critchley, H. O. D., & Kelly, R. W. (2003). Hormonal contraception can suppress natural antimicrobial gene transcription in human endometrium. *Fertility and Sterility*, *79*(4), 856–863. https://doi.org/10.1016/S0015-0282(02)04930-0

Fomin, A. B., Kalashnikov, V. V., Nikeforova, V. G., & Tarasova, A. I. (1982). Inbreeding of Russian Trotters and crossbreeding with American Trotters. *Nauchn. Tr. Vses. Nauchno-Issled. Inst. Konevodstva, 3-10.*, 3–10.

Fortner, D. M., Troy, R. G., & Brow^, D. A. (1994). *A stem/loop in U6 RNA defines a conformational switch required for pre-mRNA splicing*.

Fritz, S., Capitan, A., Djari, A., Rodriguez, S. C., Barbat, A., Baur, A., Grohs, C., Weiss, B., Boussaha, M., Esquerré, D., Klopp, C., Rocha, D., & Boichard, D. (2013). Detection of Haplotypes Associated with Prenatal Death in Dairy Cattle and Identification of Deleterious Mutations in GART, SHBG and SLC37A2. *PLoS ONE*, *8*(6). https://doi.org/10.1371/journal.pone.0065550

Furre, S., & Norsk Hestesenter. (2017). *Årsrapport nasjonale hesteraser*.

Gabay, J. E., Scott, R. W., Campanelli, D., Griffith, J., Wilde, C., Marra, M. N., Seeger, M., & Nathan, C. F. (1989). Antibiotic proteins of human polymorphonuclear leukocytes. *Proceedings of the National Academy of Sciences*, *86*(14), 5610–5614. https://doi.org/10.1073/pnas.86.14.5610

Gabigs, T. G., Mantel, P. L., Rosli, R., & Crean, C. D. (1994). *Tiia JOUHYAI. OF Bloim;lml. CiiawsrnY Requiem: A Novel Zinc Finger Gene Essential for Apoptosis in Myeloid Cells** (Vol. 269, Issue 47).

Gamboa, S., Machado-Faria, M., & Ramalho-Santos, J. (2009). Seminal traits, suitability for semen preservation and fertility in the native Portuguese horse breeds Puro Sangue Lusitano and Sorraia: Implications for stallion classification and assisted reproduction. *Animal Reproduction Science*, *113*(1–4), 102–113. https://doi.org/10.1016/j.anireprosci.2008.06.004

Ganz, T. (2003). Defensins: antimicrobial peptides of innate immunity. *Nature Reviews Immunology*, *3*(9), 710–720. https://doi.org/10.1038/nri1180

Gibbs, J. R., & Singleton, A. (2006). Application of genome-wide single nucleotide polymorphism typing: Simple association and beyond. In *PLoS Genetics* (Vol. 2, Issue 10, pp. 1511–1517). https://doi.org/10.1371/journal.pgen.0020150

Gómez, M. D., Sánchez, M. J., Bartolomé, E., Cervantes, I., Poyato-Bonilla, J., Demyda-Peyrás, S., & Valera, M. (2020). Phenotypic and genetic analysis of reproductive traits in horse populations with different breeding purposes. *Animal*, *14*(7), 1351–1361. https://doi.org/10.1017/S1751731120000087

Gómez, M. D., Valera, M., Molina, A., Gutiérrez, J. P., & Goyache, F. (2009). Assessment of inbreeding depression for body measurements in Spanish Purebred (Andalusian) horses. *Livestock Science*, *122*(2–3), 149–155. https://doi.org/10.1016/j.livsci.2008.08.007

Gottschalk, M., Sieme, H., Martinsson, G., & Distl, O. (2016a). Analysis of breed effects on semen traits in light horse, warmblood, and draught horse breeds. *Theriogenology*, *85*(8), 1375–1381. https://doi.org/10.1016/j.theriogenology.2015.11.030

Gottschalk, M., Sieme, H., Martinsson, G., & Distl, O. (2016b). Heritability of semen traits in German Warmblood stallions. *Animal Reproduction Science*, *170*, 10–14. https://doi.org/10.1016/j.anireprosci.2016.03.004

Greenberg, D. A. (1986). The effect of proband designation on segregation analysis. *American Journal of Human Genetics*, *39*(3), 329–339.

Greenberg, D. A., & Abreu, P. C. (2001). Determining trait locus position from multipoint analysis: Accuracy and power of three different statistics. *Genetic Epidemiology*, *21*(4), 299–314. https://doi.org/10.1002/gepi.1036

Greiser, T., Sieme, H., Martinsson, G., & Distl, O. (2019). Genetic parameters and estimated breeding values for traits of raw and frozen-thawed semen in German Warmblood stallions. *Animal Reproduction Science*, *210*. https://doi.org/10.1016/j.anireprosci.2019.106194

Griffin. (2014a). *Genetic Conditions and Testing*. Irongateequine.Com. https://www.irongateequine.com/education/genetic-condition-testing

Griffin, N. G. (2014b). *GENETIC ANALYSIS OF COMPLEX AND MENDELIAN DISEASES* [Doctor of Philosopy]. University of North Carolina.

Grønntun, I. (2020). *Conservaon breeding in real life*. https://www.fjordhorseinternational.org/FjHI/index.php/10-norway/341-conservation-breeding-in-real-life

Gueneau, L., Fish, R. J., Shamseldin, H. E., Voisin, N., Tran Mau-Them, F., Preiksaitiene, E., Monroe, G. R., Lai, A., Putoux, A., Allias, F., Ambusaidi, Q., Ambrozaityte, L., Cimbalistienė, L., Delafontaine, J., Guex, N., Hashem, M., Kurdi, W., Jamuar, S. S., Ying, L. J., … Reymond, A. (2018). KIAA1109 Variants Are Associated with a Severe Disorder of Brain Development and Arthrogryposis. *American Journal of Human Genetics*, *102*(1), 116–132. https://doi.org/10.1016/j.ajhg.2017.12.002

Guthrie, B. L., Introini, A., Roxby, A. C., Choi, R. Y., Bosire, R., Lohman-Payne, B., Hirbod, T., Farquhar, C., & Broliden, K. (2015). Depot Medroxyprogesterone Acetate Use Is Associated With Elevated Innate Immune Effector Molecules in Cervicovaginal Secretions of HIV-1-Uninfected Women. *Journal of Acquired Immune Deficiency Syndromes (1999)*, *69*(1), 1–10. https://doi.org/10.1097/QAI.0000000000000533

Häfliger, I. M., Seefried, F. R., & Drögemüller, C. (2021). Reverse genetic screen for deleterious recessive variants in the local simmental cattle population of Switzerland. *Animals*, *11*(12). https://doi.org/10.3390/ani11123535

Hanlon, D. W., Stevenson, M., Evans, M. J., & Firth, E. C. (2012). Reproductive performance of Thoroughbred mares in the Waikato region of New Zealand: 2. Multivariable analyses and sources of variation at the mare, stallion and stud farm level. *New Zealand Veterinary Journal*, *60*(6), 335–343. https://doi.org/10.1080/00480169.2012.696240

Haseman, J. K., & Eiston, R. C. (1972). The Investigation of Linkage Between a Quantitative Trait and a Marker Locus. In *Behavior Genetics* (Vol. 2, Issue l).

Hasstedt, S. J. (1993). Variance components/major locus likelihood approximation for quantitative, polychotomous, and multivariate data. *Genetic Epidemiology*, *10*(3), 145–158. https://doi.org/10.1002/gepi.1370100302

Hemberg, E., Lundeheim, N., & Einarsson, S. (2004). Reproductive performance of Thoroughbred mares in Sweden. *Reproduction in Domestic Animals*, *39*(2), 81–85. https://doi.org/10.1111/j.1439-0531.2004.00482.x

Hoff, J. L., Decker, J. E., Schnabel, R. D., & Taylor, J. F. (2017). Candidate lethal haplotypes and causal mutations in Angus cattle. *BMC Genomics*, *18*(1). https://doi.org/10.1186/s12864-017-4196-2

Hoffmann, I., & Scherf, B. (2010). Implementing the Global Plan of Action for Animal Genetic Resources. *Animal Genetic Resources/Ressources Génétiques Animales/Recursos Genéticos Animales*, *47*, 1–10. https://doi.org/10.1017/s2078633610001050

Holly, M. K., Diaz, K., & Smith, J. G. (2017). Defensins in Viral Infection and Pathogenesis. *Annual Review of Virology*, *4*(1), 369–391. https://doi.org/10.1146/annurev-virology-101416-041734

Howard, D. M., Pong-Wong, R., Knap, P. W., & Woolliams, J. A. (2017). Use of haplotypes to identify regions harbouring lethal recessive variants in pigs. *Genetics Selection Evolution*, *49*(1). https://doi.org/10.1186/s12711-017-0332-3

Høyseth, M. (2017). *Masteroppgave 2017 60 stp Fakultet for Biovitenskap Institutt for Husdyr-og Akvakulturvitenskap*.

Huber, F. M., Greenblatt, S. M., Davenport, A. M., Martinez, C., Xu, Y., Vu, L. P., Nimer, S. D., & Hoelz, A. (2017). Histone-binding of DPF2 mediates its repressive role in myeloid differentiation. *Proceedings of the National Academy of Sciences*, *114*(23), 6016–6021. https://doi.org/10.1073/pnas.1700328114

Huggett, M. (2009, December 1). *The Norwegian Fjord: A Horse for All Ages*. Canadian Horse Journal. https://www.horsejournals.com/popular/breed-profiles/norwegian-fjord-horse-all-ages

Jeng, E. E., Bhadkamkar, V., Ibe, N. U., Gause, H., Jiang, L., Chan, J., Jian, R., Jimenez-Morales, D., Stevenson, E., Krogan, N. J., Swaney, D. L., Snyder, M. P., Mukherjee, S., & Bassik, M. C. (2019). Systematic Identification of Host Cell Regulators of Legionella pneumophila Pathogenesis Using a Genome-wide CRISPR Screen. *Cell Host & Microbe*, *26*(4), 551-563. https://doi.org/10.1016/j.chom.2019.08.017

Jin, G., & Weinberg, A. (2019). Human antimicrobial peptides and cancer. *Seminars in Cell & Developmental Biology*, *88*, 156–162. https://doi.org/10.1016/j.semcdb.2018.04.006

Jones, D. E., & Bevins, C. L. (1992). Paneth cells of the human small intestine express an antimicrobial peptide gene. *The Journal of Biological Chemistry*, *267*(32), 23216–23225.

Jones, D. E., & Bevins, C. L. (1993). Defensin-6 mRNA in human Paneth cells: implications for antimicrobia peptides in host defense of the human bowel. *FEBS Letters*, *315*(2), 187–192. https://doi.org/10.1016/0014-5793(93)81160-2

Kalbfleisch, T. S., Rice, E. S., DePriest, M. S., Walenz, B. P., Hestand, M. S., Vermeesch, J. R., O'Connell, B. L., Fiddes, I. T., Vershinina, A. O., Saremi, N. F., Petersen, J. L., Finno, C. J., Bellone, R. R., McCue, M. E., Brooks, S. A., Bailey, E., Orlando, L., Green, R. E., Miller, D. C., … MacLeod, J. N. (2018). Improved reference genome for the domestic horse increases assembly contiguity and composition. *Communications Biology*, *1*(1). https://doi.org/10.1038/s42003-018-0199-z

Kandels-Lewis, S., & Séraphin, B. (1993). Involvement of U6 snRNA in 5' Splice Site Selection. *Science*, *262*(5142), p. 2035. https://doi.org/10.1126/science.8266100

Kane, M. S., Diamonstein, C. J., Hauser, N., Deeken, J. F., Niederhuber, J. E., & Vilboux, T. (2019). Endosomal trafficking defects in patient cells with KIAA1109 biallelic variants. *Genes and Diseases*, *6*(1), 56–67. https://doi.org/10.1016/j.gendis.2018.12.004

Katila, T., Reilas, T., Nivola, K., Peltonen, T., & Virtala, A. M. (2010). A 15-year survey of reproductive efficiency of Standardbred and Finnhorse trotters in Finland - descriptive results. *Acta Veterinaria Scandinavica*, *52*(1). https://doi.org/10.1186/1751-0147-52-40

Keller, L., & Waller, D. M. (2002). Inbreeding effects in wild populations. *Trends in Ecology & Evolution*, *17*(5), 230–241. https://doi.org/10.1016/S0169-5347(02)02489-8

Kiani, A. K., John, P., Bhatti, A., Zia, A., Shahid, G., Akhtar, P., Wang, X., Demirci, F. Y., & Kamboh, M. I. (2015). Association of 32 type 1 diabetes risk loci in Pakistani patients. *Diabetes Research and Clinical Practice*, *108*(1), 137–142. https://doi.org/10.1016/j.diabres.2015.01.022

Kikuno, R. (1999). Prediction of the Coding Sequences of Unidentified Human Genes. XIV. The Complete Sequences of 100 New cDNA Clones from Brain Which Code for Large Proteins in vitro. *DNA Research*, *6*(3), 197–205. https://doi.org/10.1093/dnares/6.3.197

Kim, C., Gajendran, N., Mittrücker, H.-W., Weiwad, M., Song, Y.-H., Hurwitz, R., Wilmanns, M., Fischer, G., & Kaufmann, S. H. E. (2005). Human α-defensins neutralize anthrax lethal toxin and protect against its fatal consequences. *Proceedings of the National Academy of Sciences*, *102*(13), 4830–4835. https://doi.org/10.1073/pnas.0500508102

King, A. E., Kelly, R. W., Sallenave, J.-M., Bocking, A. D., & Challis, J. R. G. (2007). Innate Immune Defences in the Human Uterus during Pregnancy. *Placenta*, *28*(11–12), 1099–1106. https://doi.org/10.1016/j.placenta.2007.06.002

King, A. E., Paltoo, A., Kelly, R. W., Sallenave, J.-M., Bocking, A. D., & Challis, J. R. G. (2007). Expression of Natural Antimicrobials by Human Placenta and Fetal Membranes. *Placenta*, *28*(2–3), 161–169. https://doi.org/10.1016/j.placenta.2006.01.006

King, R. C., Mulligan, P. K., & Stansfield, W. D. (2013). *A Dictionary of Genetics*. Oxford University Press. https://doi.org/10.1093/acref/9780199766444.001.0001

Kirk, K. M., & Cardon, L. R. (2002). The impact of genotyping error on haplotype reconstruction and frequency estimation. *European Journal of Human Genetics*, *10*(10), 616–622. https://doi.org/10.1038/sj.ejhg.5200855

Kjöllerström, H. J., Gama, L. T., & Oom, M. M. (2015). Impact of inbreeding on fitness-related traits in the highly threatened Sorraia horse breed. *Livestock Science*, *180*, 84–89. https://doi.org/10.1016/j.livsci.2015.08.001

Klemetsdal, G., & Johnson, M. (1989). Effect of Inbreeding on Fertility in Norwegian Trotter. In *Livestock Production Science* (Vol. 21).

Klotman, M. E., & Chang, T. L. (2006). Defensins in innate antiviral immunity. *Nature Reviews Immunology*, *6*(6), 447–456. https://doi.org/10.1038/nri1860

Koch, A. A., & Rampersaud, E. (2022). Determining the Genetic Component of a Disease. In W. K. Scott & M. D. Ritchie (Eds.), *Genetic Analysis of Complex Disease* (3rd ed., pp. 36–53). John Wiley & Sons.

Kruglyak, L., Daly, M. J., & Lander, E. S. (1995). Rapid Multipoint Linkage Analysis of Recessive Traits in Nuclear Families, Including Homozygosity Mapping. In *Am. J. Hum. Genet* (Vol. 56).

Kuhl, J., Stock, K. F., Wulf, M., & Aurich, C. (2015). Maternal lineage of Warmblood mares contributes to variation of gestation length and bias of foal sex ratio. *PLoS ONE*, *10*(10). https://doi.org/10.1371/journal.pone.0139358

Kumar, K., Bellad, A., Bellad, A., Prasad, P., Girimaji, S. C., Muthusamy, B., & Muthusamy, B. (2020). KIAA1109 gene mutation in surviving patients with Alkuraya-Kučinskas syndrome: A review of literature. *BMC Medical Genetics*, *21*(1). https://doi.org/10.1186/s12881-020-01074-2

Kuo, M. T., Wei, Y., Yang, X., Tatebe, S., Liu, J., Troncoso, P., Sahin, A., Ro, J. Y., Hamilton, S. R., & Savaraj, N. (2006). Association of fragile site-associated (FSA) gene expression with epithelial differentiation and tumor development. *Biochemical and Biophysical Research Communications*, *340*(3), 887–893. https://doi.org/10.1016/j.bbrc.2005.12.088

Lai, Y., & Gallo, R. L. (2009). AMPed up immunity: how antimicrobial peptides have multiple roles in immune defense. *Trends in Immunology*, *30*(3), 131–141. https://doi.org/10.1016/j.it.2008.12.003

Lalouel, J. M. (1984). Segregation analysis: a gene or not a gene. In C. Carter, R. Havlik, M. Feinleib, L. H. Kuller, R. Elston, & D. C. Rao (Eds.), *Genetic epidemiology of coronary heart disease. Past, present, and future.* (pp. 217–243). Alan R. Liss, Inc. https://doi.org/10.1161/01.atv.4.5.510

Lalouel, J. M., & Morton, N. E. (1981). Complex segregation analysis with pointers. *Human Heredity*, *31*(5), 312–321. https://doi.org/10.1159/000153231

Lander, E. S., & Botstein, D. (1987). Homozygosity Mapping: A Way to Map Human Recessive Traits with the DNA of Inbred Children. In *Source: Science* (Vol. 236, Issue 4808).

Langlois, B., & Blouin, C. (2004). Statistical analysis of some factors affecting the number of horse births in France. *Reproduction Nutrition Development*, *44*(6), 583–595. https://doi.org/10.1051/rnd:2004055

Laseca, N., Anaya, G., Peña, Z., Pirosanto, Y., Molina, A., & Demyda-Peyrás, S. (2021). Impaired reproductive function in equines: From genetics to genomics. In *Animals* (Vol. 11, Issue 2, pp. 1–15). MDPI AG. https://doi.org/10.3390/ani11020393

Laseca, N., Demyda-Peyrás, S., Valera, M., Ramón, M., Escribano, B., Perdomo-González, D. I., & Molina, A. (2022). A genome-wide association study of mare fertility in the Pura Raza Español horse. *Animal*, *16*(3). https://doi.org/10.1016/j.animal.2022.100476

Laseca, N., Perdomo-González, D. I., Valera, M., Molina, A., Sanchez-Guerrero, M. J., Azcona, F., Pirosanto, Y., & Demyda-Peyrás, S. (2020). Foaling number are highly affected by the genomic homozygosity in the Pura Raza Spanish mares. *Journal of Equine Veterinary Science*, *89*, 103098. https://doi.org/10.1016/j.jevs.2020.103098

Lehrer, R. I., & Lu, W. (2012). α-Defensins in human innate immunity. In *Immunological Reviews* (Vol. 245, Issue 1, pp. 84–112). https://doi.org/10.1111/j.1600-065X.2011.01082.x

Leroy, G. (2014). Inbreeding depression in livestock species: Review and meta-analysis. In *Animal Genetics* (Vol. 45, Issue 5, pp. 618–628). Blackwell Publishing Ltd. https://doi.org/10.1111/age.12178

Leroy, G., Phocas, F., Hedan, B., Verrier, E., & Rognon, X. (2015). Inbreeding impact on litter size and survival in selected canine breeds. *Veterinary Journal*, *203*(1), 74–78. https://doi.org/10.1016/j.tvjl.2014.11.008

Lesser, C. F., & Guthrie, C. (1993). Mutations in U6 snRNA that Alter Splice Site Specificity: Implications for the Active Site. *Science*, *262*(5142), 1982–1988. https://doi.org/10.1126/science.8266093

Levinson, P., Choi, R. Y., Cole, A. L., Hirbod, T., Rhedin, S., Payne, B., Guthrie, B. L., Bosire, R., Cole, A. M., Farquhar, C., & Broliden, K. (2012). HIV-Neutralizing Activity of Cationic Polypeptides in Cervicovaginal Secretions of Women in HIV-Serodiscordant Relationships. *PLoS ONE*, *7*(2), e31996. https://doi.org/10.1371/journal.pone.0031996

Lim, Y., Mantalaris, A., Yap, M. G. S., & Wong, D. C. F. (2010). Simultaneous targeting of requiem & Alg-2 in Chinese hamster ovary cells for improved recombinant protein production. *Molecular Biotechnology*, *46*(3), 301–307. https://doi.org/10.1007/s12033-010-9304-3

Lipkin, E., Mosig, M. O., Darvasi, A., Ezra, E., Shalom, A., Friedmann, A., & Soller, M. (1998). *Quantitative Trait Locus Mapping in Dairy Cattle by Means of Selective Milk DNA Pooling Using Dinucleotide Microsatellite Markers: Analysis of Milk Protein Percentage*.

Liu, Z., Jaitner, J., Reinhardt, F., Pasman, E., Rensing, S., & Reents, R. (2008). Genetic evaluation of fertility traits of dairy cattle using a multiple-trait animal model. *Journal of Dairy Science*, *91*(11), 4333–4343. https://doi.org/10.3168/jds.2008-1029

Louahchi, S., Allam, I., Berkani, L., Boucharef, A., Abdesemed, A., Khaldoun, N., Nebbab, A., Ladjouze, A., & Djidjik, R. (2016). Association study of single nucleotide polymorphisms of IL23R and IL17 in rheumatoid arthritis in the Algerian population. *Acta Reumatologica Portuguesa*, *41*(2), 151–157.

MacLean, C. J., Morton, N. E., & Yee, S. (1984). Combined analysis of genetic segregation and linkage under an oligogenic model. *Computers and Biomedical Research, an International Journal*, *17*(5), 471–480. https://doi.org/10.1016/0010-4809(84)90013-2

Mahon, G. A. T., & Cunningham, E. P. (1982). *INBREEDING AND THE INHERITANCE OF FERTILITY IN THE THOROUGHBRED MARE*.

Malheiros de Souza, J. R., Gonçalves, P. B. D., Bertolin, K., Ferreira, R., Ribeiro, A. S. S., Ribeiro, D. B., Rechsteiner, S. M. da E. F., Araújo, R. R. de, Pimentel, C. A., & Pimentel, A. M. H. (2020). Age-Dependent Effect of Foal Heat Breeding on Pregnancy and Embryo Mortality Rates in Thoroughbred Mares. *Journal of Equine Veterinary Science*, *90*. https://doi.org/10.1016/j.jevs.2020.102982

Malnic, B., Godfrey, P. A., Buck, L. B., & Howard, § ‡. (2004). *The human olfactory receptor gene family*. www.pnas.orgcgidoi10.1073pnas.0307882100

Maltecca, C., Tiezzi, F., Cole, J. B., & Baes, C. (2020). Symposium review: Exploiting homozygosity in the era of genomics—Selection, inbreeding, and mating programs. In *Journal of Dairy Science* (Vol. 103, Issue 6, pp. 5302–5313). Elsevier Inc. https://doi.org/10.3168/jds.2019-17846

Mantovani, R., Folla, F., Pigozzi, G., Tsuruta, S., & Sartori, C. (2020). Genetics of lifetime reproductive performance in italian heavy draught horse mares. *Animals*, *10*(6), 1–14. https://doi.org/10.3390/ani10061085

McDowell, K. J., Powell, D. G., & Baker, C. B. (1992). Effect of book size and age of mare and stallion on foaling rates in thoroughbred horses. *Journal of Equine Veterinary Science*, *12*(6), 364–367. https://doi.org/10.1016/S0737-0806(06)81363-8

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., Flicek, P., & Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biology*, *17*(1). https://doi.org/10.1186/s13059-016-0974-4

McQuillan, R., Leutenegger, A.-L., Abdel-Rahman, R., Franklin, C. S., Pericic, M., Barac-Lauc, L., Smolej-Narancic, N., Janicijevic, B., Polasek, O., Tenesa, A., MacLeod, A. K., Farrington, S. M., Rudan, P., Hayward, C., Vitart, V., Rudan, I., Wild, S. H., Dunlop, M. G., Wright, A. F., … Wilson, J. F. (2008). Runs of Homozygosity in European Populations. *The American Journal of Human Genetics*, *83*(3), 359–372. https://doi.org/10.1016/j.ajhg.2008.08.007

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). https://doi.org/10.1186/s12864-015-1977-3

Meyer, K., Hammond, K., Parnell, P. F., Mackinnon, M. J., & Sivarajasingam, S. (1990). Estimates of Heritability and Repeatability for Reproductive Traits in Australian Beef Cattle. In *Livestock Production Science* (Vol. 25).

Milone, R., Gnazzo, M., Stefanutti, E., Serafin, D., & Novelli, A. (2020). A new missense mutation in DPF2 gene related to Coffin Siris syndrome 7: Description of a mild phenotype expanding DPF2-related clinical spectrum and differential diagnosis among similar syndromes epigenetically determined. *Brain and Development*, *42*(2), 192–198. https://doi.org/10.1016/j.braindev.2019.10.007

Morel, M. C. G. D., & Gunnarsson, V. (2000). *A survey of the fertility of Icelandic stallions*.

Morris, L. H. A., & Allen, W. R. (2002). Reproductive efficiency of intensively managed T h o r o u g h b r e d mares in Newmarket. In *EQUINE VETERINARY JOURNAL Equine vet. J* (Vol. 34, Issue 1).

Morton, N. E., & MacLean, C. J. (1974). Analysis of family resemblance. 3. Complex segregation of quantitative traits. *American Journal of Human Genetics*, *26*(4), 489–503.

Moyers, B. T., Morrell, P. L., & McKay, J. K. (2018). Genetic costs of domestication and improvement. *Journal of Heredity*, *109*(2), 103–116. https://doi.org/10.1093/jhered/esx069

Mucha, S., Wolc, A., & Szwaczkowski, T. (2012). Bayesian and REML analysis of twinning and fertility in Thoroughbred horses. *Livestock Science*, *144*(1–2), 82–88. https://doi.org/10.1016/j.livsci.2011.10.015

Mucosa. (2012). Farlex Partner Medical Dictionary. https://medical-dictionary.thefreedictionary.com/mucosa

Müller-Unterberg, M., Wallmann, S., & Distl, O. (2017). Effects of inbreeding and other systematic effects on fertility of Black Forest Draught horses in Germany. *Acta Veterinaria Scandinavica*, *59*(1). https://doi.org/10.1186/s13028-017-0338-4

Nandolo, W., Utsunomiya, Y. T., Mészáros, G., Wurzinger, M., Khayadzadeh, N., Torrecilha, R. B. P., Mulindwa, H. A., Gondwe, T. N., Waldmann, P., Ferenčaković, M., Garcia, J. F., Rosen, B. D., Bickhart, D., Van Tassell, C. P., Curik, I., & Sölkner, J. (2018). Misidentification of runs of homozygosity islands in cattle caused by interference with copy number variation or large intermarker distances. *Genetics Selection Evolution*, *50*(1). https://doi.org/10.1186/s12711-018-0414-x

Nicholas, F. W. (1987). *Veterinary Genetics* (F. W. Nicholas, Ed.; 1st ed.). Oxford University Press.

Norsk Hestesenter. (n.d.). *Fjordhest*. Nhest.no. Retrieved August 6, 2022, from https://www.nhest.no/fjordhest.476359.no.html

Norsk Hestesenter. (2020). *Nøkkeltallsrapport*.

Norsk Hestesenter. (2021a). *Handlingsplan for bevaring og utvikling av de nasjonale hesteraser 2021-2030*.

Norsk Hestesenter. (2021b). *Nøkkeltallsrapport*.

Olsen, H. F. (2011). *Genetic variation and management of the Norwegian horse breeds Genetisk variasjon og forvaltning av de saernorske hesterasene* [PhD, NMBU]. www.umb.no,

Olsen, H. F., & Klemetsdal, G. (2010). Management to ensure effective population size in a breeding programme for the small Norwegian horse breeds a simulation study. *Acta Agriculturae Scandinavica A: Animal Sciences*, *60*(1), 60–63. https://doi.org/10.1080/09064700903567799

Olsen, H. F., Tenhunen, S., Dolvik, N. I., Våge, D. I., & Klemetsdal, G. (2020). Segment-based coancestry, additive relationship and genetic variance within and between the Norwegian and the Swedish Fjord horse populations. *Acta Agriculturae Scandinavica A: Animal Sciences*, *69*(1–2), 118–126. https://doi.org/10.1080/09064702.2019.1711155

Ordonez, S. R., Veldhuizen, E. J. A., van Eijk, M., & Haagsman, H. P. (2017). Role of Soluble Innate Effector Molecules in Pulmonary Defense against Fungal Pathogens. *Frontiers in Microbiology*, *8*. https://doi.org/10.3389/fmicb.2017.02098

Ouellette, A. J. (2011). Paneth cell α-defensins in enteric innate immunity. *Cellular and Molecular Life Sciences*, *68*(13), 2215–2229. https://doi.org/10.1007/s00018-011-0714-6

Palmer, E., & Chavatte-Palmer, P. (2020). Contribution of Reproduction Management and Technologies to Genetic Progress in Horse Breeding. *Journal of Equine Veterinary Science*, *89*. https://doi.org/10.1016/j.jevs.2020.103016

Patil, A. A., Cai, Y., Sang, Y., Blecha, F., & Zhang, G. (2005). Cross-species analysis of the mammalian β-defensin gene family: presence of syntenic gene clusters and preferential expression in the male reproductive tract. *Physiological Genomics*, *23*(1), 5–17. https://doi.org/10.1152/physiolgenomics.00104.2005

Pausch, H., Schwarzenbacher, H., Burgstaller, J., Flisikowski, K., Wurmser, C., Jansen, S., Jung, S., Schnieke, A., Wittek, T., & Fries, R. (2015). Homozygous haplotype deficiency reveals deleterious mutations compromising reproductive and rearing success in cattle. *BMC Genomics*, *16*(1). https://doi.org/10.1186/s12864-015-1483-7

Perdomo-González, D. I., Molina, A., Sánchez-Guerrero, M. J., Bartolomé, E., Varona, L., & Valera, M. (2021). Genetic inbreeding depression load for fertility traits in Pura Raza Española mares. *Journal of Animal Science*, *99*(12). https://doi.org/10.1093/jas/skab316

Perdomo-González, D. I., Sánchez-Guerrero, M. J., Molina, A., & Valera, M. (2020). Genetic structure analysis of the pura raza español horse population through partial inbreeding coefficient estimation. *Animals*, *10*(8), 1–13. https://doi.org/10.3390/ani10081360

Phocas, F., Bloch, C., Chapelle, P., Becherel, F., Renand, G., & Menissier, F. (1998). Developing a breeding objective for a French purebred beef 1 cattle selection programme. In *Livestock Production Science* (Vol. 57).

Pirosanto, Y., Valera, M., Molina, A., Dorado, J., & Demyda-Peyrás, S. (2020). 23 Sperm quality of Pure Spanish stallions is affected by inbreeding coefficient and age. *Reproduction, Fertility and Development*, *32*(2), 137. https://doi.org/10.1071/RDv32n2Ab23

Plaza-Izurieta, L., Castellanos-Rubio, A., Irastorza, I., Fernandez-Jimenez, N., Gutierrez, G., & Bilbao, J. R. (2011). Revisiting genome wide association studies (GWAS) in coeliac disease: replication study in Spanish population and expression analysis of candidate genes. *Journal of Medical Genetics*, *48*(7), 493–496. https://doi.org/10.1136/jmg.2011.089714

Qiao, Z., Jiang, Y., Wang, L., Wang, L., Jiang, J., & Zhang, J. (2019). Mutations in KIAA1109, CACNA1C, BSN, AKAP13, CELSR2, and HELZ2 Are Associated With the Prognosis in Endometrial Cancer. *Frontiers in Genetics*, *10*. https://doi.org/10.3389/fgene.2019.00909

Quayle, A. J., Porter, E. M., Nussbaum, A. A., Wang, Y. M., Brabec, C., Yip, K. P., & Mok, S. C. (1998). Gene expression, immunolocalization, and secretion of human defensin-5 in human female reproductive tract. *The American Journal of Pathology*, *152*(5), 1247–1258.

Reynolds, E. G. M., Neeley, C., Lopdell, T. J., Keehan, M., Dittmer, K., Harland, C. S., Couldrey, C., Johnson, T. J. J., Tiplady, K., Worth, G., Walker, M., Davis, S. R., Sherlock, R. G., Carnie, K., Harris, B. L., Charlier, C., Georges, M., Spelman, R. J., Garrick, D. J., & Littlejohn, M. D. (2021). Non-additive association analysis using proxy phenotypes identifies novel cattle syndromes. *Nature Genetics*, *53*(7), 949–954. https://doi.org/10.1038/s41588-021-00872-5

Rice, T., Després, J., Daw, E., Gagnon, J., Borecki, I., Pérusse, L., Leon, A., Skinner, J., Wilmore, J., Rao, D., & Bouchard, C. (1997). Familial resemblance for abdominal visceral fat: the HERITAGE family study. *International Journal of Obesity*, *21*(11), 1024–1031. https://doi.org/10.1038/sj.ijo.0800511

Risch, N. (1990). Linkage strategies for genetically complex traits. II. The power of affected relative pairs. *American Journal of Human Genetics*, *46*(2), 229–241.

RNA, U6 SMALL NUCLEAR, 1; RNU6-1. (n.d.). OMIM. Retrieved March 11, 2023, from https://www.omim.org/entry/180692

Robertson, M. J., Kent, K., Tharp, N., Nozawa, K., Dean, L., Mathew, M., Grimm, S. L., Yu, Z., Légaré, C., Fujihara, Y., Ikawa, M., Sullivan, R., Coarfa, C., Matzuk, M. M., & Garcia, T. X. (2020). Large-scale discovery of male reproductive tract-specific genes through analysis of RNA-seq datasets. *BMC Biology*, *18*(1), 103. https://doi.org/10.1186/s12915-020-00826-z

Robinson, D. L. (2008). Days to calving in artificially inseminated cattle: Alternative models and analyses. *Livestock Science*, *117*(1), 34–42. https://doi.org/10.1016/j.livsci.2007.11.010

Rochmann, J. (2016). *CURRENT BREEDING AND GENETIC HEALTH STATUS OF THE NORWEGIAN FJORD HORSE*.

Rodrigues, J. A., Gonçalves, A. R., Antunes, L., Bettencourt, E. V., & Gama, L. T. (2020). Genetic and Environmental Factors Influencing Gestation Length in Lusitano Horses. *Journal of Equine Veterinary Science*, *84*. https://doi.org/10.1016/j.jevs.2019.102850

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.

Sabeva, I. (2011). Longevity and life reproductive efficiency in Arabian broodmares. In *Apostolov 66 Archiva Zootechnica* (Vol. 14).

Sahana, G., Nielsen, U. S., Aamand, G. P., Lund, M. S., & Guldbrandtsen, B. (2013). Novel harmful recessive haplotypes identified for fertility traits in Nordic Holstein cattle. *PLoS ONE*, *8*(12). https://doi.org/10.1371/journal.pone.0082909

Sahebi, L., Dastgiri, S., Ansarin, K., Sahebi, R., & Mohammadi, S. A. (2013). Study Designs in Genetic Epidemiology. *ISRN Genetics*, *2013*, 1–8. https://doi.org/10.5402/2013/952518

Sairanen, J., Nivola, K., Katila, T., Virtala, A. M., & Ojala, M. (2009). Effects of inbreeding and other genetic components on equine fertility. *Animal*, *3*(12), 1662–1672. https://doi.org/10.1017/S1751731109990553

Samper, J. C. (2001). *Breeding Mares with Frozen Semen in Private Practice*. https://www.researchgate.net/publication/265450901

Sangeeta, K., & Yenugu, S. (2019). Male reproductive tract antimicrobial expression in the extremes of ages of rats. *Gene*, *710*, 218–232. https://doi.org/10.1016/j.gene.2019.05.053

Schubert, M., Jónsson, H., Chang, D., Der Sarkissian, C., Ermini, L., Ginolhac, A., Albrechtsen, A., Dupanloup, I., Foucal, A., Petersen, B., Fumagalli, M., Raghavan, M., Seguin-Orlando, A., Korneliussen, T. S., Velazquez, A. M. V., Stenderup, J., Hoover, C. A., Rubin, C. J., Alfarhan, A. H., … Orlando, L. (2014). Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(52), E5661–E5669. https://doi.org/10.1073/pnas.1416991111

Schutte, B. C., & McCray, P. B. (2002). β-Defensins in Lung Host Defense. *Annual Review of Physiology*, *64*(1), 709–748. https://doi.org/10.1146/annurev.physiol.64.081501.134340

Scott, W. K., Ritchie, M. D., & Haines, J. L. (2022). Designing a Study for Identifying Genes in Complex Traits. In *Genetic Analysis of Complex Disease* (Vol. 3, pp. 1–11). John Wiley & Sons.

Selsted, M. E., & Ouellette, A. J. (2005). Mammalian defensins in the antimicrobial immune response. *Nature Immunology*, *6*(6), 551–557. https://doi.org/10.1038/ni1206

Semple, F., & Dorin, J. R. (2012). β-Defensins: Multifunctional Modulators of Infection, Inflammation and More? *Journal of Innate Immunity*, *4*(4), 337–348. https://doi.org/10.1159/000336619

Shafee, T. M. A., Lay, F. T., Phan, T. K., Anderson, M. A., & Hulett, M. D. (2017). Convergent evolution of defensin sequence, structure and function. *Cellular and Molecular Life Sciences*, *74*(4), 663–682. https://doi.org/10.1007/s00018-016-2344-5

Shamseldin, H. E., Kurdi, W., Almusafri, F., Alnemer, M., Alkaff, A., Babay, Z., Alhashem, A., Tulbah, M., Alsahan, N., Khan, R., Sallout, B., Al Mardawi, E., Seidahmed, M. Z., Meriki, N., Alsaber, Y., Qari, A., Khalifa, O., Eyaid, W., Rahbeeni, Z., … Alkuraya, F. S. (2018). Molecular autopsy in maternal–fetal medicine. *Genetics in Medicine*, *20*(4), 420–427. https://doi.org/10.1038/gim.2017.111

Sharma, R., Gupta, S., Agarwal, A., Henkel, R., Finelli, R., Parekh, N., Saleh, R., Arafa, M., Ko, E., Zini, A., Tadros, N., Shah, R., Ambar, R. F., Elbardisi, H., Sengupta, P., Martinez, M., Boitrelle, F., Simopoulou, M., Vogiatzi, P., … Colpi, G. M. (2022). Relevance of Leukocytospermia and Semen Culture and Its True Place in Diagnosing and Treating Male Infertility. *The World Journal of Men's Health*, *40*(2), 191. https://doi.org/10.5534/wjmh.210063

Sharma, S., Dhaliwal, G. S., & Dadarwal, D. (2010). Reproductive efficiency of Thoroughbred mares under Indian subtropical conditions: A retrospective survey over 7 years. *Animal Reproduction Science*, *117*(3–4), 241–248. https://doi.org/10.1016/j.anireprosci.2009.05.011

Sharmaa, A., Lee, J. S., Dang, C. G., Sudrajad, P., Kim, H. C., Yeon, S. H., Kang, H. S., & Lee, S. H. (2015). Stories and challenges of genome wide association studies in livestock - a review. In *Asian-Australasian Journal of Animal Sciences* (Vol. 28, Issue 10, pp. 1371–1379). Asian-Australasian Association of Animal Production Societies. https://doi.org/10.5713/ajas.14.0715

Shrine, N., Portelli, M. A., John, C., Soler Artigas, M., Bennett, N., Hall, R., Lewis, J., Henry, A. P., Billington, C. K., Ahmad, A., Packer, R. J., Shaw, D., Pogson, Z. E. K., Fogarty, A., McKeever, T. M., Singapuri, A., Heaney, L. G., Mansur, A. H., Chaudhuri, R., … Sayers, I. (2019). Moderate-to-severe asthma in individuals of European ancestry: a genome-wide association study. *The Lancet. Respiratory Medicine*, *7*(1), 20–34. https://doi.org/10.1016/S2213-2600(18)30389-8

Shust, G. F., Cho, S., Kim, M., Madan, R. P., Guzman, E. M., Pollack, M., Epstein, J., Cohen, H. W., Keller, M. J., & Herold, B. C. (2009). ORIGINAL ARTICLE: Female Genital Tract Secretions Inhibit Herpes Simplex Virus Infection: Correlation with Soluble Mucosal Immune Mediators and Impact of Hormonal Contraception. *American Journal of Reproductive Immunology*, *63*(2), 110–119. https://doi.org/10.1111/j.1600-0897.2009.00768.x

Spencer, J. D., Hains, D. S., Porter, E., Bevins, C. L., DiRosario, J., Becknell, B., Wang, H., & Schwaderer, A. L. (2012). Human Alpha Defensin 5 Expression in the Human Kidney and Urinary Tract. *PLoS ONE*, *7*(2), e31712. https://doi.org/10.1371/journal.pone.0031712

Spencer, T. E. (2013). Early pregnancy: Concepts, challenges, and potential solutions. *Animal Frontiers*, *3*(4), 48–55. https://doi.org/10.2527/af.2013-0033

Stock, K. F., Quinn Brady, K., Christiansen, K., Viklund, Å., Cervantes, I., Ricard, A., Ducro, B., & Janssens, S. (2015). *Breeding objectives and practices of sport horse studbooks: results of a worldwide inventory*.

Sullivan, J. J., Turner, P. C., Self, L. C., Gutteridge, H. B., & Bartlett, D. E. (1975). Survey of reproductive efficiency in the Quarter-horse and Thoroughbred. *Journal of Reproduction and Fertility. Supplement*, *23*, 315–318.

Svinarich, D. M., Gomez, R., & Romero, R. (1997). Detection of Human Defensins in the Placenta. *American Journal of Reproductive Immunology*, *38*(4), 252–255. https://doi.org/10.1111/j.1600-0897.1997.tb00511.x

Tanabe, H., Yuan, J., Zaragoza, M. M., Dandekar, S., Henschen-Edman, A., Selsted, M. E., & Ouellette, A. J. (2004). Paneth Cell α-Defensins from Rhesus Macaque Small Intestine. *Infection and Immunity*, *72*(3), 1470–1478. https://doi.org/10.1128/IAI.72.3.1470-1478.2004

Tando, T., Ishizaka, A., Watanabe, H., Ito, T., Iida, S., Haraguchi, T., Mizutani, T., Izumi, T., Isobe, T., Akiyama, T., Inoue, J. I., & Iba, H. (2010). Requiem protein links RelB/p52 and the Brm-type SWI/SNF complex in a noncanonical NF-κB pathway. *Journal of Biological Chemistry*, *285*(29), 21951–21960. https://doi.org/10.1074/jbc.M109.087783

Tang, D., Kang, R., Coyne, C. B., Zeh, H. J., & Lotze, M. T. (2012). PAMPs and DAMPs: signal 0s that spur autophagy and immunity. *Immunological Reviews*, *249*(1), 158–175. https://doi.org/10.1111/j.1600-065X.2012.01146.x

The National Committee for Research Ethics in Science and Technology (NENT). (2016). *Guidelines for Research Ethics in Science and Technology* (2nd ed.). https://www.forskningsetikk.no/globalassets/dokumenter/4-publikasjoner-som-pdf/60126_fek_guidelines_nent_digital.pdf

Thomas, D. C. (2004). Statistical Methods in Genetic Epidemiology. In D. Thomas (Ed.), *Statistical Methods in Genetic Epidemiology* (Vol. 01). Oxford University Press.

Tindall, E. A., Hoang, H. N., Southey, M. C., English, D. R., Hopper, J. L., Giles, G. G., Severi, G., & Hayes, V. M. (2010). The 4q27 locus and prostate cancer risk. *BMC Cancer*, *10*(1), 69. https://doi.org/10.1186/1471-2407-10-69

Todd, E. T., Hamilton, N. A., Velie, B. D., & Thomson, P. C. (2020). The effects of inbreeding on covering success, gestation length and foal sex ratio in Australian thoroughbred horses. *BMC Genetics*, *21*(1). https://doi.org/10.1186/s12863-020-00847-1

Todd, E. T., Ho, S. Y. W., Thomson, P. C., Ang, R. A., Velie, B. D., & Hamilton, N. A. (2018). Founder-specific inbreeding depression affects racing performance in Thoroughbred horses. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-24663-x

Todd, E. T., Thomson, P. C., Hamilton, N. A., Ang, R. A., Lindgren, G., Viklund, Å., Eriksson, S., Mikko, S., Strand, E., & Velie, B. D. (2020). A genome-wide scan for candidate lethal variants in Thoroughbred horses. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-68946-8

Tollner, T. L., Bevins, C. L., & Cherr, G. N. (2012). Multifunctional glycoprotein DEFB126—a curious story of defensin-clad spermatozoa. *Nature Reviews Urology*, *9*(7), 365–375. https://doi.org/10.1038/nrurol.2012.109

Tollner, T. L., Venners, S. A., Hollox, E. J., Yudin, A. I., Liu, X., Tang, G., Xing, H., Kays, R. J., Lau, T., Overstreet, J. W., Xu, X., Bevins, C. L., & Cherr, G. N. (2011). A Common Mutation in the Defensin *DEFB126* Causes Impaired Sperm Function and Subfertility. *Science Translational Medicine*, *3*(92). https://doi.org/10.1126/scitranslmed.3002289

United Nations Conference on Environment and Development. (1992).

Utsunomiya, Y. T., Milanesi, M., Barbato, M., Utsunomiya, A. T. H., Sölkner, J., Ajmone-Marsan, P., & Garcia, J. F. (2020). Unsupervised detection of ancestry tracks with the GHap r package. *Methods in Ecology and Evolution*, *11*(11), 1448–1454. https://doi.org/10.1111/2041-210X.13467

Utsunomiya, Y. T., Milanesi, M., Utsunomiya, A. T. H., Ajmone-Marsan, P., & Garcia, J. F. (2016). GHap: An R package for genome-wide haplotyping. *Bioinformatics*, *32*(18), 2861–2862. https://doi.org/10.1093/bioinformatics/btw356

Valadkhan, S., & Manley, J. L. (2001). Splicing-related catalysis by protein-free snRNAs. In *NATURE* (Vol. 413). www.nature.com

Valera, M., Blesa, F., Santos, R. Dos, & Molina, A. (2006). Genetic study of gestation length in andalusian and arabian mares. *Animal Reproduction Science*, *95*(1–2), 75–96. https://doi.org/10.1016/j.anireprosci.2005.09.008

van Buiten, A., Remmen, J. L. A. M., & Colenbrander, B. (1998). Fertility of shetland pony stallions used in different breeding systems: A retrospective study. *Veterinary Quarterly*, *20*(3), 100–103. https://doi.org/10.1080/01652176.1998.9694849

Van Niekerk, F. E., & Van Niekerk, C. H. (1998). *The effect of dietary protein on reproduction in the mare. VII. Embryonic development, early embryonic death, foetal losses and their relationship with serum progestagen*.

VanRaden, P. M., Olson, K. M., Null, D. J., & Hutchison, J. L. (2011). Harmful recessive effects on fertility detected by absence of homozygous haplotypes. *Journal of Dairy Science*, *94*(12), 6153–6161. https://doi.org/10.3168/jds.2011-4624

Variant Effect Predictor - Command Line VEP. (n.d.). Retrieved February 15, 2023, from https://www.ensembl.org/info/docs/tools/vep/script/VEP_script_documentation.pdf

Vasileiou, G., Vergarajauregui, S., Endele, S., Popp, B., Büttner, C., Ekici, A. B., Gerard, M., Bramswig, N. C., Albrecht, B., Clayton-Smith, J., Morton, J., Tomkins, S., Low, K., Weber, A., Wenzel, M., Altmüller, J., Li, Y., Wollnik, B., Hoganson, G., … Reis, A. (2018). Mutations in the BAF-Complex Subunit DPF2 Are Associated with Coffin-Siris Syndrome. *American Journal of Human Genetics*, *102*(3), 468–479. https://doi.org/10.1016/j.ajhg.2018.01.014

Vasko, A., Drivas, T. G., & Schrier Vergano, S. A. (2021). Genotype-phenotype correlations in 208 individuals with coffin-siris syndrome. *Genes*, *12*(6). https://doi.org/10.3390/genes12060937

Verstreken, P., Ohyama, T., Haueter, C., Habets, R. L. P., Lin, Y. Q., Swan, L. E., Ly, C. V., Venken, K. J. T., De Camilli, P., & Bellen, H. J. (2009). Tweek, an Evolutionarily Conserved Protein, Is Required for Synaptic Vesicle Recycling. *Neuron*, *63*(2), 203–215. https://doi.org/10.1016/j.neuron.2009.06.017

Visscher, P. M., Brown, M. A., McCarthy, M. I., & Yang, J. (2012). Five Years of GWAS Discovery. *The American Journal of Human Genetics*, *90*(1), 7–24. https://doi.org/10.1016/j.ajhg.2011.11.029

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). “Welcome to the tidyverse.” *Journal of Open Source Software*, 4(43), 1686. doi:10.21105/joss.01686.

Wira, C. R., Fahey, J. V., Sentman, C. L., Pioli, P. A., & Shen, L. (2005). Innate and adaptive immunity in female genital tract: cellular responses and interactions. *Immunological Reviews*, *206*(1), 306–335. https://doi.org/10.1111/j.0105-2896.2005.00287.x

Wolc, A., & Olori, V. E. (2009). *Genetics of hatchability-egg quality from the perspective of a chick*. https://www.researchgate.net/publication/228651859

Wong, D. C. F., Wong, K. T. K., Nissom, P. M., Heng, C. K., & Yap, M. G. S. (2006). Targeting early apoptotic genes in batch and fed-batch CHO cell cultures. *Biotechnology and Bioengineering*, *95*(3), 350–361. https://doi.org/10.1002/bit.20871

Woods, G. L., Baker, C. B., Baldwin, J. L., Ball, B. A., Bilinski, J., Cooper, W. L., Ley, W. B., Mank, E. C., & Erb, H. N. (1987). Early pregnancy loss in brood mares. *Journal of Reproduction and Fertility. Supplement*, *35*, 455–459.

Wooliams, J. A., & Oldenbroek, J. K. (2017). Genetic diversity issues in animal populations in the genomic era. In J. K. Oldenbroek (Ed.), *Genomic management of animal genetic diversity*. Wageningen Academic Publishers. https://doi.org/10.3920/978-90-8686-850-6

Xu, D., & Lu, W. (2020). Defensins: A Double-Edged Sword in Host Immunity. *Frontiers in Immunology*, *11*. https://doi.org/10.3389/fimmu.2020.00764

Yamaguchi, Y., Nagase, T., Makita, R., Fukuhara, S., Tomita, T., Tominaga, T., Kurihara, H., & Ouchi, Y. (2002). Identification of Multiple Novel Epididymis-Specific β-Defensin Isoforms in Humans and Mice. *The Journal of Immunology*, *169*(5), 2516–2523. https://doi.org/10.4049/jimmunol.169.5.2516

Yang, M., Lai, T. Y. Y., Tam, P. O. S., Chiang, S. W. Y., Chan, C. K. M., Luk, F. O. J., Ng, T. K., & Pang, C.-P. (2011). CFH 184G as a genetic risk marker for anterior uveitis in Chinese females. *Molecular Vision*, *17*, 2655–2664.

Yarbrough, V. L., Winkle, S., & Herbst-Kralovetz, M. M. (2015). Antimicrobial peptides in the female reproductive tract: a critical component of the mucosal immune barrier with physiological and clinical implications. *Human Reproduction Update*, *21*(3), 353–377. https://doi.org/10.1093/humupd/dmu065

Young, J. M., Friedman, C., Williams, E. M., Ross, J. A., Tonnes-Priddy, L., & Trask, B. J. (2002). Different evolutionary processes shaped the mouse and human olfactory receptor gene families. In *Human Molecular Genetics* (Vol. 11, Issue 5).

Yu, H., Dong, J., Gu, Y., Liu, H., Xin, A., Shi, H., Sun, F., Zhang, Y., Lin, D., & Diao, H. (2013). The Novel Human β-Defensin 114 Regulates Lipopolysaccharide (LPS)-mediated Inflammation and Protects Sperm from Motility Loss. *Journal of Biological Chemistry*, *288*(17), 12270–12282. https://doi.org/10.1074/jbc.M112.411884

Zaballos, Á., Villares, R., Albar, J. P., Martínez-A, C., & Márquez, G. (2004). Identification on Mouse Chromosome 8 of New β-Defensin Genes with Regionally Specific Expression in the Male Reproductive Organ. *Journal of Biological Chemistry*, *279*(13), 12421–12426. https://doi.org/10.1074/jbc.M307697200

Zasloff, M. (2002). Antimicrobial peptides of multicellular organisms. *Nature*, *415*(6870), 389–395. https://doi.org/10.1038/415389a

Zhai, Y. J., Feng, Y., Ma, X., & Ma, F. (2023). Defensins: defenders of human reproductive health. In *Human reproduction update* (Vol. 29, Issue 1, pp. 126–154). NLM (Medline). https://doi.org/10.1093/humupd/dmac032

Zhang, W., Chronis, C., Chen, X., Zhang, H., Spalinskas, R., Pardo, M., Chen, L., Wu, G., Zhu, Z., Yu, Y., Yu, L., Choudhary, J., Nichols, J., Parast, M. M., Greber, B., Sahlén, P., & Plath, K. (2019). The BAF and PRC2 Complex Subunits Dpf2 and Eed Antagonistically Converge on Tbx3 to Control ESC Differentiation. *Cell Stem Cell*, *24*(1), 138-152.e8. https://doi.org/10.1016/j.stem.2018.12.001

Zhang, X., & Firestein, S. (2002). The olfactory receptor gene superfamily of the mouse. *Nature Neuroscience*, *5*(2), 124–133. https://doi.org/10.1038/nn800

8 Appendices

Table 8.1: Contains the significant haploblocks with one or fewer observed homozygotes. The columns in the table consist of BLOCK (haplotype block identification), CHR (Chromosome), Size (the size of the haploblock in kbp), BP1, and BP2 (Start and end position of the haploblock). Further, the ALLELE (Alleles of SNPs observed), FREQ (frequency of haploblock in population), O.HOM (Observed homozygotes), O.HET (Observed heterozygotes), E.HOM (Expected homozygotes), and Padj (Bonferroni corrected P-values).

BLOCK	CHR	Size	BP1	BP2	ALLELE	FREQ	O.HOM	O.HET	E.HOM	Padj
B160101	1	10	160100001	160110000	AGCATGTGGCAC	0.31	0	200	30.5	2.591e-7
B160102	1	10	160101001	160111000	AGCATGTGGCAC	0.31	0	200	30.5	2.591e-7
B160103	1	10	160102001	160112000	AGCATGTGGCAC	0.31	0	200	30.5	2.591e-7
B160104	1	10	160103001	160113000	ATGTGGCAC	0.31	0	200	30.5	2.591e-7
B160105	1	10	160104001	160114000	TGTGGCAC	0.31	0	200	30.5	2.591e-7
B160106	1	10	160105001	160115000	TGGCAC	0.31	0	200	30.5	2.591e-7
B160107	1	10	160106001	160116000	GGCAC	0.31	0	200	30.5	2.591e-7
B160108	1	10	160107001	160117000	GGCAC	0.31	0	200	30.5	2.591e-7
B160109	1	10	160108001	160118000	GGCAC	0.31	0	200	30.5	2.591e-7
B160110	1	10	160109001	160119000	CAC	0.31	0	203	31.4	1.031e-7
B106260	2	10	106259001	106269000	TCAATAA	0.30	0	196	29.3	8.667e-7
B106261	2	10	106260001	106270000	TCAATAATT	0.30	0	196	29.3	8.667e-7
B222	12	300	13260001	13560000	CTAGT	0.29	1	188	27.5	3.949e-5
B223	12	300	13320001	13620000	CTAGTAG	0.29	1	188	27.5	3.949e-5
B673	12	100	13440001	13540000	CTAGT	0.29	1	188	27.5	5.867e-5
B674	12	100	13460001	13560000	CTAGT	0.29	1	188	27.5	5.867e-5
B675	12	100	13480001	13580000	CTAGT	0.29	1	188	27.5	5.867e-5
B2699	12	40	13490001	13530000	CTAG	0.29	1	188	27.5	1.171e-4
B2700	12	40	13495001	13535000	CTAG	0.29	1	188	27.5	1.171e-4
B2701	12	40	13500001	13540000	TAGT	0.29	1	188	27.5	1.171e-4
B676	12	100	13500001	13600000	TAGTA	0.29	1	188	27.5	5.867e-5
B787	12	100	15720001	15820000	ACCGA	0.29	0	187	26.7	4.872e-6
B788	12	100	15740001	15840000	CCGACC	0.27	0	178	24.2	5.958e-5
B789	12	100	15760001	15860000	CCGACC	0.27	0	178	24.2	5.958e-5
B3157	12	40	15780001	15820000	CCGA	0.29	0	187	26.7	9.728e-6
B790	12	100	15780001	15880000	CCGACC	0.27	0	178	24.2	5.958e-5
B3158	12	40	15785001	15825000	CCGAC	0.29	0	187	26.7	9.728e-6
B3159	12	40	15790001	15830000	CGACC	0.27	0	178	24.2	1.190e-4
B3160	12	40	15795001	15835000	CGACC	0.27	0	178	24.2	1.190e-4
B3161	12	40	15800001	15840000	CGACC	0.27	0	178	24.2	1.190e-4
B791	12	100	15800001	15900000	CGACC	0.27	0	178	24.2	5.958e-5
B3162	12	40	15805001	15845000	CGACC	0.27	0	178	24.2	1.190e-4
B15807	12	10	15806001	15816000	CG	0.29	0	188	26.9	9.010e-6
B15808	12	10	15807001	15817000	CGA	0.29	0	188	26.9	9.010e-6
B15809	12	10	15808001	15818000	CGA	0.29	0	188	26.9	9.010e-6
B15810	12	10	15809001	15819000	CGA	0.29	0	188	26.9	9.010e-6
B3163	12	40	15810001	15850000	GACC	0.27	0	179	24.4	9.062e-5
B3164	12	40	15815001	15855000	GACC	0.27	0	179	24.4	9.062e-5

B1462	12	100	29220001	29320000	AAGTA	0.29	0	189	27.2	2.746e-6
B1462	12	100	29220001	29320000	AAGCA	0.30	0	198	29.9	1.931e-7
B5857	12	40	29280001	29320000	GCA	0.44	0	291	64.5	3.408e-22
B29311	12	10	29310001	29320000	GCA	0.44	0	291	64.5	4.201e-22
B29312	12	10	29311001	29321000	GCAA	0.42	0	272	56.4	1.460e-18
B1314	13	100	26260001	26360000	TTCAACG	0.28	0	185	26.1	8.590e-6
B1315	13	100	26280001	26380000	TTCAACG	0.28	0	185	26.1	8.590e-6
B5260	13	40	26295001	26335000	CAACG	0.28	0	186	26.4	1.293e-5
B5261	13	40	26300001	26340000	CAACG	0.28	0	186	26.4	1.293e-5
B1316	13	100	26300001	26400000	CAACG	0.28	0	186	26.4	6.474e-6
B5262	13	40	26305001	26345000	CAACG	0.28	0	186	26.4	1.293e-5
B74588	18	10	74587001	74597000	CCA	0.33	1	211	34.6	1.540e-7
B74589	18	10	74588001	74598000	CCA	0.33	1	211	34.6	1.540e-7
B74590	18	10	74589001	74599000	CCAC	0.29	1	191	28.4	6.200e-5
B1687	27	100	33720001	33820000	GA	0.35	0	231	40.7	3.978e-12
B1688	27	100	33740001	33840000	GA	0.35	0	231	40.7	3.978e-12
B1689	27	100	33760001	33860000	GA	0.35	0	231	40.7	3.978e-12
B6757	27	40	33780001	33820000	GA	0.35	0	231	40.7	7.942e-12
B1690	27	100	33780001	33880000	GA	0.35	0	231	40.7	3.978e-12
B6758	27	40	33785001	33825000	GA	0.35	0	231	40.7	7.942e-12
B6759	27	40	33790001	33830000	GA	0.35	0	231	40.7	7.942e-12
B6760	27	40	33795001	33835000	GA	0.35	0	231	40.7	7.942e-12
B6761	27	40	33800001	33840000	GA	0.35	0	231	40.7	7.942e-12
B1691	27	100	33800001	33900000	GAA	0.31	0	201	30.8	7.755e-8
B6762	27	40	33805001	33845000	GA	0.35	0	231	40.7	7.942e-12
B33807	27	10	33806001	33816000	GA	0.35	0	231	40.7	9.791e-12
B33808	27	10	33807001	33817000	GA	0.35	0	231	40.7	9.791e-12
B33809	27	10	33808001	33818000	GA	0.35	0	231	40.7	9.791e-12
B33810	27	10	33809001	33819000	GA	0.35	0	231	40.7	9.791e-12
B33811	27	10	33810001	33820000	GA	0.35	0	231	40.7	9.791e-12
B6763	27	40	33810001	33850000	GA	0.35	0	231	40.7	7.942e-12
B33812	27	10	33811001	33821000	GA	0.35	0	231	40.7	9.791e-12
B33813	27	10	33812001	33822000	GA	0.35	0	231	40.7	9.791e-12
B33814	27	10	33813001	33823000	GA	0.35	0	231	40.7	9.791e-12
B33815	27	10	33814001	33824000	GA	0.35	0	231	40.7	9.791e-12

Table 8.2: SNPs that were within the haploblocks. Within the table, the columns are CHR (chromosome number), SNP (SNP identification), BP (bp position), A1, and A2 (The minor and major allele, respectively). Further, the GENO column is sorted with the minor homozygote (A1A1), followed by the heterozygote (A1A2), and the major allele homozygote (A2A2). PLINK calculated the P value as the deviation from Hardy-Weinberg expectations ('P') and the expected homozygote animals ('E.HOM').

CHR	SNP	BP	A1	A2	GENO	OHET	EHET	P
1	AX-103967617	160102376	C	A	35/131/161	0.4006	0.4258	0.2989
1	AX-104296832	160102463	A	G	67/155/106	0.4726	0.4929	0.5013
1	AX-103543159	160102487	A	C	46/138/143	0.422	0.456	0.1828
1	AX-104270442	160103444	G	A	7/94/227	0.2866	0.2751	0.5497
1	AX-104774384	160104175	C	T	0/21/307	0.06402	0.06197	1
1	AX-104683849	160104234	A	G	0/30/297	0.09174	0.08753	1
1	AX-104393776	160105928	C	T	35/133/160	0.4055	0.4274	0.3662
1	AX-103289852	160108059	A	G	66/156/105	0.4771	0.4929	0.5753
1	AX-104370876	160108263	A	G	1/29/297	0.08869	0.09031	0.5258
1	AX-104607802	160109287	C	T	0/199/122	0.6199	0.4278	1.116e-20
1	AX-103522164	160109357	C	A	0/22/305	0.06728	0.06502	1
1	AX-103103787	160109928	A	C	1/28/299	0.08537	0.08728	0.5007
2	AX-103711548	106263332	C	T	0/1/327	0.003049	0.003044	1

2	AX-104314691	106265474	T	C	1/18/309	0.05488	0.05912	0.2581
2	AX-104340845	106266479	G	A	12/109/207	0.3323	0.3233	0.7331
2	AX-104003928	106267079	G	A	14/120/194	0.3659	0.3494	0.5264
2	AX-104521486	106267713	C	T	0/1/327	0.003049	0.003044	1
2	AX-103702712	106268127	A	G	15/259/49	0.8019	0.4945	3.3699e-31
2	AX-104191493	106268639	G	A	0/13/314	0.03976	0.03897	1
2	AX-104867635	106269154	C	T	1/18/309	0.05488	0.05912	0.2581
2	AX-104699852	106269311	C	T	18/121/189	0.3689	0.3641	0.8803
12	AX-103850196	13499879	T	C	0/14/313	0.04281	0.0419	1
12	AX-103598998	13504171	T	C	9/241/74	0.7438	0.4799	1.095e-25
12	AX-103811510	13522616	C	A	0/24/303	0.07339	0.0707	1
12	AX-103203710	13525377	G	T	16/242/65	0.7492	0.4885	4.533e-23
12	AX-103230535	13535803	C	T	0/1/327	0.003049	0.003044	1
12	AX-104330539	13592517	G	A	0/21/306	0.06422	0.06216	1
12	AX-104162460	15733153	T	A	1/23/302	0.07055	0.07375	0.3801
12	AX-103905155	15789779	A	C	1/22/302	0.06769	0.07112	0.3564
12	AX-104766198	15809758	T	C	0/22/306	0.06707	0.06482	1
12	AX-104429419	15815391	G	T	21/166/138	0.5108	0.4352	0.002106
12	AX-104580592	15816191	C	A	2/37/289	0.1128	0.1172	0.3611
12	AX-103335321	15823002	T	C	1/38/289	0.1159	0.1145	1
12	AX-103195739	15829875	A	C	0/31/295	0.09509	0.09057	1
12	AX-104160918	29220764	G	A	53/111/160	0.3426	0.4455	3.701e-5
12	AX-104495525	29230538	G	A	29/83/213	0.2554	0.3397	2.653e-5
12	AX-103064397	29317733	A	G	0/15/312	0.04587	0.04482	1
12	AX-103275076	29317857	T	C	26/227/71	0.7006	0.4904	4.736e-15
12	AX-104611438	29319554	C	A	0/75/252	0.2294	0.2031	0.01258
13	AX-103615447	26291055	C	T	8/103/217	0.314	0.297	0.3564
13	AX-103740082	26293353	C	T	8/104/216	0.3171	0.2989	0.3555
13	AX-104657286	26309037	A	C	21/152/155	0.4634	0.4165	0.04717
13	AX-104531770	26330075	G	A	22/178/121	0.5545	0.4524	6.844e-5
13	AX-103544073	26330960	G	A	0/1/327	0.003049	0.003044	1
13	AX-103126844	26331981	T	C	12/124/189	0.3815	0.3517	0.1557
13	AX-103866287	26332675	A	G	9/107/211	0.3272	0.3092	0.3716
18	AX-103980123	74590141	C	T	12/267/47	0.819	0.4942	1.256e-35
18	AX-103333636	74590886	T	C	3/38/287	0.1159	0.1251	0.1669
18	AX-104804861	74596153	G	A	24/68/228	0.2125	0.2968	2.365e-6
18	AX-102958495	74598729	T	C	0/51/276	0.156	0.1438	0.2393
27	AX-104754829	33814845	G	A	22/209/96	0.6391	0.4744	1.783e-10
27	AX-104376096	33815537	G	A	80/115/131	0.3528	0.4878	5.431e-7
27	AX-104958326	33888828	G	A	89/118/113	0.3688	0.4972	3.733e-6

Table 8.3: Significant SNPs within haploblocks related to a known gene, including the expected homozygotes and separated columns for observed genotypes. The first three columns are the chromosome, position, and name of the SNP, and the following two columns are the alleles.

CHR	SNP	BP	A1	A2	GENO	OHET	EHET	P	O.A1A1	O.A1A2	O.A2A2	E.A1A1
2	AX-103711548	106263332	C	T	0/1/327	0.003049	0.003044	1	0	1	327	0,5
2	AX-104314691	106265474	T	C	1/18/309	0.05488	0.05912	0.2581	1	18	309	10,0

2	AX-104340845	106266479	G	A	12/109/207	0.3323	0.3233	0.7331	12	109	207	66,5
2	AX-104003928	106267079	G	A	14/120/194	0.3659	0.3494	0.5264	14	120	194	74,0
2	AX-104521486	106267713	C	T	0/1/327	0.003049	0.003044	1	0	1	327	0,5
2	AX-103702712	106268127	A	G	15/259/49	0.8019	0.4945	3.3699e-31	15	259	49	149,5
2	AX-104191493	106268639	G	A	0/13/314	0.03976	0.03897	1	0	13	314	7,5
2	AX-104867635	106269154	C	T	1/18/309	0.05488	0.05912	0.2581	1	18	309	10,0
2	AX-104699852	106269311	C	T	18/121/189	0.3689	0.3641	0.8803	18	121	189	78,5
12	AX-104766198	15809758	T	C	0/22/306	0.06707	0.06482	1	0	22	306	11,0
12	AX-104429419	15815391	G	T	21/166/138	0.5108	0.4352	0.002106	21	166	138	107,0
12	AX-104580592	15816191	C	A	2/37/289	0.1128	0.1172	0.3611	2	37	289	20,5
12	AX-104160918	29220764	G	A	53/111/160	0.3426	0.4455	3.701e-5	53	111	160	112,5
12	AX-104495525	29230538	G	A	29/83/213	0.2554	0.3397	2.653e-5	29	83	213	73,5
12	AX-103064397	29317733	A	G	0/15/312	0.04587	0.04482	1	0	15	312	8,5
12	AX-103275076	29317857	T	C	26/227/71	0.7006	0.4904	4.736e-15	26	227	71	143,5
12	AX-104611438	29319554	C	A	0/75/252	0.2294	0.2031	0.01258	0	75	252	38,5
18	AX-104804861	74596153	G	A	24/68/228	0.2125	0.2968	2.365e-6	24	68	228	66,0
18	AX-102958495	74598729	T	C	0/51/276	0.156	0.1438	0.2393	0	51	276	26,5
27	AX-104754829	33814845	G	A	22/209/96	0.6391	0.4744	1.783e-10	22	209	96	127,5
27	AX-104376096	33815537	G	A	80/115/131	0.3528	0.4878	5.431e-7	80	115	131	139,5
27	AX-104958326	33888828	G	A	89/118/113	0.3688	0.4972	3.733e-6	89	118	113	156,0

Table 8.4: Genes and loci associated with the SNPs within the significant haploblocks as estimated by Ensembl. The table describes the input SNP chromosome and physical position (bp) before the SNP's strand and allele (major). Further, if known, the gene's symbol is the Ensembl equine genomic gene ID and the possible consequences predicted from the SNP.

Chr	Position	Str	Allele	Symbol	Gene	Consequence
1	160102376	1	A	-	ENSECAG00000038214	downstream_gene
1	160102376	-1	A	-	ENSECAG00000048033	intron,non_coding_transcript
1	160102463	1	G	-	ENSECAG00000038214	downstream_gene
1	160102463	-1	G	-	ENSECAG00000048033	intron,non_coding_transcript
1	160102487	1	C	-	ENSECAG00000038214	downstream_gene
1	160102487	-1	C	-	ENSECAG00000048033	intron,non_coding_transcript
1	160103444	1	A	-	ENSECAG00000031454	upstream_gene
1	160103444	1	A	-	ENSECAG00000038214	downstream_gene
1	160103444	-1	A	-	ENSECAG00000048033	intron,non_coding_transcript, downstream_gene
1	160104175	1	T	-	ENSECAG00000031454	upstream_gene
1	160104175	-1	T	-	ENSECAG00000048033	intron,non_coding_transcript, downstream_gene
1	160104234	1	G	-	ENSECAG00000031454	upstream_gene
1	160104234	-1	G	-	ENSECAG00000048033	intron,non_coding_transcript
1	160104234	-1	G	-	ENSECAG00000048033	downstream_gene
1	160105928	1	T	-	ENSECAG00000031454	upstream_gene
1	160105928	-1	T	-	ENSECAG00000048033	intron,non_coding_transcript, downstream_gene
1	160108059	1	G	-	ENSECAG00000031454	downstream_gene
1	160108059	-1	G	-	ENSECAG00000048033	intron,non_coding_transcript, downstream_gene
1	160108263	1	G	-	ENSECAG00000031454	downstream_gene
1	160108263	-1	G	-	ENSECAG00000048033	intron,non_coding_transcript, downstream_gene
1	160109287	1	T	-	ENSECAG00000031454	downstream_gene
1	160109287	-1	T	-	ENSECAG00000048033	intron,non_coding_transcript
1	160109357	1	A	-	ENSECAG00000031454	downstream_gene
1	160109357	-1	A	-	ENSECAG00000048033	intron,non_coding_transcript
1	160109928	1	C	-	ENSECAG00000031454	downstream_gene
1	160109928	-1	C	-	ENSECAG00000048033	intron,non_coding_transcript
2	106263332	-1	T	KIAA1109	ENSECAG00000017899	intron
2	106265474	-1	C	KIAA1109	ENSECAG00000017899	intron
2	106266479	-1	A	KIAA1109	ENSECAG00000017899	intron
2	106267079	-1	A	KIAA1109	ENSECAG00000017899	intron
2	106267713	-1	T	KIAA1109	ENSECAG00000017899	intron
2	106268127	-1	G	KIAA1109	ENSECAG00000017899	intron
2	106268639	-1	A	KIAA1109	ENSECAG00000017899	intron
2	106269154	-1	T	KIAA1109	ENSECAG00000017899	intron
2	106269311	-1	T	KIAA1109	ENSECAG00000017899	intron
12	13522616	-1	A	-	ENSECAG00000007667	downstream_gene
12	13522616	-1	A	-	ENSECAG00000057635	downstream_gene
12	13525377	-1	T	-	ENSECAG00000007667	non_coding_transcript_exon, downstream_gene
12	13535803	-1	T	-	ENSECAG00000057635	intron,non_coding_transcript, upstream_gene
12	15809758	-1	C	OR5AS1	ENSECAG00000007033	downstream_gene

12	15815391	-1	T	OR5AS1	ENSECAG00000007033	upstream_gene
12	15816191	-1	A	OR5AS1	ENSECAG00000007033	upstream_gene
12	29220764	1	A	DPF2	ENSECAG00000001650	intron
12	29230538	1	A	DPF2	ENSECAG00000001650	intron
12	29317733	1	G	DPF2	ENSECAG00000001650	intron
12	29317857	1	C	DPF2	ENSECAG00000001650	intron
12	29319554	1	A	DPF2	ENSECAG00000001650	intron
13	26291055	-1	T	-	ENSECAG00000010204	intron
13	26291055	-1	T	-	ENSECAG00000048502	intron
13	26293353	-1	T	-	ENSECAG00000010204	intron
13	26293353	-1	T	-	ENSECAG00000048502	intron
13	26309037	-1	C	-	ENSECAG00000010204	intron
13	26309037	1	C	-	ENSECAG00000019564	intron
13	26330075	-1	A	-	ENSECAG00000010204	intron
13	26330075	1	A	-	ENSECAG00000019564	intron
13	26330960	-1	A	-	ENSECAG00000010204	intron
13	26330960	1	A	-	ENSECAG00000019564	intron
13	26331981	-1	C	-	ENSECAG00000010204	intron
13	26331981	1	C	-	ENSECAG00000019564	intron
13	26332675	-1	G	-	ENSECAG00000010204	intron
13	26332675	1	G	-	ENSECAG00000019564	intron
18	74590141	-1	T	-	ENSECAG00000052876	intron,non_coding_transcript
18	74590886	-1	C	-	ENSECAG00000052876	intron,non_coding_transcript
18	74596153	-1	A	U6	ENSECAG00000026419	downstream_gene
18	74596153	-1	A	-	ENSECAG00000052876	intron,non_coding_transcript
18	74598729	-1	C	U6	ENSECAG00000026419	upstream_gene
18	74598729	-1	C	-	ENSECAG00000052876	intron,non_coding_transcript
27	33814845	-1	A	DEFA22	ENSECAG00000027869	downstream_gene
27	33815537	-1	A	DEFA22	ENSECAG00000027869	downstream_gene
27	33888828	1	A	DEFA5L	ENSECAG00000043705	downstream_gene

Table 8.5: Candidate genes related to problems in gonadal or sexual development in equines.

Chromosome	Gene	Name	Position	Approach	Reference
Y	SRY	Sex determining region	Y	Mutations/ Deletions	Pailhoux et al., 1995 Raudsepp et al., 2010 Villagomez et al., 2011
7	HSD17B6	Hydroxysteroid 17-beta dehydrogenase 6	3,935,674-3,938,482	CNV	Gosh et al., 2014
11	SOX9	SRY-Box Transcription Factor 9	9,224,053-9,229,840	Mutations/ Deletions	Kent et al., 1996
12	SF1	Splicing factor 1	28,619,898-28,632,463	Mutations/ Deletions	Kent et al., 1996
12	AR	Androgen receptor	26,039,218-26,041,649	Mutations/ Deletions	Villagomez et al., 2020 Revay et al., 2015
29	UCMA	Upper zone of growth plate and cartilage matrix associated	22,681,823-22,691,596	CNV	Gosh et al., 2014
29	CBRr	Campylobacter bile resistance regulator	32,837,886-32,838,194	CNV	Gosh et al., 2014
29	PHYH	Phytanoyl-CoA 2-Hydroxylase	22,540,934-	CNV	Gosh et al., 2014

			22,563,145		
29	AKR1C	Aldo-keto reductase family 1 member C	29,700,000-29,900,000	CNV	Gosh et al., 2014 Gosh et al., 2020

Table 8.6: Genes related to oocyte development in equines.

Chromosome	Gene	Name	Position	Approach	Reference
3	BMPR1B	Bone morphogenetic protein receptor-1B	44,402,722-44,692,141	CNV	Gosh et al., 2014, Doan et al., 2012
4	ADCY1	Adenylate cyclase 1	16,027,025-16,171,067	ROH	Gurgul et al., 2019
7	PRKACA	Protein kinase cAMP-activated catalytic subunit alpha	46,048,251-46,065,141	ROH	Gurgul et al., 2019
8	ANAPC5	Anaphase promoting complex subunit 5	24,310,740-24,348,719	ROH	Gurgul et al., 2019
8	ANAPC7	Anaphase promoting complex subunit 7	23,907,492-23,927,512	ROH	Gurgul et al., 2019
9	LRRC6	Leucine-rich repeat containing 6	75,402,662-75,588,283	Candidate gene	Gottschalk et al., 2016
15	ATP6V1E2	ATPase H+ transporting V1 subunit E2	53,416,247-53,416,927	Candidate gene	Gottschalk et al., 2016

Table 8.7: Genes related to gametic interaction and embryo development in equines.

Chromosome	Gene	Name	Position	Approach	Reference
1	MFGE8	Milk fat globule EGF and factor V/VIII domain containing	95,221,735-95,253,405	CNV Candidate gene	Sole et al., 2019 Gottschalk et al., 2016
3	FRAS1	Fraser extracellular matrix complex subunit 1	59,404,529-59,818,746	ROH	Metzger et al., 2015
4	ZPBP	Zona pellucida binding protein	19,776,870-19,907,352	CNV ROH	Gosh et al., 2014, Sole et al., 2019 Ablondi et al., 2019
6	LY49B	Killer cell lectin-like receptor	39,335,921-39,347,553	ROH	Todd et al., 2020
8	UBBP4	Ubiquitin B pseudogene 4	24,467,333-24,468,548	CNV	Sole et al., 2019
10	SP-1	Sp1 transcription factor	14,480,982-14,485,022	CNV	Gosh et al., 2014

10	BSP2	Binder of sperm 2	14,481,079-14,506,004	CNV	Gosh et al., 2014
10	SULT2A1	Sulfotransferase family, cytosolic, 2A, dehydroepiandrosterone (DHEA)-preferring, member 1	18,124,115-18,322,483	CNV	Gosh et al., 2014
10	BSPH1	Binder of sperm protein homolog 1	18,375,988-18,377,065	CNV	Gosh et al., 2014
10	ELSPBP1	Epididymal Sperm Binding Protein 1	18,397,898-18,416,427	CNV	Gosh et al., 2014
13	PLOD3	Procollagen-lysine,2-oxoglutarate 5-dioxygenase 3	9,454,913-9,462,562	CNV	Sole et al., 2019
28	KITLG	KIT ligand	15,726,503-15,807,871	ROH Candidate gene	Metzger et al., 2015 Mau et al., 2004

Table 8.8: Genes related to sperm quality traits in equines.

Chromosome	Gene	Name	Position	Approach	Reference
1	HERC4	HECT and RLD domain containing E3 ubiquitin protein ligase 4	56,815,617-56,954,039	Candidate gene	Gottschalk et al., 2016
1	MFGE8	Milk fat globule EGF and factor V/VIII domain containing	95,221,735-95,253,405	CNV Candidate gene	Sole et al., 2019 Gottschalk et al., 2016
4	SPATA48	Spermatogenesis associated 48	19,909,625-19,963,732	ROH	Ablondi et al., 2019
5	MIER1	MIER1 transcriptional regulator	91,061,840-91,119,716	Candidate gene	Gottschalk et al., 2016
5	L1TD1	LINE1 type transposase domain containing	95,020,437-95,025,347	CNV	Gosh et al., 2014
8	IFT81	Intraflagellar transport protein 81 homolog	24,053,812-24,132,668	CNV	Gosh et al., 2014
8	YES1	YES proto-oncogene 1, Src family tyrosine kinase	44,273,501-44,304,857	ROH	Metzger et al., 2015
13	FKBP6	FKBP prolyl isomerase 6	11,350,401-11,378,073	CNV Candidate gene	Sole et al., 2019 Raudsepp et al., 2012, Schrimpf et al., 2015
18	DNAH7	Dynein axonemal heavy chain 7	71,435,145-71,669,919	CNV	Sole et al., 2019
20	ZNF331	Zinc finger protein 331	28,318,795-28,329,094	CNV	Gosh et al., 2014
20	CRISP3	Cysteine-rich secretory protein 2	ECA20: 48,708,574-48,761,076	Candidate gene	Gottschalk et al., 2016, Giese et al., 2002, Hamann et al., 2007
20	CRISP1	Cysteine-rich secretory protein 1	48,856,838-48,887,485	Candidate gene	Gottschalk et al., 2016
22	SPATA25	Spermatogenesis associated 25	35,747,531-35,748,590	ROH	Metzger et al., 2015

24	ADAM20	ADAM metallopeptidase domain 20	16,539,958-16,547,675	CNV	Gosh et al., 2014
25	SOHLH1	Spermatogenesis and oogenesis-specific basic helix-loop-helix 1	38,791,446-38,797,028	CNV	Sole et al., 2019
28	GLIPR1L1	GLIPR1 like 1	4,284,550-4,323,990	Candidate gene	Gottschalk et al., 2016

Table 8.9: Genes related to fertility traits in equines.

Chromosome	Gene	Name	Position	Approach	Reference
2	GJA4	Gap junction protein alpha 4	22,443,340-22,444,341	Candidate gene	El-Sheikh Ali et al., 2020
3	HTRA3	HtrA serine peptidase 3	117,767,023-117,803,782	Candidate gene	Laseca et al., 2022
3	CXCL2	Chemokine (C-X-C motif) ligand 1 (melanoma growth stimulating activity, alpha)	63,470,014-63,586,176	Candidate gene	El-Sheikh Ali et al., 2020
4	INHBA	Inhibin subunit beta A	12,760,757-12,808,658	Candidate gene	Giesecke et al., 2010
4	CFTR	CF transmembrane conductance regulator	74,741,421-74,918,780	Candidate gene	Schrimpf et al., 2016
5	PTGS2	Prostaglandin-endoperoxide synthase 2	20,490,127-20,497,264	Candidate gene	El-Sheikh Ali et al., 2020
5	S100A8	S100 calcium-binding protein A8	40,744,248-40,744,667	Candidate gene	El-Sheikh Ali et al., 2020
5	S100A9	S100 calcium-binding protein A9	40,778,743-40,821,668	Candidate gene	El-Sheikh Ali et al., 2020
5	OVGP1	Oviductal glycoprotein 1	53,508,181-53,522,618	Candidate gene	Schrimpf et al., 2016
5	SPATA1	Spermatogenesis associated 1	76,122,099-76,165,463	Candidate gene	Giesecke et al., 2009
5	PTGER3	Prostaglandin E receptor 3	87,780,622-87,951,028	Candidate gene	El-Sheikh Ali et al., 2020
6	PLCz1	Phospholipase C zeta 1	46,812,109-46,852,694	Candidate gene	Schrimpf et al., 2014
7	MMP1	Matrix metallopeptidase 1	13,098,650-13,176,364	Candidate gene	El-Sheikh Ali et al., 2020
7	SP17	Sperm autoantigenic protein 17	34,254,555-34,264,346	Candidate gene	Giesecke et al., 2011
7	RLN	Relaxin 3 RLN 3	46,105,165-46,106,720	Candidate gene	El-Sheikh Ali et al., 2020
7	RETN	Resistin	5,460,957-5,462,512	Candidate gene	El-Sheikh Ali et al., 2020
7	FSHB	Follicle-stimulating hormone beta subunit	98,422,248-98,424,267	Candidate gene	Giesecke et al., 2011

8	SPIRE1	Spire type actin nucleation factor 1	40,539,360-40,796,466	Candidate gene	Laseca et al., 2022
9	FBXO43	F-box protein 43	45,973,733-45,985,463	Candidate gene	Schrimpf et al., 2016
10	NECTIN-2	Nectin cell adhesion molecule 2	15,518,024-15,671,728	Candidate gene	Laseca et al., 2022
10	APOE	apolipoprotein E	15,713,215-15,715,198	Candidate gene	Laseca et al., 2022
10	KLC3	Kinesin light chain 3	16,083,814-16,100,636	Candidate gene	Laseca et al., 2022
10	ERCC1	ERCC excision repair 1, endonuclease non-catalytic subunit	16,144,062-16,154,444	Candidate gene	Laseca et al., 2022
10	RSPH6A	Radial spoke head 6 homolog A	16,453,159-16,471,388	Candidate gene	Laseca et al., 2022
10	FOXA3	Forkhead box A3	16,507,566-16,513,710	Candidate gene	Laseca et al., 2022
11	ACE	Angiotensin I converting enzyme	15,802,359-15,822,526	Candidate gene	Giesecke et al., 2011
13	FKBP6	FKBP prolyl isomerase 6	11,350,401-11,378,073	Candidate gene	Raudsepp et al., 2012 Schrimpf et al., 2015
13	PAQR4	Progestin and adipoQ receptor family member 4	41,020,286-41,022,630	Candidate gene	Laseca et al., 2022
13	PRSS21	Serine protease 21	41,174,424-41,178,932	Candidate gene	Laseca et al., 2022
13	PDPK1	3-phosphoinositide-dependent protein kinase 1	41,420,452-41,494,281	Candidate gene	Laseca et al., 2022
13	PKD1	Polycystin 1, transient receptor potential channel interacting	41,880,905-41,926,116	Candidate gene	Schrimpf et al., 2016 Laseca et al., 2022
13	MEIOB	Meiosis specific with OB-fold	42,141,232-42,165,354	Candidate gene	Laseca et al., 2022
13	NME3	NME/NM23 nucleoside diphosphate kinase 3	42,228,271-42,229,120	Candidate gene	Laseca et al., 2022
13	IFT140	Intraflagellar transport 140	42,375,191-42,459,360	Candidate gene	Laseca et al., 2022
16	FOXP1	Forkhead box P1	20,353,146-20,717,328	Candidate gene	Schrimpf et al., 2016
20	TCP11	T-complex 11	36,147,583-36,279,044	Candidate gene	Schrimpf et al., 2016
21	PRLR	Prolactin receptor	31,054,801-31,107,331	Candidate gene	Giesecke et al., 2010b
21	TSSK6	Testis-specific serine kinase	4,554,495-4,555,316	Candidate gene	Schrimpf et al., 2016
22	P53	P53 and DNA damage regulated 1	23,560,441-23,566,500	Candidate gene	De Leon et al., 2012

22	P13	Peptidase inhibitor 3	35,155,086-35,157,165	Candidate gene	El-Sheikh Ali et al., 2020
25	NOTCH1	Notch receptor 1	38,056,617-38,104,337	Candidate gene	Schrimpf et al., 2016
28	APOBEC3Z1B	Apolipoprotein B mRNA-editing enzyme-catalytic polypeptide-like 3Z1b	37,062,159-37,065,847	Candidate gene	El-Sheikh Ali et al., 2020

Table 8.10: Overview of known equine recessive disorders, the genes, location, and mutation cause for the disease.

Chromosome	Disorder	Genes	Causative Mutation
1	Hereditary Equine Regional Dermal Asthenia (HERDA)	PPIB	c.115G>A
1	Hydrocephalus	B3GALNT2	c.1423C>T
1	Lavender Foal Syndrome	MYO5A	c.4249del
2	Cerebellar Abiotrophy (CA)	MUTYH	c.541-13539C>T
2	Fragile Foal Syndrome	PLOD1	c.2032G>A
4	Hypoparathyroidism	RAPGEF5	c.2624C>A
4	Myotonia	CLCN1	c.1775A>C
5	Junctional Epidermolysis Bullosa (JEB)	LAMC2	c.1372dup
8	Junctional Epidermolysis Bullosa (JEB)	LAMA3	e.24-27del
9	Severe Combined Immunodeficiency (SCID)	DNAPK	c.9478_9482del
11	Immune-mediated myositis	MYH1	g.53345548T>C
17	Overo Lethal White Syndrome (OLWS)	EDNRB	c.353_354delinsAG
26	Foal immunodeficiency syndrome	SLC5A3	c.1352C>T
26	Glycogen Branching Enzyme Deficiency (GBED)	GBE1	c.102A>C

Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway