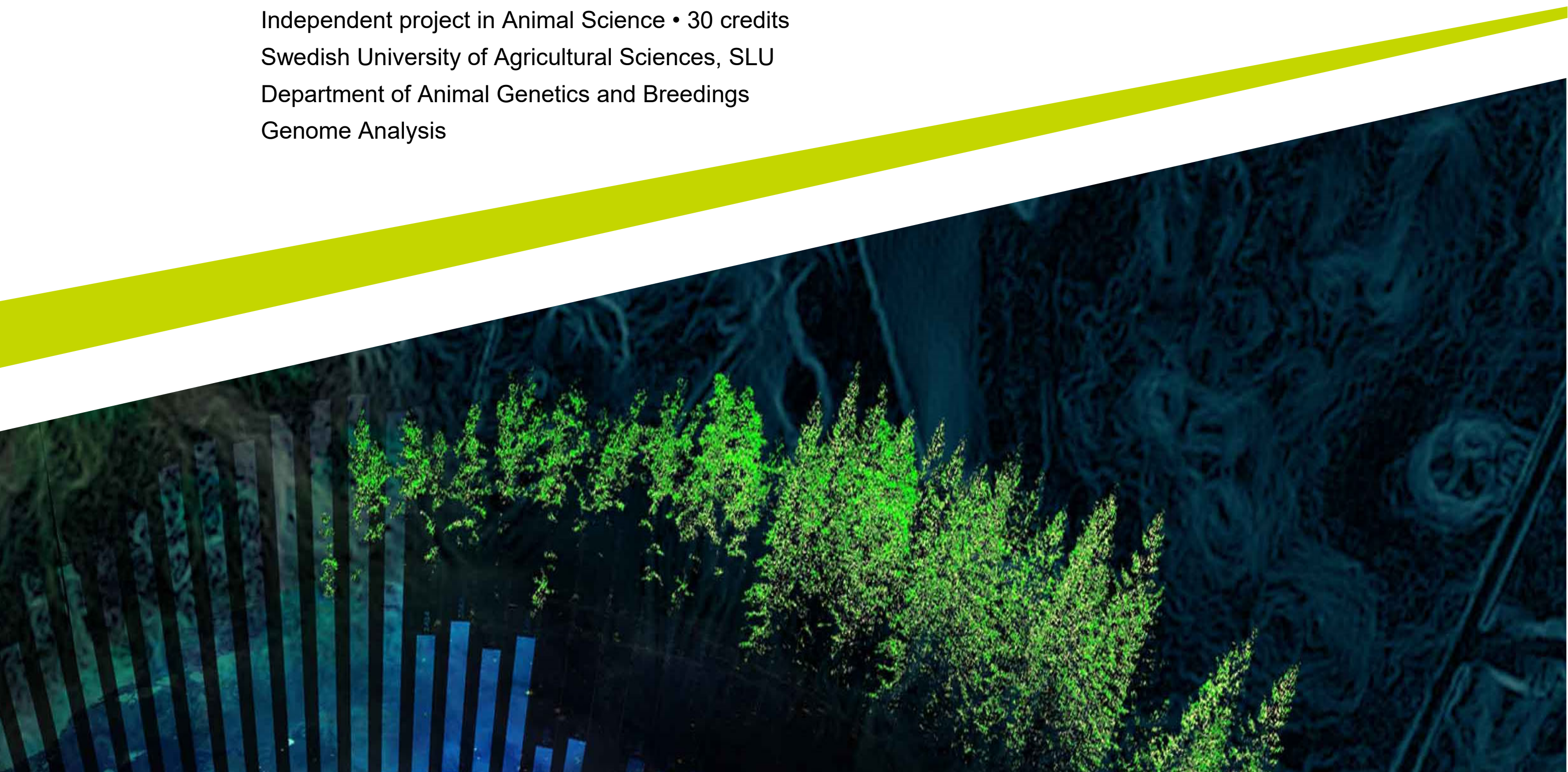
Genomic Analysis of a Hypermobility Syndrome in Gotland Ponies

Francesco Picchi

Independent project in Animal Science • 30 credits
Swedish University of Agricultural Sciences, SLU
Department of Animal Genetics and Breedings
Genome Analysis



Genomic Analysis of a Hypermobility Syndrome in Gotland Ponies

Francesco Picchi

Supervisor: **Sofia Mikko, Swedish University of Agricultural Science, Dept Animal Biosciences**

Assistant supervisor: Maria Lina Elena Longeri, Università degli Studi di Milano, Dipartimento di Medicina Veterinaria e Scienze Animali

Examiner: Martin Johnsson, Swedish University of Agricultural Science, Dept of Animal Bioscien

Credits: 30 ECTS

Level: Second cycle

Course title: Husdjursvetenskap

Course code: EX0870:30035

Programme/education: independent project in Animal Science

Course coordinating dept: Anna Maria Johansson

Place of publication:

Year of publication:

Cover picture:

Copyright: All featured images are used with permission from the copyright owner.

Keywords: Gotland Ponies, Hypermobility, Axiom® Equine Genotyping Array 670K, Genome mapping, Association analysis, Linkage analysis.

Swedish University of Agricultural Sciences

Department Department of Animal Genetics and Breedings

Abstract

The study investigates the prevalence and potential genetic basis of hypermobility disease among a family of Gotland ponies at Lojsta Hed on Gotland Island. It hypothesizes that this condition, inherited in an autosomal recessive manner, may be connected to a single mutation or structural variant. When sire *276* was selected to breed all the mares at Lojsta Hed, unhealthy foals were born, diagnosed with hydrocephalus and hyperflexible joints. There was a significant increase in the percentage of mares not giving birth (52%) compared to previous years with different sires (20%). This difference may be attributed to hypermobility disease, which is associated with abortion and stillbirths, similar to Warmblood Fragile Foal Syndrome (WFFS). The study compares the breeding outcomes of sire *276* at Lojsta Hed with his performance on other farms, revealing a huge contrast in foal birth percentages (52% vs. 16.2%). This suggests a potential inbreeding effect or environmental influence at Lojsta Hed. The samples included foals and their parents to identify carriers of the suspected recessive allele. After extracting the samples' DNA from the blood, they were genotyped using the Axiom® Equine Genotyping Array containing 670,806 markers. To map the genomic loci, both association and linkage analyses were used. Linkage analysis was employed to find the location of disease genes by noticing that genes close together on the same chromosome tend to be inherited together during meiosis. An association analysis was also performed to study the relationship between genotypes and phenotypes in healthy (controls) and affected horses (cases). The analyses identified significant markers on chromosome 9, at positions between 17,8 and 18,5 cM. These markers are notably within the *TRPA1* gene, linked to neurological and skeletal muscle conditions in humans and is hypothesized to be associated with the observed hypermobility in the foals. They showed a correlation with the recessive disease genotypes, suggesting a potential genetic basis for the disease. Further research with larger sample sizes and unrelated Gotland ponies is necessary to identify the causative gene in that region of hypermobility disease

Keywords: Gotland Ponies, Hypermobility, Axiom® Equine Genotyping Array 670K, Genome mapping, Association analysis, Linkage analysis.

Table of contents

List of tables 6

List of figures 7

Abbreviations 9

1. Aim of the study 11

2. Introduction 12

3. Background 13
3.1 Gotland Ponies 13
3.1.1 History 13
3.1.2 Today's herd 13
3.2 The pathology 14
3.2.1 Hypermobility 14
3.2.2 Hydrocephalus 14
3.3 Genome analysis 15
3.3.1 Genomic markers 15
3.3.2 SNP micro-array 15
3.3.3 Equine Genotyping Array 15
3.4 Genomic mapping 16
3.4.1 Linkage Analysis 16
3.4.2 Association studies 17
3.4.3 Phasing 18

4. Material and Methods 19
4.1 Animals 19
4.1.1 Genotyping 20
4.1.2 Genotype calling 20
4.2 Statistical analysis 20
4.2.1 Quality Control Statistical Analysis 20
4.2.2 Phasing 21
4.2.3 Case/control study 22
4.2.4 A case-control study with samples from another breed 22
4.2.5 A case-control study with only the heterozygous markers of the Sire 276 ... 23
4.2.6 Family-based association 23

4.2.7 Linkage Analysis ... 24
4.3 Ensembl Gene Databases ... 25
5. Results ... 26
5.1 Comparison of the born foals rate form the Sire and previous stallions at Lojsta Head ... 26
5.2 Genotyping and Heterozygosity Rates in Axiom Analysis Suite ... 27
5.3 Quality control (QC) ... 29
5.4 Association Study ... 29
5.4.1 Case/control study ... 29
5.4.2 A case-control study with samples from another breed ... 35
5.4.3 A case-control study with only the heterozygous markers of the Size 276 ... 36
5.5 Family-based Associaton ... 38
5.6 Linkage analysis ... 40
5.7 The most sigifcant markers ... 42
6. Discussion ... 43
Conclusions ... 48
References ... 49
Popular science summary ... 53
Appendix ... 54

List of tables

Table 1 The study cohort with the 41 genotyped Gotland horses. All seventeen foals have different mares, but the same sire. Three of the mares had a foal who did not survive, and one mare died before the parturition. 19

Table 2. The recessive model in the linkage analysis to map the genetic loci on relatives for the co-segregation of a chromosomal region and a trait of interest .... **Fel! Bokmärket är inte definierat.**

Table 3..The recessive modlel in the linkage analysis to map the genetic loci on relatives for the co-segregation of a chromosomal region and a trait of interest.. The Sire 276 and the other two stallions chosen the years before at Lojsta Hed. 26

Table 4. Classification of the 670,806 markers from Equine Genotyping Array670k in base of The Probeset Metric Summary. 28

Table 5. List of the best 20 markers obtained from the association-adjusted analysis in the 41 samples...

Table 6. Table of the Family Trio with the father, the three cases foals and their three mares for tdt. ....**Fel! Bokmärket är inte definierat.**

Table 7. The Transmission Disequilibrium Test (TDT) for the Gotland ponies samples. It’s the adjust text file with the best marlers for UNADJ and FC values. 38

Table 8. The parent orgin analysis (tdt poo) file.. 39

Table 9. The Linkage analysis of non-parametric model in chromosome 9 with the lowest LOD score (0-10 and 67-91 cM) and the highest score (11-66 cM) in the centri Morgan position of the chromosome. 40

Table 10. The four markers on chromosome 9 (AX-104091145, AX-103221544, AX-103979373, AX-103993194) have parameters to be significant candidates for the variations causing hypermobility in the foals 46

List of figures

Figure 1.The graph is a scatter plot, with the call rate values on the y-axis and the heterozygosity rate on the x-axis for each sample. 27

Figure 2. The Manhattan plot displays the association study results for each marker in the 41 Gotland ponies samples. 30

Figure 3. Genotype table of the most significant markers in chromosome 9 from the association study of Gotland Ponies samples. The 4 markers A) AX-103979373. B) AX-104091145. C) AX-103221544. D) AX-103993194 have the correct correlation of the genotype for a recessive disease. The Sire: Blue,3 foals cases: Red, 3 Dams:Gold 3 Mares of died foals: Acquamarine, Other samples: White. 32

Figure 4. A) Manhattan Plot about the -log10(p-value) of the association fisher study and their genomic position in chromosome 9 for the 41 samples. B) Table with the most significant markers in case-control study for p-value in chromosome 9. SNP: genetic marker, BP: Base pair position, CHR: Chromsome, UNADJ: unadjusted p-value, GC: Genomi Control p-value, Gene: Gene associated to the marker. 33

Figure 5. Example of one of the markers in chromosome 9 with four homozygous (Affected/Affected) sample (3 cases and dam). The Sire: Blue,3 foals cases: Red, 3 Dams:Gold 3 Mares of died foals: Acquamarine, Other samples: White. 34

Figure 6.A) Manhattan plot of the association fisher study for the 95 samples, where unrelated samples from other horse breeds are included to Goltand ponies samples. B) Table with the most significant markers in the case-control study. The classification is based on an Unadjusted p-value (UNADJ). Every marker has his ID variant with the base pairs position (Bp) in the chromosome (CHR). 35

Figure 7. ProbeSet Summary of the genotype analysis for the markers of 95 samples in chromosome 9. The Sire: Blue,3 foals cases: Red, 3 Dams:Gold 3 Mares of died foals: Acquamarine, Other samples: White. 36

Figure 8. ProbeSet Summary of the genotype analysis for the markers of 95 samples in chromosome 9. The Sire: Blue,3 foals cases: Red, 3 Dams:Gold 3 Mares of died foals: Acquamarine, Other samples: White. 36

Figure 9.A) Manhattan plot of the association fisher study for the 41 Gotland Ponies samples B) Table with the most significant markers in the case-control study considerin only the heterozygous markers of the Size 276. The classification is based on an Unadjusted p-value (UNADJ). Every marker has his ID variant with the base pairs position (Bp) in the chromosome (CHR). 37

Figure 10. ProbeSet Summary of the genotype analysis for the markers of 41 samples in chromosomes 12 and 11. The Sire: Blue,3 foals cases: Red, 3 Dams:Gold 3 Mares of died foals: Acquamarine, Other samples: White. A) Example of one of the markers in chromosome 12 with five homozygous (Affected/Affected) sample (3 cases, one dam and one mare). B) Genotype of the samples for the markers in chromosome 38

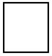
Figure 11. The Linkage analysis of non-parametric model (free-model) for chromosome 9. The x-axes is the LOD score and the y-axes is the Centri-Morgan positon in the chromosome 9. 41

Abbreviations

B3GALNT2	b1,3-N-acetylgalactosaminyl transferase
CSF	Cerebrospinal Fluid
CH	Congenital Hydrocephalus
CNVs	Copy Number Variations
ECA	Equus Caballus chromosome
GWAS	Genome-Wide Association Studies
GC	Genomic Control value
IBD	Identity by Detention Mapping
LD	Linkage Disequilibrium
LOD	Logarithm of Odds
MAF	Minor Allele Frequency
PLOD1	2-oxoglutarate 5-dioxygenase 1
POO	Parent origin analysis
QC	Quality Control
QTL	Quantitative Trait Loci
SNP	Single Nucleotide Polymorphism
TDT	Transmission Disequilibrium Test
TRPA1	Transient receptor potential cation channel subfamily A member 1
UNADJ	Unajusted p-value

1. Aim of the study

This study aims to determine if there is a genetic cause to the hypermobile foals born from the same sire at Lojsta Hed. The hypothesis is that there is a single mutation or structural variant inherited in an autosomal recessive model.



2. Introduction

Gotland Pony foals with hydrocephalus and hyperflexible joints were born at Lojsta Hed (Gotland Island). All mares within the cohort were bred with the same sire. In the same year, there was also a significant increase in the percentage of mares not giving birth. These observations led to the hypothesis that a single mutation or structural variant, inherited in an autosomal recessive model, was responsible. Genome mapping of the samples was conducted to identify the chromosomal location of the disease genes responsible for the foals' hypermobility and the mares' gestational issues.

3. Background

3.1 Gotland Ponies

3.1.1 History

The Gotland Pony is a Swedish horse breed on Gotland Island in the Baltic Sea since early Iron Age. At the beginning of the 1800s, the Gotland Ponies were common in all the large forest area of Gotland, but the number decreased to around 200 samples from the middle of the century. This was an effect due to agricultural intensification as well as ponies being sold to work abroad in mines. In 1992, the population was almost reduced to extinction, with only seven mares and a few young horses left. The Society brought the horses received to Lojsta Hed, a 650-hectare-forest pasture in Gotland Island. To resolve the problem of inbreeding, the breeders imported Welsh pony stallions to Gotland in the middle of '1900. This brought new blood to the breed and resulted in a lower inbreeding level. Its environmental adaptation, together with a feral history of the breed, has created a very hardy breed of healthy horses. (Sofie M Viksten, 2023)

3.1.2 Today's herd

Today there are approximately 5,000 Gotland ponies around the world. At Lojsta Hed in Gotland Island, a herd of 50 mares and around 30-35 foals live in a semi-feral state. All the members of the herd belong to six mares' families and stay outdoors all year round, however, during the winter, they are provided with supplementary forage. In November the foals are separated from the mothers, and they are sold or kept in the herd. Each year a new breeding stallion is chosen to stay with the herd. This is to control the breeding and avoid inbreeding. The stallion is released to the herd at the beginning of the summer and brought home in the autumn. The supervision of the horses is carried out by an association formed by fifteen Gotland families and the Gotland Rural Economy and Agricultural Society. (Gotlandsrusset – Swedish Breeders' Association)

3.2 The pathology

3.2.1 Hypermobility

Horse mobility issues are complex conditions connected to joint hyperlaxity and musculoskeletal pain. These issues involve increased joint movement beyond normal ranges, with symptoms varying in severity. In Warmblood horses, an autosomal monogenetic disease called Warmblood Fragile Foal Syndrome (WFFS) causes joint hyper-elasticity, hypotonia, and mechanical instability of the affected tissues (Ablondi M, 2022). The majority of foals homozygous for WFFS are believed to be lost due to abortion during late gestation. This autosomal recessive disease is fully penetrant, linked to a lethal missense mutation in the procollagen-lysine, 2-oxoglutarate 5-dioxygenase 1 gene (*PLOD1*). A study conducted by (Monthoux C, 2015) concluded that the WFFS homozygous genotype is incompatible with foal viability, leading to immediate euthanasia of affected foals after birth. In humans, mutations of the lysyl hydroxylase 1 gene (*PLOD1*) cause a rare connective tissue disorder known as Ehlers-Danlos Syndrome. This pathology is classified as a hypermobility disorder characterized at birth by severe muscular hypotonia and hypermobile joints. Research by Stevens (Stevens et al., 2013) and Hedberg (Hedberg, Oldfors, and Darin, 2013) demonstrated that dystrophy-dystroglycanopathy with muscle-like hypermobility could be connected to congenital hydrocephalus mutations.

3.2.2 Hydrocephalus

Congenital hydrocephalus (CH) is the abnormal accumulation of cerebrospinal fluid in the ventricles of the brain. The result is an increase of the volume in the cranial due to an inadequate passage of cerebrospinal fluid (CSF) from its point of production to the systemic circulation (Kousi and Katsanis, 2016). CH is also associated with dystocia and can bring fatal complications for the mare at parturition. Sometimes foals are stillborn and it's necessary to be fototomized at the parturition. Hydrocephalus can be hereditary or acquired due to infection or trauma, however, no scientific literature has ever reported acquired hydrocephalus in horses. (Schmidt and Ondreka, 2019). This disorder is inherited in an autosomal recessive way and associated with a nonsense mutation in the gene b1,3-N-acetylgalactosaminyl transferase *(B3GALNT2).* Congenital hydrocephalus is declared in breeds like Friesian horses, Warmblood horses, American miniature horses, American Quarter horses, Thoroughbred, Standardbred, and Belgian draft horses (Hayashi, Leifer and Cohen, 2000). In Angus cattle has been reported Arthrogryposis multiplex, called "Curly Calf Syndrome" (Agerholm *et al.*, 2015), where there's the presence of hydrocephalus-associated pathologies such as fixed joints in all limbs and the affected calves are stillborn.

3.3 Genome analysis

3.3.1 Genomic markers

Genome markers are DNA sequences important to identify an inherited disease with the responsible gene. They are genetic variations with known physical locations on chromosomes that can be present in different forms, ranging from Single Nucleotide Polymorphisms (SNPs) to structural variants such as Copy number variations (CNVs). Single-nucleotide polymorphisms (SNPs) are single base position variations in the DNA that follow the principle of Mendelian inheritance. A variation is classified as SNP if the minor allele frequency is >1%. They have an important role in genomic mapping to identify genes or Quantitative Trait Loci (QTL) for a specific trait (Zhang *et al.*, 2012).

3.3.2 SNP micro-array

Detection of SNPs is based on oligonucleotide hybridization analysis on microarrays. The method is a collection of probes bound to a solid surface of glass, plastic, or silicon (chip) carrying many different oligonucleotides (SNPs) in a high-density array. The principle of the method is the hybridization between two DNA strands. One single strand labelled with a fluorescence marker from the sample pipetted onto the surface of the chip and one single strand from the chip (Brown, 2018). There will be hybridization if the two strands are complementary to each other. The hybridization is detected by fluorescence, and through a laser scan is possible to obtain the intensity ratio expressed in a graphic. The position of the emission of the fluorescence signal onto the array (grid) indicates which SNPs have been hybridized (Laframboise, 2009). After scanning, high-quality SNP calls provide information to identify a single or structural mutation. Bioinformatics has made it possible to sequence all genomes and to use that sequence data to identify variations in the genome. This method has high-quality calls, with a very high call rate (capacity to call an SNP when it's present) and accuracy (SNPs identified are called correctly).

3.3.3 Equine Genotyping Array

Single Nucleotide Polymorphism (SNPs) has an important role in mapping diseases and traits in equine populations. One of the main applications of SNP-chips is mapping a trait of interest by genotyping a large cohort of individuals. (Schaefer and McCue, 2020). After the sequence of the whole equine genome of Twilight horse in 2006, the first generation of high-throughput genotyping arrays (SNP-chips) has been developed with 54,602 SNPs (EquineSNP50 Genotyping BeadChip, Illumina Ltd, San Diego, California, USA) from seven horses representing seven different breeds. The second one with 65,157 SNPs (Axiom

Equine Genotyping Array, Thermo Fisher Scientific Inc, Waltham, Massachusetts, USA) where there has been an increase of the number of SNPs and breeds used to achieve information about the SNPs (Finno and Bannasch, 2014). Affymetrix-based Axiom extended the coverage and number of available short-read-based whole genome sequences (Schaefer *et al.*, 2017). The MNEc670K Axiom array, developed by Affymetrix, created 670,805 markers from the whole genome sequencing of 154 horses representing 24 different breeds. Affymetrix has chosen criteria to filter markers with an algorithm concerning the quality control (QC) of SNP, breed representation, linkage disequilibrium, and genome coverage. Thanks to a larger SNP set, there have been studies and analyses about more complex studies, for example, the evaluation of an HMGA2 variant for pleiotropic effects on height and metabolic traits in ponies (Norton et al., 2019). The equine genotyping array has an important role in the characterization and comparison of populations of individuals in single or different breeds.

3.4 Genomic mapping

Genomic mapping is based on the principle of inheritance of Mendel to identify the chromosomal position of the DNA region where is located the molecular cause of a phenotype variation. (Brown, Terence A. 2002). Linkage and Association analysis are the two main approaches to map genetic loci.

3.4.1 Linkage Analysis

Linkage analysis is a powerful tool to detect disease genes to a specific chromosome. It’s necessary to know the pedigree of samples that have been genotyped with known genetic markers and with the wild type and the mutant risk-associated form. Linkage analysis is a method to test the markers for their allelic cosegregation with the casual allele of a trait gene within the pedigrees. A panel of selected genetic markers is analysed to observe the segregation of alleles during meiosis, to study the co-inheritance of marker alleles, and to estimate the chromosomal location of a trait locus (Elston, 2017). The recombination of genetic material via crossover events, results in a unique set of chromosomes for each offspring. Linkage analysis is based on observation of the relatives from a family-based pedigree structure to map genetic loci. The closer two loci are to each other on the same chromosome, the higher the probability that they will remain linked and not undergo recombination. They’ll be genetically linked and so transmitted together during meiosis from parents to offspring. When two loci are close enough on the same homologous chromosome, the probability of recombination (recombination factor) of alleles in two loci during meiosis is close to zero. The most interesting parameter is the recombination factor (theta) for each pair of loci

in affected and unaffected individuals. Theta is the estimation of the ratio between the number of crossover events and the number of meiotic events in two loci. The value depends on the distance of the two loci, with a range from 0.0 (linked) and 0.5 (unlinked). For the measure of linkage statistical support of theta, there is the logarithm of the odds (LOD). The value explains the correlation between the probability of the variant locus being in a specific position and the probability that it is in an unlinked part of the genome (Barrett and Teare, 2011). A value of 3 or more corresponds to a significance level of 0,0001 and it's considered the minimum value for a significant result.

3.4.2 Association studies

An association study is formed by case and control samples where cases are animals with mutated phenotypes and controls are wild type. The association mapping aims to identify the markers with a different frequency for a particular trait in cases compared to the controls. The mapping relies on Linkage Disequilibrium (LD) with some combination of variants at close loci will stay together in the genome more often than by chance cause of common history (Slatkin, 2008). In a case-control study, there could be a problem connected to stratification, where a stratified population contains several subgroups that are genetically separated (Price *et al.*, 2010) In this situation the association between a trait and a DNA region could be given by the underlying genetic structure of the population and not by a causative gene in that region. The consequence is connected to the creation of a false-positive signal of association. The main difference is that a Linkage analysis looks for chromosomal regions shared by family members who are affected by the disease in a particular study. An association mapping tries to find alleles in the same chromosome with Linkage Disequilibrium (LD) to a casual variant in a case-control study (Hodge, 1993). Different categories of association studies can include population-based cohort and family-based studies. GWAS is a population-based cohort study for the identification of the relationship between a phenotype and one or more DNA variants, that are evenly spaced and distributed throughout the genome (Tam *et al.*, 2019). Previous studies in GWAS have shown that casual variants influence specific traits, and the population chosen for a study comes from a randomly population-based cohort (Jiang *et al.*, 2019). The aim is to describe the association of genotypes with phenotypes by testing the differences in the allele frequency of genetic variants (Uffelmann *et al.*, 2021). Genetic and phenotypic observations used in GWAS are often derived from a population-based cohort where individuals are assumed to be a random draw from the population. The most common markers for a GWAS study are single nucleotide polymorphisms (SNPs), selected in genomic risk loci by statistically significant association (Oti, Ballouz and Wouters, 2011). The family-based method aims to identify mutations responsible for an inherited disease, with success for mapping Mendelian traits (Oti,

Ballouz and Wouters, 2011). Typical family designs include parent-offspring trios. These studies segregate the trait in a family data, where the controls are the relatives and offsprings with wild type and the cases are relatives with mutated phenotype (Pulst, 1999) Thanks to the Transmission Disequilibrium Test (TDT) is possible to measure the association in the presence of genetic linkage between a genetic marker and a trait/pathology. It's a research for the identification of alleles shared by family members, who are affected by an inherited disease (Bang and Davidian, 2010). The main function is to do a comparison of the frequency of an allele that is transmitted from a heterozygous parent to an affected offsprings (Ott, Kamatani and Lathrop, 2011). Studies based on familial cases have more power than unrelated samples to detect genetic effects given an equivalent number of sampling units. A family-based design can by-pass the possible stratification problem but is more difficult to collect an elevated sample size of the family respect to associated-based design.

3.4.3 Phasing

Phasing has a relevant impact with the increasing of the number of genetic markers to identify alleles that are located on the same chromosome. In response to enormous amount of genotype data, computationally methods were developed to handle whole-chromosome data efficiently (Hofmeister, 2023). The aim concerns to distinguish the two inherited copies of each chromosome in the haplotype. Haplotype phase has different applications, including imputation of genetic variation and calling genotypes in sequence data. Computational phasing can be estimated in both related and unrelated individuals, with different choices of parameters and algorithms (Browning, 2011). In a family data with pedigree information, for determining the haplotype phase is needed a method that provides information about Mendelian inheritance . A parent-offspring pair has to share one allele at every marker that is identical by descendent. Computational phasing in unrelated individuals is based on the estimation sets of common haplotypes that can explain the genotyped data. Haplotype frequency information across families is more informative for phase estimation than the haplotype frequency information obtained from unrelated individuals.

4. Material and Methods

4.1 Animals

Blood samples were previously collected, and DNA was extracted for genotyping from 41 Gotland Ponies (17 foals, 21 mares, and 3 sires). Three out of the seventeen foals were diagnosed with hyperflexible joints, but recovered and survived. Three of the mares gave birth to foals with hydrocephalus and ataxia syndromes, who died shortly after birth, why no samples were collected. The study cohort comprised the 41 horses, including the seventeen surviving foals, their dams and sire (Table 1). Samples from two other stallions were collected as control samples. A complete list of the included horses with their ID, is shown in Appendix 1.

Table 1. The study cohort with the 41 genotyped Gotland Ponies horses. All seventeen foals have different mares, but the same sire 276.

Study cohort	*Number of samples*
Normal foals	14
Hyperflexible foals	3
Mares	21
Sire of the 17 foals	1
Control stallions	2
Total number of samples	41

The Swedish horse registry site "Blå basen" provides pedigree information on Gotland ponies and provided the number of mares covered and foals born among Gotland pony sires. The study focused on stallion 276 during the year he bred mares at Lojsta Hed, as well as the preceding 7 years at other farms. For comparison, two other Gotland pony sires, Sire C and Sire Q, were selected as controls to evaluate foaling rates (Table 1). They bred mares at Lojsta Hed for two and three years, respectively, before the study period of sire 276.

4.1.1 Genotyping

Genotyping was performed using the Axiom® Equine Genotyping Array containing 670,806 markers according to the standard protocol Axiom™ 2.0 Assay Manual Workflow User Guide (P/N 702990, ThermoFisher Scientific, Life Technologies, Carlsbad, CA 92008 USA). Cases were defined by the Gotland ponies showing signs of hyper-flexible joints (Appendix 1). Foals and mares without any clinical symptoms were used as controls. In total there were three cases (8,57%) and 38 controls (91,43%). The Sire 276 he was born in a normal farm, so he wasn't directly related to the mares at Lojsta Head.

4.1.2 Genotype calling

File samples were imported to the Axiom Analysis Suite software 4.0 together with the EquCab3 cluster file (Axiom_MNEc670P.r4). In the Axiom Analysis Suite, the "Genotyping" workflow was used to perform genotyping, with the default parameters settings. Axiom Analysis Suite created an analysis folder with three main viewer windows. From the analysis folder was possible to obtain information about the characteristics of the samples and markers like SNP Call Rate (CR), minor allele frequency (MAF), and the Hardy-Weinberg Equilibrium p-value.

The information about SNPs and samples was exported in Variant Call Format (VCF), a standardized format to store SNPs and other DNA polymorphisms. VCF files were generated for each horse using sequencing data from the Axiom Analysis Suite according to the best practice guidelines. Each horse was genotyped for 670K markers and at the moment to export the data in VCF format, all the options for "Annotation Columns" and "Probeset data Columns" were included.

4.2 Statistical analysis

4.2.1 Quality Control Statistical Analysis

PLINK, software was used for the quality control of genotyped data. The VCF files were converted to the binary file format (BED/BIM/FAM) with the option –make-bed, to have the data in the correct format for analysis. It was necessary to change the chromosome set to --horse and --allowed extra chromosome set to recognise different chromosome codes. In the conversion to bfile, the command --allow-extra-chr 0 was used for markers that cannot be assigned to a specific chromosome, due to uncertainty about the marker's chromosomal location. Assigning the chromosome '0' was possible to include as many markers as possible. The command "--double-id" was used to modify sample IDs from VCF format to PLINK IDs.

To filter out SNPs from the dataset, there was an analysis to check the quality. Two approaches with different MAF thresholds of 0.01 and 0.05 were run to check for significant changes in the number of markers passing the threshold (Appendix 2). Failing markers were removed by calculating the minor allele frequency (maf < 0.01). All variants with a minor allele frequency below this threshold were filtered out. When a marker showed MAF <0.01, it was assumed to be a mistake and removed from downstream analysis. Variants deviating from Hardy-Weinberg equilibrium (--hwe) with a p-value under the threshold of 1e-5 were filtered out. Variants and samples with missing data exceeding the provided values (--geno 0.02, --mind 0.02) were screened and removed. The --check-sex option generated an X chromosome-based sex validity report. This command compared the input dataset (fam file) sex with the X chromosome inbreeding coefficient, inferring the sex of the sample by analyzing the mean homozygosity rate across the X chromosome markers. The check-sex compared the sex information provided from ped file (ped-sex) and the sex observed in sex chromosomes (snp-sex). The Heterozygosity Rate (--het) option computed observed and expected autosomal homozygous genotype counts for each sample. The F coefficient in the generated het file estimated the coefficients for observed and expected autosomal homozygosity.

4.2.2 Phasing

A haplotype is the combination of alleles on a single chromosome, and in the case of two heterozygous loci, there are two possible haplotypes. Genotypes obtained from a SNP array are typically unphased. Software for statistical methods has to be used for inferring the sequence of alleles on each inherited chromosome. Beagle is a free download software package for phasing genotypes. The software was run using Java at the command prompt (java -jar beagle.01Mar24.d36.jar) and the argument to input was specified with the command “gt”. Beagle was based on an iterative algorithm (HMM) and accepted the files in Variant Call Format, containing genotypes for the study samples. For each sample in Variant Call Format was specified the pedigree information and genetic relationships. The method was run with default parameters and after the phasing analysis, a single-phased VCF output file was created. VCF files were phased using Beagle 5.4, without performing imputation on genotypes (Brian Browning, 2022). Phased VCF files were collected into a single data frame and queried for unique haplotypes. The phasing parameters are concerned with the burning, iterations, and phase states. After phasing with the Beagle software, the VCF files were converted to PLINK format (bed/bim/fam). During the export of genotyping data from Axiom Analysis Suite to VCF format, the phenotype data concerning the sex of the samples was missing. Before the phasing analysis, these parameters were manually added to the fam data file to include this information during the analysis process.

4.2.3 Case/control study

Genetic association tests were conducted for each genetic variant, using an appropriate model, specifically Fisher's exact test. The Fisher test was specific for small sample sizes to evaluate the non-random association between two markers. The study included three cases and 38 controls to test for association between the markers and disease status. In this study, it's already known that the samples were not a random draw from the population because they were related. To test for association between SNPs and disease status, the allelic Fisher's exact test command (--assoc fisher) generated the p-values.

The association analysis file from PLINK for the three cases and 38 controls was generated using a Fisher exact test (--assoc fisher) and produced a text file with a header line for each marker and nine fields. In this text file, the most important values were the Odds Ratio (OR), allelic test p-value (P), allelic test chi-square statistic (CHISQ), (F_A) allele frequency among cases, and (F_U) allele frequency among controls.

A Bonferroni correction was applied using the --adjust option to simplify the interpretation of the association file and identify the most significant markers. The file classified all the markers according to the best value for the Unadjusted p-value (UNADJ) and genomic control corrected p-value (GC).

Several basic multiple-testing corrections for the raw p-value were made with the adjust option. The data were routinely visualized using Manhattan plots, generated with R software using the 'qqman' packages. The Manhattan plot was formed with the -log10(p-value) on the y-axis and the 32 Equine chromosomes on the x-axis.

On the Ensembl website, the makers with the most significant results were used to determine their positions and characteristics in the horse reference genome (EquCab3.0). From their base-pair position, it was discovered their associated gene and function.

4.2.4 A case-control study with samples from another breed

The test includes other 54 horses from another breed on the two Axiom plates. In total there were 41 Gotland ponies samples with other 54 samples from different breeds. The 54 samples didn't show cases of hyperflexibility joints. They were considered negative controls since none of them should be homozygous for a candidate variant associated with hyper-flexibility in the three cases. In total, there were 92 controls and three cases, following the same protocol (qc threshold) as the first association fisher study.

4.2.5 A case-control study with only the heterozygous markers of the Sire 276

The case-control study focused solely on markers where Size 276 was heterozygous, identifying potential candidates for a recessive mutation disease. A total of 41 Gotland pony samples were included in the analysis. After selecting heterozygous markers using the Axiom Analysis Suite, the data were processed through Plink software following the same protocol as the initial association study.

4.2.6 Family-based association

In the family-based association analysis, three types of tests were conducted: Transmission Disequilibrium Test (TDT), TDT Parental Origin of Offspring (tdt_poo), and DFAM (family-based association test with unrelated individuals).

For the Transmission Disequilibrium Test, it was necessary to have samples of the same family. The tdt analysis and his implementation with parent origin analysis used trios family data, typically involving the two parents and one or more affected offspring. A family trios with the sire (276), the three foals with hypermobility and the three dams were created. There was a second analysis of the Transmission Disequilibrium Test with the sire (276), the 17 mares and 17 foals, for a total of 35 samples. The 4 mares without foals and the other two stallions used as controls in case-control association analysis, were removed because they were not a trios family. Conducting two analyses with different sample sizes was important to check if different results were obtained and to increase the power of the analysis. The Transmission Disequilibrium Test (TDT) was based on the rate of allele transmission from heterozygous parents to affected offspring. It's family allele shared, explaining the results with the odds ratio, p-value, and chi-square values. In PLINK, there was an implementation of the TDT test called parent origin analysis, where heterozygous mothers and heterozygous fathers were considered separately. The parent origin analysis produced a text file where it was considered separately the transmission from the heterozygous mother and heterozygous father. The analysis was used to locate genes associated with diseases. It included an extended version of the transmission disequilibrium test, with a cluster of unrelated individuals (dfam). In this way, it was possible to increase the overall sample size and improve the detection of the genetic variant. This test includes the 41 Gotland ponies samples with all the other horses from other breeds on the two Axiom plates, for a total of 95 samples. They need as negative controls since none of them should be homozygous for a candidate variant associated with hyperpflexibility in the three cases. In the dfam, there has been another analysis with only 41 Gotland Ponies to see the differences in the two methods.

For the family-based association test, multiple comparisons were made using the --adjust option to apply multiple testing corrections to the p-values obtained from

association tests. The test produced a file with the best SNPs associated with the trait, minimizing the risk of reporting false positives in GWAS results. For better visualization of the TDT results, a Manhattan plot was created in R using the 'qqman' package, while the 'ggplot2' package was used for the DFAM graphics. All graphs were composed with chromosome number on the x-axis and -log10(p-value) on the y-axis.

4.2.7 Linkage Analysis

Merlin software was used to map the genetic loci on family trios of Gotland Ponies for the co-segregation of a chromosomal region and a trait of interest. Merlin supported the PED and MAP formats from PLINK but required a file (DAT file) containing information about the structure of the pedigree file. The DAT file was necessary to have information about the affection status and each marker ID during the analysis. The centiMorgan position (cM) provided in the map file was obtained from the reference genome map of the horse. The Centri Morgan position has given the possibility to determine how closely two markers were associated and inherited together. In total the analyses were on the Sire 276, the 17 foals and their 17 mares, for a total of 35 samples. The linkage analysis was performed for each chromosome separately, and with the option --chr Nchromosome in PLINK software, it was possible to calculate each chromosome step by step. The PEDSTATS option in Merlin (pedstats -d file.dat -p file.ped) provided the ability to check the data and obtain a summary description of any pair of pedigree and data files. The summary description from the pedstats has given information about the integrity of the markers and the family structure. Once the input data file (-d parameter), pedigree file (-p parameter), and map file (-m parameter) were correctly constructed, the analysis was run. The non-parametric model analysis was for the investigation of linkage without specification of a disease model. In the non-parametric model (free model), both the Whittemore and Halpern NPL pairs (--pairs) and NPL all (--npl) statistics were calculated. The –npl used information from all individuals in the pedigree, while the –pairs option considered pairs of relatives and calculated shared of alleles identical by descent (IBD) between these pair.

In the second statistical method, a specific disease model was used to describe the segregation of the trait locus. The trait model parameter (--model) was the autosomal recessive model, with a low disease allel frequency of 0.0001 (Table 2). In the parametric model it was already assumed the mode of inheritance of the disease.. The parametric analysis also included option --step 3 at three equally spaced locations between each consecutive pair of markers. The genetic markers were assumed to be mapped accurately. The order and distances between each locus were specified in the '.map' file and the evidence for linkage was computed at many locations within the range of the genetic markers.

Table 2. The recessive model in the linkage analysis to map the genetic loci on relatives for the co-segregation of a chromosomal region and a trait of interest.

Affection	Disease Allele Frequency	Penetrances	Model Name
Disease	0.0001	0.0001,0.0001,1.000	Recessive_Model

4.3 Ensembl Gene Databases

The most significant markers obtained from the statistical analysis were checked if they were associated to a particular gene. A gene was considered interesting if it was associated to pathways or diseases similar to hyperflexibility joints. On the Ensembl Gene Database, it was possible to achieve information about the gene associated to a base position of the markers in the genome. It was necessary to select the horse breed (EquCab3.0) because the genomic variant and gene structures were species-specific. Once discovered the gene, it was used a gene database (Gene Cards) which provided details about its function, and related disorders of the gene.

5. Results

5.1 Comparison of the born foals rate form the Sire and previous stallions at Lojsta Head

The comparison between the Sire (276) used in this study and the other two stallions chosen the years before to breed the mares at Lojsta Hed revealed significant differences in the rates of mares who didn't give birth or had stillborn foals (Table 3). The Sire (276) bred 50 mares at Lojsta Hed, resulting in 23 born foals out of 48, without considering the two mares died during the winter. Among the 23 foals, six of them were diagnosed with hypermobility (three of whom died within nine days), and 17 foals were normal. Seven years before being selected for Lojsta Hed, this sire bred 38 Gotland mares outside of Lojsta Hed. From these breedings, there were six mares were not pregnant, four had stillborn foals, and 28 had normal foals.

The Sire C, chosen to breed the mares at Lojsta Hed for three years before this study, bred 147 mares. According to pedigree information, there were 107 normal foals and 38 non-pregnant mares. In one of the years, when it was selected for Lojsta Hed, there were two cases of stillborn foals and one foal that died within nine days. The Sire Q, serving two years at Lojsta, achieved a percentage of 80% of normal-born foals and only 17 cases of non-pregnant mares among more than 96 mares covered (Table 3).

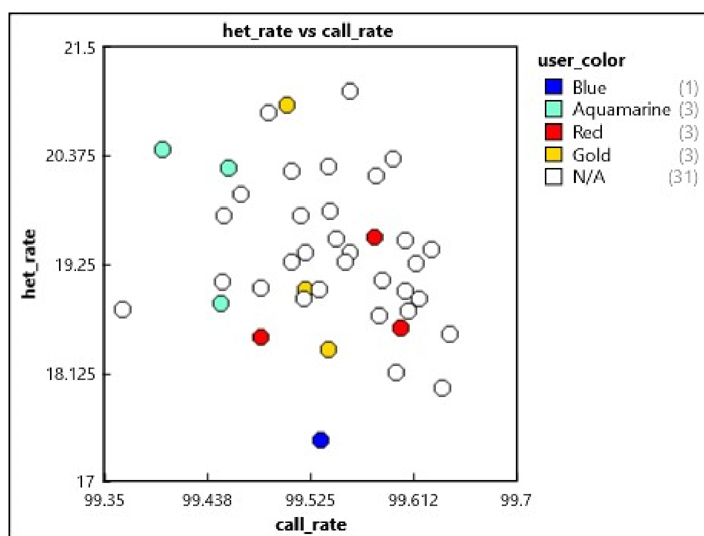
Table 3. The Sire 276 and the other two stallions chosen the years before at Lojsta Hed. Comparison of the born foals rate and the ratio of the mares who didn't give birth. 'No' is the number of samples, while the Ratio is given by the number of samples with a specific trait divided by the total number.

	Sire 276(not at Lojsta)		Sire 276 (at Lojsta)		Sire C (at Lojsta)		Sire Q (at Lojsta)	
	No	Ratio	No	Ratio	No	Ratio	No	Ratio
Mares covered	38		50		147		97	
Mares who died	1	0.026	2	0.040	2	0.013	1	0.01
Mares who gave birth to a foal	31	0.837	23	0.479	107	0.737	79	0.823

Nonpregnant mares	6	0.162	25	0.521	38	0.262	17	0.177
Resorbed	1	0.027	0	0	0	0	0	0
Normal foals born	28	0.757	17	0.354	104	0.707	78	0.804
Nonviable foals born	3	0.081	3	0.062	3	0.02	1	0.01
Foals with hyperflexibility syndrome	0	0	3	0.062	0	0	0	0

5.2 Genotyping and Heterozygosity Rates in Axiom Analysis Suite

In the Axiom Analysis Suite, all 41 samples passed the call rate threshold. The call rate ranged between 99.38% and 99.6% (Figure 1), and the threshold for a non-passing rate was when the average call rate was below 98.5%. The heterozygosity rate had the lowest value for the Sire (276) at 17.4, and the highest value for one of the mares at 21.2 (Figure 1). The blue circles represent the stallion (276) with the lowest percentage of het rate (17.4%). The three foals with hypermobility are shown in red and their three dams are labelled gold. The three mares who gave birth to the three foals who died within nine days are labeled aquamarine. The other samples are white.



Samples	Color
The Sire	Blue
3 foals cases	Red
3 Dams	Gold
3 Mares of died foals	Acquamarine
Other samples	White

Figure 1.The graph is a scatter plot, with the call rate values on the y-axis and the heterozygosity rate on the x-axis for each sample.

In the summary window and sample table there was the classification of the markers in the 6 SNP categories (Table 4).

The Axiom Analysis Guide (Axiom™ Genotyping Solution Data Analysis User Guide) suggested that the last three categories (Other SNPs, Call Rate Below Threshold, and OTV) were not recommended for the analysis. The exclusion of these three categories was essential for maintaining data quality and avoiding noise and ambiguity. Using these parameters, from the initial 670,806 markers, 642,300 SNPs were exported to VCF format, with a call rate (CR) value of at least 97.1% and one in "BestandRecommended".

Table 4. Classification of the 670,806 markers from Equine Genotyping Array670k in base of The Probeset Metric Summary.

%	SNP categories	Description
39.43	MonoHighResolution	One well-formed genotype cluster; must be homozygous.
35.075	PolyHighResolution	Well-separated genotype clusters and two alleles in the genotype calls.
21.243	NoMinorHom	Well-separated genotype clusters; one cluster is homozygous and one is heterozygous.
3.039	Other	Some issues
0.851	CallRateBelowThreshold	Low call rate
0.356	Off-Target Variant	Sequences significantly different from the sequences of the hybridization probes.

5.3 Quality control (QC)

During the phasing, there were 83 windows of 629,466 markers with a genotype rate of 1. The quality control parameters included filtering markers by minor allele frequency of 0.01, which reduced the markers to 409,030 SNPs, excluding 220,436 SNPs with a MAF below 1%. Hardy-Weinberg equilibrium tests based on a p-value of 0.00001 removed 1,680 markers. No individuals were removed due to missing data and 21,221 genetic markers were removed due to missing data . The --checksex option was an X chromosome-based sex validity report that compared the input dataset (fam file) and the sex and the X chromosome inbreeding coefficient. This command inferred the sex of the samples by examining the mean homozygosity rate across the sex in the fam file and the X-chromosome markers. The check-sex file exposed a mismatch in the sex information provided from ped file (ped-sex) and the sex observed in sex chromosomes (snp-sex) for six of the 41 samples. The pedsex It was represented by a "problem" status in one column, while the other 35 samples had an "OK" status. The pedsex for the samples with problem status was indicated as "0", suggesting the sex could not be verified due to low quality. Verification of the sex of samples in the Gotland ponies register “Blå basen” confirmed correspondence between the sex indicated in the Axiom Analysis Suite and that in the Swedish register.

The analysis for expected and observed homozygosity showed that in the het file, the expected number of homozygous (E(HOM)) was 257,100 markers for all the samples. The observed number of homozygous (O(HOM)) ranged between 243,374 and 265,894 markers, with the highest value for the stallion (276).

5.4 Association Study

5.4.1 Case/control study

After verifying the quality control of the samples and markers in Plink software, all 41 samples and 387,808 markers passed the threshold. The Manhattan plot (Figure 2) illustrated the association study, with chromosomes on the x-axis and -log10(p-value) on the y-axis. Each dot represented a SNP, indicating the strength of its association and its genomic position. A higher value on the -log10(p-value) suggested a potential candidate marker for an association.

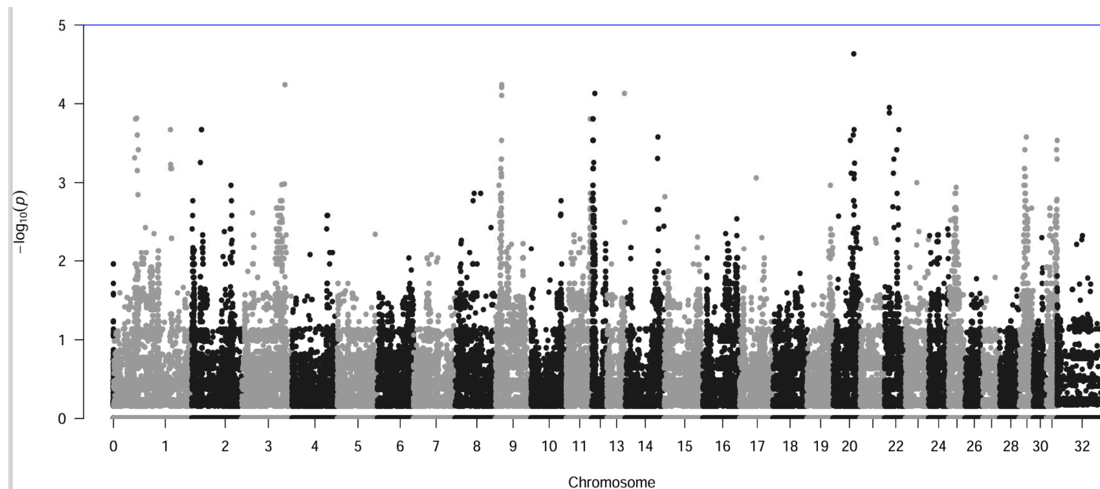


Figure 2. The Manhattan plot displays the association study results for each marker in the 41 Gotland ponies samples.

The most significant markers for the association study obtained from the association study were in the text file (Table 5). The threshold considered for both the Unadjusted p-value (UNADJ) and genomic control corrected p-value (GC) was 0.04557, resulting in 6,427 markers. The genomic control p-value was to correct the cryptic relatedness of the samples. For the Bonferroni correction, the adjusted p-value for all markers was 1. They presented the twenty best markers with UNADJ and GC values ranging between 2.665e-05 and 2.483e-04. For each of the most significant markers has been reported his position and the gene associated (Table 5).

There was a correlation between the suggestive significant SNPs in the Manhattan plot (Figure 2) and the results obtained from the association Fisher's adjusted table in PLINK (Table 5).

Table 5. List of the most significant 20 markers obtained from the association-adjusted analysis in the 41 Gotland ponies samples. Each marker (SNP) describes his base pair position in the chromosome(CHR) and the Gene Associated. The markers colored in red are the markers with the right correspondence of the genotype for a recessive disease model.

CHR	Base pair	SNP	Gene Associated
3	103,887,803	AX-103674726	Intergenic variant
9	14,476,700	AX-104091145	(*TRPA1)* transient receptor potential cation channel subfamily A member 1
9	14,476,934	AX-103221544	(*TRPA1)* transient receptor potential cation channel subfamily A member 1
12	6967973	AX-103941291	Intergenic variant

13	41660494	AX-103891793	(*ABCA17*) ATP Binding Cassette Subfamily A Member 17
9	14,483,709	AX-103979373	*(TRPA1*) transient receptor potential cation channel subfamily A member 1
22	11734766	AX-104425263	Intergenic variant
22	11418640	AX-104557013	Non coding transcript exon variant
1	55295136	AX-103438746	(*CTNNA3)* Catenin Alpha 3
1	52325361	AX-104938061	91 way GERP elements
11	57605600	AX-104721791	Intergenic variant
12	2867278	AX-103987038	*(PRR5L*) Proline Rich 5 Like
12	2915667	AX-103317894	(*PRR5L)* Proline Rich 5 Like
1	137036346	AX-104630292	ENSECAG00000059532
20	50359726	AX-103547076	Intergenic variant
22	34711094	AX-104882021	(*ADA*) adenosine deaminase
1	56,839,961	AX-104032666	(*HERC4)* HECT and RLD domain containing E3 ubiquitin protein ligase 4

In the Probeset Summary of Axiom Analysis Suite there was graphic representation of the genotype of the markers for each sample. The graphic representation of the genotype was the base to check the correspondence between the alleles of the three cases (213, 214, and 219) and the controls, focusing particularly on the heterozygous alleles of the stallion and the three dams. In an autosomal recessive disease, it was expected to obtain homozygous alleles only for the three foals with hypermobility and heterozygous alleles at least for the sire and the three dams. In the family-based analysis, also the three mares of the dead foals were included as "carriers" since they were hypothesized to be heterozygotes (Normal/Affected).

The three markers in chromosome 9 (AX-104091145, AX-103221544, AX-103979373) were the only markers from Table 5 of the association study where there was a correlation between the genotype of the samples and a recessive disease model (Figure 3 A, B, C). The analysis showed that only the three hypermobile foals were homozygous (Affected/Affected), while all three dams and the sire were heterozygous (Affected/Normal). The three mares of the death foals weren't heterozygous (Normal/Normal), so there was no correspondence. These three markers were associated with the same transient receptor potential cation channel subfamily A member 1 gene (*TRPA1*).

Also for marker AX-103993194 (Figure 3.D), the same genotyping situation was observed as in the other three markers before. The research for the position in the reference genome (Equcab3) found the marker AX-103993194 in 14276816 bp a sequence intergenic variant for the *TRPA1* gene. This marker showed the same genotype of the other three markers, but still, not all the mares of the dead foals were heterozygous.

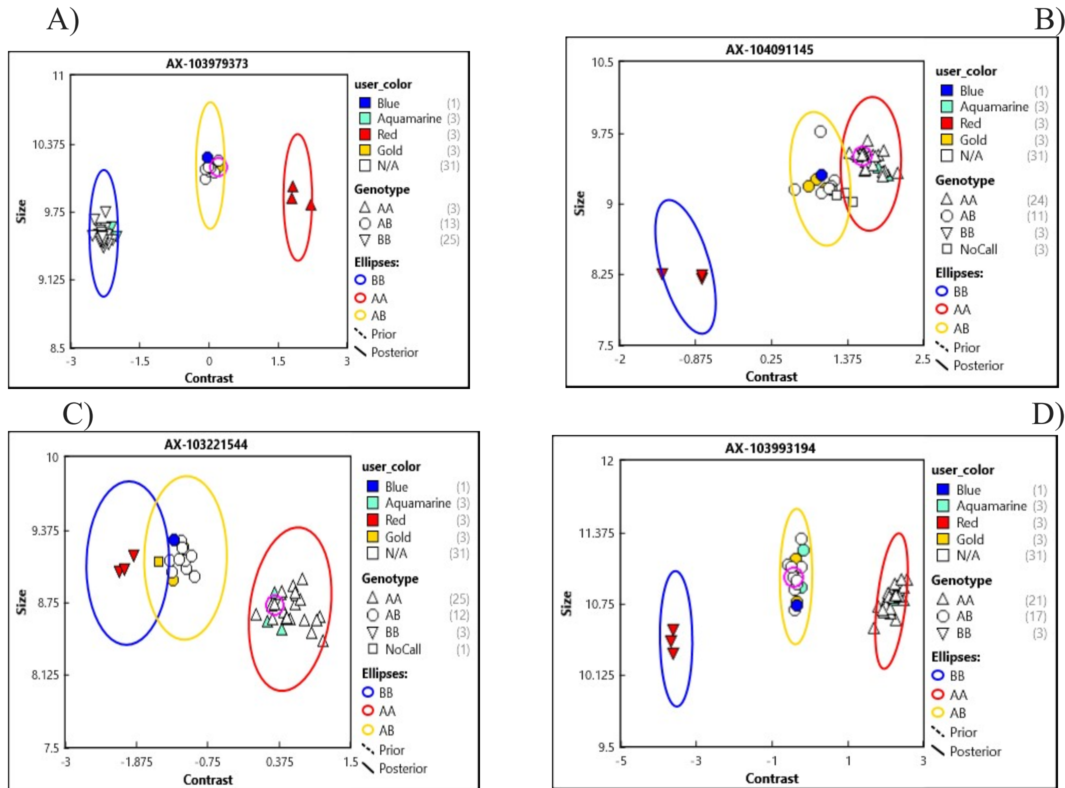


Figure 3. Genotype table of the most significant markers in chromosome 9 from the association study of Gotland Ponies samples. The 4 markers A) AX-103979373. B) AX-104091145. C) AX-103221544. D) AX-103993194 have the correct correlation of the genotype for a recessive disease. The Sire: Blue,3 foals cases: Red, 3 Dams:Gold 3 Mares of died foals: Acquamarine, Other samples: White.

PLINK was used to filter the chromosomes for the association study and to create the Manhattan plot specific for chromosome 9 (Figure 4. A) and it has been obtained from 11869 markers. The most significant markers for chromosome 9 are in Figure 4. B, where it was described their base pair positions (BP) with the gene associated. The significance of the markers for the association study was given by unadjusted p-value (UNADJ) and the genomic control p-value (GC)

SNP	BP	UNADJ	GC	Gene
AX-104091145	14476700	5.661e-05	5.661e-05	TRPA1
AX-103221544	14476934	6.178e-05	6.178e-05	TRPA1
AX-103979373	14483709	7.748e-05	7.748e-05	TRPA1
AX-103993194	14276816	0.0002883	0.0002883	TRPA1
AX-104111830	13970755	0.0005058	0.0005058	KCNB2
AX-104898282	13983014	0.0005058	0.0005058	KCNB2
AX-104732399	12273228	0.0006575	0.0006575	KCNB2
AX-103858430	12273934	0.0006575	0.0006575	KCNB2
AX-103975719	13784276	0.0008454	0.0008454	TERF1
AX-104750625	13791288	0.0008454	0.0008454	TERF1

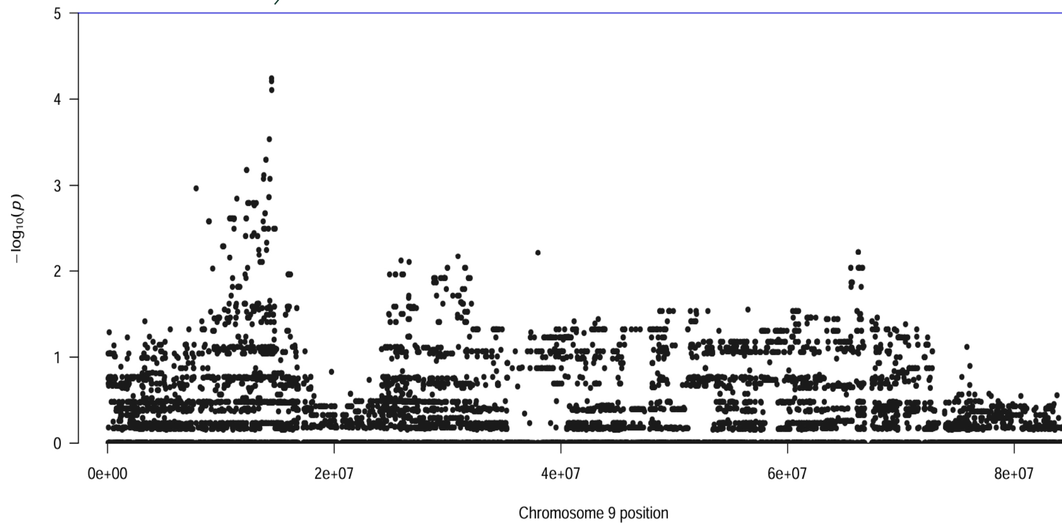


Figure 4. A) Manhattan Plot about the -log10(p-value) of the association fisher study and their genomic position in chromosome 9 for the 41 samples. B) Table with the most significant markers in case-control study for p-value in chromosome 9. SNP: genetic marker, BP: Base pair position, CHR: Chromsome, UNADJ: unadjusted p-value, GC: Genomi Control p-value, Gene: Gene associated to the marker.

The subsequent markers for the most significant p-value on chromosome 9 (AX-104111830, AX-104898282, AX-104732399, AX-103858430, AX-103975719, AX-104750625) did not show the correct correspondence of the genotype for an autosome recessive disease. In these markers, in addition to the homozygous genotype in the three cases, the mare sample (217) was as well homozygous (Affected/Affected) for all these markers (Figure 5). Also in these markers the three mares of the death mares were not heterozygous.

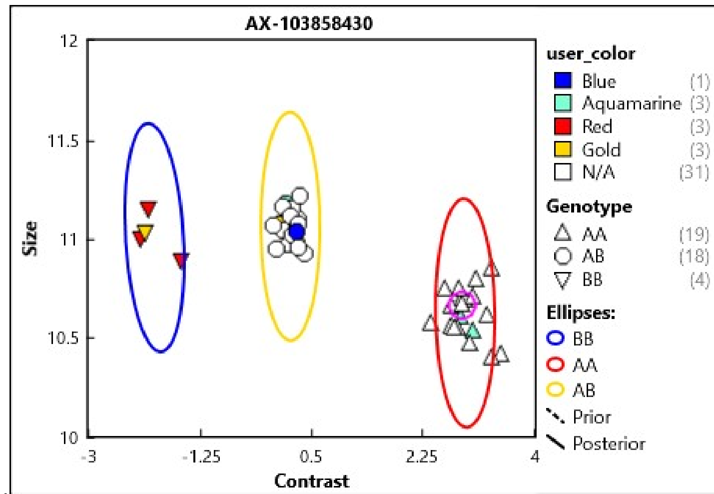


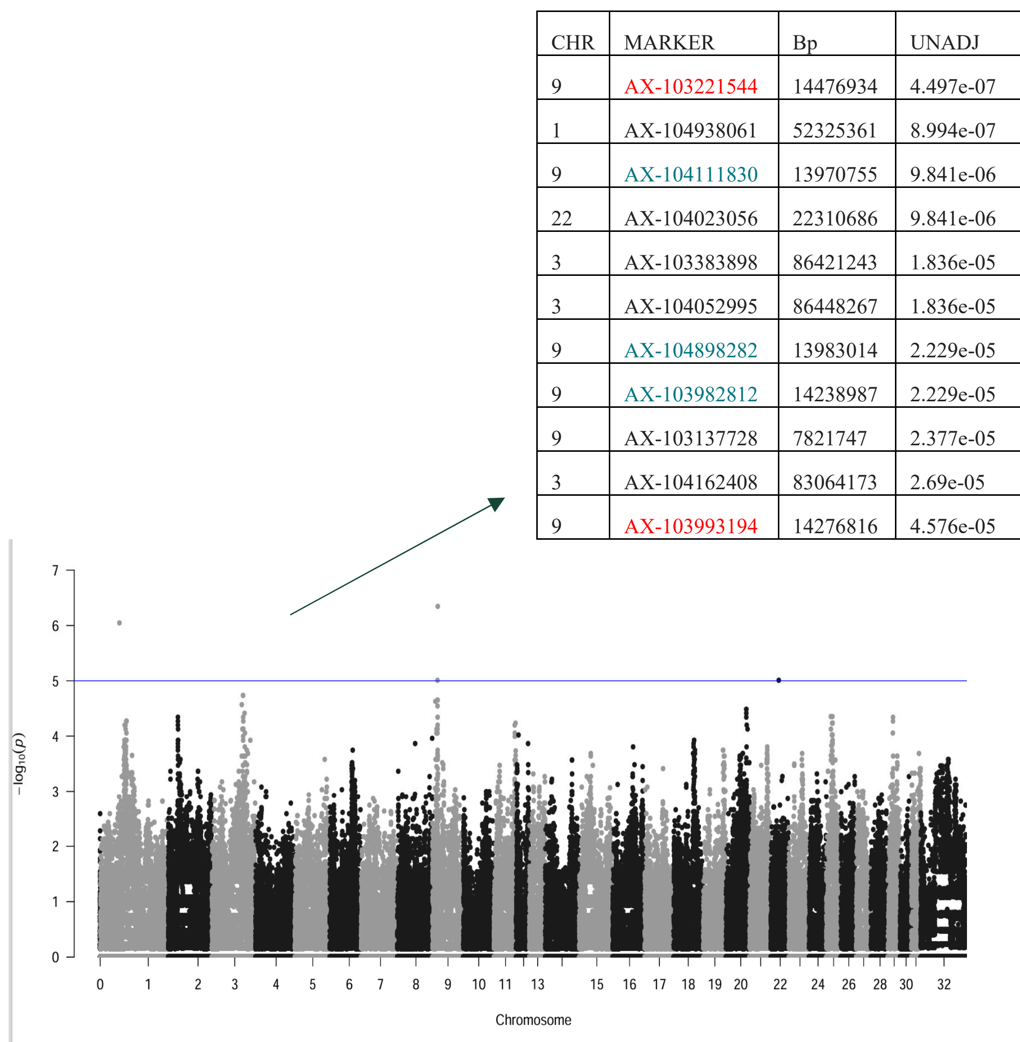
Figure 5. Example of one of the markers in chromosome 9 with four homozygous (Affected/Affected) sample (3 cases and dam). The Sire: Blue,3 foals cases: Red, 3 Dams:Gold 3 Mares of died foals: Acquamarine, Other samples: White.

In AX-104111830 and AX-104898282, both markers are associated with a variation in the gene for potassium voltage-gated channel subfamily B member 2 (*KCNB2)*. The other two markers (AX-104732399 and AX-103858430) were an in intergenic variant for *KCNB2*. The last two markers (AX-103975719 and AX-104750625) were parts of the gene telomeric repeat binding factor 1 (*TERF*1).

5.4.2 A case-control study with samples from another breed

The association analysis file from PLINK for the three cases and 92 controls was based on the 586521 genetic variants that passed filters and QC.

The Manhattan plot (Figure 6.A) showed that the best markers for the p-value (UNADJ) in a case-control study including samples from different breeds were still for chromosome 9 (Figure 6.B).



CHR	MARKER	Bp	UNADJ
9	AX-103221544	14476934	4.497e-07
1	AX-104938061	52325361	8.994e-07
9	AX-104111830	13970755	9.841e-06
22	AX-104023056	22310686	9.841e-06
3	AX-103383898	86421243	1.836e-05
3	AX-104052995	86448267	1.836e-05
9	AX-104898282	13983014	2.229e-05
9	AX-103982812	14238987	2.229e-05
9	AX-103137728	7821747	2.377e-05
3	AX-104162408	83064173	2.69e-05
9	AX-103993194	14276816	4.576e-05

Figure 6.A) Manhattan plot of the association fisher study for the 95 samples, where unrelated samples from other horse breeds are included to Goltand ponies samples. B) Table with the most significant markers in the case-control study. The classification is based on an Unadjusted p-value (UNADJ). Every marker has his ID variant with the base pairs position (Bp) in the chromosome (CHR).

The Genotype study in chromosome 9, demonstrated that in two markers (AX-103993194, AX-103221544) (Figure 7. A. B) there was still a correlation for the genotype of the cases and control in an autosome recessive disease.

In the marker AX-104091145 (Figure 7. C) there was a particular situation: the three foals with hypermobility were separated in the graphic from the other

samples, with low size and contrast parameters. The foals for this marker were considered homozygous (B/B) with the other eleven controls, but from the graph, it seemed that they were part of a different combination of alleles. This could an effect of the low quality of the cluster for the marker.

A) B) C

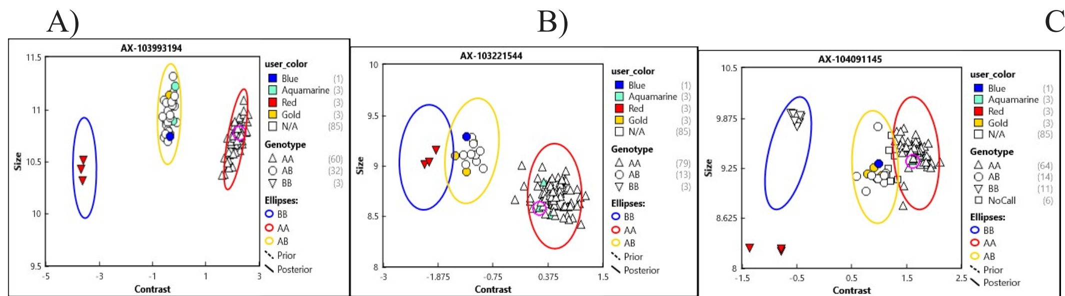


Figure 7. ProbeSet Summary of the genotype analysis for the markers of 95 samples in chromosome 9. The Sire: Blue,3 foals cases: Red, 3 Dams:Gold 3 Mares of died foals: Acquamarine, Other samples: White.

The AX-104111830, AX-103982812 and AX-104898282 in chromosome 9 from the Table 6.B with significant p-values had the genotypes with the 3 cases and still one of their mare (217) was homozygous. The sire with the other mares heterozygous for these markers. (Figure 8). A) AX-104111830. B) AX-103982812. C) AX-104898282 are the markers in chromosome 9 with four homozygous (Affected/Affected) samples (3 cases and mare). The sire and the 3 dams of the cases are heterozygous (Affected/Normal), while the three mares of the dead foals are homozygous (Normal/Normal).

A) B) C)

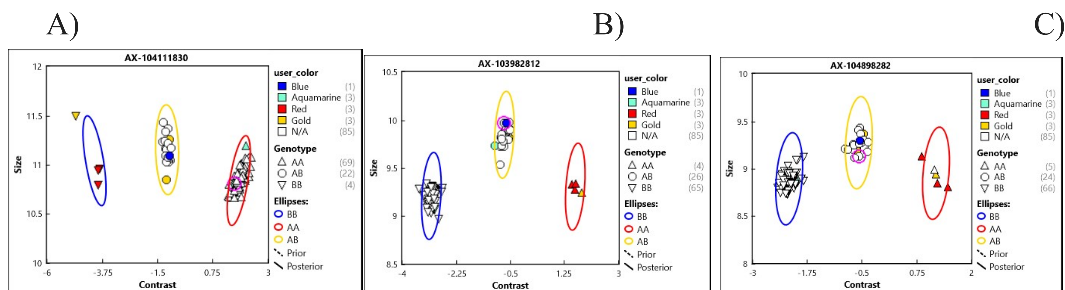


Figure 8. ProbeSet Summary of the genotype analysis for the markers of 95 samples in chromosome 9. The Sire: Blue,3 foals cases: Red, 3 Dams:Gold 3 Mares of died foals: Acquamarine, Other samples: White.

5.4.3 A case-control study with only the heterozygous markers of the Sire 276.

This case-control study analyzed 102,135 heterozygous markers of Sire 276 from 41 Gotland pony samples. A Manhattan plot (Fig9A) derived from the Fisher association analysis revealed the most significant markers on chromosomes 9 and 12 (Fig9.B). Notably, the markers on chromosome 9 (AX-103979373, AX-

103993194) were consistent with those identified in a previous study, where a correlation was established between the genotype and a recessive disease model.

Chr	ID	Bp	UNADJ
9	AX-103979373	14483709	7.748e-05
11	AX-104721791	57605600	0.000155
12	AX-103987038	2867278	0.000155
12	AX-103317894	2915667	0.000155
9	AX-103993194	14276816	0.0002883
12	AX-103724117	2846411	0.0002883
12	AX-103822649	2853885	0.0002883
12	AX-103934732	2895732	0.0002883
12	AX-103510216	2918666	0.0002883
12	AX-102964704	2918972	0.0002883
12	AX-104852864	2920693	0.0002883
12	AX-104606962	2958870	0.0002883

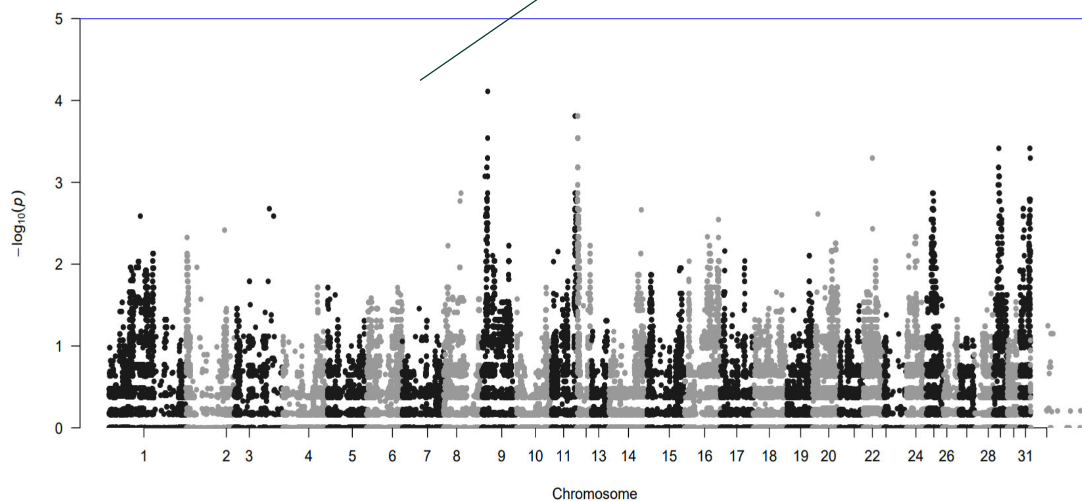


*Figure 9.*A) Manhattan plot of the association fisher study for the 41 Gotland Ponies samples B) Table with the most significant markers in the case-control study considering only the heterozygous markers of the Sire 276. The classification is based on an Unadjusted p-value (UNADJ). Every marker has his ID variant with the base pairs position (Bp) in the chromosome (CHR).

All the most significant markers on chromosome 12 exhibited the same genotype. However, a perfect correlation between the 41 Gotland pony samples and the recessive disease model was not observed (Fig 10.A). Of the samples, three affected foals were homozygous (Affected/Affected), as were Dam 217 (mother of one affected foal) and Mare 244 (mother of a normal foal), bringing the total to five homozygous samples. The Sire 276 and the other two dams were heterozygous (Affected/Normal) for all these markers. The markers on chromosome 12 were located close to base pairs 2,846,411 and 2,958,870 bp, all associated with the same

gene, proline-rich 5-like 8 (PRR5L). This gene is involved in the positive regulation of mRNA catabolic processes and the regulation of fibroblast migration. No known correlations were found between this gene and hyperflexibility disorders. Additionally, the study identified one significant marker on chromosome 11 (Fig10.B). Five samples were homozygous (Affected/Affected) for this marker, including the three affected foals and one normal foal and its mare. The Sire 276 and the three dams were heterozygous (Affected/Normal) for this marker. The marker is located at position 57,605,600 bp, but no small structural variants for this region were identified on the Ensembl.

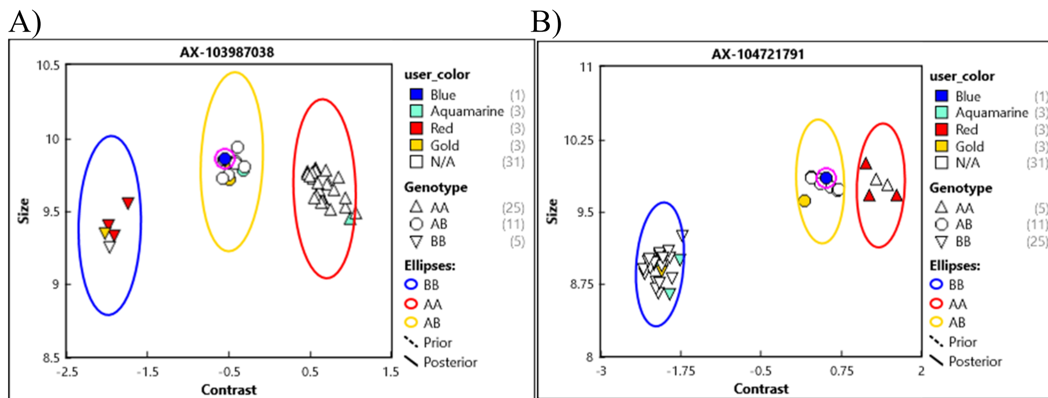


Figure 10. ProbeSet Summary of the genotype analysis for the markers of 41 samples in chromosomes 12 and 11. The Sire: Blue,3 foals cases: Red, 3 Dams:Gold 3 Mares of died foals: Acquamarine, Other samples: White. A) Example of one of the markers in chromosome 12 with five homozygous (Affected/Affected) samples (3 cases, one dam and one mare). B) Genotype of the samples for the markers in chromosome 11.

5.5 Family-based Associaton

The TDT result for the family trio of the three cases produced the same unadjusted p-value (UNADJ) of 0.5247 and genomic control (GC) p-value of.5247. for most of the chromosomes, without giving significant results. The Transmission disequilibrium test where it was included to the family trio also the other 14 normal foals and their mares (35 samples in total) indicated the best results in chromosomes 1, 9, 11, 20, and 31, each with at least 10 markers having an unadjusted p-value (UNADJ) of 0.01431 and a genomic control (GC) value of 0.09811 (Table 6).

The family-based association analysis (dfam) with all the other horses from another breed on the two Axiom plates produced the same results with the most interesting markers in 1, 9, 11, 20 and 31 chromosomes.

Table 6. The Transmission Disequilibrium Test (TDT) for the 35 Gotland ponies samples. It's the adjusted text file with the best UNADJ and FC values markers. CHR: chromosome, POSITION: Bae-pair position, SNP: Variant ID, UNADJ: unadjusted p-value, GC: genomic control p-value, BONF: Bonferroni. GENE: Gene associated, VARIANT: type of variant. The markers colored in red are the same obtained in the association study for chromosome 9.

CHR	POSITION	SNP	UNADJ	GC	BONF	GENE	VARIANT
1	66524460	AX-104592651	0.01431	0.09811	1		
9	14276816	AX-103993194	0.01431	0.09811	1	TRPA1	Intron
9	14476700	AX-104091145	0.01431	0.09811	1	TRPA1	Intron
9	14483709	AX-103979373	0.01431	0.09811	1	TRPA1	Intron
11	55598993	AX-103221338	0.01431	0.09811	1		
11	55599575	AX-104532729	0.01431	0.09811	1		
11	55616080	AX-104227942	0.01431	0.09811	1		
20	49233734	AX-104399156	0.01431	0.09811	1		
20	49242686	AX-104860745	0.01431	0.09811	1		
20	49258322	AX-103424859	0.01431	0.09811	1		
31	8279705	AX-104611904	0.01431	0.09811	1		
31	8281379	AX-104337208	0.01431	0.09811	1		
31	8298639	AX-103941760	0.01431	0.09811	1		

For chromosome 9, the markers with the best p-value and genotype from the case/control study analysis were also significant in the TDT results. However, in chromosomes 1, 11, 20, and 31, there wasn't a perfect match between the genotypes of the cases and the controls. The most significant markers were in chromosomes 1, 9, 11, 30, and 31 both for the TDT test and also for the Dfam test with the other unrelated horses of other breeds

There were only four markers (Table 7) where both paternal (T: U_PAT) and maternal transmission (T: U_MAT) counted for each allele transmission was 3. 'T' was the number of times the variant was transmitted versus untransmitted (U) from the parents. The ratio between the transmitted and untransmitted in paternal/maternal analysis was 3:0 . The Chi-square statistic (CHSIQ) separately calculated for stallion and mare had a value of three. The value of three indicated a deviation from the expected randomly trasmisson. This analysis revealed that there was the same pattern of recombination in maternal and paternal positions for these four markers. These four markers were in chromosome 9 and they were the same markers obtained with the best values in the case-control study (AX-104091145, AX-103221544, AX-103979373, AX-103993194). In the parent origin analysis, the p-value for paternal (P_PAT) and maternal (M_PAT) were above the significance threshold. The most significant p-value was 0.08326 in the first four markers of Table 7. For other variants no data was available (NA) and also the transmitted/untransmitted ratio was 0:0. It means there were no observations available to analyze transmission patterns for those markers.

Table 7. The parent origin analysis table. It represents p-values (P) and chi-square (CHISQ) for paternal and maternal transmission disequilibrium tests of genetic variants. T:U is the ratio of transmitted and untransmitted alleles from the sire and dam.

CHR	Variant ID	T:U_PAT	CHISQ_PAT	P_PAT	T:U_MAT	CHISQ_MAT	P_MAT

9	AX-103993194	3:0	3	0.08326	3:0	3	0.08326
9	AX-104091145	3:0	3	0.08326	3:0	3	0.08326
9	AX-103221544	3:0	3	0.08326	3:0	3	0.08326
9	AX-103979373	3:0	3	0.08326	3:0	3	0.08326
9	AX-104111830	3:0	3	0.08326	2:0	2	0.1573
9	AX-104294174	3:0	3	0.08326	2:0	2	0.1573
1	AX-104029597	0:0	NA	NA	0:0	NA	NA
1	AX-104093815	0:0	NA	NA	0:0	NA	NA
3	AX-103674726	0:0	NA	NA	0:0	NA	NA

The TDT, tdt parent origin analysis, and dfam analyses identified several markers of interest, particularly on chromosome 9, where significant markers were also found in the case/control study.

5.6 Linkage analysis

The most significant values were the p-value and the LOD score for every CentriMorgan of the chromosome. In the non-parametric models in different Centri-Morgan regions, the highest values were 0,602 LOD score and 0.04795 for the p-value, while the lowest values were -0.125 LOD score and 0.7759 p-value (Table 8). The first analysis was made on the chromosomes (1,9,11,20,31) where it was achieved the most significant results in the case-control study and family-based association. The 11-66 Cm showed a moderate level of evidence for linkage, with a Z-score and Delta indicating a positive association, and a p-value suggesting statistical significance. In the positions 0-10 and 67-91 cM there was no significant evidence of linkage or association with the trait.

Table 8. The Linkage analysis of the non-parametric model in chromosome 9 with the lowest LOD score (0-10 and 67-91 cM) and the highest score (11-66 cM) in the centri Morgan position of the chromosome. CHR: Chromsome, POS: Centri-Morgan position, Z-SCORE: Deviations from the Mean, DELTA: recombination facto, LOD: Logarithm of Odds, P: P-value,

CHR	POS (cM)	ZSCORE	DELTA	LOD	P
9	0-10 and 67-91	-0.57	-0.577	-0.12	0.7759
9	11-66	1.73	1.72	0.602	0.04795

The most interesting results were in chromosome 9, which is composed by 91 cM and the LOD score of 0,602 with a p-value 0.04795 from 10.9 cM to 66 cM (Figure 11). Other chromosomes achieved the same results for LOD score, but for a shorter Centri Morgan region respect to chromosome 9. .

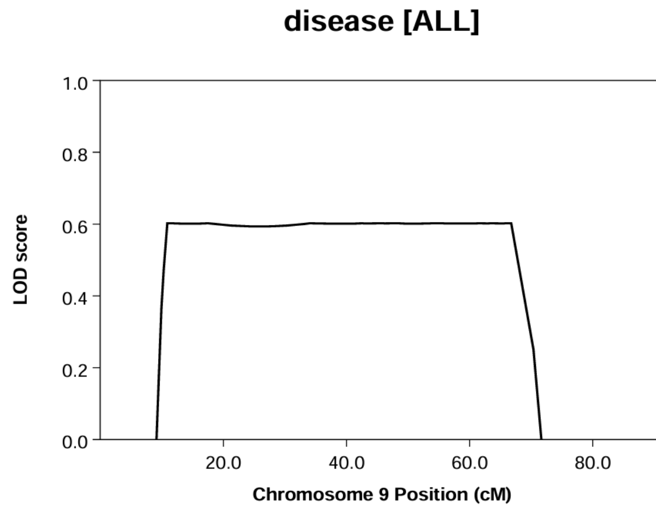


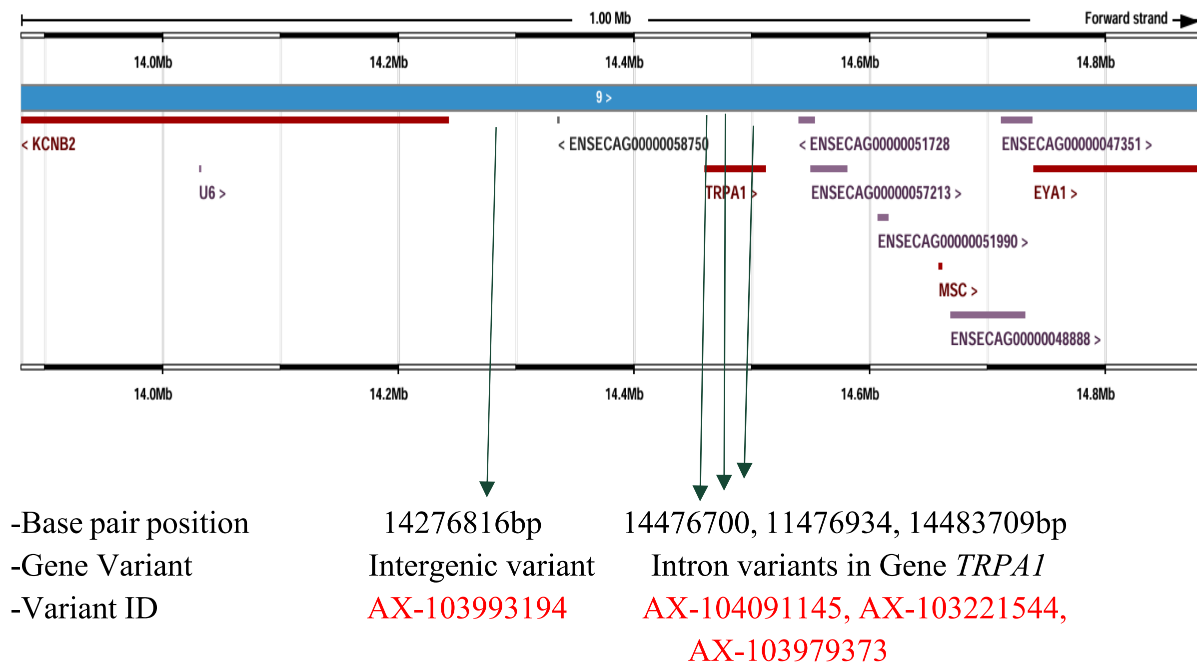
Figure 11. The Linkage analysis of non-parametric model (free-model) for chromosome 9. The x-axes is the LOD score and the y-axes is the Centri-Morgan positon in the chromosome 9.

The most significant centri-Morgan regions were the same for non-parametric model and recessive model in linkage analysis. In chromosome 9 there recessive parametric model produced the LOD 0.0001 ALPHA 1.000 HLOD 0.0001, while in the other centri morgan regions there are 0 values for these three parameters. ALPHA was the proportion of the families that are linked to the trait for a recessive model, the value of 1,000 means that all the families are linked to the markers. HLOD was the alpha lod, where the lod score considered also the proportion of linked families.

The most sigifcant markers

The most significant makers for the statistical analysis were in chromosome 9. On chromosome 9, the first four markers (AX-104091145, AX-103221544, AX-103979373, AX-103993194) had the right correlation with the genotypes of foals, dams and sire for an autosomal recessive trait. These markers had a close location on the chromosome (14476700bp, 14476934bp, 14483709bp, 14276816bp). They were also associated with the same gene. The map from the Ensembl gene database (Table 9) described the base pair position of markers in chromosome 9.

Table 9. The map from the Ensembl gene database. AX-103993194 in 14276816bp is the intergenic variant before Gene TRPA1. AX-104091145, AX-103221544 and AX-103979373 in 14476700, 11476934, 14483709bp are intron variants for Gene TRPA1.



6. Discussion

Every year, all the foals on Lojsta Hed come from the same stallion, so the hypermobility disease could be connected to a single mutation or structural variant inherited in an autosomal recessive manner. In the year when the stallion (sire 276) was selected to breed the mares on Lojsta Hed, 52% of mares didn't give birth. This percentage is higher than the cases of non-born foals registered in the two stallions (sire C, sire Q) chosen in the previous years at Lojsta Hed, with percentages of 26% and 17%, respectively. This aspect could be explained by hypermobility disease, which is also associated with cases of abortion and stillborn foals. An autosomal recessive disease called Warmblood Fragile Foal Syndrome (WFFS) is connected to horse mobility issues in Warmblood horses. The majority of homozygous foals for WFFS are lost due to abortion during late gestation, which could explain the high percentage of mares that didn't give birth. The number of mares covered and the number of foals born to sire 276 in the seven years before being selected for Lojsta Hed on "normal" farms has been calculated. The calculation discovered a percentage of 16.2%, which is lower than the 52% during the period at Lojsta Hed. This difference of percentage could be connected to a lower inbreeding on the farms compared to mares at Lojsta Hed or the disease could be caused by an environmental effect and not a genetic disease.

The Gotland ponies selected were a family group (trios), with a founder (sire 276), the mares, and the foals. The three mares with the died foals with the status "carrier" were also included since they are hypothesized to be heterozygotes (Affected/Normal). Another two stallions and one mare that didn't give birth were added to the samples or the analysis as controls, to increase the number of samples and achieve more significant results.

In an autosomal recessive study, it's necessary to have the homozygous alleles mutation (Affected/Affected) for the markers connected to the pathology to have the trait. In this study, the three surviving affected foals were the cases (Affected/Affected), and we expected that the sire and their dams be heterozygous. The other foals and the mares have to be heterozygous (Affected/Normal) or homozygous (Normal/Normal) for those markers. The family-based analysis included the three mares with the dead foals with the status "carrier" since they are hypothesized to be heterozygotes (Affected/Normal). Of course, some of the other mares could be carriers as well but gave birth to foals with genotypes

Normal/Normal or Normal/Affected while the cases are homozygous Affected/Affected.

The sire 276 has the lowest percentage of heterozygosity (17.42%) compared to other samples of Gotland Ponies, which is around 20%. These low levels could be connected to Gotland's relatively small population size and lower genetic diversity compared to larger, more diverse populations. A stallion and a mare with lower heterozygosity have a higher proportion of homozygous gene loci, which could play an important role in the potential for recessive genetic diseases.

In the Gotland Ponies, the level of inbreeding was elevated due to the small effective population size and to avoid this problem during the quality control process of the markers, different parameters were used. It was selected a maf of 0.01. In this way, the threshold was 1% of the frequency at which the second common allele occurs in the population. A low threshold of minor allele frequency was important to increase the sensitivity to detect rare genetic variants. These variants could have a relevant role in specific traits or diseases. The difference in markers between the threshold of 0.01 and 0.05 maf was about 64,631 markers. The value of 0.01 has given the possibility to work with more markers and to have more possibilities to find the mutation for the disease.

The Gotland Ponies samples are part of the same family, and in this situation, the association between a trait and a DNA region could be given by the underlying genetic structure of the population and not by a causative gene in that region. The consequence is connected to the creation of a false-positive signal of association. Other samples from different breeds were included in the second case-control study to address the stratification problem. Also, with samples from different breeds used as controls, the difference compared to the three cases could be due to the genetic structure of different populations and not by the causative gene. The analysis with the family-based association study (TDT) and the linkage analysis was necessary to resolve the stratification and false positive problems.

The Transmission disequilibrium test (tdt) and his implementation with unrelated samples (dfam) identified several markers of interest, particularly on chromosome 9, where significant markers were also detected in the case/control study. This indicates potential regions of the genome that could be associated with the traits of interest. The disease was supposed to be an autosomal recessive pathology, and the markers connected to the pathology have to be homozygous for the cases and heterozygous for their mares and the stallion. On chromosome 9, the first four markers (AX-104091145, AX-103221544, AX-103979373, AX-103993194) had the right correlation with the genotypes of foals, mares, and stallion for an autosomal recessive trait. These markers had a close location on the chromosome (14476700bp, 14476934bp, 14483709bp, 14276816bp). This aspect suggested the possibility of the presence of a haplotype in that portion of chromosome 9. Another case-control study focused on heterozygous markers of

Size 276. Since a recessive disease model was expected, the markers in the founder (Size 276) were supposed to be carriers (affected/normal) of the disease. The most significant markers identified in chromosome 9 (AX-103979373 and AX-103993194) were consistent with those from previous analyses, further strengthening the results.

One negative aspect is that in these four markers, the mares of the foals who died within 9 days from hypermobility were homozygous (Normal/Normal), while they were expected to be carriers since they were hypothesized to be heterozygotes (Affected/Normal). The three foals who shortly died after birth due to hydrocephalus and ataxia had a more severe disease compared to the other three foals who survived during recovery. This difference in the severity of the disease could be connected to different mutations in the genome between the foals who survived and those who died. This could explain why the three mares of the dead foals weren't carriers (Affected/Normal) for the four markers on chromosome 9.

These four markers on chromosome 9 were part of the *TRPA1* gene. The marker AX-104091145 at position 14,476,700 bp had an SNP with ambiguity code R. The ambiguity code R means that the base at this position could be a purine (A/G), but there was no certainty about which one it is. This SNP was an intron variant, a non-coding region with an important role in gene regulation and expression. The AX-103221544 and AX-103979373 were intron variants, while the AX-103993194 was an intergenic variant for the *TRPA1* gene.

The gene connected to this marker, *TRPA1*, is a protein-coding gene known as the Transient Receptor Potential Cation Channel. It enables monoatomic ion channel activity, enables protein binding, is involved in monoatomic ion transport, and is involved in transmembrane transport. It is involved in a rare genetic disease called Cramp-Fasciculation Syndrome, which affects neuronal and skeletal muscle conditions. The main effects are painful muscle cramps, which are also connected to hyperlexia and muscle stiffness similar to neuromyotonia. The foals in this study were diagnosed with hypermobility, which could be similar to the effects of this syndrome. The difference is that Cramp-Fasciculation Syndrome is classified as a rare human genetic disease, with most cases being idiopathic and occurring in adult patients. There's no prior study about the correlation between Cramp-Fasciculation Syndrome and cases of abortion. It is also related to a group of Episodic Ataxia (EA) inherited in an autosomal dominant manner and caused by genetic changes in various genes. This condition affects the nervous system, causing problems with movement and coordination, which can begin in childhood or adulthood.

Other markers on chromosome 9 (AX-104111830, AX103982812, AX-104898282) had genotypes with the three cases and still, one of their mares (217) was homozygous, and the stallion with the other two mares was heterozygous for these markers. These markers achieved a good p-value result in the case-control analysis with only the Gotland Ponies samples and also with the second case-

control association where samples from different breeds were included. Their locations are very close to each other (13970755bp and 13983014bp), which could suggest a haplotype. They were associated to the gene *KCNB2,* involved in the Potassium Voltage-Gated Channel Subfamily B Member 2. No diseases similar to hypermobility and ataxia were connected to a mutation in this gene. The mare (217), despite being only a carrier, could be an (Affected/Affected) case and it could be necessary to check if the mare has a hypermobility syndrome.

These markers on chromosome 9 (AX-104091145, AX-103221544, AX-103979373, AX-1039931949) and (AX-104111830, AX103982812, AX-104898282) have a similar CentiMorgan position between 17,801 and 18,544 cM. This could be connected to the principle that genes located close to each other on the same chromosome are likely to be inherited together. The linkage analysis for the non-parametric and recessive parametric model demonstrated that in the region between 11 and 67 cM, there were the most significant results for chromosome 9. The LOD score is the estimation position of the disease locus on the genome. A value of 3 or more corresponds to a significance level of 0.0001 and is considered the minimum value for significant results. In this study, the LOD score was lower than the expected value of 3, which could be caused by the low number of case samples (3). An increase in the number of cases from Gotland Ponies could help achieve more significant results. It could also be important to improve the unrelated samples from the general populations of Gotland Ponies to discover if a DNA region is given by the underlying genetic structure of the population and not by a causative gene in that region.

Table 10. The four markers on chromosome 9 (AX-104091145, AX-103221544, AX-103979373, AX-103993194) have parameters to be significant candidates for the variations causing hypermobility in the foals

p-value	Significant p-value in association and family-based analysis
Genotype	Match in the genotype of the three cases with the stallion and the three mares. Their genotypes followed the rules of an autosomal recessive disease
Parent origin analysis	Parent origin analysis (POO) demonstrated that the allele had both parents' origin.
Base pair position	Close position on the same chromosome, suggesting a haplotype
LOD score	The higher LOD score and p-values in the linkage analysis was in the CentiMorgan region of chromosome 9.
Gene Associated	*TRPA1* gene, linked to Cramp-Fasciculation Syndrome and Episodic Ataxia, which affect neuronal and skeletal muscle conditions

Some aspects were cons for these markers, because the genotype of the three mares of the foals who died with hydrocephalus and ataxia were homozygous (Normal/Normal). They were expected to be carriers (Affected/Normal) for the trait. In the linkage analysis, the LOD score was lower (0.6) than the value of 3, which is considered the threshold level to have significant results. Further validation and functional studies are required to understand the biological significance of these markers.
There could be a possibility of a different inheritance pattern than autosomal recessive. The disease could be connected to a novel dominant allele and more than one causative variant. In the option, it could be necessary to make another analysis where it's considered a different parametric model.
The methods chosen for Genome Analysis were based on previous studies to achieve the most significant results. PLINK was ideal for handling large-scale genotype and phenotype data analysis. The data obtained from the Axiom analysis suite, which included all markers and horse samples, were provided in Variant Call Format (VCF). This format was easily converted into PLINK-compatible files (MAP/PED), making the data ready for analysis. While PLINK was used exclusively for association analysis, a different tool was required for linkage analysis. Although several software options were available for linkage analysis, Merlin was identified to be the most suitable for this study. R packages like kinship and paramlink2 were also available for performing parametric linkage analysis, both of which were designed to calculate LOD scores. However, the main issue was that these packages required a different dataset format than what was produced from the quality control in PLINK (MAP/PED). This led to the selection of Merlin, which could directly use the PLINK-format data without requiring further conversion. For haplotype phasing, which identifies alleles co-located on the same chromosome, Beagle and Shape were both viable options. While both methods could work with the VCF data from the Axiom Analysis Suite, Beagle demonstrated higher accuracy, particularly for larger sample sizes and markers.
This study highlights aspects of genomic data analysis and research, which raise important social, ethical, and sustainability considerations. The results of genomic analysis may have social impacts, such as the prioritization of certain genetic traits over others, potentially affecting animal welfare, breed diversity. Additionally, the results could influence how animal populations are managed. It could be important to exclude the animal carriers or homozygous for this trait, to avoid other cases.
There are also concerns related to data privacy, particularly regarding the genetic data of the animals involved in the study. These considerations are crucial in ensuring the responsible and ethical use of genomic data in research and breeding programs.

Conclusions

The discussion delves into an analysis of hypermobility disease in Gotland ponies, particularly focusing on the genetic factors associated with it. The hypermobility disease in Gotland ponies is likely inherited in an autosomal recessive manner, suggesting a potential genetic mutation linked to the condition. The stallion (276) chosen for breeding at Lojsta Hed, appears to have a higher rate of non-born foals compared to previous years, potentially due to the hypermobility disease. Several markers on chromosome 9, particularly those associated with the TRPA1 gene, have shown significant correlation with the hypermobility disease. These markers suggest potential regions of interest for further genetic investigation. Further validation and functional studies are necessary to fully understand the biological significance of these markers. the discussion provides valuable insights into the genetic basis of hypermobility disease in Gotland ponies, highlighting the need for continued research to unravel the complexities of this condition and develop effective breeding strategies to mitigate its impact on the population. There could be a possibility of a different inheritance pattern than autosomal recessive. The disease could be connected to a novel dominant allele and more than one causative variant. The practical implications of the results are significant for preventing hyper-flexibility in Gotland Ponies. The study found that 52% of the mares did not give birth. This issue is particularly important from an economic perspective, as it could result in fewer foals available for farmers to sell each year.

References

Ablondi M, Johnsson M, Eriksson S, Sabbioni A, Viklund ÅG, Mikko S.(2022)’Performance of Swedish Warmblood fragile foal syndrome carriers and breeding prospects.’Genet Sel Evol. 2022 Jan 21;54(1):4

Agerholm, J.S. et al. (2015) ‘Virus-induced congenital malformations in cattle’, Acta Veterinaria Scandinavica, 57(1), p. 54. Available at: https://doi.org/10.1186/S13028-015-0145-8.

Bailey-Wilson, J.E. and Wilson, A.F. (2011) ‘Linkage Analysis in the Next-Generation Sequencing Era’, Human Heredity, 72(4), p. 228. Available at: https://doi.org/10.1159/000334381.

Balagu-Dobón, L., Cáceres, A. and González, J.R. (2022) ‘Fully exploiting SNP arrays: a systematic review on the tools to extract underlying genomic structure’, Briefings in bioinformatics, 23(2). Available at: https://doi.org/10.1093/BIB/BBAC043.

Bang, H. and Davidian, M. (2010) ‘Experimental statistics for biological sciences.’, Methods in molecular biology (Clifton, N.J.), 620, pp. 3–104. Available at: https://doi.org/10.1007/978-1-60761-580-4_1.

Barrett, J.H. and Teare, M.D. (2011) ‘Linkage Analysis’, Methods in Molecular Biology, 760, pp. 19–33. Available at: https://doi.org/10.1007/978-1-61779-176-5_2/TABLES/1.

Brian Brownin (2022) ‘Beagle 5.4 Manual’, Department of Medicine Division of Medical Genetics University of Washington.

Brown, T.A. (2002) ‘Genomes’, UK: Wiley-Liss Manchester, p. 572. Available at: https://www.ncbi.nlm.nih.gov/books/NBK21128/ (Accessed: 7 March 2024).

Browning, S., Browning, B.(2011) Haplotype phasing: existing methods and new developments. Nat Rev Genet 12, 703–714

Ducro, B.J. et al. (2011) ‘A nonsense mutation in B3GALNT2 is concordant with hydrocephalus in Friesian horses’. Available at: https://doi.org/10.1186/s12864-015-1936-z.

Elston, R.C. (ed.) (2017) ‘Statistical Human Genetics’, 1666. Available at: https://doi.org/10.1007/978-1-4939-7274-6.

Finno, C.J. and Bannasch, D.L. (2014) ‘Applied equine genetics’, Equine Vet J, 46(5), pp. 538–544. Available at: https://doi.org/10.1111/evj.12294.

Gorlova, O.Y. et al. (2022) ‘SNP characteristics and validation success in genome wide association studies’, Human Genetics, 141, pp. 229–238. Available at: https://doi.org/10.1007/s00439-021-02407-8.

Gotlandsrusset – Swedish Breeders' Association (no date). Available at: https://www.gotlandsruss.se/gotlandsrusset/ (Accessed: 7 March 2024).

Hayashi, N., Leifer, D.W. and Cohen, A.R. (2000) 'Chronologic changes of cerebral ventricular size in a transgenic model of hydrocephalus', Pediatric Neurosurgery, 33(4), pp. 182–187. Available at: https://doi.org/10.1159/000055950.

Hedberg, C., Oldfors, A. and Darin, N. (2013) 'B3GALNT2 is a gene associated with congenital muscular dystrophy with brain malformations', European Journal of Human Genetics 2014 22:5, 22(5), pp. 707–710. Available at: https://doi.org/10.1038/ejhg.2013.223.

Hodge, S.E. (1993) 'Linkage Analysis versus Association Analysis: Distinguishing between Two Models That Explain Disease-Marker Associations', Am. J. Hum. Genet, 53(3), pp. 367–384.

Hofmeister, R.J., Ribeiro, D.M., Rubinacci, S. et al. (2023) 'Accurate rare variant phasing of whole-genome and whole-exome sequencing data in the UK Biobank'. Nat Genet 55, 1243–1249.

Jiang, L. et al. (2019) 'A resource-efficient tool for mixed model association analysis of large-scale data', Nature genetics, 51(12), pp. 1749–1755. Available at: https://doi.org/10.1038/S41588-019-0530-8.

Kolb, D.S. and Klein, C. (2019) 'Brief Communication Communication brève Congenital hydrocephalus in a Belgian draft horse associated with a nonsense mutation in B3GALNT2', Can Vet J, 60, pp. 197–198. Available at: http://library.wur.nl/ (Accessed: 7 March 2024).

Kousi, M. and Katsanis, N. (2016) 'The Genetic Basis of Hydrocephalus', Annual review of neuroscience, 39, pp. 409–435. Available at: https://doi.org/10.1146/ANNUREV-NEURO-070815-014023.

Laframboise, T. (2009) 'SURVEY AND SUMMARY Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances', Nucleic Acids Research, 37(13), pp. 4181–4193. Available at: https://doi.org/10.1093/nar/gkp552.

Mccoy, A.M. and Mccue, M.E. (2013) 'Validation of imputation between equine genotyping arrays'. Available at: https://doi.org/10.1111/age.12093.

Mendell, J.R., Boué, D.R. and Martin, P.T. (2006) 'The Congenital Muscular Dystrophies: Recent Advances and Molecular Insights', Pediatric and developmental pathology : the official journal of the Society for Pediatric Pathology and the Paediatric Pathology Society, 9(6), p. 427. Available at: https://doi.org/10.2350/06-07-0127.1.

Monthoux C, de Brot S, Jackson M, Bleul U, Walter J. (2015) 'Skin malformations in a neonatal foal tested homozygous positive for Warmblood Fragile Foal Syndrome.' BMC Vet Res. 2015;11:12.

Norton, E.M. et al. (2019) 'Evaluation of an HMGA2 variant for pleiotropic effects on height and metabolic traits in ponies', Journal of Veterinary Internal Medicine, 33(2), pp. 942–952. Available at: https://doi.org/10.1111/JVIM.15403.

Oti, M., Ballouz, S. and Wouters, M. a (2011) 'In Silico Tools for Gene Discovery', Methods in Molecular Biology. Edited by Array, 760(i), pp. 175–187. Available at: https://doi.org/10.1007/978-1-61779-176-5.

Ott, J., Kamatani, Y. and Lathrop, M. (2011) 'Family-based designs for genome-wide association studies', Nature reviews. Genetics, 12(7), pp. 465–474. Available at: https://doi.org/10.1038/NRG2989.

Price, A.L. et al. (2010) 'New approaches to population stratification in genome-wide association studies', Nature Publishing Group [Preprint]. Available at: https://doi.org/10.1038/nrg2813.

Pulst, S.M. (1999) 'Genetic linkage analysis', Archives of neurology, 56(6), pp. 667–672. Available at: https://doi.org/10.1001/ARCHNEUR.56.6.667.

Schaefer, R.J. et al. (2017) 'Developing a 670k genotyping array to tag ~2M SNPs across 24 horse breeds', BMC Genomics, 18(1). Available at: https://doi.org/10.1186/s12864-017-3943-8.

Schaefer, R.J. and McCue, M.E. (2020) 'Equine Genotyping Arrays', The Veterinary clinics of North America. Equine practice, 36(2), pp. 183–193. Available at: https://doi.org/10.1016/J.CVEQ.2020.03.001.

Schmidt, M. and Ondreka, N. (2019) 'Hydrocephalus in Animals'. Available at: https://doi.org/10.1007/978-3-319-27250-4_36.

Schurink, A. et al. (2018) 'Copy number variations in Friesian horses and genetic risk factors for insect bite hypersensitivity', BMC Genetics, 19(1), pp. 1–13. Available at: https://doi.org/10.1186/S12863-018-0657-0/FIGURES/3.

Slatkin, M. (2008) 'Linkage disequilibrium — understanding the evolutionary past and mapping the medical future', Nature Reviews Genetics 2008 9:6, 9(6), pp. 477–485. Available at: https://doi.org/10.1038/nrg2361.

Solé, M. et al. (2019) 'Inter- and intra-breed genome-wide copy number diversity in a large cohort of European equine breeds', BMC genomics, 20(1). Available at: https://doi.org/10.1186/S12864-019-6141-Z.

Stevens, E. et al. (2013) 'ARTICLE Mutations in B3GALNT2 Cause Congenital Muscular Dystrophy and Hypoglycosylation of a-Dystroglycan', The American Journal of Human Genetics, 92, pp. 354–365. Available at: https://doi.org/10.1016/j.ajhg.2013.01.016.

Svenska Hästavelsförbundets webbtjänst: Blå Basen (no date). Available at: https://www.blabasen.se/sh/SokHast?a=visa_hast&id=374522 (Accessed: 7 March 2024).

Tam, V. et al. (2019) 'Benefits and limitations of genome-wide association studies', Nature reviews. Genetics, 20(8), pp. 467–484. Available at: https://doi.org/10.1038/S41576-019-0127-1.

Uffelmann, E. et al. (2021) 'Genome-wide association studies', Nature Reviews Methods Primers 2021 1:1, 1(1), pp. 1–21. Available at: https://doi.org/10.1038/s43586-021-00056-9.

Viksten SM, Hartmann E, Schneller K, Steen M. (2023)'Welfare of extensively managed Swedish Gotland ponies. Anim Welf'. 2023 Feb 23;32:e21.

Zhang, H. et al. (2012) 'Progress of genome wide association study in domestic animals', Journal of Animal Science and Biotechnology, 3(1), pp. 1–10. Available at: https://doi.org/10.1186/2049-1891-3-26/TABLES/1.

Popular science summary

Recent research has focused on the prevalence and potential genetic causes of hypermobility disease among Gotland ponies at Lojsta Hed on Gotland Island. This condition, believed to be inherited in an autosomal recessive manner, might be linked to a specific mutation or structural variant in the DNA. The study began when a stallion, known as sire 276, was selected to breed all the mares at Lojsta Hed. Unfortunately, this decision led to a troubling outcome: many foals were born with severe health issues, including hydrocephalus (an accumulation of cerebrospinal fluid in the brain) and hyperflexible joints. The percentage of mares that failed to give birth skyrocketed to 52%, compared to just 20% in previous years with different sires. This significant increase raised concerns that hypermobility disease might be causing higher rates of abortion and stillbirth, akin to conditions seen in other breeds, such as Warmblood Fragile Foal Syndrome (WFFS). The researchers collected DNA samples from the affected foals and their parents to identify carriers of the suspected recessive allele responsible for hypermobility disease. Using the Axiom® Equine Genotyping Array, which includes over 670,000 genetic markers, they conducted both linkage and association analyses to map the disease's genetic loci. Linkage analysis revealed that genes located close together on the same chromosome are often inherited together, aiding in locating the disease genes. The association analysis compared genotypes of healthy control ponies to those affected by hypermobility disease. Significant genetic markers were found on chromosome 9, specifically between positions 17.8 and 18.5 centimorgans. These markers overlap with the TRPA1 gene, which in humans is linked to neurological and skeletal muscle conditions. This suggests a possible connection to the hypermobility seen in the affected foals.

The study’s findings point to a genetic basis for hypermobility disease in these ponies. However, it’s necessary further investigation with larger sample sizes and unrelated Gotland ponies to pinpoint the causative gene in the identified region. Understanding this condition could lead to improved breeding practices and the overall health of Gotland ponies.

Appendix

Appendix 1.List of the 41 Gotland ponies samples genotyped in Axiom Analysis Suite. The samples were classified by their ID, sex (Male/Female), their mother and father ID and phenotype. The three foal cases are colored by red.

ID	Sex	FatherID	MotherID	Phenotype
209	F	NA	NA	Mother of dead foal with hydrocephalus and ataxia
210	F	NA	NA	Mother of dead foal with hydrocephalus and ataxia
211	F	NA	NA	Mother of dead foal with hydrocephalus and ataxia
212	F			Mother of an hypermobility foal
213	F	276	212	Hypermobile foal
214	M	276	217	Hypermobile foal
215	F	276	216	Normal foal
216	F	NA	NA	Mother of a normal foal
217	F	NA	NA	Mother of an hypermbolity foal
218	F	NA	NA	Mother of an hypermobility foal
219	F	276	218	Hypermobile foal
220	M	NA	NA	Other Sire
221	F	NA	NA	Mother of a normal foal
222	F	276	221	Normal foal
223	F	NA	NA	Mother of a normal foal
224	F	276	223	Normal foal
225	M	276	226	Normal foal
226	F	NA	NA	Mother of a normal foal
227	F	NA	NA	Mother of a normal foal
228	M	276	227	Normal foal
229	M	276	230	Normal foal
230	F	NA	NA	Mother of a normal foal
231	F	NA	NA	Mare without foal
232	F	NA	NA	Mother of a normal foal
233	M	276	232	Normal foal

234	F	NA	NA	Mother of a normal foal
235	F	276	234	Normal foal
236	F	NA	NA	Mother of a normal foal
237	M	276	236	Normal foal
238	F	NA	NA	Mother of a normal foal
239	F	276	238	Normal foal
240	F	NA	NA	Mother of a normal foal
241	F	276	240	Normal foal
242	F	NA	NA	Mother of a normal foal
243	M	276	242	Normal foal
244	F	NA	NA	Mother of a normal foal
245	F	276	244	Normal foal
246	F	NA	NA	Mother of a normal foal
247	M	276	246	Normal foal
248	M	NA	NA	Other foal
276	M	NA	NA	Father of all 17 foals

Appendix 2. Plink settings for the Quality Control analysis of the samples and markers.

Plink Settings	Code	Carachteristic
Allow extra chr	--allow-extra-chr 0	set to recognise different chromosome codes.
Horse	--horse	chromosome set for horse species (32 chromosomes).
Double ID	--double-id	to modify sample IDs from VCF format to PLINK IDs.
Minor Allele Frequency	--maf 0.01	All variants with a minor allele frequency below 1% were filtered out.
Hardy-Weinberg Equilibrium p-value	--hwe 1e-5	p-value under the threshold of 1e-5 were removed.
Geno	--geno 0.02	Variants with missing data exceeding the 2%.

Mind	--mind 0.02	Samples with missing data exceeding the 2%.
Check sex	--check sex	compared the input dataset (fam file) sex with the X chromosome inbreeding coefficient.
Heterozygosity Rate	--het	observed and expected autosomal homozygous genotype counts for each sample.

Publishing and archiving

Approved students' theses at SLU are published electronically. As a student, you have the copyright to your own work and need to approve the electronic publishing. If you check the box for **YES**, the full text (pdf file) and metadata will be visible and searchable online. If you check the box for **NO**, only the metadata and the abstract will be visible and searchable online. Nevertheless, when the document is uploaded it will still be archived as a digital file.

If you are more than one author you all need to agree on a decision. Read about SLU's publishing agreement here: https://www.slu.se/en/subweb/library/publish-and-analyse/register-and-publish/agreement-for-publishing/.

☒ YES, I/we hereby give permission to publish the present thesis in accordance with the SLU agreement regarding the transfer of the right to publish a work.

☐ NO, I/we do not give permission to publish the present work. The work will still be archived and its metadata and abstract will be visible and searchable.