Polymorphism in Myostatin gene and athletic performance in Nordic horse breeds

Agnese Viluma

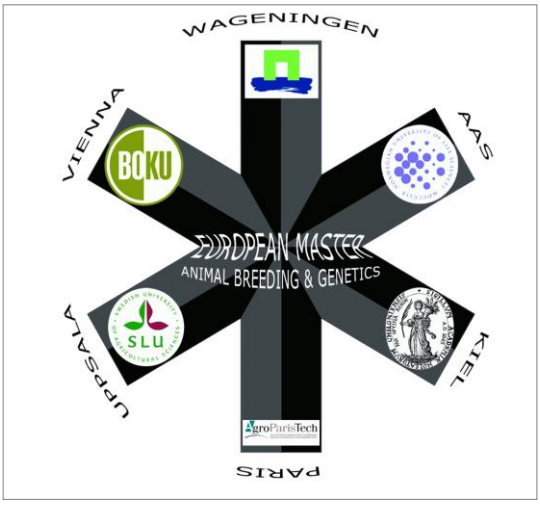
Norwegian University of Life Sciences
Department of Animal and Aquiculture Sciences
Master Thesis 30 credits 2012

European Master in Animal Breeding and Genetics

POLYMORPHISM IN MYOSTATIN GENE AND ATHLETIC PERFORMANCE IN NORDIC HORSE BREEDS

Agnese Viluma
977103

THESIS ANIMAL BREEDING AND GENETICS (EX0556)
Jun 2012



Department of Animal Breeding and Genetics

SUPERVISORS
Gabriella Lindgren, SLU
Lisa Andersson, SLU
Dag Inge Våge, UMB

Education and Culture
Erasmus Mundus

Swedish University of Agricultural Sciences
Faculty of Veterinary Medicine and Animal Science
Department of Animal Breeding and Genetics

Title:	Polymorphism in myostatin gene and athletic performance in Nordic horse breeds
Author:	Agnese Viluma
Supervisor(s) at SLU:	Gabriella Lindgren
	Lisa Andersson
Supervisor at UMB	Dag Inge Våge
Examinator at SLU:	Leif Andersson

Credits:	30 ECTS
Level of work:	Advanced; A2E; (E-level)
Course title:	Degree Project for MSc Thesis in Animal Science
Course code:	EX0556
Programme:	European Master in Animal Breeding and Genetics

Keywords:	myostatin, horse, athletic performance, Coldblooded Trotter, Icelandic Horse, Nordic breeds

Uppsala 2012

CONTENTS

ABBREVIATIONS

BLUP	Best linear unbiased prediction
bp	Base pairs
BRD	Best race distance
DNA	Deoxyribonucleic acid
ECA	Equine chromosome
HWE	Hardy-Weinberg equilibrium
LD	Linkage disequilibrium
MAF	Minor allele frequency
miRNA	Micro Ribonucleic acid
MSTN	Myostatin
PCR	Polymerase chain reaction
UTR	Untranslated region
WRD	Win-race distance

ABSTRACT

Athletic performance including working ability, riding, racing and endurance has played a leading role in horse breeding possibly since their domestication. Even though traits of athletic performance are considered as complex and dependant on environment, management and training, it has been shown that heritability of some performance traits can be considerably high. Skeletal muscle system seems to be a key factor to exceptional athletic performance and numerous studies in other species (dogs, humans, sheep, cattle etc.) have shown that polymorphisms in *MSTN* gene can be responsible for phenotypic changes in muscle tissues. Several studies focusing on horse MSTN gene region have been made and some significant associations discovered in Thoroughbred populations. The main objective of this study is to estimate the allele frequencies of the g.66493737 SNP, which is the strongest associated SNP with the optimim race distance, and some additional SNPs (two in a promoter region and one downstream the gene) in *MSTN* gene of Nordic horse breeds and correlate them with performance trait data to create a deeper understanding of the gene's influence on performance ability in horses.

For analysis 25 horse breeds and four Donkey and Przewalski's horse individuals were selected. Four SNP marker positions (g.65868604G>T, g.66493737C>T, g.66495696G>A, g.66495826G>A) and one insertion (Ins227) were chosen based on previously published research results. For genotyping SNPs real time PCR TaqMan approach and touchdown PCR for detection of Ins227 was used. Phenotypic data of performance traits were correlated with inferred haplotypes of Icelandic Horse and Coldblooded Trotters using R software. Results revealed segregation in all SNP marker positions, having variable allele frequencies between breeds and lacking g.66495696G homozygous individuals. No Ins227 allele was detected, except those in Thoroughbred horses and one heterozygous individual of Swedish Warmblood. Haplotype association resulted in some significant haplotype effects on proportions (p-value<0.05) and tolt - running walk - ability (p-value<0.05) in Icelandic horse. More significant effect was detected for Coldblooded Trotters where both haplotypes having mutation in proximal promoter region had a significant effect on total BLUP value and BLUP values for Position, Earnings and Time. Haplotype having g.66495696G allele shoved the most significant association having p-values from 0.029 for Start status BLUP value to 0.001 for Position and Earnings. Finally, we can conclude that tested polymorphisms in *MSTN* gene are widespread in Nordic Horse breeds and there are indications that some of them might have a functional effect on horse performance traits. In addition, we have obtained specific directions of further research possibilities.

Keywords: myostatin, horse, athletic performance, Coldblooded Trotter, Icelandic Horse, Nordic breeds

1. INTRODUCTION

In modern horse (*Equus ferus caballus*) breeding athletic performance has been considered as the most important criteria for selection and horses with superior athletic performance are of the highest value. They have been selected for working, riding, racing and endurance performance since their domestication and possess a range of functional and structural adaptations for exceptional athletic performance (Schröder et al., 2011). Traits of athletic performance are of complex nature and are highly dependent on such factors as environment, management and training (Bower et al., 2012). However, since the end of 20th century it has been accepted that there are underlying genetic factors and the heritability of various measures of racing performance have been considered as high (Gaffney & Cunningham; 1988). In addition, studies in humans, rats and mice have revealed a large number of genes that are involved in elite athletic performance (Schröder et al. 2011). Thoroughbreds is one of the most investigated horse populations and the racing performance of Thoroughbreds is believed to be polygenic - being influenced by genes that contribute to the wide range of anatomical, metabolic and physiological adaptations (Bower et al., 2012). Myostatin has been proved to be one of the genes with an important role in horse athletic performance and polymorphism of this gene (g.66493737) has been associated with the best race distance in Thoroughbred horses (Hill et. al., 2010a).

Myostatin (*MSTN*) or transforming growth factor-8 (TGF-8) is a member of TGF-B superfamily and was discovered by McPherron et al. (1997) during the search of new TGF-B superfamily members. Genomic location of *MSTN* gene is on equine chromosome 18 (ECA18) 66,490208 - 66,495,180 reversed strand and it has only one transcript that consists of 3 exons and in total is 1128 bp long (Ensembl release 65 - Dec 2011, EquCab2 assembly). *MSTN* consists of 376 amino acids and contains all characteristic features of members of the TGF-B superfamily and is expressed almost exclusively in cells of skeletal-muscle lineage throughout embryonic development as well as in adult animals (McPherron et al., 1997). McCroskery et al. (2003) showed that in adult animals *MSTN* is expressed in satellite cells (myogenic progenitors responsible for postnatal muscle growth) and it maintains the satellite cell division. Respectively, when stimulated by damage of the muscle, satellite cells are activated to reenter the cell cycle and express myogenic regulatory factors.

Myostatin is a negative regulator of muscle growth that is responsible for muscle hypertrophy in a range of mammalian species (mouse, cattle, dog, sheep and human) and the function of myostatin has been highly conserved among vertebrates (McPherron and Lee, 1997). Huang et al. (2011) reports that mature *MSTN* is a disulfide-linked dimer of C-terminal domain and is 100% identical among human, mouse, rat, pig, chicken, turkey and dog; and after summarizing the recent research of endocrinology, he states that there is evidence that myostatin may reduce muscle mass by decreasing protein synthesis and that myostatin inhibits protein synthesis in both myoblasts and myotubes. In addition, recent findings have proved that myostatin has an effect on some signaling pathways and reveal a novel role of it in protein metabolism - regulation of protein synthesis (Huang et al., 2011).

Despite conservative nature of myostatin function the mutations that influence phenotypic changes in different species are individual. For human (Schuelke et al. 2004), cattle (Jeanplong et al., 2001), sheep (Boman et al., 2009; Boman & Våge 2009) and dogs (Mosher et al., 2007) increased muscle

mass phenotype is caused by the mutation in the coding part (respectively loss-of-function mutation has been reported in human, cattle and sheep, and premature stop codon in dogs). Clop et al. (2006) has shown that mutation in the *MSTN* 3' UTR in sheep creates an illegitimate target site for at least two miRNAs that are strongly expressed in skeletal muscle and results in miRNA-mediated translational downregulation and reduction in *MSTN* concentration. The mechanism how polymorphism in *MSTN* gene influences phenotype of the horse is still unknown. However, as all SNPs that have been associated with phenotypic variation are in the intronic or promoter regions and are not considered to be linked with coding variants, it has been suggested that phenotype variation has been influenced by regulation of *MSTN* gene expression (Tozaki et al., 2011).

Until now several studies has described various polymorphisms in the equine *MSTN* gene, most of them focusing on Thoroughbreds (Hill et al., 2010a; Hill et al., 2010b; Tozaki et al., 2010; Tozaki et al., 2011; Tozaki et al., 2012; Binns et al. 2010; Bower et al., 2012) While in some species it has been proved that mutations in the *MSTN* gene contributes to muscle hypertrophy, resequencing of the *MSTN* gene in Thoroughbred horses revealed a novel *MSTN* sequence polymorphism that was associated with the best race distance among elite racehorses (Hill et al., 2010a). Soon afterwards two independent GWA-studies identified 3 additional SNPs (g.65809482T>C; g.65868604G>T; g.66539967A>G) on ECA18 in the Japanese Thoroughbred population (Tozaki et al., 2010). Together with the previously characterized SNP g.66493737C>T these were highly associated with performance measures like lifetime earnings and performance ranks. Binns et al. (2010) identified a locus on ECA18 that was genome-wide significant for the optimum race distance and the two SNPs with the highest statistical significance in his study (BIEC2-417274 - located at 65868604 bp and BIEC2-417495 - located at 67186093 bp) enclosed the *MSTN* gene. The following GWA study of Hill et al. (2010b) confirmed the sequence variant g.66493737C>T as the most powerful in predicting the best race distance in Thoroughbred horses (Ireland, Great Britain and New Zeeland populations). Tozaki et al. (2011) has continued the investigation of all four previously identified polymorphisms. Win-race distance (WRD) among all winning racehorses and the best race distance (BRD) among elite racehorses where used as phenotypic traits and significant associations were observed for all four SNPs. The g.65809482T>C, g.65868604G>T and g.66493737C>T, or a combination of these three, were suggested as genetic diagnostic markers for racing performance indicators of WRD and BRD. Afterwards the same four SNPs were correlated with body composition traits (body weight, withers height, chest circumference, cannon circumference and body weight/withers height) and was described that body weight/withers height showed statistically significant differences depending on g.65868604G>T, g.66493737C>T and g.66539967A>G genotypes in males during the training period (Tozaki et al., 2012). A study of *MSTN* gene polymorphisms was carried out in horse breeds of different morphological types (Dall'Olio et al., 2010), and seven SNPs were identified: two in the promoter region, four in intron 1 and one in intron 2. The allele frequencies of the polymorphisms in the promoter region and haplotype analysis indicated that these polymorphisms could be associated with variability of morphology traits in studied horse breeds.

The ancestral wild type of the most associated SNP has been considered to be the g.66493737T allele. According to Bower et al. (2012), no g.66493737C allele variant was found in donkey (*Equusasinus*)

or zebra (*Equusgrevyi* and *Equusquaggaboehmi*) chromosome and it is known that wild grazing equids are suited for traveling long distances in search of food and water (Levine, 2005). After genotyping various horse breeds (French trotter; Irish Drought Horse; Quarter horse; Standardbreds; Icelandic horses, as well as horse breeds from British Isles, Middle East, North Africa and Asia) it has been concluded that the C-allele is not restricted to the Thoroughbred and Thoroughbred derived populations. It is not a new mutation and seems to occur at variable frequencies depending on the selection pressure on the population (Bower et al., 2012). After analyzing and comparing C and T allele haplotypes in Thoroughbred and Shetland samples, Bower et al. (2012) claims that results are consistent with a single introduction of the C-allele at the foundation stages of Thoroughbreds. Most likely it has been introduced by a local British mare as the prized foundation stallions were homozygous for the T-allele.

Bjørnstad et al. (2003) showed that all the northern European breeds had considerably smaller genetic distances to the Mongolian native horse than to Thoroughbred or Standardbred. Also the distance between the Mongolian native horse and the Norwegian breeds were similar to the distance between Norwegian breeds and Icelandic breed (presumably separated for about 1000 years). Bjørnstad et al. suggested (2003) that the Mongolian horse has had a major effect on a wide range of breeds and that central Asia could have been a center of dispersal of horses through trading and human migration – activities that is thought to be tightly connected with the migration of livestock.

Considering that *MSTN* genotypes can be used to predict the genetic potential for athletic performance in Thoroughbred horses, it is useful to investigate if this polymorphism is similarly associated with athletic performance in other horse breeds. In addition, as the close interactions between British Isles and Vikings might have influenced the exchange of genetic material of horses, information on polymorphisms of Nordic horse breeds might contribute essential information to broader understanding of how and where the g.664933737 mutation has originated. The main objective of this study is to estimate the allele frequencies of the g.66493737 SNP, which is the strongest associated SNP, and some additional SNPs (two in a promoter region and one downstream the gene) in *MSTN* gene of Nordic horse breeds and correlate them with performance trait data to create a deeper understanding of the gene’s influence on performance ability in horses.

2. MATERIALS AND METHODS

2.1. Study animals

Twenty five horse breeds from all over the world (Table 1.) were used to obtain an overview in which populations selected markers are segregating. However, the main focus of study has been on Northern European breeds.

Table 1. Characteristics of analyzed breeds

Breeds / Species	Nr of Horses	Origin (Hendricks, 1995)	Aptitude (Hendricks, 1995)
Swedish Ardennes	43	Sweden	Heavy draft work
Swedish Warmblood	65	Sweden	Riding horse
North Swedish Draft	41	Sweden	Heavy draft, farm work
Coldblooded Trotter	129	Sweden	Racing, riding
Gotlands Pony	32	Sweden	Riding, light draft, racing
Northlands Pony	22	Norway	Riding
Dølehest	16	Norway	Draft work, trotting racer
Norwegian Fjord	32	Norway	Riding, light draft
Icelandic Horse	318	Iceland	Riding horse; Meet horse
Faeroe Island Horse	14	Faroe Iceland	Riding horse
Connemara	12	British Isles	Riding, light draft
Shetlands Pony	69	British Isles	Riding, light draft
Thoroughbred	42	British Isles	Racing, trotting
Bashkir	20	Ural Mountains	Riding, light draft
Welsh Pony	21	British Isles	Riding, light draft
Miniatur Horse	8	British Isles	Bred as pets
Kentucky Mountain Saddle Horse	23	United States	Riding horse
Paso Fino	25	Puerto Rico	Riding horse
Missouri Fox Trotter	17	United States	Riding horse
Rocky Mountain Horse	17	United States	Riding horse
Tennessee Walker	20	United States	Riding horse
Standardbred	96	United States	Harness racing, riding
Arab	28	Middle East	Riding horse
French Trotter	41	France	Trotting racing
Mongolian Horse	11	Mongolia	Riding, pack, draft
Przewalski's Horse	4	Asia (western Mongolia)	Wild, undomesticated
Donkey	4	Domesticated in Egypt or Mesopotamia	Draft or pack work

The numbers of genotyped animals per breed are variable and depend on availability of samples from each breed. The largest genotyped populations were Icelandic Horse (318 individuals), Coldblooded Trotters and Standardbred, because for these breeds performance data were available for association

study. In order to follow the origin of particular mutations, 4 Donkeys, 11 Mongolian and 4 Przewalski's Horses were also genotyped.

Extracted DNA samples from the biobank at the Animal Genetics Laboratory, SLU, Sweden were available for almost all genotyped animals. Majority of the tested animals were unrelated. Except one group of Icelandic horses coming from one farm in Iceland and except samples from Bashkir, Faeroese, Fjord, Connemara and Welsh Pony, which might be to some extent related. Additional samples of Mongolian Horses, Dølehest and Northlands Pony were received from Knut Røed (The Norwegian School of Veterinary Science).

The phenotype information and estimated breeding values of Coldblooded trotters and Standardbreds were obtained from the database of Swedish Trotting Association. Breeding field test scores of Icelandic horses were acquired from the Icelandic horse registry World Fengur (www.worldfengur.com).

2.3. Molecular analysis

DNA from hair samples (Mongolian horse) was extracted by digestion with proteinase K. Roots of 4 – 6 hair were placed into proteinase K and buffer D solution and incubated at 55 ºC for 3 hours. After inactivation of the proteinase K at 95 ºC for 10 min, hair were discarded and extracted DNA stored in eppendorf tubes.

Horse DNA samples were genotyped at 4 SNP positions and one 227 bp insertion (Ins227) in the *MSTN* gene region. Schematic overview of marker positions is shown in Figure 1.

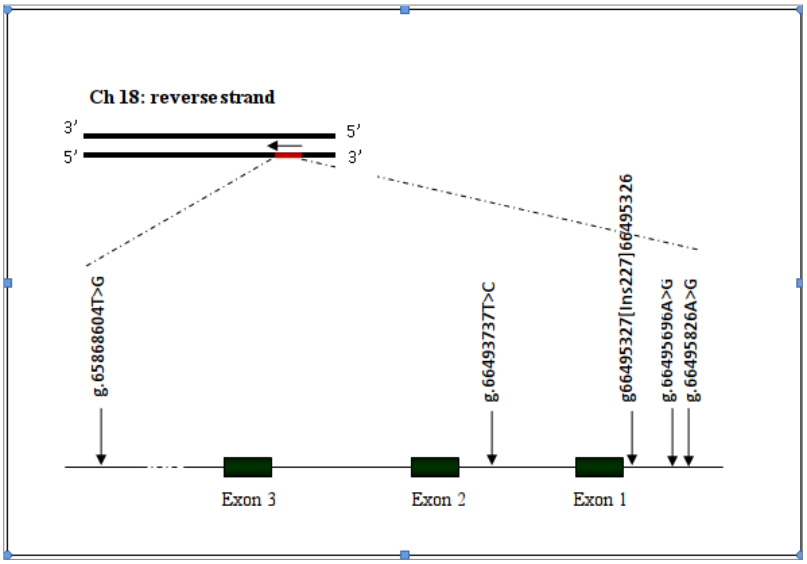


Figure 1. Relative location of analyzed markers.

Genotyping of SNP markers was performed by real-time PCR TaqMan assay. The TaqMan method is used to detect PCR products using real-time PCR instruments (in this study ABI PRISM 7900 HT sequence detection system for 384-well format from Applied Biosystems). It uses a fluorescent probe to enable the detection of a specific PCR product as it accumulates during PCR cycles to perform allelic discrimination. To generate the fluorescent signal specific hybridization between probe and target is required. At the 5' end of each probe there is a dye and at the 3' end a quencher, so that characteristic fluorescent signal is emitted only when dye is separated from quencher by the polymerase during the strand elongation (www.appliedbiosystems.com).

Primers for genotyped SNP positions and probe sequences for both alleles are given in Appendix 1. SNP positions of interest were chosen based on previous studies on *MSTN* gene in horses – all known SNP positions were summarized and the four most interesting selected. Features of chosen SNP positions are given in Table 2.

Table 2. Literature based knowledge of selected SNPs

SNP position (marker name)	Breed(s)	Nr of animals tested	Statistically significant differences depending on genotype	Reference
g.66495826A>G (PR5826)	Mostly Italian horse breeds	396	Indication that might have an influence on morphological type	(Dall'Olio et al., 2010)
g.66495696A>G (PR5696)	mostly Italian horse breeds	396	Indication that might have an influence on morphological type	(Dall'Olio et al., 2010)
g.66493737C>T (PR3737)	Japanese TB	91	Body weight/wither hight (in males and females)	(Tozaki et al.,2011)
	Japanese TB	1710	Performance rank, life time earnings, WRD	(Tozaki et al., 2012)
	Thoroughbreds	148	Most associated SNP with best race distance	(Hill et al., 2010a)
	Mostly Italian horse breeds	396	Not associated	(Dall'Olio et al., 2010)
g.65868604G>T (PR8604)	Japanese TB	3927	Performance measures (earnings of the race track)	(Tokezaki et al., 2010)
	Japanese TB	91	Body weight/wither hight (in males and females)	(Tozaki et al., 2011)
	Japanese TB	1710	Performance rank, life time earnings, WRD	(Tozaki et al., 2012)
	American TB	189	Individual optimum race performance	(Binns et al., 2010)

TB – Thoroughbred; WRD – win-race distance

Ins227 has been identified as horse-specific repetitive DNA sequence element by Hill et al. (2010b) and it is located on chromosome 18 between g.66495327 and g66495326. For genotyping Ins227 a PCR approach was used. Primers form previous study on Thoroughbred horse (Hill et al, 2010b) were used. Respectively, forward 5'-ATCAGCTCACCCTTGACTGTAAC-3' and reverse 5'-TCATCTCTCTGGACATCGTACTG-3'. Fragments were separated by gel electrophoresis: the fragment length of the allele without insertion is 600 bp, and the length of Ins227bp allele is 827 bp. A touch down PCR was used with the following thermal cycling program: 95ºC for 10 min, 11 cycles of 95ºC for 30 s, 64ºC for 30 s (every cycle decreasing by 1ºC), 72ºC for 1 min and 30 cycles of 95ºC for 30 s, 55ºC for 30 s and 72ºC for 1 min.

To make sure that it is possible to detect the Ins227 allele, 29 Thoroughbred individuals were analyzed. The following other breeds of interest were chosen: Icelandic Horse, Gotland's pony, Swedish Ardennes, North Swedish Draught, Dole, Shetlands Pony, Coldblooded Trotters, Swedish Warmbloods, Arabs and Standardbreds.

2.4. Data analysis

Frequencies of SNP alleles were calculated by MS Excel software. Deviations from Hardy-Weinberg equilibrium were obtained using Hardy-Weinberg equilibrium calculator including analysis for ascertainment bias (Rodriguez et al., 2009).

The actual number of haplotypes within the region of interest is much smaller than the number of all possible haplotypes, so that haplotype phasing helps to reduce and systemize SNP data (Lin&Zeng, 2006). According to Hagenblad et al. (2004) single marker association studies ignore the linkage information of markers, so that haplotype association method which include information of LD are of greater statistical power. However, standard genotyping techniques cannot distinguish the two homologous chromosomes of an individual and haplotype (sequence of nucleotides on a single chromosome) cannot be directly observed. Instead only combination of two haplotypes was obtained (Lin & Zeng, 2006). The software Phase2.1.1. (Stephens et al. 2001; Stephens and Scheet 2005) were used to construct haplotypes from unphased SNP data, which implements a Bayesian statistical method for reconstructing haplotypes from population genotype data of unrelated individuals. This software allowed missing data. Alternative input file format function (-f1) was used, allowing input files where genotypes are listed on a single line, locus by locus. When haplotypes of each individual are inferred, haplotypes can be considered as alleles for a single multi-allelic marker. All association methods based on single markers can then be applied to analyze haplotypes (Liu et al., 2008).

2.5. Phenotypic associations

The phenotype and genotype associations have been done in two ways:

- Broad approach – comparing allele frequency differences among different sample groups
- Narrow approach – associating values of performance traits with haplotypes.

2.5.1. Frequency differences

All genotyped breeds have been divided into four groups according to the main aptitude of the breed (based on information in Table 1):

- heavy draft;
- light draft / riding;
- racing;
- long distance riding.

Descriptive statistics of allele frequencies have been calculated with MS Excel software and the analysis of molecular variance (AMOVA) with help of free software ARLEQUINv.3.5. (http://cmpg.unibe.ch/software/arlequin3/).

From the Icelandic horse sample two groups were formed – one consisting of horses selected for riding (189 individuals) and one group of horses bred for meat and hormone production (45 individuals from one farm).

2.5.2. Haplotype association

Haplotype association with both, BLUP values and direct measurements of evaluation results was done in Icelandic Horses and Coldblooded Trotters.

As genotyped individuals of animals used for association are unrelated, a population-based haplotype-association method was used. The one way of doing it is to perform the analysis under the regression framework, where haplotypes can be treated as categorical variables (Liu et al., 2008).
According to Schaid (2004) haplotype-association methods based on a generalized linear model (GLM) describes how the haplotypes influence the mean of the trait. It provides a possibility to estimate score statistics for haplotype effect, where null hypothesis corresponds to no haplotype effect. Thus this method can be adapted to many traits – binary, quantitative and survival.

For this purpose a package of R software Haplo Stats (version 1.5.0) developed by Sinnwell and Schaid (2011) were used. The most frequent haplotype (in this study also the wild type haplotype) was used as a base haplotype and coefficients showing deviation from the average value of base haplotype were obtained together with standard errors and p-values.

3. RESULTS

3.1. SNP marker segregation in different horse breeds

To obtain the broader view of how the most associated best race distance SNP (Hill et al., 2010a) and 3 additional SNPs in *MSTN* gene region are segregating, horse sample of more than 1000 individuals belonging to 25 different breeds were genotyped. As we couldn't determine absolutely all genotypes,

missing alleles of markers (2% of data) were inferred by PHASE.2.1.1. software using information of other known marker alleles of individual and population data. Genotype and allele frequencies in all 4 SNP positions are shown in Appendix 2 and general information on marker segregation is summarized in Table 3.

In a whole sample we were able to detect segregation and homozygote state for minor allele was present in all markers, except PR5696, where in some genotyping attempts two separated groups of heterozygotes were observed (Figure 2).

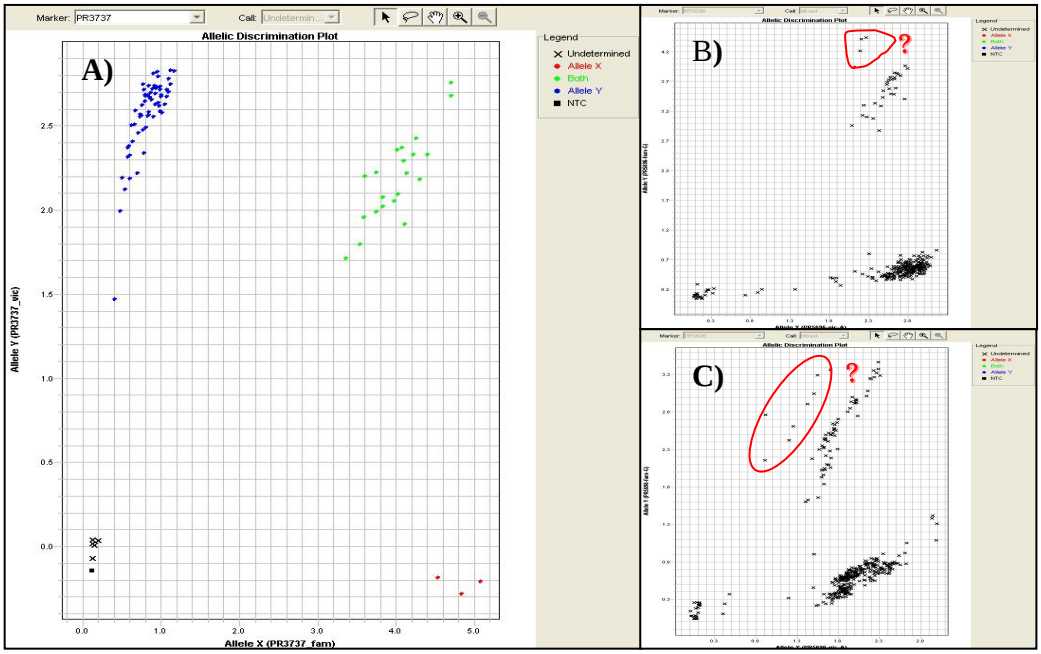


Figure 2. TaqMan genotyping plots: A, Marker PR3737; B and C, Marker PR5696.

From the figure 2 we can see that all three genotypes were detected for marker PR3737 - homozygous individuals are placed along the axis (colored in blue and red) and heterozygotes are located in the middle (green). Only two genotypes were detected for marker PR5696. Furthermore, heterozygotes seem to form two groups, where one is showing higher values for G allele than the other.

The most associated best race distance SNP (PR3737) is segregating in 18 breeds of 25 tested with MAF≥5% and is fixed only in 5 breeds. The g.66493737C allele has been detected in all tested Nordic breeds, except Faeroe Island Horse. Marker PR8604, which is located ~600 kb downstream the *MSTN* gene is segregating in 80% of analyzed breeds. The only breeds where there was no g.65868604G allele identified were Gotlands pony, Faeroese Iceland Horse and French Trotters. Segregation of both promoter region SNP markers (PR5826 and PR5696) is relatively less common. With MAF≥5% segregation has been observed only in the half of analyzed breeds. There are 7 breeds in which all four SNP markers are polymorphic: Swedish Ardennes, Icelandic Horse, Bashkir, Norwegian Fjord Horse, Nordlands Pony, as well as Mini and Mongolian Horses. However the

sample size of Minis and Mongolian horses is so small, that even one occurrence of rare allele results in MAF≥5%.

Table 3: Overview of marker segregation in 25 different horse breeds

Breed	N	PR5826		PR5696		PR3737		PR8604	
		Alleles	$P_{(G)}$	Alleles	$P_{(G)}$	Alleles	$P_{(C)}$	Alleles	$P_{(G)}$
NSD	41	A/G	0.01	A/G	0.07	T/C	0.21	T/G	0.29
Sw Ardennes	43	A/G	0.10	A/G	0.28	T/C	0.20	T/G	0.22
Dølehest	16	A	0.00	A	0.00	T/C	0.31	T/G	0.16
SW	65	A/G	0.01	A/G	0.03	T/C	0.16	T/G	0.10
Gotlands Pony	32	A	0.00	A/G	0.02	T/C	0.47	T	0.00
Icelandic Horse	318	A/G	0.30	A/G	0.05	T/C	0.09	T/G	0.20
Shetlands Pony	69	A/G	0.08	A	0.00	T/C	0.43	T/G	0.11
Bashkir	20	A/G	0.08	A/G	0.12	T/C	0.06	T/G	0.10
FIH	14	A/G	0.21	A/G	0.04	T	0.00	T	0.00
N. Fjord	32	A/G	0.05	A/G	0.17	T/C	0.28	T/G	0.16
Connemara	12	A	0.00	A/G	0.33	T	0.00	T/G	0.38
Welsh Pony	21	A	0.00	A/G	0.08	T	0.00	T/G	0.21
Nordlands Pony	22	A/G	0.08	A/G	0.25	T/C	0.85	T/G	0.23
Thoroughbred	42	A/G	0.01	A/G	0.04	T/C	0.79	T/G	0.44
CBT	129	A/G	0.17	A/G	0.16	T/C	0.02	T/G	0.05
French Trotters	41	A/G	0.02	A	0.00	T	0.00	T	0.00
Standardbred	96	A/G	0.04	A/G	0.03	T/C	0.01	T/G	0.01
Arab Horse	28	A	0.00	A/G	0.11	T	0.00	T/G	0.09
KY	23	A/G	0.09	A/G	0.04	T/C	0.07	T/G	0.16
Rocky Mountain	17	A/G	0.13	A	0.00	T/C	0.21	T/G	0.06
Paso Fino	25	A/G*	0.04	A	0.00	T/C	0.22	T/G	0.02
T. Walker	20	A/G	0.05	A/G	0.03	T/C	0.18	T/G	0.28
MFT	17	A	0.00	A/G	0.03	T/C	0.06	T/G	0.13
Miniature Horse	8	A/G	0.25	A/G	0.06	T/C	0.13	T/G	0.13
Mongolian	11	A/G	0.09	A/G	0.05	T/C	0.05	T/G	0.05

Breed abbreviations: NSD-North Swedish Draft; SW-Swedish Warmblood; FIH-Faeroe Iceland Horse; KY-Kentucky Mountain Saddle Horse; CBT-Coldblooded Trotters; N. Fjord – Norwegian Fjord; T. Walker – Tennessee Walker; MFT – Missuri Fox Trotter

The highest frequency of CC homozygotes (not considering Thoroughbreds) is in Gotlands (0.19) and Shetlands (0.18) Pony samples, followed by Norwegian Fjord horse (0.13). Homozygous state of g.65868604G is even more rare, where the highest frequency of GG genotype has been observed in Thoroughbred sample (0.15) and followed by Tennese Walker sample (0.10). Nordic breeds with the highest occurrence of g.65868604G homozygote where Nordlands Pony (0.08) and North Swedish Draft (0.07). Regarding two promoter region SNPs it is interesting to note that only the g.66495826G allele has been observed in homozygous state in 6 populations and the frequency of this genotype varies from 0.02 in Coldblooded Trotters to 0.13 in Minihorse.

Hardy-Weinberg deviation test (full results in Appendix 3) showed that deviations from HWE in most of the breeds and markers are not significant (p-val>0.05), with exception of Thoroughbreds in PR3737 marker (p-val<0.002), Ardennes in PR5826 marker (p-val<0.05) and PR5696 marker (p-val<0.01), Coldblooded trotters in PR5696 marker (p-val<0.05), Bashkir in PR5826 (p-val<0.01) and Connemara in PR8604 (p-val<0.05) marker position. The Total sample Deviation from HWE is significant in all tested markers, but the highest χ^2 value (49.42) is for Pr5696 marker.

All SNP markers were phased using the PHASE2.1.1. program (Table 4).

Table 4. Haplotype frequencies of genotyped breeds

Breed	n	TTAA	GTAA	TCAA	TTGA	TTAG	GCAA	GTGA	GTAG	TCAG	TCGA	GCGA
NSD	41	0.40	0.29	0.22	0.07	0.01	0	0	0	0	0	0
Dole	16	0.53	0.16	0.31	0	0	0	0	0	0	0	0
Ardennes[?]	43	0.53	0.01	0.11	0.08	0.07	0	0.12	0.01	0	0	0.08
SW	65	0.71	0.10	0.16	0.02	0.01	0	0	0	0	0	0
Gotlands	32	0.52	0	0.47	0.02	0	0	0	0	0	0	0
Shetlands	69	0.49	0	0.35	0	0.07	0.09	0	0.01	0	0	0
Bashkir[?]	20	0.78	0.05	0.03	0.07	0.04	0	0.04	0	0	0	0
Faeroese	14	0.75	0	0	0.04	0.21	0	0	0	0	0	0
Fjord[?]	32	0.42	0.08	0.20	0.17	0.05	0.08	0	0	0	0	0
Connemara	12	0.33	0.38	0	0.29	0	0	0	0	0	0	0
Welsh	21	0.74	0.19	0	0.07	0	0	0	0	0	0	0
Nordlands	22	0.32	0.23	0.16	0.25	0.05	0	0	0	0	0	0
TB	42	0.14	0.02	0.38	0.01	0	0.40	0	0	0.01	0.02	0
French	41	0.98	0	0	0	0.02	0	0	0	0	0	0
CBT	129	0.60	0.05	0.02	0.16	0.17	0	0	0	0	0	0
SB	96	0.84	0.02	0.01	0.07	0.07	0	0	0	0	0	0
Icelandic	318	0.39	0.18	0.09	0.04	0.29	0	0	0	0	0	0
Arabs	28	0.80	0.09	0	0.11	0	0	0	0	0	0	0
KY	23	0.65	0.15	0.07	0.04	0.09	0	0	0	0	0	0
RH	17	0.68	0.06	0.15	0	0.12	0	0	0	0	0	0
Paso Fino	25	0.72	0.02	0.22	0	0.04	0	0	0	0	0	0
TW[?]	20	0.53	0.23	0.13	0	0.08	0.05	0	0	0	0	0
MF	17	0.76	0.12	0.06	0.03	0.03	0	0	0	0	0	0
Mini[?]	8	0.56	0	0.13	0.06	0.13	0	0	0.13	0	0	0
Mongolian	11	0.77	0.05	0.05	0.05	0.09	0	0	0	0	0	0

[?] Breeds with uncertain phasing; NSD - North Swedish Draft; SW - Swedish Warmblood; TB – Thoroughbred; CBT – Coldblooded Trotter; SB – Standardbred; KY - Kentucky Mountain Horse; RH - Rocky Mountain Horse; TW – Tennessee Walker; MF – Missouri Foxtrotter;

Since marker PR8604 is located relatively far from the other three markers, it was not possible to infer haplotypes with a certainty for some breeds.

The phasing analysis provided five common haplotypes – one wild type (TTAA) and a separate haplotype per each SNP mutation (GTAA; TCAA; TTGA and TTAG). The most frequent in all breeds, except Thoroughbreds, is TTAA haplotype, which frequencies varies from 0.32 in Nordlands Pony to 0.98 in French Trotters. Haplotype TCAA has been observed in 20 of analyzed breeds and its frequency varies from 0.01 in Standardbreds to 0.47 in Gotlands Pony. There is a trend that breeds tend to have high frequency of only one promoter region haplotype. One exception is Coldblooded trotter sample, where frequencies of both promoter region haplotypes are almost equal. Connemara and Nordlands Pony has the highest frequency of TTGA haplotype.

Haplotypes that have had at least one recombination event between markers are less frequent. The most remarkable is Thoroughbred population which has a considerable GCAA haplotype frequency. As distance between PR8604 and PR3737 is approximately 600 kb, recombination event between them is not unusual. The only breed where we can certainly confirm the presence of this haplotype is Shetlands pony (homozygous individuals for both positions were detected) which taking in account the relatedness among Shetlands pony and Thoroughbred (Bower, 2012) could be reasonably explained. For Fjord Horses and Tennessee Walkers we can't assure with certainty that this haplotype is present. The same situation is for GTGA and GCGA haplotype in Ardennes.

3.2. Detection of Ins227

First, to make sure that it is possible to detect the Ins227 allele, we analyzed 29 Thoroughbred samples and we were able to detect all possible variants of genotypes (Figure 3).

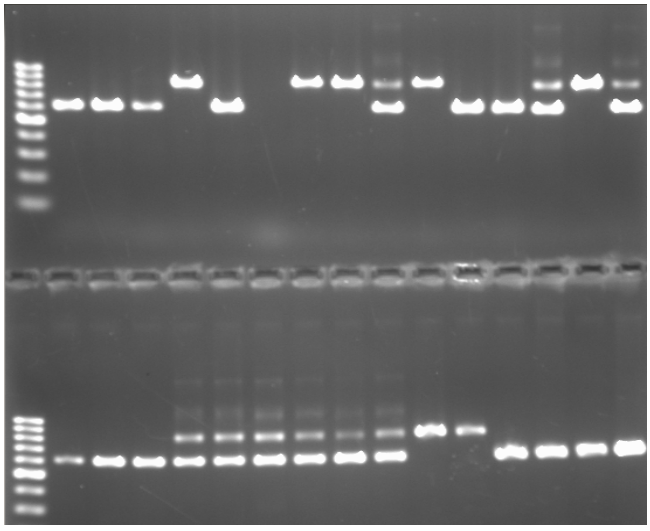


Figure 3 Ins227 polymorphism in Thoroughbred Horses

In 29 individuals we detected 23 Ins227 alleles, both in heterozygous (9 individuals) and homozygous state (7 individuals). After analyzing 100 samples of Icelandic Horses, no Ins227 alleles were detected (Figure 4).

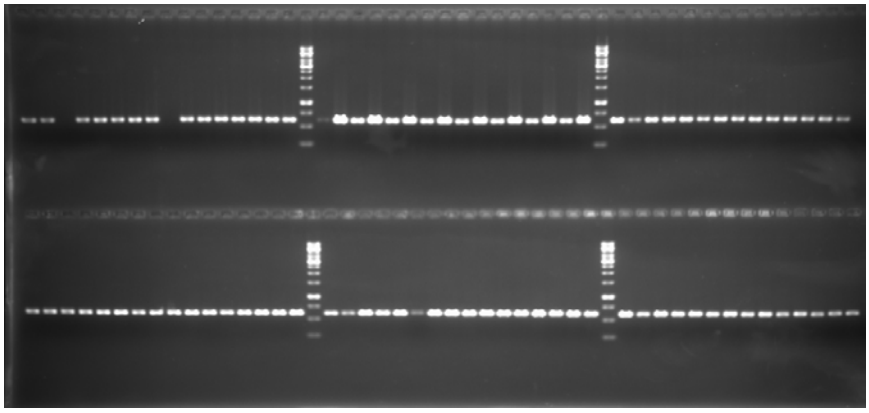


Figure 4 Absence of Ins227 in Icelandic Horse individuals

One heterozygous animal with Ins227 allele was detected in Swedish Warmblood horse (25 individuals tested). However we didn't detect Ins227 allele in the sample size of 55 individual of different breeds: Swedish Ardens (7), North Swedish Draft (7), Dole (4), Shetlands (4), CBT (14), Arabs (9) and SB (10). Thus, in following association studies this polymorphism is not used.

2.3. Phenotypic associations

2.3.1. Comparison of allele and haplotype frequencies between groups

Genotyped horse breeds were grouped in four different classes depending on what is the main aptitude of the breed. Accordingly: heavy draft (Swedish Ardennes, North Swedish Draft and Dølehest), light draft/riding (Swedish Warmblood, Gotland's Pony, Shetlands Pony, Bashkir, Faeroe Island Horse, Norwegian Fjord, Connemara, Welsh and Nordalnds Pony), racing (Coldblooded Trotters, Standardbred, Thoroughbred and French trotters) and long distance riding (Icelandic Horse, Arabian Horse, Kentucky Mountain Saddle Horse, Paso Fino, Missouri Fox Trotter, Rocky Mountain Horse and Tennessee Walker) . The mean values of minor allele frequencies of each group were calculated (Table 5)

Table 5. Mean values of minor allele frequencies for grouped breeds

Breed group	PR5826G	PR5696G	PR3737C	PR8604G
Heavy draft	0.03	0.11	0.25	0.23
Light draft/Riding	0.06	0.12	0.19	0.15
Racing	0.06	0.06	0.20	0.12
Long dist. riding	0.09	0.04	0.12	0.13

If we compare the allele frequencies of rare alleles, it seems that Heavy draft horses tend to have higher values (except in marker PR5826). However, the variance within groups are much larger than variance between groups and we can say that there is no significant difference (p-value>0.05) between them.

As for some of the breeds it was impossible to infer the most likely haplotype of all four SNPs, only the three closest markers were phased to see the group differences of haplotypes, accordingly PR3737, PR5696 and PR5826. In haplotype level, differences of haplotype frequencies among the breed groups have been observed (Table 6). For example, frequency of the wild type allele in horses that are bred for long distance riding is higher, but the frequency of haplotype TGA is lower than for those of other groups. Haplotype with the most associated SNP with best race distance in Thoroughbred horses (Hill et al., 2010) appears to be more common in heavy draft horses and the high frequency of racing horse group is only due to Thoroughbreds. And these results may support the hypothesis that this mutation is connected to explosiveness of muscle.

Table 6. Mean values of haplotype frequencies of different breed groups

Breed group	TAA(1)	CAA(2)	TAG(3)	TGA(4)	CGA(5)	CAG(6)
Heavy draft	0.61	0.24	0.04	0.11	0.01	0.00
Light draft/ riding	0.67	0.17	0.05	0.11	0.00	0.00
Racing	0.68	0.20	0.06	0.05	0.01	0.01
Riding long distances	0.77	0.11	0.09	0.04	0.00	0.00

However, as variance within groups is still larger than between groups, we can't claim that there is a statistically significant difference (p-value>0.05).

The sample of analyzed Icelandic horses consisted of both horses selected for riding ability (189individuals) and those also used for meat and hormone production (45 individuals). As horses come from the same population, we would expect that any difference in allele frequencies between groups we spot might be due to different breeding goals and selection strategies. Minor allele frequencies are summarized in Table 7.

Table 7. Different minor allele frequencies in horses selected for Riding/Racing and meat production

	PR5826G	PR9656G	PR3737C	PR8604G
Riding	0.30	0.01	0.09	0.20
Meat	0.51	0.16	0.02	0.15

We can see that horses bred for meat production has higher minor allele frequencies in both promoter SNP markers (PR5826 and PR9656), but horses selected for riding has slightly higher frequencies in

rest of the markers (PR3737 and PR8604). However, the majority of horses possessing PR9656G allele come from the same farm and thus a possible high frequency of this allele is due to relatedness.

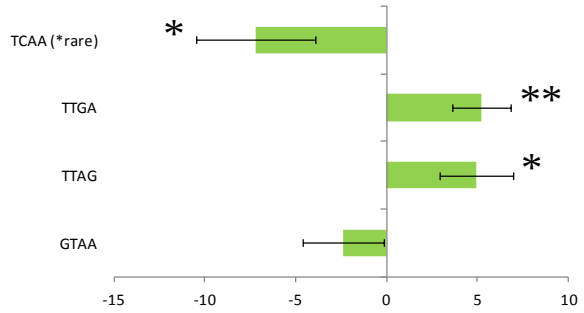
2.3.2. Associations with phenotypic data

Data for phenotypic correlations were available for Icelandic Horses, Coldblooded trotters and Standardbred horses. Luckily, Icelandic horses and Coldblooded trotters were polymorphic in all tested SNP loci and thus could be correlated with phenotypic data. In total genotype information from 186 Icelandic horses and from 73 Coldblooded Trotters were linked with phenotypic data values. Even though there was some segregation detected in Standardbred sample as well, the MAF did not exceed 5% in any of the four loci.

Field evaluation scores and according BLUP values were associated with inferred haplotypes by use of generalized linear model and haplo.stats package in R software. Results of what provided us with coefficient describing the effect of haplotype in comparison with base haplotype, standard error, t-statistic and p-value (Appendix 4). One significant association for Icelandic horses was detected for both BLUP and direct value for 'Proportions' trait (p-value=0.014), where haplotype TTAG has a slight negative effect (-2.1) on a trait in comparison to base haplotype (TTAA). The only performance trait that has an acceptable significance level in association with haplotypes was BLUP value for Tolt (ambling gate of Icelandic Horse), where GTAA haplotype has a slight positive effect (2.4) over the base haplotype (p-value=0.047).

For Coldblooded Trotters, first direct data of 'Start position', 'Position in competition', 'Earnings' and two different time measures were used for haplotype association. No significant effect was detected (Appendix 5). As the number of horses having direct values was very small and these values are influenced by number of environmental factors we have no information about it was not possible to capture the difference in haplotype effects.

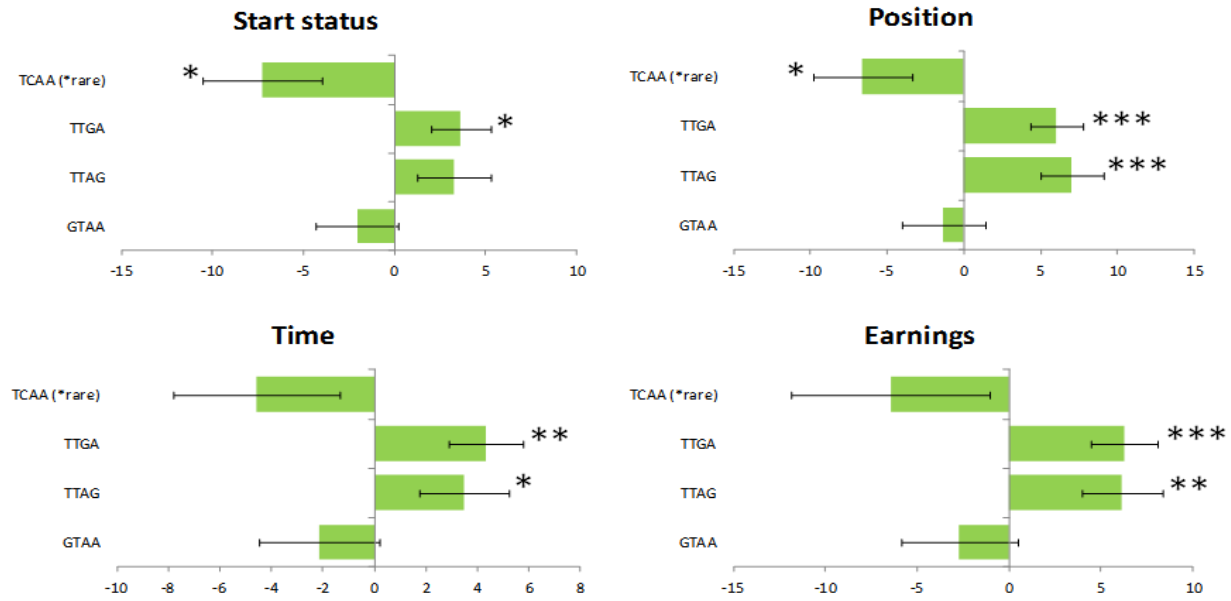
Further, BLUP values of 'Start status', 'Position in competition', 'Earnings', 'Time' and the total BLUP value was associated with haplotypes (Appendix 5). In this case numerous significant haplotype effects were detected, including the BLUP value for the total score (Figure 5).



* p-value < 0.05, **p-value < 0.01, 0 value on x-axis – base haplotype (TTAA), effect expressed in points of BLUP value

Figure 5. The effects of haplotypes on the Total BLUP value of Coldblooded Trotters.

The most significant effect on total score has been observed for TTGA haplotype (5.2) which contains G allele in SNP PR5696. However the largest deviation from base haplotype (TTAA) is for the rare haplotype (TCAA). This haplotype is not common in CBT and have been linked with phenotypic values only for 2 horses. Effects of haplotypes on other performance traits are shown in Figure 6.



* p-value < 0.05, **p-value < 0.01, ***p-value=0.001

Figure 6. The effects of haplotypes on the Performance trait BLUP values of CBT.

Haplotype TTGA keeps its significant effect for all of them, reaching p-value of 0.001 for 'Position in competition' and 'Earnings'. Haplotype TTAG has remained significant in all traits except 'Start status', but in general it has slightly smaller deviation from base haplotype and higher p-values. Rare haplotype has still significant effect on Start status and Position in competition.

4. DISCUSSION

According to Bower et al. (2010) the most associated SNP PR3737 has originated from a native British mare at the foundation of Thoroughbred population. Knowing the common history of British Isles and Nordic countries (during Viking era) we hypothesized that Nordic Horse breeds might be carrying the mutant allele in *MSTN* gene region. After screening several Nordic and British Isles horse breeds we observed that in the majority of Nordic breeds marker PR3737 is segregating with MAF>5%, which confirm our hypothesis. The fact that this mutation is present in old and relatively pure breeds as Icelandic and Fjord horses, and Nordlands and Gotlands Pony confirms g.66493737C being an old mutation. In addition, considering the high number of Nordic breeds that are segregating

in this position, we can guess that the frequency of mutant allele was considerably high when Nordic breed populations were formed. Absence of this mutation in Faroese Island Horse, which is known to be one of the purest breed that has not been altered by crossbreeding (Hendricks, 1995), might be explained by the sever bottle neck event in 1960 (Faroe Pony, Wikipedia, free encyclopedia).

Our results are to some extent consistent with those described by Bower et al. (2012) having some slight differences. Results from this study support the previous statement of Bower et.al. that PR3737T allele is a wild type allele, as it was fixed in all tested Donkeys. However, we couldn't detect PR3737C allele in Connemara and Arabian horse samples. The frequency of mutant allele in Icelandic horse sample of our study is twice as small (0.09) than that of Bowers (0.17), but our sample size is markedly larger and thus the estimate can be regarded as more precise. We detected similar allele frequencies for Shetlands pony with somewhat lower frequency of homozygous CC individuals. Results for Standardbred and French trotter horses are consistent in both studies. The frequency of the downstream gene marker PR8604 hasn't been described in other breeds than Thoroughbreds and this is therefore new information. Two promoter markers have previously been described only in the Dall'Ollio et al. (2010) study, but the breeds tested differ from those in our study, except Thoroughbreds. They have similar allele frequencies with those in our study for both markers. One can point out that frequencies in both markers are alike for Italian Trotters and French Trotters, as well as for Noric and Swedish Ardennes.

Since horse groups of different aptitude has very variable allele frequencies it is hard to draw any conclusions from the comparison. Allele frequencies of markers can be easily influenced by the history of a particular population. Horse breeds like Icelandic, Faroe Island horse and Nordlands Pony has suffered more or less severe bottleneck events. Others like, Dole and North Swedish draft horse, have been extensively crossbred. Allele frequencies within groups can be easily influenced by other factors than the main aptitude and there are breeds that tend to be more universal type than others.

It is interesting that Gotlands pony has a remarkable segregation in PR3737 loci, but it is completely fixed in marker PR5826 and PR8604 and has MAF<0.05 in PR5696 locus. According to Hendricks (1995) Gotlands pony has descended from the ancient northern European forest horse and has existed on Gotlands Island since Stone Age. To improve the breed two Wales pony stallions have been crossed into it, but the pedigree was closed in 1971 after which only registered Gotland ponies have been accepted (Gotlands Pony, Wikipedia, free encyclopedia). It suggests that mutation in PR3737 locus might be coming from northern European forest horse and mutations in other markers occurred later or were not present at the foundation population. Our genotyped sample of Welsh pony is fixed for T allele in PR3737 locus, but has a 0.08 frequency of G allele in PR5696 locus, and this might explain the minor segregation of PR5696 locus in Gotlands Pony.

Marker PR5696 seems to be intriguing. In all the tested breeds the g.66495696G allele has only been observed in the heterozygous state (more than 1000 horses tested), causing a severe deviation from Hardy Weinberg equilibrium for this marker. For example, in Swedish Ardennes population the g.66495696G allele (0.28) is almost three times more frequent than the g.66495826G allele (0.10),

which gives 5% homozygous individuals. The uncertainty is whether there is no homozygous state of this allele or we have failed to identify it. According to Dall'Olio et al. (2010) marker PR5696 is located within a TATA box like motif in a proximal promoter region and it has been identified in homozygous state in a few Bardigiano, Haflinger, Noric, Rapid Heavy Draft and Uruguayan Creole horses (in total 396 horses tested), which has been classified as brachymorphic and mesomorphic breeds (both powerful, compact and with massive muscling). The similar situation, where no homozygous individuals were detected with TaqMan approach, was described by Sundström et al. (2010). Using other methods like Sanger sequencing and PCR approach they were able to prove that lack of identified homozygotes is because tested SNP was representing a second transposed copy present somewhere else in the genome. However, we cannot claim that to be true in our research without further investigation and we have to consider that lack of homozygous individuals in tested populations might also be due to lethality of phenotype.

Previous studies (Botstein&Risch,2003; Lin&Zeng, 2006) have shown that in complex disease mapping methods which are using haplotype information result in more power and accuracy. Performance traits can be looked at in a similar way, because they also are influenced by many genetic and environmental factors. However, even after clustering all four SNP markers there were no performance trait data of Icelandic horse that could be associated with genotype information with p-value < 0.01. Icelandic horse is known as universal horse – famous for its strength and endurance, but nowadays being used not only for long distance riding, but also for five-gated competition, as a family horse, and in dressage (Hendricks, 1995) thus having very broad breading purpose.

Coldblooded trotters have been selected for trotting since 1924 and are known as exceptionally talented for trotting and are very fast (Hendricks, 1995). We managed to capture significant associations of some haplotypes with BLUP measures of performance traits. Better results were obtained when correlating BLUP values rather than direct trait measurements, because BLUP value is already corrected for environmental effects, which is not the case with direct observations. The fact that we could not strongly associate performance traits of Icelandic horse, but succeeded in Coldblooded trotters, could be explained with differences in breeding work. Coldblooded trotters are bred and strongly selected for their success on the racing track (earnings, time, position), but Icelandic horses are more of a universal riding horse that is not strongly selected for best race time, success in competition or earnings. Results in Coldblooded Trotters suggest that both promoter region markers may contribute to the variability of phenotype. According to the Haplo Stats manual (Sinnwell and Schaid, 2011) it is not advised to make any inference of effect of rare haplotype as it may yield misleading results. Finally, we cannot claim that we have found the causative SNP that is influencing the racing ability, but we have a strong indication that further research of these haplotypes is necessary.

The haplotype that has the strongest association in Coldblooded trotters has a mutation in PR5696 position. It is interesting to note that this allele has been absent in all Icelandic horses that have performance data. However, if we compare horses bred for riding and those bred for meat and hormone production, the frequency of G allele in PR5696 marker is remarkably higher in meat

horses. It is important to say that horses from the meat farm harboring G allele in PR5696 position are born in one year, so that it is possible that those individuals are offspring from one stallion. Unexpectedly, these results lead us to discovery of a family where PR5696G allele being passed through generations starting from a mare named Skessa born in 1986. It is known that heterozygouse stallion (grandson of Skessa) in PR5696 marker has been used in a farm where high PR5696G allele frequency has been spotted.

In this study we couldn't prove any significant association of performance traits and the most associates SNP of Thoroughbred horse (Hill et al., 2010a). For further investigation of PR3737 marker and its influence on performance we should focus on association study in Shetlands and Gotlands pony breeds, as they both are segregating for PR3737 marker and Gotlands Ponies has been known as excellent trotters and used for racing (Hendrics, 1995).

A further investigation is necessary for the promoter region. In this case it would be useful to focus on breeds like Swedish Ardennes, Coldblooded Trotters, Connemara horses or Icelandic horses of previously described family as they have high frequency of PR5696G allele. First of all, Sanger sequencing of heterozygous individuals having PR5696G allele is essential to confirm true genotypes of two heterozygote groups. Followed by more detailed analysis of polymorphisms in both haplotypes associated with performance traits (TTGA and TTGA). To further understand mechanism how performance traits are influenced by a genotype, association of pulling data in Swedish Ardennes would help to investigate if any of haplotypes is associated with a pulling capacity.

In conclusions, we can say that tested polymorphisms in *MSTN* gene are widespread in Nordic Horse breeds. As well there are some indications that some of them might have a functional effect on horse performance traits and indicates directions for further research possibilities.

REFERENCES

Binns M.M., Boehler D.A. and Lambert D.H. (2010) Identification of the myostatin locus (*MSTN*) as having a major effect on optimum racing distance in the Thoroughbred horse in the USA. *Animal Genetics*, 41 (Suppl. 2), 154 – 158

Boman I.A., Våge D.I. (2009) An insertion in the coding region of the myostatin (*MSTN*) gene affects carcass conformation and fatness in the Norwegian Spaelsau (*Ovis aries*). *BMC Res Notes*, Jun 8;2:98.

Boman I.A., Klemetsdal G., Blichfeldt T., Nafstad O., Våge D.I. (2009) A frameshift mutation in the coding region of the myostatin gene (MSTN) affects carcass conformation and fatness in Norwegian White Sheep (*Ovis aries*). *Anim Genet*, Aug; 40 (4): 418-22.

Botstein D. and Risch N. (2003). Discovering genotypes underlying human phenotypes: Past successes for mendelian disease, future approaches for complex disease. *Nat. Genet*. 33 (Suppl.), 228–237.

Bjørnstad G., Nilsen N.Ø., Røed K.H. (2002) Genetic relationship between Mongolian and Norwegian horses? *Animal Genetics*, 34, 55-58

Bower M.A., McGivney B.A., Campana M.G. et al. (2012) The genetics and history of speed in the Thoroughbred racehorse. *Nature communications*, 3:643, DOI: 10.1038/ncomms1644 Available from http://www.nature.com/ncomms/journal/v3/n1/full/ncomms1644.html [16 June 2012]

Clop A., Marcq F., Takeda H., Pirottin D., Tordoir X. et al. (2006) A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nature Genetics*, Vol. 38 (7) July, 813 – 818

Dall'Olio S., Fontanesi L., Costa L.N., Tassinari M., Minieri L., Falaschini A. (2010) Analysis of Horse *Myostatin* Gene and Identification of Single Nucteotide Polymorphisms in Breeds of Different Morphological Types. *Journal of Biomedicine and Biotechnology*, Volume 2010, Article ID 542945, 11 pages

Gaffney B. and Cunningham E.P. (1988) Estimation of genetic trend in racing performance of thoroughbred horses. *Nature*, Vol. 332 (21) April, 722-724

Hagenblad J., Tang C.L., Werner J.M.J., Zhao K. , Zheng H.G., Marjoram P., Weigel D., and Nordborg M. (2004) Haplotype structure and phenotypic associations in the chromosomal regions surrounding two *Arabidopsis thaliana* flowering time loci. *Genetics*, 168:1627–1638.

Hendricks, B. (1995) International Encyclopedia of Horse Breeds. University of Oklahoma Press: NORMAN

Hill E.W., Gu J., Eivers S.S., Fonseca R.G., McGivney B.A., Govindarajan P., Orr N., Katz L.M., MacHugh D. (2010a) A Sequence Polymorphism in MSTN Predicts Sprinting Ability and Racing Stamina in Thoroughbred Horses. *PLoS ONE* 5(1): e8645. doi;10.1371/journal.pone.0008645

Hill E.W., Mcgivney B. A., Gu J., Whiston R., and Machugh D. E. (2010b). A genome-wide SNP-association study confirms a sequence variant (g . 66493737C > T) in the equine myostatin (MSTN) gene as the most powerful predictor of optimum racing distance for Thoroughbred racehorses A genome-wide SNP-association study confirms. *BMC Genomics*, 552 (October).

Huang Z., Chen X., Chen D. (2011) Myostatin: A novel insight into its role in metabolism, signal pathways, and expression regulation. *Cellular Signalling*, 23, 1441-1446

Jeanplong F., Sharma M., Somers W.G., Bass J.J., Kambadur R. (2001) Genomic organization and neonatal expression of the bovine myostatin gene. *Molecular Cell Biochemistry*, 220 (1-2) April, 31-7

Levine M. A. (2005) Domestication and early history of the horse in The Domestic Horse (eds. Mills, D. & McDonnell, S.) 5-22 (Cambridge University Press, London)

Lin D. Y. and Zeng D. A. (2006) Likelihood-Based Inference on Haplotype Effects in Genetic Association Studies. *Journal of the American Statistical Association*, Vol. 101, No. 473, (March) Theory and Methods.

Liu N., Zhang K., Zhao H., Haplotype-Association Analysis. From: *Advances in Genetics*, Vol. 60 Copyright 2008, Elsevier Inc. All rights reserved.

McCroskery S., Thomas M., Maxwell L., Sharma M., Kambadur R. (2003) Myostatin negatively regulates satellite cell activation and self-renewal. *Journal of Cell Biology*, 162:1135-1147

McPherron A.C., Lawier A.M., Lee S.J. (1997) Regulation of skeletal muscle mass in mice by a new TGF-β superfamily member. *Nature*, Vol. 387 1 May, 83 - 90

McPherron A.C., Lee S.J. (1997) Double muscling in cattle due to mutations in the myostatin gene. *Proceedings of the National Academy of Sciences of the United States of America*, Vol.94. pp. 12457-12461

Mosher D.S., Quignon P, Bustamante C.D., Sutter N.B., Mellersh C.S., et al. (2007) A Mutation in the Myostatin Gene Increases Muscle Mass and Enhances Racing Performance in Heterozygote Dogs. *PLoS Genet*, 3(5): e79. doi:10.1371/journal.pgen.0030079

Sinnwell J.P. and Schaid D.J. (2011) Haplo Stats (version 1.5.0). Statistical Methods for Haplotypes When Linkage Phase is Ambigouse. Mayo Clinic Division of Health Sciences Research: Rochester MN USA. 62 p. Available from http://cran.r-project.org/web/packages/haplo.stats/haplo.stats.pdf

Schaid, D. J. (2004). Evaluating associations of haplotypes with traits. *Genet. Epidemiol.* 27, 348–364.

Schröder W., Klostermann A., Distl O. (2011) Candidate genes for physical performance in the horse. *The Veterinary Journal*, 190, 39-48

Schuelke M., Wagner K.R., Stolz L.E.et al. (2004) Myostatin Mutation Associated with Gross Muscle Hypertrophy in a Child. *The New England Journal of Medicine*, 350;26, 2683 – 2688

Stephens M., Smith N. J., Donnelly P. (2001) A new statistical method for haplotype reconstruction from population data. American Journal of Human Genetics 68, 978-989.

Stephens M. and Scheet P. (2005) Accounting for decay of linkage disequilibrium in haplotype inference and missing data imputation. *American Journal of Human Genetics*, 76, 449-462.

Tozaki T., Hill E. W., Hirota K., Kakoi H., Gawahara H., Miyake T., Sugita S. et al. (2012) A cohort study of racing performance in Japanese Thoroughbred racehorses using genome information on ECA18. *Animal genetics*, 43(1), 42-52. doi:10.1111/j.1365-2052.2011.02201.x

Tozaki T., Miyake T., Kakoi H., et al. (2010) A genome-wide association study for racing performances in Thoroughbred clarifies a candidate region near the MSTN gene. *Animal Genetics*, 41 (suppl.2), 28 - 35

Tozaki T., Sato F., Hill E. W., Miyake T., Endo Y., Kakoi H., Gawahara H. et al. (2011) Sequence variants at the myostatin gene locus influence the body composition of Thoroughbred horses. *The Journal of veterinary medical science* / the Japanese Society of Veterinary Science, 73(12), 1617-24. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21836385

Rodriguez S., Gaunt T.R. and. Day I.N.M. (2009) Hardy-Weinberg Equilibrium Testing of Biological Ascertainment for Mendelian Randomization Studies. *American Journal of Epidemiology Advance Access*, published on January 6

Wikipedia, The Free Encyclopedia. Faroe Pony. Available from: http://en.wikipedia.org/wiki/Faroe_pony [16 June 2012]

Wikipedia, The Free Encyclopedia. Gotlands Pony. Available from: http://en.wikipedia.org/wiki/Gotland_Pony [16 June 2012]

ACKNOWLEDGEMENTS

First and foremost my gratitude to all my supervisors Gabriella Lindgren, Lisa Andersson and Dag Inge Våge, who guided and supported me during my thesis project.

Thanks to Knut Røed, The Norwegian School of Veterinary Science, and Freyja Imsland, Department of Medical Biochemistry and Microbiology, Uppsala University, for valuable sample contribution and special thanks to Freyja for sharing her knowledge and interesting ideas.

My work would not be imaginable without all my colleges at IMBIM laboratory and EM-ABG friends, who always where there not only to help with practical advices and new ideas, but also to cheer up and make my days more fun.

Last but not the least, to my family and Sergio Vela Avitúa for believing in me and encouraging me to challenge my abilities.

APPENDIX

Appendix 1

Primer sequence information of four studied SNP markers

Position (Assay name)	Forward primer	Reverse primer
g.66493737T>C (PR3737)	TCCAGGACTATTTGATAGCAGAGTCA	GACACAACAGTTTCAAAATATTGTTCTCCTT
g.66495696A>G (PR5696)	CATACTTAAAAAAGTTGAAATGTTACTTCCTCAGAA	CAGCACTAAAAATAATTTAATACAACAAATACAATTCTTT
g.66495826A>G (PR5826)	CTGTATAATTAGGGAAATATAAATTTAAATTAGGAAGACCTG	GTCAGGAAAACAAGTTTCTCAAATTATAGTTGAA
g.65868604T>G (PR8604)	GCAGTGATTTGTTCTGACCAAGACT	TGAGATAATGCAGGCAGGAAAAGAA

Designed probes for allele discrimination of four studied SNP markers

Position (Assay name)	Probe for allele 1 (VIC)	Probe for allele 2 (FAM)
g.66493737T>C (PR3737)	ATGCACCAAATAATTT	ATGCACCAAGTAATTT
g.66495696A>G (PR5696)	TGTTTGCCTAAATAATATAAAAT	TTGCCTAAATAACATAAAAT
g.66495826A>G (PR5826)	ACCTTTTACTAGTATCACAATC	CTTTTACTAGTACCACAATC
g.65868604T>G (PR8604)	ACCCATACTTAACATACATG	ACCCATACTTAAAATACATG

Appendix 2

Genotype and allele frequencies at 4 SNP loci in the MSTN gene region

		PR3737					PR8604					
		Genotype freq.			Allele freq.			Genotype freq.			Allele freq.	
Breeds / species	**Nr**	**CC**	**CT**	**TT**	**C**	**T**	**Nr**	**TT**	**TG**	**GG**	**T**	**G**
NSD	42	0.07	0.29	0.64	0.21	0.79	41	0.49	0.44	0.07	0.71	0.29
Swedish Ardennes	43	0.07	0.26	0.67	0.20	0.80	45	0.56	0.44	0.00	0.78	0.22
Dølehest	16	0.13	0.38	0.50	0.31	0.69	16	0.69	0.31	0.00	0.84	0.16
SW	63	0.03	0.25	0.71	0.16	0.84	67	0.82	0.15	0.03	0.90	0.10
Gotlands Pony	32	0.19	0.56	0.25	0.47	0.53	32	1.00	0.00	0.00	1.00	0.00
Shetlands Pony	65	0.18	0.49	0.32	0.43	0.57	68	0.81	0.16	0.03	0.89	0.11
Bashkir	16	0.00	0.13	0.88	0.06	0.94	35	0.83	0.14	0.03	0.90	0.10
FIH	14	0.00	0.00	1.00	0.00	1.00	13	1.00	0.00	0.00	1.00	0.00
Fjord Horse	32	0.13	0.31	0.56	0.28	0.72	31	0.71	0.26	0.03	0.84	0.16
Connemara	12	0.00	0.00	1.00	0.00	1.00	12	0.25	0.75	0.00	0.63	0.38
Welsh Pony	14	0.00	0.00	1.00	0.00	1.00	19	0.58	0.42	0.00	0.79	0.21
Nordlands Pony	24	0.04	0.21	0.75	0.15	0.85	24	0.63	0.29	0.08	0.77	0.23
Thoroughbred	36	0.78	0.03	0.19	0.79	0.21	34	0.26	0.59	0.15	0.56	0.44
CBT	129	0.00	0.02	0.96	0.02	0.98	132	0.91	0.09	0.00	0.95	0.05
French Trotters	42	0.00	0.00	1.00	0.00	1.00	45	1.00	0.00	0.00	1.00	0.00
Standard Bread	92	0.00	0.01	0.99	0.01	0.99	80	0.98	0.03	0.00	0.99	0.01
Arab Horse	28	0.00	0.00	1.00	0.00	1.00	27	0.81	0.19	0.00	0.91	0.09
Icelandic Horse	313	0.00	0.17	0.83	0.09	0.91	313	0.64	0.34	0.03	0.80	0.20
KY	23	0.00	0.13	0.87	0.07	0.93	22	0.68	0.32	0.00	0.84	0.16
Rocky Mountain H.	12	0.00	0.42	0.58	0.21	0.79	16	0.88	0.13	0.00	0.94	0.06
Paso Fino	25	0.04	0.36	0.60	0.22	0.78	25	0.96	0.04	0.00	0.98	0.02
Tennese Walker	19	0.00	0.37	0.63	0.18	0.82	20	0.55	0.35	0.10	0.73	0.28
Missuri Fox Trotter	17	0.00	0.12	0.88	0.06	0.94	15	0.73	0.27	0.00	0.87	0.13
Mini Horse	8	0.00	0.25	0.75	0.13	0.88	8	0.75	0.25	0.00	0.88	0.13
Mongolian Horse	11	0.00	0.09	0.91	0.05	0.95	10	0.90	0.10	0.00	0.95	0.05
Donkey	4	0.00	0.00	1.00	0.00	1.00	4	1.00	0.00	0.00	1.00	0.00
Przewalski's Horse	4	0.00	0.00	1.00	0.00	1.00	4	1.00	0.25	0.00	0.88	0.12

Appendix 2 continued

		PR5826						PR5696				
		Genotype freq.			Allele freq.			Genotype freq.			Allele freq.	
Breeds / species	Nr	AA	AG	GG	A	G	Nr	AA	AG	GG	A	G
NSD	41	0.98	0.02	0.00	0.99	0.01	41	0.85	0.15	0.00	0.93	0.07
Swedish Ardennes	44	0.84	0.11	0.05	0.90	0.10	45	0.44	0.56	0.00	0.72	0.28
Dølehest	16	1.00	0.00	0.00	1.00	0.00	16	1.00	0.00	0.00	1.00	0.00
SW	66	0.98	0.02	0.00	0.99	0.01	64	0.94	0.06	0.00	0.97	0.03
Gotlands Pony	32	1.00	0.00	0.00	1.00	0.00	32	0.97	0.03	0.00	0.98	0.02
Shetlands Pony	66	0.83	0.17	0.00	0.92	0.08	66	1.00	0.00	0.00	1.00	0.00
Bashkir	18	0.89	0.06	0.06	0.92	0.08	34	0.76	0.24	0.00	0.88	0.12
FIH	14	0.64	0.29	0.07	0.79	0.21	14	0.93	0.07	0.00	0.96	0.04
Fjord Horse	32	0.91	0.09	0.00	0.95	0.05	32	0.66	0.34	0.00	0.83	0.17
Connemara	12	1.00	0.00	0.00	1.00	0.00	12	0.33	0.67	0.00	0.67	0.33
Welsh Pony	13	1.00	0.00	0.00	1.00	0.00	19	0.84	0.16	0.00	0.92	0.08
Nordlands Pony	24	0.83	0.17	0.00	0.92	0.08	24	0.50	0.50	0.00	0.75	0.25
Thoroughbred	35	0.97	0.03	0.00	0.99	0.01	37	0.92	0.08	0.00	0.96	0.04
CBT	130	0.68	0.30	0.02	0.83	0.17	131	0.68	0.32	0.00	0.84	0.16
French Trotters	43	0.95	0.05	0.00	0.98	0.02	42	1.00	0.00	0.00	1.00	0.00
Standard Bread	83	0.93	0.07	0.00	0.96	0.04	93	0.94	0.06	0.00	0.97	0.03
Arab Horse	28	1.00	0.00	0.00	1.00	0.00	28	0.79	0.21	0.00	0.89	0.11
Icelandic Horse	312	0.50	0.41	0.09	0.70	0.30	314	0.91	0.09	0.00	0.95	0.05
KY	23	0.83	0.17	0.00	0.91	0.09	23	0.91	0.09	0.00	0.96	0.04
Rocky Mountain H.	15	0.73	0.27	0.00	0.87	0.13	17	1.00	0.00	0.00	1.00	0.00
Paso Fino	25	0.92	0.08	0.00	0.96	0.04	25	1.00	0.00	0.00	1.00	0.00
Tennese Walker	20	0.90	0.10	0.00	0.95	0.05	20	0.95	0.05	0.00	0.98	0.03
Missuri Fox Trotter	17	0.94	0.06	0.00	0.97	0.03	16	0.94	0.06	0.00	0.97	0.03
Mini Horse	8	0.63	0.25	0.13	0.75	0.25	8	0.88	0.13	0.00	0.94	0.06
Mongolian Horse	11	0.82	0.18	0.00	0.91	0.09	11	0.91	0.09	0.00	0.95	0.05
Donkey	4	1.00	0.00	0.00	1	0.00	4	1.00	0.00	0.00	1.00	0.00
Przewalski's Horse	4	1.00	0.00	0.00	1	0.00	4	1.00	0.00	0.00	1.00	0.00

Breed abriviations: NSD-North Swedish Draft; SW-Swedish Warmblood; FIH-Faeroe Iceland Horse; KY-Kentucky Mountain Saddle Horse; CBT-Coldblooded Trotters;

Heavy draft | Light draft/riding | Racing | Riding long distances

Appendix 3

SNP marker deviations from HWE in tested horse breeds

Marker	PR3737							
	Common Homozygosity		Heterozygosity		Rare Homozygosity		χ^2	p val
Breed	O	E	O	E	O	E		
TB	7	1.56	1	11.88	28	22.56	30.19	p<0.001
Arabs	28	NaN	0	NaN	0	NaN	NaN	NS
SW	45	44.59	16	16.83	2	1.59	0.15	NS
FT	42	NaN	0	NaN	0	NaN	NaN	NS
CBT	121	125.05	5	4.9	0	0.05	0.05	NS
NSD	27	25.93	12	14.14	3	1.93	0.96	NS
Ardennes	29	27.68	11	14	3	1.68	1.61	NS
Russ	8	9.03	18	15.94	6	7.03	0.54	NS
Icelandic	155	155.62	33	31.76	1	1.62	1.01	NS
Shetlands	21	21.06	32	31.88	12	12.06	0	NS
KY	20	20.1	3	2.8	0	0.1	0.11	NS
Mini	6	6.13	2	1.75	0	0.13	0.16	NS
RH	7	7.52	5	4	0	0.52	0.83	NS
Paso Fino	15	15.21	9	8.58	1	1.21	0.06	NS
TW	12	12.64	7	5.71	0	0.64	0.97	NS
MF	15	15.6	2	1.88	0	0.06	0.07	NS
Mongolian	10	10.02	1	0.95	0	0.02	0.02	NS
Dole	8	7.56	6	6.88	2	1.56	0.26	NS
Bashkir§	14	14.06	2.000	1.880	0.000	0.06	0.07	NS
Faeroese§	14	NaN	0	NaN	0	NaN	NaN	NS
Fjord§	18	16.53	10	12.94	4	2.53	1.65	NS
Connemara§	12	NaN	0	NaN	0	NaN	NaN	NS
SB	91	91	1	0.99	0	0	0	NS
Welsh§	14	NaN	0	NaN	0	NaN	NaN	NS
Nordlands	18	17.51	5	5.98	1	0.51	0.64	NS
Total	**757**	**644.4**	**181**	**195.27**	**63**	**55.86**	**21.63**	**p<0.001**

§ - samples that might consist of related individuals; NSD - North Swedish Draft; SW - Swedish Warmblood; TB – Thoroughbred; CBT – Coldblooded Trotter; SB – Standard Bread; KY - Kentucky Mountain Horse; RH - Rocky Mountain Horse; TW – Tennese Walker; MF – Missuri Foxtrotter; FT - French Trotter

O - observed amount; E - Expected amount

NS - deviation from HW equilibrium is not signifficant

Appendix 3 continued

Marker	PR5826							
	Common Homozygosity		Heterozygosity		Rare Homozygosity		χ^2	p val
Breed	O	E	O	E	O	E		
TB	34	34.01	1	0.99	0	0.01	0.01	NS
Arabs	28	NaN	0	NaN	0	NaN	NaN	NS
SW	65	65	1	0.99	0	0	0	NS
FT	41	41.02	2	1.95	0	0.02	0.02	NS
CBT	86	86.81	38	36.38	3	3.81	0.25	NS
NSD	40	40.01	1	0.99	0	0.01	0.01	NS
Ardennes	37	35.46	5	8.08	2	0.46	6.39	p<0.05
Russ	32	NaN	0	NaN	0	NaN	NaN	NS
Icelandic	92	91.98	79	79.04	17	16.98	0.18	NS
Shetlands	55	55.46	11	10.08	0	0.46	0.55	NS
KY	19	19.17	4	3.65	0	0.17	0.21	NS
Mini	5	4.5	2	3	1	0.5	0.89	NS
RH	11	11.27	4	3.47	0	0.27	0.36	NS
Paso Fino	23	23.04	2	1.92	0	0.04	0.04	NS
TW	18	18.5	2	1.9	0	0.05	0.06	NS
MF	16	16.01	1	0.97	0	0.01	0.02	NS
Mongolian	9	9.09	2	1.82	0	0.09	0.11	NS
Dole	16	NaN	0	NaN	0	NaN	NaN	NS
Bashkir§	16	15.12	1	2.75	1	0.13	7.29	p<0.01
Faeroese§	9	8.64	4	4.71	1	0.64	0.32	NS
Fjord§	29	29.07	3	2.86	0	0.07	0.08	NS
Connemara§	12	NaN	0	NaN	0	NaN	NaN	NS
SB	77	77.11	6	5.78	0	0.11	0.12	NS
Welsh§	13	NaN	0	NaN	0	NaN	NaN	NS
Nordlands	20	20.17	4	3.67	0	0.17	0.2	NS
Total	**803**	**701.44**	**173**	**175**	**25**	**24**	**17.11**	**p<0.001**

§ - samples that might consist of related individuals; NSD - North Swedish Draft; SW - Swedish Warmblood; TB – Thoroughbred; CBT – Coldblooded Trotter; SB – Standard Bread; KY - Kentucky Mountain Horse; RH - Rocky Mountain Horse; TW – Tennese Walker; MF – Missuri Foxtrotter; FT - French Trotter

O - observed amount; E - Expected amount

NS - deviation from HW equilibrium is not signifficant

Appendix 3 continued

Marker	PR8604							
	Common Homozygosity		Heterozygosity		Rare Homozygosity		χ^2	p val
Breed	O	E	O	E	O	E		
TB	9	10.62	20	16.76	5	6.62	1.27	NS
Arabs	22	22.23	5	4.54	0	0.23	0.28	NS
SW	55	53.73	10	12.54	2	0.73	2.74	NS
FT	45	NaN	0	NaN	0	NaN	NaN	NS
CBT	118	118.23	11	10.53	0	0.23	0.26	NS
NSD	20	20.51	18	16.98	3	3.51	0.15	NS
Ardennes	25	27.22	20	15.56	0	2.22	3.67	NS
Russ	32	NaN	0	NaN	0	NaN	NaN	NS
Icelandic	120	121.44	63	60.12	6	7.44	0.43	NS
Shetlands	55	53.83	11	13.35	2	0.83	2.1	NS
KY	15	15.56	7	5.86	0	0.56	0.79	NS
Mini	6	6.13	2	1.75	0	0.13	0.16	NS
RH	14	14.06	2	1.88	0	0.06	0.07	NS
Paso Fino	24	24.01	1	0.98	0	0.01	0.01	NS
TW	11	10.51	7	7.98	2	1.51	0.3	NS
MF	11	11.27	4	3.47	0	0.27	0.36	NS
Mongolian	9	9.03	1	0.95	0	0.03	0.03	NS
Dole	11	11.39	5	4.22	0	0.39	0.55	NS
Bashkir§	29	28.35	5	6.3	1	0.35	1.49	NS
Faeroese§	14	NaN	0	NaN	0	NaN	NaN	NS
Fjord§	22	21.81	8	8.39	1	0.81	0.07	NS
Connemara§	3	4.69	9	5.63	0	1.69	4.32	p<0.05
SB	78	78.01	2	1.98	0	0.01	0.01	NS
Welsh§	11	11.84	8	6.32	0	0.84	1.35	NS
Nordlands	15	14.26	7	8.48	2	1.26	0.73	NS
Total	**695**	**622.04**	**196**	**187.93**	**22**	**26.04**	**9.53**	**p<0.005**

§ - samples that might consist of related individuals; NSD - North Swedish Draft; SW - Swedish Warmblood; TB – Thoroughbred; CBT – Coldblooded Trotter; SB – Standard Bread; KY - Kentucky Mountain Horse; RH - Rocky Mountain Horse; TW – Tennese Walker; MF – Missuri Foxtrotter; FT - French Trotter

O - observed amount; E - Expected amount

NS - deviation from HW equilibrium is not signifficant

Appendix 3 continued

Marker	PR5696							
	Common Homozygosity		Heterozygosity		Rare Homozygosity		χ^2	p val
Breed	O	E	O	E	O	E		
TB	34	34.06	3	2.88	0	0.06	0.01	NS
Arabs	22	22.32	6	5.36	0	0.32	0.4	NS
SW	60	60.06	4	3.88	0	0.06	0.07	NS
FT	42	NaN	0	NaN	0	NaN	NaN	NS
CBT	86	89.45	42	35.11	0	3.45	4.93	p<0.05
NSD	35	35.22	6	5.56	0	0.22	0.26	NS
Ardennes	20	23.47	25	18.06	0	3.47	6.66	p<0.01
Russ	27	27.01	1	0.98	0	0.01	0.01	NS
Icelandic	185	185.02	4	3.96	0	0.02	0.02	NS
Shetlands	66	NaN	0	NaN	0	NaN	NaN	NS
KY	21	21.04	2	1.91	0	0.04	0.05	NS
Mini	7	7.03	1	0.94	0	0.03	0.04	NS
RH	17	NaN	0	NaN	0	NaN	NaN	NS
Paso Fino	25	NaN	0	NaN	0	NaN	NaN	NS
TW	19	19.01	1	0.98	0	0.01	0.01	NS
MF	15	15.02	1	0.97	0	0.02	0.02	NS
Mongolian	10	10.02	1	0.95	0	0.02	0.02	NS
Dole	16	NaN	0	NaN	0	NaN	NaN	NS
Bashkir§	26	26.47	8	7.06	0	0.47	0.6	NS
Faeroese§	13	13.02	1	0.96	0	0.02	0.02	NS
Fjord§	21	21.95	11	9.11	0	0.95	1.38	NS
Connemara§	4	5.33	8	5.33	0	1.33	3	NS
SB	87	87.1	6	5.81	0	0.1	0.1	NS
Welsh§	16	16.12	3	2.76	0	0.12	0.14	NS
Nordlands	12	13.5	12	9	0	1.5	2.67	NS
Total	**886**	**732.22**	**146**	**121.57**	**0**	**12.22**	**49.42**	**p<0.001**

§ - samples that might consist of related individuals; NSD - North Swedish Draft; SW - Swedish Warmblood; TB – Thoroughbred; CBT – Coldblooded Trotter; SB – Standard Bread; KY - Kentucky Mountain Horse; RH - Rocky Mountain Horse; TW – Tennese Walker; MF – Missuri Foxtrotter; FT - French Trotter

O - observed amount; E - Expected amount

NS - deviation from HW equilibrium is not signifficant

Appendix 4

Haplotype association with phenotype in Icelandic Horses

Base haplotype: TTAA

Order of markers: PR8604 PR3737 PR5696 PR5826

Association results for different traits:

Height	coef	se	t.stat	pval
(Intercept)	-0.04638	0.24172	-0.19186	0.848
GTAA	0.12914	0.25293	0.51059	0.61
TCAA	0.31526	0.31988	0.98555	0.326
TTAG	0.02615	0.21198	0.12337	0.902
TTGA (*rare)	0.53407	0.44954	1.18804	0.236

Slowtolt	coef	se	t.stat	pval
(Intercept)	102.6478	0.9814	104.589	0
GTAA	0.4195	1.0195	0.4115	0.681
TCAA	-0.4252	1.3054	-0.3257	0.745
TTAG	0.4988	0.8579	0.5814	0.562
TTGA (*rare)	0.4847	1.7738	0.2732	0.785

Walk	coef	se	t.stat	pval
(Intercept)	100.293	0.75493	132.8501	0
GTAA	0.05777	0.79219	0.07293	0.942
TCAA	0.75687	0.99855	0.75796	0.449
TTAG	0.04946	0.66331	0.07457	0.941
TTGA (*rare)	1.10177	1.37849	0.79926	0.425

Proportion	coef	se	t.stat	pval
(Intercept)	105.1695	0.9831	106.9795	0
GTAA	-0.3697	1.0119	-0.3653	0.715
TCAA	-1.4346	1.311	-1.0943	0.275
TTAG	-2.1153	0.8549	-2.4744	0.014
TTGA (*rare)	1.3369	1.8246	0.7327	0.465

Tolt	coef	se	t.stat	pval
(Intercept)	104.7948	1.2259	85.4811	0
GTAA	2.5448	1.2745	1.9968	0.047
TCAA	-0.2615	1.6314	-0.1603	0.873
TTAG	0.1653	1.072	0.1542	0.878
TTGA (*rare)	0.9298	2.2065	0.4214	0.674

Appendix 4 continued

Trot	coef	se	t.stat	pval
(Intercept)	105.7948	1.1009	96.1018	0
GTAA	1.0807	1.1478	0.9415	0.348
TCAA	-1.2489	1.4609	-0.8549	0.394
TTAG	-0.6998	0.9641	-0.7259	0.469
TTGA (*rare)	-0.2211	2.0008	-0.1105	0.912

Pace	coef	se	t.stat	pval
(Intercept)	99.179	2.0419	48.5729	0
GTAA	-0.7732	2.1253	-0.3638	0.716
TCAA	-0.818	2.7114	-0.3017	0.763
TTAG	-1.8635	1.7867	-1.043	0.298
TTGA (*rare)	-1.6147	3.7121	-0.435	0.664

Gallop	coef	se	t.stat	pval
(Intercept)	105.2424	1.1078	95.0032	0
GTAA	1.4075	1.1411	1.2334	0.219
TCAA	-1.4624	1.482	-0.9868	0.325
TTAG	-0.926	0.9642	-0.9604	0.338
TTGA (*rare)	-0.963	1.9718	-0.4884	0.626

Form	coef	se	t.stat	pval
(Intercept)	106.2411	1.3387	79.3592	0
GTAA	2.1455	1.3878	1.546	0.124
TCAA	-2.3796	1.7817	-1.3356	0.183
TTAG	-0.3688	1.1685	-0.3156	0.753
TTGA (*rare)	0.5573	2.4542	0.2271	0.821

Riding	coef	se	t.stat	pval
Intercept)	104.2778	1.37173	76.01905	0
GTAA	1.34241	1.41959	0.94564	0.346
TCAA	-1.08264	1.82968	-0.59171	0.555
TTAG	-1.13566	1.19679	-0.94893	0.344
TTGA (*rare)	-0.24019	2.45881	-0.09768	0.922

Totalindex	coef	se	t.stat	pval
(Intercept)	105.3057	1.42684	73.80329	0
GTAA	1.63844	1.47758	1.10887	0.269
TCAA	-1.61763	1.9016	-0.85067	0.396
TTAG	-1.57813	1.24513	-1.26744	0.207
TTGA (*rare)	0.11161	2.57496	0.04334	0.965

Appendix 5

Haplotype association with phenotype of Coldblooded Trotters

Base haplotype: TTAA

Order of markers: PR8604 PR3737 PR5696 PR5826

Association results for different traits:

a) Direct values

Start	coef	se	t.stat	pval
(Intercept)	34.20155	5.69448	6.00609	0
GTAA	12.32558	10.85157	1.13583	0.263
TTAG	-0.45711	8.50836	-0.05372	0.957
TTGA	-2.66408	7.2987	-0.36501	0.717
TCAA (*rare)	32.18125	19.44285	1.65517	0.106

Position	coef	se	t.stat	pval
(Intercept)	12.98045	2.51175	5.16789	0
GTAA	1.76443	4.41563	0.39959	0.692
TTAG	0.13188	3.64699	0.03616	0.971
TTGA	-0.85625	3.24704	-0.2637	0.793
TCAA (*rare)	14.01955	9.44688	1.48404	0.146

Earnings	coef	se	t.stat	pval
(Intercept)	3.97E+05	1.34E-12	2.96E+17	0
GTAA	4.17E+04	1.34E-12	3.11E+16	0
TTAG	1.50E+05	1.34E-12	1.12E+17	0
TTGA	6.41E+04	2.84E-18	2.26E+22	0
TCAA (*rare)	1.49E+05	1.18E-19	1.26E+24	0

time/km1	coef	se	t.stat	pval
(Intercept)	1281.73044	9.14868	140.1001	0
GTAA	5.01447	17.66125	0.28392	0.778
TTAG	-3.35004	13.73628	-0.24388	0.809
TTGA	-0.17569	11.70815	-0.01501	0.988
TCAA (*rare)	-15.09015	30.18813	-0.49987	0.62

time/km2	coef	se	t.stat	pval
(Intercept)	1276.133	9.0148	141.5592	0
GTAA	-11.798	16.9326	-0.6968	0.491
TTAG	-23.2691	13.9115	-1.6727	0.104
TTGA	-2.9581	11.4626	-0.2581	0.798
TCAA (*rare)	-28.2235	26.36	-1.0707	0.292

Appendix 5 continued

b)BLUP values

Start_status	coef	se	t.stat	pval
(Intercept)	115.241	1.067	108.037	0
GTAA	-2.06	2.159	-0.954	0.344
TTAG	3.294	1.949	1.69	0.096
TTGA	3.639	1.562	2.33	0.023
TCAA (*rare)	-7.281	3.261	-2.233	0.029

Total	coef	se	t.stat	pval
(Intercept)	116.517	1.103	105.608	0
GTAA	-2.374	2.254	-1.054	0.296
TTAG	4.948	2.024	2.445	0.017
TTGA	5.201	1.614	3.222	0.002
TCAA (*rare)	-7.182	3.297	-2.178	0.033

Position	coef	se	t.stat	pval
(Intercept)	117.0138	1.128	103.7332	0
GTAA	-1.3538	2.7064	-0.5002	0.619
TTAG	7.0263	2.051	3.4259	0.001
TTGA	6.0303	1.6831	3.5829	0.001
TCAA (*rare)	-6.6144	3.2177	-2.0556	0.044

Earn_BLUP	coef	se	t.stat	pval
(Intercept)	117.3286	1.2076	97.1598	0
GTAA	-2.7146	3.1845	-0.8525	0.397
TTAG	6.1583	2.2047	2.7933	0.007
TTGA	6.2373	1.815	3.4366	0.001
TCAA (*rare)	-6.4391	5.3969	-1.1931	0.237

Time_BLUP	coef	se	t.stat	pval
(Intercept)	115.7861	0.9582	120.8427	0
GTAA	-2.1461	2.3311	-0.9206	0.361
TTAG	3.4745	1.7337	2.0041	0.049
TTGA	4.3336	1.4312	3.0278	0.004
TCAA (*rare)	-4.5965	3.23	-1.423	0.16