Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Faculty of Veterinary Medicine and Animal Science
Department of Animal Breeding and Genetics

The role of *Myostatin* in Coldblooded Trotter harness racing performance

Katrine Petäjistö

Examensarbete / Swedish University of Agricultural Sciences
Department of Animal Breeding and Genetics

495

Uppsala 2016

Master's Thesis, 30 hp

Agricultural Science programme – Animal Science

Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Faculty of Veterinary Medicine and Animal Science
Department of Animal Breeding and Genetics

The role of Myostatin in Coldblooded Trotter harness racing performance

Myostatins påverkan på prestation hos kallblodstravare

Katrine Petäjistö

Supervisors:
Gabriella Lindgren, SLU, Department of Animal Breeding and Genetics
Brandon Velie, SLU, Department of Animal Breeding and Genetics
Kim Jäderkvist Fegraeus, SLU, Department of Animal Breeding and Genetics

Examiner:
Sofia Mikko, SLU, Department of Animal Breeding and Genetics

Credits: 30 hp
Course title: Degree project in Animal Science
Course code: EX0558
Programme: Agricultural Science programme - Animal Science
Level: Advanced, A2E

Place of publication: Uppsala
Year of publication: 2016
Cover Picture: Micke Gustafsson, Kanal 75
Name of series, no: Examensarbete / Swedish University of Agricultural Sciences, Department of Animal Breeding and Genetics, 495
On-line publicering: http://stud.epsilon.slu.se

Keywords: horse, myostatin, coldblooded trotter, harness racing, performance

1. Sammanfattning

Myostatin (MSTN) är en gen som begränsar massan av skelettmuskler genom att reglera både antalet och tillväxten av muskelfibrer. Mutationer i den här genen resulterar i en dubbelt-musklad fenotyp noterad hos ett flertal boskapsraser, hundar och människor. Förutom att påverka det fysiska utseendet hos en individ, har mutationer i *MSTN* också associerats med kapplöpningsprestation hos både hundar och hästar. Studier har visat att *MSTN* påverkar både hastighet och distans hos fullblod, den mest använda rasen inom hästkapplöpning. I Skandinavien är travsport mer vanligt förekommande än kapplöpning med fullblod och förutom varmblodstravare är kallblodstravare en av de vanligast förekommande raserna inom travsporten. Den här studien undersökte om det fanns några associationer mellan tre genetiska markörer i *MSTN* och prestation hos kallblodiga travhästar. Kallblodiga travhästar som tävlat inom travsport (n=482) genotypades för tre sekvenspolymorfier tidigare associerade med prestation hos engelskt fullblod (g.65868604*G>T*, g.66493737*C>T* and g.66495826*A>G*). Genotyperna testades statistiskt för associationer med livstids tävlingsprestationsresultat för: antal och proportion av starter, vinster, placeringar (1-3), icke placeringar samt prissumma, prissumma per start och tider för volt- och autostart. Utöver det, utvärderades genotyperna för associationer med prestationsresultat (prissumma, prissumma per start och tider för volt- och autostart) erhållen i olika åldrar: 3 års ålder (n=266), mellan 3 och 6 års ålder (n=469) och 6 år eller äldre (n=361). Det var inga signifikanta associationer ($p>0.05$) för någon av egenskaperna och de tre SNPs, förutom associeringen mellan prestation och *MSTN* hos 3- till 6 åringar. Vid 3-6 års ålder hade GG för SNP PR5826 hästar signifikant högre prissumma än hästarna som var AG eller AA ($P<0.05$). Dock var där endast tre hästar som var GG för SNP PR5826 som hade prestationsresultat för perioden 3 till 6 års ålder. Fullblod har större variation i allel frekvens för g.66493737*C>T* (en stark indikator av prestation hos engelskt fullblod) med en frekvens på 0.51 av C-allelen, jämfört med kallblodstravare som går mot fixering av T-allelen. Prestation bedöms lika hos fullblod och kallblodstravare men tävlingstyperna skiljer sig i både gångart och startmetod. Skillnaden i allelfrekvens kan tyda på att C-allelen för den *MSTN* varianten har en större påverkan på galopp än på trav. Den här studien lyfter fram hur väsentligt det är att studera möjliga associationer, mellan gener och egenskaper, hos olika raser som används för samma syften.

2. Abstract

Myostatin (MSTN) is a gene that limits the skeletal muscle mass by regulating both the number and growth of muscle fibres. Mutations in this gene result in the double-muscled phenotype seen in several livestock breeds, dogs and humans. In addition to affecting the physical appearance of an individual, it has also been associated with racing performance in both dogs and horses. Studies have shown that *MSTN* influences both speed and best race distance in Thoroughbreds, the breed most commonly used in gallop racing. In Scandinavia, harness racing is more common compared to gallop racing and apart from Standardbreds, the Coldblooded trotter is one of the main breeds used in harness racing. This study investigated the association with *MSTN* and harness racing performance in Coldblooded trotter. Coldblooded trotters used for harness racing (n=482) were genotyped for three sequence polymorphisms previously associated with performance in racehorses (g.65868604*G>T*, g.66493737*C>T* and g.66495826*A>G*). The genotypes were evaluated for association with lifetime racing performance results for: number and proportion of starts, victories, placings (1-3) and unplaced as well as earnings, earnings per start and race times for volt- and autostart. Additionally, the genotypes were evaluated for association with performance results (earnings, earnings per start and race times for volt- and autostart) obtained at different ages: 3 years of age (n=266), between 3 and 6 years of age (n= 469) and 6 years or older (n=361). There were no significant associations ($p>0.05$) between any of the traits and the three SNPs except for the association with performance and *MSTN* was seen in 3 to 6 years old. At 3-6 years of age GG horses for SNP PR5826 had significantly higher earnings ($P>0.05$) than the AG and AA horses. However, there were only three horses that were homozygous GG for SNP PR5826 that had performance results for the age period 3 to 6 years. Thoroughbreds have larger variation in allele frequency at g.66493737*C>T* (strong predictor of racing performance in horses) with a frequency of 0.51 at the C-allele, compared to Coldblooded trotters that are close to fixation in the T-allele. The measurements of performance are similar in Thoroughbreds and Coldblooded trotters but the race types differ in both gait and starting method. The difference in allele frequency could indicate that the C-allele at that *MSTN* variant has a greater influence on gallop than trotting ability. This study highlights the importance of studying possible associations between genes and traits in different breeds even if they are used for similar purposes.

3. Literature review

3.1 Myostatin

Myostatin (MSTN) is a protein in the transforming growth factor beta (TGF-β) superfamily. The TGF-β superfamily plays critical roles in tissue growth and differentiation, regeneration and repair as well as embryonic developments. *MSTN* acts as a repressor in the development and regulation of differentiation and proliferative growth of skeletal muscle. It acts to limit the skeletal muscle mass in several mammalian species, by regulating both the number and the growth of muscle fibres (Binns et al, 2010). *MSTN* inhibits cell cycle progression mediated by the p21 gene, which results in interruption of myoblast proliferation (Thomas et al., 2000). It is highly expressed in both adult and developing skeletal muscle.

McPherron et al (1997) identified *MSTN* by degenerating oligonucleotides corresponding to conserved regions among known family members of the TGF-β superfamily on mouse genomic DNA. They explored the biological function of *MSTN* by disrupting the *MSTN* gene through gene targeting in mice. The myostatin-null mice were significantly larger than the wild-type, by 30%, and showed a dramatic increase in skeletal muscle mass due to an increased number of muscle fibres and thickness of the fibres (McPherron et al., 1997). The individual muscles of *MSTN* null mice weighed two to three times more than the muscles of the wild-type mice, due to an increase in the number of muscle fibres without a corresponding increase in the amount of fat (McPherron & Lee, 1997).

3.1.1 Myostatin in livestock

The double-muscled phenotype has been seen in several livestock breeds since it was first documented in 1807 (Culley, 1807). This phenotype has attracted attention from beef producers because the double-muscled animals are characterized by a 20 % increase in muscle mass due to skeletal muscle hyperplasia (increase in the number of muscle fibres) (Grobet et al., 1997). As a result, double-muscling phenotype has come to be a well-studied phenomenon in livestock.

3.1.1.1 Cattle

This double-muscling phenotype has shown to have a high frequency in the cattle breeds Belgian Blue and Piedmontese (McPherron & Lee, 1997; Grobet et al., 1997; Kambadur et al., 1997). Sequence analysis has revealed mutations in the heavy-muscled cattle of both Belgian Blue and Piedmontese. The Belgian Blue cattle have an 11-bp deletion in exon 3 of *MSTN*, which results in a frame-shift mutation that eliminates virtually the entire active region of the *MSTN* protein. The Piedmontese cattle have instead a G-A transition in the same region as Belgian Blue that changes a cysteine residue to a tyrosine, which give a negative effect on the activity of *MSTN* (McPherron & Lee, 1997). In Belgian Blue, heterozygotes for the 11bp deletion display very mild abnormalities in muscle phenotype so the locus has been termed "partially recessive" because there is some effects of a single copy of the allele. However, the animal needs to be homozygous to express the double muscle phenotype. The increased musculature of double-muscled cattle results only from hyperplasia of muscle fibre. Compared to mice, where the increased musculature results from hyperplasia and hypertrophy of the muscle fibre (Kambadur et al., 1997). The Belgian Blue and Piedmontese cattle have an increase ability to convert feed into lean muscle and have a higher percentage of the most desired cuts of meat, compared to normal cattle. However, in the Belgian Blue there are some problems associated with the phenotype. The breed is less stress tolerate, the fertility rate is

lower and there is a reduction in calf viability, which has hindered the exploitation of the hypertrophy by classical genetic selection (Casas et al., 1998).

3.1.1.2 Sheep

Texel sheep are known for their exceptional meatiness. The *MSTN* allele of Texel sheep is characterized by a G to A transition in the untranslated region of the trailer sequence (3' UTR) that creates a target site for microRNAs that are highly expressed in skeletal muscle. The transition of G to A causes a transitional inhibition of the *MSTN* gene, which contributes to the muscular hypertrophy seen in Texel sheep (Clop et al., 2006).

3.1.2 Myostatin in humans

The double-muscled phenotype caused by a lack of *MSTN* has also been seen in several mammalian species also including humans. Schuelke et al (2004) reported the identification of a *MSTN* mutation in a human child with muscle hypertrophy. At birth, the child appeared extraordinary muscular with protruding muscles in his thighs and upper arms. When the child was six days old muscle hypertrophy was verified by an ultrasonography. The motor and mental development was normal and at four and a half years of age he continued to increase in muscle bulk and strength. At four and a half years of age the child was able to hold two 3 kg dumbbells in a horizontal suspension with his arms extended. The family members of this child have been reported to be unusually strong and the mother was a former professional athlete and appeared muscular but not to the extent observed in her child. Schuelke et al (2004) studied the child and found that he had a G-A transition at nucleotide g. IVS1+5 at the splice donor site in intron 1 in both alleles. The results strongly indicated that the child had a loss-of-function mutation in the *MSTN* thus suggesting that inactivation of *MSTN* has the same effect in humans as seen in mice and livestock (McPherron & Lee, 1997; Grobet et al., 1997; Clop et al., 2006).

3.1.3 Myostatin in dogs

In addition to affecting the physical appearance of an individual, mutations in *MSTN* have also been associated with racing performance in dogs (Mosher et al., 2007). Mosher et al (2007) reported in their study a mutation in *MSTN* causing double-muscle phenotype in the whippet dog breed. They discovered that the mutation caused a 2-bp deletion in the exon 3 of the MSTN gene, which removes nucleotides that in turn lead to a premature stop codon at the amino acid instead of the normal cysteine. Dogs with two copies of the mutation exhibit the double-muscled phenotype known as the "bully" whippet. Dogs with only one copy of the mutation where not as muscular as the "bully" whippet but were more muscular than the wild-type genotype and also significantly faster. The results of the study highlighted the utility of performance enhancing polymorphism and marked the first time a mutation in *MSTN* had been quantitatively linked to increased athletic performance. Greyhounds and Whippets share a common ancestral gene pool but no greyhound tested has carried the mutation (Mosher et al., 2007).

3.2 Myostatin in horses

When it was published that *MSTN* was associated with performance in the Whippet dog breed (Mosher et al., 2007), the association with *MSTN* and performance continued to be studied in other animals used in racing. The English Thoroughbred is the most common breed used in horse racing (The Swedish Horse Racing Authority, 2013) and has therefore come to be of interest in several studies regarding *MSTN* and performance (Hill et al., 2010a; Bower et al.,

2012; Tozaki et al, 2011a; Tozaki et al, 2011b). In horses, *MSTN* is located at the distal end of chromosome 18 (ECA18) and consists of three exons and two intron regions (Grobet et al., 1997).

3.2.1 Myostatin and racing performance

In 2010, Hill et al (2010a) described for the first time an identification of a gene variant in Thoroughbred racehorses that was predictive of genetic potential and athletic phenotype. By using a sample of Thoroughbreds, a *MSTN* sequence polymorphism (g.66493737*C>T [PR3737]*, located in the first intron of the *MSTN* gene and strongly associated ($P<4.85 \times 10^{-8}$) with best race distance among elite racehorses, was identified. The horses in the study were divided into those who won their best races over short distance (1000-1600m) and those who won their best races over long distance (1400-2400m). The C allele for the SNP was twice as common in the group competing preferably in short distance races requiring exceptional speed compared to horses performing optimally in longer distance races that requires more stamina. Additionally, the CC genotype was most common in the short distance group and absent in the long distance group. Hill et al (2010a) stated, that the CC genotype was more suited for fast, short-distance races while horses with CT genotype competed more favourably in middle-distance races and the TT horses had more stamina. Moreover, their evaluation of racecourse performance and stallion progeny performance showed that CC and CT horses are predicted to be more successful two-year-old racehorses than TT horses. CC horses had also significantly greater muscle mass that TT horses at two-years of age.

Shortly after Hill et al (2010a) had identified the *MSTN* sequence polymorphism (PR3737) strongly associated with best race distance, another study by Hill et al (2010b) was published. In the later study, a genome-wide SNP-association study was performed to confirm the sequence variant in the first intron (PR3737) of the equine *MSTN* gene for being the most powerful predictor of optimum racing distance for Thoroughbred race horses. In their cohort-based association test they evaluated the genotypic variation at 40 977 SNPs between horses suited for to short distance (1000-1600m) and middle-long distance (1400-2400m) races. The results of the study showed that the SNP PR3737 was the superior variant in prediction of distance aptitude in Thoroughbred racehorses.

Tozaki et al (2011a) performed a study on racing performance in Japanese thoroughbred horses using genome information on the equine chromosome 18 (ECA18), similar to the study Hill et al (2010a) performed. Tozaki et al (2011a) analysed genotypes at four SNP, SNP PR3737 included. The findings of their study corroborated with the findings in Hill et al (2010a) study, that *MSTN* is associated with best racing distance in Thoroughbreds. Apart from SNP PR3737 the SNP g.65868604G>T *[PR8604]*, located in the promotor region of ECA18, was also associated with race distance. Apart from race distance, a strong relation with performance rank (ranking of horses based on earnings from first place finishes) in SNP PR8604 and lifetime earnings in female horses for both SNPs was observed. The GG genotype for SNP PR8604 had the highest average of lifetime earnings compared to the TG and TT genotypes. For SNP PR3737, the CC genotype had highest average of lifetime earnings compared to CT and TT horses.

Apart from being an optimum predictor of race distance, SNP PR3737 has also shown to be associated with speed in Thoroughbred racehorses. Hill et al (2011) found that CC and CT horses outperformed TT horses at all measured speed variables in the study, which indicated that at least one C-allele is required to improve speed.

3.2.2 Myostatin and body composition

A significant association has been found between *MSTN* genotype and body composition measurements and the SNPs PR3737 and PR8604. Tozaki et al (2011b) found significant association between *MSTN* and body weight/withers height in Thoroughbreds. In the study, they tested four SNPs, including PR3737 and PR8604, in the *MSTN* ECA18 and all were significantly associated with body weight/withers height. The CC horses had significantly greater height-to-body mass ratio than CT and TT horses at SNP PR3737 (Tozaki et al, 2011b).

Velie et al (2015) studied the possible role of *MSTN* in horses outside of racing performance in Icelandic horses, which is a breed both used for meat production and riding. A significant difference ($P < 0.001$) was found in genotype frequency at SNP PR5826 between horses used for meat production and horses used for riding. According to Velie et al (2015), the different genotypes at SNP PR5826 are likely to result in phenotypes less desirable for riding or more desirable for meat production. Thus, the study suggests that *MSTN* also has a valuable role in horses outside of racing performance.

3.3 Allele and genotype frequencies of *MSTN* among different horse breeds

As several studies have shown, the SNP PR3737 in the *MSTN* gene is the most associated SNP with performance in Thoroughbred racehorses (Hill et al., 2010a; Hill et al., 2010b; Hill et al., 2011). Selective breeding for speed in Thoroughbred racehorses has resulted in a high frequency (0.51) of the C-variant at this SNP (PR3737). When looking at other breeds there is a large variation in allele frequencies in this particular SNP (Bower et al., 2011).

Bower et al (2011) wanted to investigate the genetic origin and history of speed in the Thoroughbred, so to determine the ancestral SNP PR3737 allele in equid they genotyped the SNP in 40 donkeys and 2 zebras. The genotyping showed that both species lacked the C allele and by that the authors ascertained that the T-allele was the wild type. To evaluate the effect of recent selection of different competition pressures they compared genotypes from recently derived breeds with recorded Thoroughbred influences: Quarter horses (a breed known for short distance races and activities requiring short bursts of speed) and Standardbreds (harness racing). The results revealed that the C-allele occurred at a frequency of 0.51 in Thoroughbreds, it was absent in Standardbreds and most common in the Quarter Horse at a frequency of 0.90. Whereas, the allele frequency of Egyptian Arabian horses, a breed used for endurance exercise, have an excess of TT genotype at a frequency of 0.90 (Hill et al., 2010a) and Spanish trotters, used in harness racing, a T-allelic frequency of 0.98 (Negro et al., 2016). Thus, Bower et al (2011) concluded that the C-allele was not restricted to the Thoroughbred and Thoroughbred-derived populations and that it seemed to occur at variable frequencies depending on selection pressure in a population.

Polasik et al (2015) studied the g.66493737C>T polymorphism in four different polish horse breeds; Holstein, Polish Noble Half-breed, Polish Heavy Draft and Polish Konik. The results showed that CC genotype was present only in the Holstein breed, which is known to have an English Thoroughbred ancestor and is used for show jumping and dressage. The CT genotype was most frequent in the Polish Noble Half-breed, a breed commonly used for show jumping and dressage. The TT genotype was most frequent in the Polish Heavy Draft, classified as heavy draught horse, and Polish Konik, classified as general type horse used for light agriculture and as a horse for children.

3.4 Known effect of *DMRT3* on performance

The success of a racehorse can be measured in different ways such as how often a horse wins, how much money it earns or how fast it runs. However, several environmental factors as well as genes affect these traits. In addition to *MSTN, DMRT3* is a gene that has been shown to affect performance in horses. Previous studies have shown a single base-pair mutation, a change from cytosine (C) to adenine (A), in the *DMRT3* gene affects the ability to show ambling and lateral gaits in several horse breeds (Andersson et al., 2012). The mutation in *DMRT3* has shown to influence riding traits in multiple breeds as well as have a positive impact on trotting performance in Coldblooded and Standardbred trotters (Jäderkvist et al., 2014; Jäderkvist et al., 2015).

3.5 Horse racing

There are two major types of horse racing in the world, gallop racing and harness racing. The horses in harness racing race in a trotting or pacing gait and pull a two-wheeled cart called a sulky. Several breeds have evolved for harness racing purposes in different countries. In Sweden and Norway harness racing is more common than gallop racing with Thoroughbreds and the main breeds used are the Standardbred and Coldblooded trotter (Thiruvenkadan et al., 2009). The two breeds are today pure trotting racehorses and compete under similar conditions but in separate races (Revold et al., 2010). The Standardbred is the most common breed used in trotting in Scandinavia.

3.5.1 The Coldblooded Trotter

The Coldblooded trotter is a breed commonly used for harness racing in Sweden and Norway. It originates from the North-Swedish Horse and the Norwegian Døle horse. These breeds were used as draft horses but when harness racing became more popular the breeds were subdivided into one heavier and one lighter and faster breed respectively. To get faster trotting horses these heavy breeds were bred with a lighter breed (e.g. Standardbred) resulting in the Norwegian Coldblooded trotter and the Swedish Coldblooded trotter (Hendricks, 2007). Due to exchange of breeding animals between the two countries the Norwegian and Coldblooded trotter are close and the studbook is now joint. The Swedish Coldblooded trotters are together with the Norwegian and Finnish Coldblooded trotters the only Coldblooded trotters in the world. The Coldblooded trotter is a tractable and robust breed that is also commonly used for riding and is very appreciated for its energy, sustainability and reliability (The Swedish Trotting Association, 2015).

3.5.2 Performance in Coldblooded trotters

Estimated breeding values (EBV) for Coldblooded trotters have been developed to genetically evaluate the horses so that the most successful can be selected for breeding (The Swedish Trotting Association, 2014). EBVs in Coldblooded trotters are estimated using an animal model that is based on total number of starts, percentage of placings (1-3), earnings, earnings per start, best record and starting status. The model also includes fixed effects for gender and birth year for the horse. EBVs are calculated and updated yearly (The Swedish Trotting Association, 2014).

4. Introduction

The gene *Myostatin (MSTN)* is a well-studied gene in domestic animals. It acts to limit the skeletal muscle mass by regulating both the number and the growth of muscle fibres (Binns et al, 2010; McPherron et al., 1997; McPherron & Lee., 1997; Thomas et al., 2000). Mutations in this gene result in the double-muscled phenotype seen in several livestock breeds, such as Belgian Blue cattle and Texel sheep (Clop et al., 2006; McPherron & Lee, 1997; Grobet et al., 1997). Similar phenotypes are also seen in Whippet dogs and humans (Mosher et al., 2007; Schuelke et al., 2004). In addition to affecting the physical appearance of an individual, mutations in *MSTN* have also been associated with racing performance in both dogs and horses (Mosher et al., 2007; Hill et al, 2011; Hill et al, 2010). Dogs with only one copy of the mutation, causing a 2-bp deletion, are more muscular than the wild-type genotype and significantly faster (Mosher et al., 2007). In horses, associations between *MSTN* and racing performance have been identified in Thoroughbreds, the breed most commonly used in horse racing. Studies have shown that *MSTN* influences both speed (Hill et al, 2011) and best race distance (Hill et al, 2010) in this breed. In Sweden and Norway, harness racing is more common compared to Thoroughbred gallop racing. In harness racing, horses race in a trotting or pacing gait and several breeds have evolved for harness racing purposes in different countries (Thiruvenkadan et al., 2009). The Standardbred and Coldblooded trotter are the two main breeds used in harness racing in Scandinavia. The Standardbreds have been bred solely for their trotting abilities while the Coldblooded trotter origin from a heavier rural carriage horse, like the North-Swedish horse or the Døle horse (Thiruvenkadan et al., 2009). The two breeds are today pure trotting racehorses and compete under similar conditions but in separate races (Revold et al., 2010). The Standardbred is the most common breed used in trotting in Scandinavia.

In Thoroughbreds, previous studies have shown a large variation in allele frequency of *MSTN* variants that are strongly associated with racing performance. In Standardbreds, there is very little variation in allele frequency of the same variants. In Coldblooded trotters, the allele frequency in the same *MSTN* variant is until this study unknown. Regarding the different breeding history of the Standardbreds and Coldblooded trotters the breeds differ and for the present study it was hypothesized that there could be a larger variation in allele frequency in Coldblooded trotters compared to Standardbreds. There are no previous studies on associations between *MSTN* and performance in harness racing. As genetic variants in *MSTN* previously been associated with racing performance in Thoroughbreds it was of interest to investigate what role *MSTN* have on performance in another breed used for racing. The measurements of performance are similar in Thoroughbreds and Coldblooded trotters, both breeds are supposed to run fast, win and earn money. However, the race types and the breeds differ.

4.1 Aim of the study

The aim of this study was to investigate the association of *MSTN* with racing performance in the Norwegian-Swedish Coldblooded trotter.

5. Material and methods

5.1 Population

A sample of 1000 Coldblooded trotters born between 2000 and 2009, originating from Sweden and Norway, were randomly selected for this study using custom scripts written in the statistical software R (R Development Core Team, 2015). Out of the sample, horses with a minimum of one start were extracted (n=621). Of the 621 raced horses, 486 DNA samples were used for SNP genotyping. The sample of successfully genotyped horses (n=482) consisted of 46 intact males, 188 mares, 247 geldings and 1 cryptorchid representing offspring from 109 sires and 431 dams. The average number of foals per sire was four while the average number of foals per dam was one. The maximum number of offspring per sire was 51 while the maximum offspring per dam was three.

5.2 Racing Performance

For this study an association analysis was performed using lifetime racing performance results as of December 31, 2015 (number and proportion of starts, victories, placings (1-3) and unplaced as well as earnings, earnings per start and race times for volt- and autostart). Additionally, BLUP values and inbreeding coefficient was analysed. A corresponding analysis based on performance results (earnings, earnings per start and race time (volt and auto)) obtained at different ages: 3 years of age (n=266), between 3 and 6 years of age (n= 469) and 6 years or older (n=361) was also included in the study.

The performance data for each horse were obtained from the Swedish Trotting Association and the Norwegian Trotting Association.

5.2.1 Finish position

Number of victories (Wins) was calculated as the total number of times a horse finished a race in first place. Number of placings was calculated as the total number of times a horse finished a race in first, second or third place. The number of times a horse was Unplaced was calculated as the total number of times a horse failed to finish in first, second or third place even if prize money was distributed for less than third place. In addition, the frequencies of races a horse participated in that resulted in a finish position of first (Wins. freq); first, second or third (Plac. freq) were calculated.

5.2.2 Race times

Race times for two different start methods (volt – and autostart) were included in the study. Voltstart is a starting method where the horses start in pens that disposes 20 meter of volt space and trot in circular patterns to then hit the starting line as a group (The Swedish Trotting Association, 2016). Autostart is a starting method where a car is used to set the starting line. The car is placed 260 m before the start line and gradually increases the speed so it hits the start line in 52 km/hour (The Swedish Trotting Association, 2016). The best race time for each horse was used from the respective race types. The best race time of a horse was measured as the fastest racing time for 1 km, in seconds. Distances for races with auto starts ranged from 1600 to 2720 m. Distance in races with voltstart ranged from 1600 to 2180 m.

5.2.3 Earnings

Cumulative earnings (Earnings) were calculated as the total amount of prize money a horse had earned between January 2003 and November 2015. Earnings per start (EPS) were

calculated as the average amount of prize money earned per start. The majority of horses had earnings in Swedish currency (SEK); however, for horses with earnings in Norwegian currency (NOK) the earnings were converted to Swedish kronor by multiplying the NOK earnings with an estimated average exchange rate of 0.95, estimated from the average exchange rate from 2003 to 2015 (EUROINVESTOR, 2016).

5.2.4 Best linear unbiased prediction (BLUP)

BLUP values for each horse were provided from the Swedish- and Norwegian Trotting Association. BLUP is a measurement that predicts the expected or documented breeding value of a horse. In Coldblooded trotters, it is estimated using an animal model that is based on total number of starts, percentage of placings (1-3), earnings, earnings per start, best record and starting status. The model also includes fixed effects for gender and birth year for the horse. BLUPs are calculated and updated yearly (The Swedish Trotting Association, 2014). For this study, the latest updated BLUP values for each horse was used.

5.2.5 Inbreeding coefficient

The inbreeding for a horse is given as inbreeding coefficient in percent (The Swedish Trotting Association, 2014). Inbreeding coefficients for each horse were provided from the Swedish- and Norwegian Trotting Association.

5.3 SNP Genotyping

DNA samples for the horses were obtained from the Animal Genetics Laboratory, Swedish University of Agricultural Sciences, Uppsala, Sweden and from the Animal Genetics Laboratory, Norwegian University of Life Sciences, Norway. Genomic DNA was extracted from hair and blood samples. DNA from hair samples was prepared from hair roots using a standard hair-preparation procedure where 7 µL proteinase K (20 mg/mL; Merck KgaA, Darmstadt, Germany) and 100 µL 5 % Chelex Resin (Bio-Rad Laboratories, Hercules, CA) were added to each sample. The mix was incubated for 1 hour in 56°C and the proteinase K was inactivated for 10 min in 95°C. For the DNA preparation from blood samples 350 µL of blood was used and isolated by the Qiasymphony instrument (Qiagen, Hilden, Germany).

Based on previous associations with performance rank three SNPs (g.65868604*G*>*T* [PR8604], g.66493737*C*>*T* [PR3737] and g.66495826*A*>*G* [PR5826]) within *MSTN* were genotyped (Figure 1). Single nucleotide polymorphism genotyping was performed with the StepOnePlus Real-Time PCR System (Life Technologies [Thermo Fisher Scientific], Waltham, MA) using Custom TaqMan Genotyping assays (Applied Biosystems by Life Technologies [Thermo Fisher Scientific]).

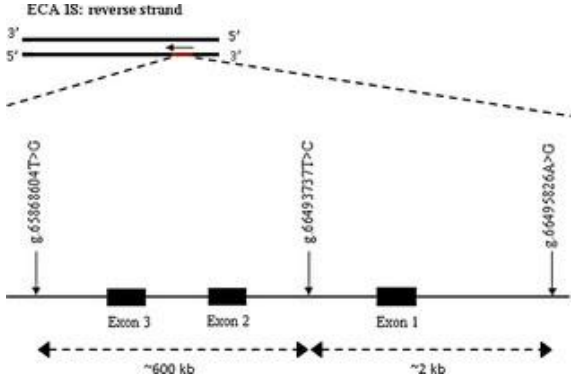

Figure 1. *Relative locations of the analysed SNP markers in horse MSTN (Velie et al, 2015).*

5.4 Statistical Analysis

Data were structured for analyses using custom scripts written in the statistical software R (R Development Core Team, 2015). Normally distributed traits were assessed by the Shapiro-Wilk test. Traits that were not normally distributed were log transformed (log_{10}). These traits included number of starts, number and proportion of victories, placings and unplaced, earnings, EPS and race times. However, race times were transformed by using log(race time - 68.2) / log(exp(1)). BLUP values are by default normally distributed. Horses that had no race time for auto (n=166) or volt starts (n=19) had therefore no race time phenotypes for the respective race type and were subsequently excluded from the analyses concerning race time.

Horses that were not successfully genotyped (n=4) for all three SNPs were excluded from the study. Hardy-Weinberg equilibrium (HWE) of the genotyped SNPs was evaluated using the R package "SNPassoc" (González et al, 2007).

5.4.1 Models

Fixed effects and covariates for the performance traits were assessed with ANOVA type 3. Factors included in the models were: sex, birth year, country of birth and *DMRT3* genotype. The *DMRT3* "Gait keeper" gene is a gene that has been shown to affect performance in harness racing horses. A mutation in this gene alters the pattern of locomotion in horses and has been shown to have a positive impact on harness racing performance (Jäderkvist et al., 2014; Jäderkvist et al., 2015). As such, the model for all traits included the fixed effect of *DMRT3*. All traits except BLUP included fixed effect of sex, country of birth and birth year. All significant fixed effects and covariates that were included in the final association analysis models were appropriate.

5.4.2 Association Analysis

For the association analysis the "SNPassoc" package in R was used. "The SNPassoc package contains facilities for data manipulation, tools for exploratory data analysis, convenient graphical facilities, and tools for assessing genetic association for both quantitative and categorical (case-control) traits in whole genome approaches" (González et al. 2007, p. 2). The SNPassoc package can be used to perform whole genome association studies as well associations with only one SNP and different genetic models can be implemented. With the package, traits can be adjusted for fixed effects and covariates and the association between a single SNP and the dependent variable is calculated under five different genetic models testing different modes of inheritance: Codominant, Dominant, Recessive, Overdominant and log-Additive.

Using the SNPassoc package, the traits were adjusted for the significant fixed effects and covariates. Associations between the traits and the SNPs were evaluated for all SNPs separately. The associations were calculated under different genetic models testing different modes of inheritance.

6. Results

6.1 Distributions and summary statistics for the genotyped sample

The distribution of gender and country of birth for all genotyped samples (n=482) can be found in Supplementary Table 1 and the distribution of birth year is found in Supplementary Table 2. Summary statistics including averages, medians, 1st and 3rd quartiles, min and max values for all traits for all genotyped samples (n=482) is presented in Supplementary Table 3. Summary statistics divided by country of birth (Sweden; n=311 and Norway; n=171) can be found in Supplementary Tables 4 and 5.

6.2 Genotypes

All individuals included in the analyses (n=482) were successfully genotyped for all three SNPs. Genotype and allele frequencies are presented in Table 1. All SNPs deviated from HWE (P>0.4) (Table 1). Min, max, median and averages for all the traits analysed stratified by genotypes are found in Supplementary Table 6.

Table 1. *Genotype and allele frequencies and p-value for Hardy-Weinberg equilibrium for the three SNPs.*

SNP	**PR3737 (n=482)**			**PR8604 (n=482)**			**PR5826 (n=482)**		
Genotype	CC	CT	TT	GG	TG	TT	AA	AG	GG
Frequency	-	23 (5%)	459 (95%)	-	57 (12%)	425 (88%)	390 (81%)	88 (18%)	4 (1%)
Allele	C	T		G	T		A	G	
Frequency	23 (2%)	941 (98%)		57 (6%)	907 (94%)		868 (90%)	96 (10%)	
HWE p-value		1			0.4			1	

* Percentages seen in parentheses.

6.3 Models

Fixed effects and covariates used in the models are listed with p-values in Table 2. Significant p-values are in bold. Only significant (P<0.05) fixed effects and covariates were included in the final models.

Table 2. *P-values for fixed effects and covariates included in the models of racing performance traits in coldblooded trotters. Significant (P<0.05) fixed effects and covariates were used in the final models*

	Fixed effect				**Covariates**			
Trait	**Sex**	**DMRT3**	**Birthyear**	**Country of Birth**	**Number of starts**	**Earnings**	**Race time (volt)**	**Race time (auto)**
Starts	**0.024**	**0.038**	0.094	0.180	-	**< 0.001**	**< 0.001**	**< 0.001**
Wins	**< 0.001**	0.242	0.763	0.487	**< 0.001**	-	**<0.001**	0.431
Wins. freq	**< 0.001**	0.058	0.252	0.271	-	-	0.942	**0.005**
Placings	**< 0.001**	**0.008**	0.748	**0.041**	**< 0.001**	-	**< 0.001**	0.442
Plac. freq	**< 0.001**	**0.002**	0.320	0.063	-	-	0.076	**< 0.001**
Unplaced	**< 0.001**	0.487	0.515	0.956	**< 0.001**	-	0.503	**0.001**
Unplac.freq	**0.001**	0.513	0.740	0.789	-	-	**< 0.001**	0.104
Earnings	**< 0.001**	0.060	0.295	0.136	**< 0.001**	-	**< 0.001**	**< 0.001**
EPS	**< 0.001**	**0.042**	0.254	**0.018**	-	-	**< 0.001**	**< 0.001**
Race time (auto)	**0.023**	0.467	0.258	0.786	**< 0.001**	-	-	-
Race time (volt)	0.052	0.305	0.428	0.195	0.050	-	-	-
Blup	-	0.700	-	-	-	-	-	-

-, a fixed effect or covariate that was not included in the analysis.
Values in bold are the traits that were significant (p < 0.05).

6.4 Association analyses

P-values for the association analysis, performed with SNPassoc package in R (González et al., 2007), are presented in Table 4. The association analysis tested for associations in different modes of inheritance. Since there were only two genotypes present for SNP PR3737 and PR8604, only codominant mode of inheritance was analysed. There were no significant associations (P<0.05) for any of the traits and the three SNPs.

Table 3. *P-values from the association analysis*

	Starts	Wins	Wins. freq	Placings	Plac. freq	Unplaced	Unplac. Freq	Earnings	EPS	Race time (auto)	Race time (volt)	BLUP
PR3737												
Codominant	0.647	0.336	0.158	0.445	0.312	0.701	0.579	0.975	0.741	0.956	0.361	0.431
PR8604												
Codominant	0.663	0.422	0.185	0.562	0.662	0.579	0.409	0.492	0.289	0.171	0.519	0.097
PR5826												
Codominant	0.282	0.234	0.156	0.908	0.968	0.969	0.743	0.542	0.362	0.289	0.525	0.897
Dominant	0.881	0.229	0.896	0.717	0.802	0.973	0.442	0.355	0.176	0.369	0.404	0.915
Recessive	0.113	0.158	0.062	0.859	0.997	0.812	0.841	0.438	0.492	0.271	0.549	0.640
Overdominant	0.826	0.372	0.569	0.681	0.798	0.928	0.462	0.447	0.226	0.241	0.322	0.999
Log Additive	0.630	0.152	0.779	0.763	0.815	0.984	0.445	0.301	0.156	0.547	0.514	0.841

6.4.1 Association analysis based on performance results obtained at different ages

The association analysis on performance results obtained at different ages showed no significant differences between the genotypes for any of the SNPs and the analysed traits. The only significant association found was on earnings (P<0.05) in 3 to 6 year olds where the GG genotype in SNP PR5826 had significantly higher earnings compared to AA and AG. However, there were only three horses that were homozygous GG for SNP PR5826 that had performance results for the age period 3 to 6 years.

7. Discussion

Previous studies have shown that *MSTN* is associated with racing performance in Thoroughbred racehorses (Hill et al., 2010a; Hill et al., 2011; Tozaki et al., 2011a). However, no significant associations between *MSTN* and performance in Coldblooded trotter were apparent in this study. Allele frequencies of the SNPs investigated in this study demonstrate what appears to be the gradual movement towards fixation of three known *MSTN* variants in the Coldblooded trotter. When looking at the allele frequencies it seems that the T-allele in both PR3737 and PR8604 is moving towards fixation. In PR5826 all three genotypes are present but only four horses out of 482 were homozygous GG and when looking at the allele frequencies for this SNP, the A allele is clearly the major allele. There are not many studies presenting allele and genotype frequencies of the SNPs PR8604 and PR5826 but SNP PR3737 has been studied in several horse breeds (Bower et al., 2011; Hill et al., 2010a). When looking at the allele frequencies of PR3737 in other breeds the frequency of the T-allele is high/close to fixation in Standardbreds, Spanish trotter, Egyptian Arabian, the Polish Heavy Draft and the Polish Konik horses. Since the Standardbreds and Spanish trotter are both used in harness racing it is quite expected that the allele frequencies would be similar in the Coldblooded trotter. However, the Egyptian Arabian horse is quite different from a Coldblooded trotter. The Arabian is used for endurance competitions where it runs long distances. Nonetheless, the allele frequencies are similar to the trotters and the TT genotype is correlated with stamina (Hill et al., 2010a). This implies that stamina is a desired trait in

trotters. The Polish heavy draft horse also had a similar allele frequency like the trotters and if discussing the fact that they were bred to pull a carriage, stamina seems to be of importance. As mentioned, the Coldblooded trotter originates from the North-Swedish horse and the Norwegian Døle horse, which are both historically used as draught horses so it is likely that endurance is a desired trait in those breeds as well (Bohlin & Rönningen, 1975).

In Thoroughbreds, the C-allele at SNP PR3737 occurs at a frequency of 0.51 and in the Quarter Horse it occurs at a frequency of 0.90 (Bower et al., 2011). Hill et al (2011) indicates that at least one C-allele is required to improve speed in Thoroughbreds. Speed is also desired in Coldblooded trotters but as shown, the T-allele is still almost fixed in this breed. When pointing out the differences of Thoroughbreds and Coldblooded trotters the gait is the one major difference. The Thoroughbred race in galloping gait while the Coldblooded trotter race in trotting gait and if the trotter falls into gallop in a race this can lead to disqualification (The Swedish Trotting Association, 2016). The Quarter Horses, which are used in fast, short distance races, are almost fixed for the C-allele and they also race in gallop. Hence, the C-allele seems to have greater influence on gallop compared to the influence on trotting ability. François et al (2016) studied the association of *MSTN* with gaits and conformations traits in the Icelandic horse. The result of their study indicated that *MSTN* might play a supporting role in a horse's ability to perform gaits of a certain quality. This is due to the influence *MSTN* has on muscle fibre type. The potential link, between *MSTN* and a horse's ability to perform gaits of a certain quality, was beyond the scope of this study as well as François et al (2016) study but it may be of interest in future studies.

The C-allele has previously been associated with higher Type 2B gluteus muscle fibre proportions, which are fast twitching muscle fibres needed for short bursts of energy. Apart from different gaits, one difference between harness and gallop racing is the start methods. The Coldblooded trotter starts either with volt start or auto start which both means that they hit the starting line with speed. However, the Thoroughbreds start the race from standing still and thus need to be more explosive. The starting method used in harness racing does not require the trotters to be that explosive. Instead, it seems that the Coldblooded trotters need to be sustainable and have stamina. The TT-genotype has shown to be more favourable in longer distance races (1400-2400m) in Thoroughbreds (Hill et al., 2010a). The T-allele is also associated with Type 1 gluteus muscle fibre proportions, which are slowly contracting fibres that have a high oxidative capacity that is beneficial for endurance (Petersen et al., 2014). The horses in this study did not compete in races shorter than 1600 m, which is considered long distance according to the distance measurements in Hill et al (2010a) study on Thoroughbreds. Since the Coldblooded trotters appears to be moving towards fixation in the T-allele it clearly seems that endurance is a desired trait in Coldblooded trotters competing in distances over 1600 m. This study did not analyse the association of best race distance and *MSTN* but since there are Coldblooded trotters that perform better in longer distances than others so there could be an association with *MSTN*. However, it is difficult to compare the genotypes since the frequency of one allele is so high.

Even though this study did not show any significant associations between *MSTN* and Coldblooded trotter harness racing performance the fixation of the T-allele still implies that *MSTN* must have or had an impact on harness performance since the selection is going in one direction and it is going in the same direction in other trotter breeds (e.g. the Standardbred and the Spanish trotter). Bower et al (2011) ascertained in their study that the TT was the wild-type. However, there were a few CT genotypes at SNP PR3737 in this study, which means that there could have been more before but some qualities with horses carrying T-allele have

been more desired by breeders. If there would have been greater variation in genotype frequencies between the SNPs the differences would have been easier to measure but since there is one major allele this must mean that it is the most desired one. Why that is so need to be studied further. There was greater variation in the SNP PR5826 where there was a significant association in earnings ($P<0.05$) in 3 to 6 year olds where the GG genotype in SNP PR5826 had significantly higher earnings compared to AA and AG. However, there were only three horses with GG genotype, which is not enough representative for the genotype.

This study only analysed three SNPs that have previously been associated with performance in Thoroughbreds. Since there are more SNPs in the *MSTN* gene there could be other SNPs that are more associated with performance for Coldblooded trotter harness racing than the ones examined in this study. Performance in trotters has shown to be influenced by other genes such as the *DMRT3* and it is safe to assume that performance in Coldblooded harness racing is polygenic. Still, further studies on performance and harness racing are required to further explore potential associations. By knowing which factors (genetic and environmental) that influence the traits of the horse, a more successful breeding and a faster genetic progress could be achieved.

8. Conclusions

This study showed no significant associations between three known variants in *MSTN* and Coldblooded trotter harness racing performance even if it has been previously associated with Thoroughbred racing performance. That *MSTN* is associated with Thoroughbred racing and not harness racing could be due to the fact that the race types differ in both gait and starting method. Hence, gallop is more influenced by *MSTN* that trot is and the fibre type composition associated with *MSTN* is more favourable in Thoroughbred races than in harness races due to the difference in starting method. This study highlights the importance of studying associations in different breeds even if they are used for similar purposes.

9. Acknowledgements

I would like to thank following people that had contributed to this project:
My supervisors Gabriella Lindgren, Brandon Velie and Kim Jäderkvist Fegraeus for all their help and support during this project.

Chameli Lawrence, my master student colleague, for her support and for the collaboration and giving me access to the *DMRT3* genotypes from her project.

Sofia Mikko for being my examiner.

My family and friends who has supported me during this whole process. Especially my partner Nilas Sundberg who has helped and encouraged me throughout the project.

10. Study Contributions

I, the author, structured the data for analyses with custom scripts written in the statistical software R, performed the preparation and genotyping of DNA, and performed all statistical analyses. I also presented the results from the study to a research group at the Swedish University of Agricultural Sciences involved in a project on performance in Coldblooded trotters. The results from the study was also presented, by the author, at the Students World Championship 2016 at the breeding station Flyinge, Sweden.

Gabriella Lindgren, Brandon Velie and Kim Jäderkvist Fegraeus, scientists at the department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Sweden, conceived and designed the experiment for this study. The Animal Genetics Laboratory, Swedish University of Agricultural Sciences, Uppsala, Sweden and the Animal Genetics Laboratory, Norwegian University of Life Sciences, Norway contributed DNA, materials and analysis tools. Thorvaldur Árnason contributed with performance data from the Swedish- and Norwegian Trotting Association. Gabriella Lindgren, Brandon Velie and Kim Jäderkvist Fegraeus supervised the project and contributed with input on my thesis and guidance throughout the whole project.

11. References

Andersson, L. S, Larhammar, M., Memic, F., Wootz, H., Schwochow, D., Rubin, C-J., Patra, K., Arnason, T., Wellbring, L., Hjälm, G., Imsland, F., Petersen, J. L., McCue, M. E., Mickelson, J. R., Cothran, G., Ahituv, N., Roepstorff, L., Mikko, S., Vallstedt, A., Lindgren, G., Andersson, L. and Kullander, K. (2012) Mutations in DMRT3 affect locomotion in horses and spinal circuit function in mice, *Nature*, vol. 488, pp. 642-646.

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*, vol. 41 (Suppl. 2), pp. 154-158.

Bohlin, O. and Rönningen, K. (1975) Inbreeding and relationship within the North-Swedish Horse. *Acta Agriculturae Scandinavica*, vol. 25 (2), pp. 121-125.

Bower, M. A., McGivney, B. A., Campana, M. G., Gu, J., Andersson, L. S., Barrett, E., Davis, C. R., Mikko, S., Stock, F., Voronkova, V., Bradley, D. G., Fahey, A. G., Lindgren, G., MavHugh, D. E., Sulimova, G. and Hill, E. (2012) The genetic origin and history of speed in the Thoroughbred racehorse. *Nature Communications*, 3:643, DOI: 10.1038/ncomms1644.

Casas, E., Keele, J. W., Shackelford, S. D., Koohmarate, M., Sonstegard, T. S., Smith, T. P. L., Kapes, S. M. and Stone, R. T. (1998) Association of the muscle hypertrophy locus with carcass traits in beef cattle, *J. Anim. Sci*, vol. 76 (2), pp. 468-473.

Clop, A., Marcq, F., Takeda, H., Pirottin, D., Tordoir, X., Bibé, B., Bouix, J., Caiment, F., Elsen, J-M., Eychenne, F., Larzul, C., Laville, E., Meish, F., Milenkovic, D., Tobin, J., Charlier, C. and Georges, M. (2006) A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nature Genetics*, vol. 38, pp. 813-818.

Culley, G. (1807) *Observations on Livestock: Containing Hints for Choosing and Improving the Best Breeds of the Most Useful Kinds of Domestic Animals*, 4th ed. London: G. Woodfall.

EUROINVESTOR (2016-06-03) *Omvandlare.* Available: http://www.valuta.se [2016-06-03]

François, L., Jäderkvist Fegraeus, K., Eriksson, S., Andersson, L. S., Tesfayonas, G. Y., Viluma, A., Imsland, F., Buys, N., Mikko, S., Lindgren, G. and Velie, B. D. (2016) Conformation Traits and Gaits in the Icelandic Horse are Associated with Genetics Variants in *Myostatin (MSTN). Journal of Heredity*, vol. 00, pp. 1-7.

González, J. R., Armengol, L., Solé, X., Guíno, E., Mercader, J. M., Estivill, X. and Moreno, V. (2007) *SNPassoc: an R package to perform whole genome association studies.* Available: http://davinci.crg.es/estivill_lab/tools/SNPassoc/SupplementaryMaterial.pdf [2016-05-13]

Grobet, L., Martin, L. J. R., Poncelet, D., Pirottin, D., Brouwers, B., Riquet, J., Schoeberlein, A., Dunner, S., Ménissier, F., Massabanda, J., Fires, R., Hanset, R. and Georges, M. (1997) A deletion in the bovine Myostatin gene causes the double-muscled phenotype in cattle, *Nature genetics*, vol. 17, pp. 71-74.

Hendricks, B. L. (2007) *International Encyclopedia of Horse Breeds*, Norman: University of Oklahoma Press.

Hill, E. W., Fonseca, R. G., McGivney, B. A., Gu, J., MacHugh, D. E. and Katz, L. M. (2011) *MSTN* genotype (g.66493737C/T) association with speed indices in Thoroughbred racehorses. *J Appl Physiol*, doi: 10.1152/japplphysiol.00793.2011.

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*, *v*ol. 5(1), Available: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0008645 (2016-06-06)

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, *BMC Genomics*, vol. 11, pp. 552.

Kambadur, R., Sharma, M., Smith, T. P. L. and Bass, J. J. (1997) Mutations in Myostatin (GDF8) in Double-Muscled Belgian Blue and Piedmontese Cattle. *Cold Spring HArbor Laboratory Press*, vol. 7, pp. 910-915.

Jäderkvist, K., Andersson, L. S., Johansson, A. M., Árnason, T., Mikko, S., Eriksson, S., Andersson, L. and Lindgren, G. (2014) The DMRT3 'Gait keeper' mutation effects performance of Nordic and Standardbred trotters, *J. Anim. Sci*, vol. 92, pp. 4279-4286.

Jäderkvist, K., Johansson, L., Mäenpää, M., Mykkänen, A., Andersson, L. S., Velie, B. D., Andersson, L., Árnason, T. and Lindgren, G. (2015) Different DMRT3 Genotypes Are Best Adapted for Harness Racing and Riding in Finnhorses, *Journal of Heredity*, vol. 2015, pp. 734-740.

McPherron, A. C. and Lee, S-J. (1997) Double muscling in cattle due to mutations in the myostatin gene. *Proc. NAtl. Acad. Sci. USA*, vol. 94, pp. 12457-12461.

McPherron, A. C., Lawler, A. M., Lee, S-J. (1997) Regulation of skeletal muscle mass in mice by a new TGF-β superfamily member, *Letters to nature*, vol. 387, pp. 83-90.

Mosher, D. S., Quignon, P., Bustamante, C. D., Sutter, N. B., Mellersh, C. S., Parker, H. G. and Ostrander, E. A. (2007) A Mutation in the Myostatin Gene Increases Muscle Mass and Enhances Racing Performance in Heterozygote dogs. *PLOS Genetics*, vol. 3(5): e79. Doi: 10.1371/journal.pgen.0030079.

Petersen, J. L., Valberg, S. J., Mickelson, J. R. And McCue, M. E. (2014) Haplotype diversity in the equine myostatin gene with focus on variants associated with race distance propensity and muscle fiber type proportions. *Animal Genetics*, vol. 45, pp. 827-835

Polasik, D., Pikula, R., Gawlik, J., Ochman, J. and Terman, A. (2015) Analysis of the myostatin gene(*MSTN*) polymorphismin four breeds of horses. *Folia Pomer. Univ. Technol. Stetin., Agric., Aliment., Pisc., Zootech*, vol. 320(35)3, pp. 81-86.

R Development Core Team (2015). *R: A Language and Enviroment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Revold, T., Larsen, S. and Ihler, C. F. (2010) Prediction of early race starts in Norwegian-Swedish Coldblooded Trotters, *Acta Vet Scand*, vol. 52:53.

Schuelke, M., Wagner, K. R., Stolz, L. E., Hübner, C., Riebel, T., Kömen, W., Braun, T., Tobin, J. F. and Lee,.S-J. (2004) Myostatin Mutation Associated with Cross Muscle Hypertrophy in a Child. *The New England Journal of Medicine*, vol. 350, pp. 2682-2688.

The Swedish Horse Racing Authority (2013-10-16) *Hästen – Det engelska fullblodet.* Available: https://www.svenskgalopp.se/artikel/hasten;jsessionid=mnyPXYgMzF1ltBpLDdg1cvdJrdGCtKTKGbhSWrPZ4vhWHftS7FtX!1975696350?defaultMenuId=true [2016-06-08]

The Swedish Trotting Association (2014-02-12) *Avelsindex.* Available: https://www.travsport.se/artikel/avelsindex [2016-05-10]

The Swedish Trotting Association (2015-02-03) *Hästarna.* Available: https://www.travsport.se/artikel/hastarna [2016-05-23]

The Swedish Trotting Association (2016) *Tävlingsreglemente.* Available: https://www.travsport.se/polopoly_fs/1.575!/menu/standard/file/tavlingsreglemente_2016_webb.pdf [2016-05-09]

Thiruvenkadan, A. K., Kandasamy, N., Panneerselvam, S. (2009) Inheritance of racing performance of trotter horses: An overview. *Livestock Science*, vol. 124, pp. 163-181.

Thomas, M., Langley, B., Berry, C., Sharma, M., Kirk, S., Bass, J. and Kambadur, R. (2000) Myostatin, a Negative Regulator of Muscle Growth, Functions by Inhibiting Myoblast Proliferation, The Journal of Biological Chemistry, vol. 257, pp. 40235-40243.

Tozaki, T. Hill, E. W., Hirota, K., Kakoi, H., Gamahara, H., Miyake, T., Sugita, S., Hasegawa, T., Ishida, N., Nakano, Y. And Kurosawa, M. (2011a) A cohort study of racing performance in Japanese Thoroughbred racehorses using genome information on ECA18. *Animal Genetics*, vol. 43, pp. 42-52.

Tozaki, T., Sato, F., Hill, E. M., Miyake, T., Endo, Y., Kakoi, H., Gawahara, H., Hirota, K., Nakano, Y., Nambo, Y. and Kurosawa, M. (2011b) Sequence Variants at the *myostatin* Gene Locus Influence the Body Composition of Thoroughbred Horses, *J. Vet. Med. Sci*, vol. 73(12), pp. 1617-1624.

Velie, B. D., Jäderkvist, K., Imsland, F., Viluma, V., Andersson, L. S., Mikko, S., Eriksson, S. and Lindgren, G. (2015) Frequencies of polymorphisms in myostatin vary in Icelandic horses according to the use of the horses, *Animal Genetics*, vol. 46 (4).

12. Supplementary Table 1

Distribution of gender and country of birth over the sample

Country of Birth	Swedish (SWE)	SWE/Total %	Norwegian (NOR)	NOR/Total %	Total	%
Gelding	173	35.89%	74	15.35%	247	51.24%
Stallion*	18	3.73%	29	6.02%	47	9.75%
Mare	120	24.90%	68	14.11%	188	39.00%
Total	311	64.52%	171	35.48%	482	100 %

*Including 1 cryptorchid (Norwegian)

13. Supplementary Table 2

Distribution of birth year over the sample

Country of Birth	SWE	SWE/Total (%)	NOR	NOR/Total (%)	Total	%
2000	30	6.22%	3	0.62%	33	6.85%
2001	31	6.43%	25	5.19%	56	11.62%
2002	34	7.05%	19	3.94%	53	11.00%
2003	37	7.68%	19	3.94%	56	11.62%
2004	28	5.81%	13	2.70%	41	8.51%
2005	28	5.81%	20	4.15%	48	9.96%
2006	38	7.88%	22	4.56%	60	12.45%
2007	32	6.64%	14	2.90%	46	9.54%
2008	29	6.02%	19	3.94%	48	9.96%
2009	24	4.98%	17	3.53%	41	8.51%
Total	311	64.52%	171	35.48%	482	100 %

14. Supplementary Table 3

Summary statistics of the performance traits for the successfully genotyped sample (n=482)

Total	Min	1st Q	Median	Mean	3rd Q	Max
Starts						
Population	1	13	28	38	54	222
Stallions	2	27	67	65	84	186
Mares	1	9	22	28	42	118
Geldings	1	15	30	41	57	222
Placings (1-3)						
Population	0	2	7	11	15	101
Stallions	0	11	23	27	37	101
Mares	0	2	6	8	11	37
Geldings	0	3	7	11	15	72
Wins						
Population	0	0	2	4	5	53
Stallions	0	3	9	11	16	53
Mares	0	0	1	2	4	17
Geldings	0	0	2	4	5	40
Unplaced						
Population	0	5	14	19	27	133
Stallions	0	9	23	27	37	92
Mares	0	3	12	15	23	62
Geldings	0	7	15	20	29	133
Race time volt (seconds)						
Population	81,40	88,00	90,20	90,60	93,00	106,90
Stallions	81,40	83,98	85,60	86,72	88,95	96,60
Mares	3,70	88,70	90,50	91,57	93,80	106,90
Geldings	83,10	88,10	90,30	90,68	92,80	104,40
Race time auto (seconds)						
Population	80,10	86,50	88,60	88,97	90,90	177,40
Stallions	80,10	83,02	84,80	84,95	87,00	92,60
Mares	83,70	87,90	89,40	89,80	91,70	99,30
Geldings	80,20	86,75	88,75	89,42	91,10	177,40
Earnings (SEK)						
Population	0	31210	97730	232200	250800	3189000
Stallions	8800	130400	517700	799400	1223000	2997000
Mares	0	21880	71360	140700	167600	1826000
Geldings	0	34210	93290	193500	229000	3189000
Earnings/ start						
Population	0	1982	3529	4797	5613	57070
Stallions	1021	5810	9242	11610	15120	53540
Mares	0	1599	3158	4191	5238	57070
Geldings	0	1970	3300	3944	4944	26660
Blup						
Population	95	107	112	111.8	117	126
Stallions	106	112	117	116.3	121	126
Mares	95	108	112	112.4	117.2	126
Geldings	95	106	110	110.4	115	124
Inbreeding						
Population	0	4.67	5.71	6.05	7.11	17.01
Stallions	2.36	5.12	5.83	6.38	7.24	12.10
Mares	1.17	4.76	5.68	6.10	7.30	16.06
Geldings	0	4.57	5.74	5.95	6.95	17.01

15. Supplementary Table 4

Summary statistics of the performance traits for the genotyped horses born in Sweden (n=311)

Swedish	Min	1st Q	Median	Mean	3rd Q	Max
Starts						
Population	1	10	22	32	46	210
Stallions	2	12	28	44	78	150
Mares	1	5	17	24	37	118
Geldings	1	14	26	37	52	210
Placings (1-3)						
Population	0	2	5	9	12	72
Stallions	0	6	14	20	33	61
Mares	0	0	3	6	8	37
Geldings	0	2	5	10	14	72
Wins						
Population	0	0	1	3	4	40
Stallions	0	2	7	8	14	24
Mares	0	0	1	2	2	14
Geldings	0	0	2	4	5	40
Unplaced						
Population	0	2	6	9	12	66
Stallions	0	3	6	9	10	37
Mares	0	2	5	7	11	36
Geldings	0	3	8	9	13	66
Earnings (SEK)						
Population	0	22650	67850	178700	183700	3189000
Stallions	8800	80750	292800	599400	997900	2114000
Mares	0	5750	42250	112200	133700	1148000
Geldings	0	31200	75200	181000	193000	3189000
Earnings/start						
Population	0	1590	3152	4022	5373	26660
Stallions	1021	8023	10980	10690	14180	25220
Mares	0	908	2588	3149	5163	9726
Geldings	0	1838	3128	3934	4826	26660
Race time volt (seconds)						
Population	82.20	88.80	91.00	91.34	93.78	106.90
Stallions	82.20	84.45	88.20	88.04	91.85	96.60
Mares	83.70	89.00	92.15	92.45	94.85	106.90
Geldings	83.10	88.82	90.70	90.99	93.07	104.40
Race time auto (seconds)						
Population	80.20	87.20	89.60	89.53	92.10	99.30
Stallions	80.90	83.40	85.30	86.02	88.30	92.60
Mares	83.70	88.10	89.90	90.14	92.80	99.30
Geldings	80.20	87.20	89.60	89.61	91.90	97.20
Blup						
Population	95.00	106.00	109.00	109.7	114.00	125.00
Stallions	106.00	109.50	113.50	113.80	118.50	125.00
Mares	95.00	106.00	110.00	110.00	114.00	125.00
Geldings	95.00	105.00	109.00	109.00	113.00	124.00
Inbreeding						
Population	1.17	4.47	5.38	5.79	6.83	17.01
Stallions	2.36	5.08	5.60	6.05	6.80	10.85
Mares	1.17	4.53	5.79	6.84	16.06	2.07
Geldings	1.21	4.43	5.36	5.76	6.8	17.01

16. Supplementary Table 5

Summary statistics of the performance traits for the genotyped horses born in Norway (171)

Norwegian	Min	1st Q	Median	Mean	3rd Q	Max
Starts						
Population	1	18	42	48	67	222
Stallions	8	44	72	78	92	186
Mares	2	15	33	34	47	88
Geldings	1	17	45	50	65	222
Placings (1-3)						
Population	0	6	11	15	21	101
Stallions	3	15	27	31	41	101
Mares	0	4	10	10	14	34
Geldings	0	4	12	14	20	70
Wins						
Population	0	1	3	5	7	53
Stallions	1	5	10	13	20	53
Mares	0	1	2	3	4	17
Geldings	0	1	3	5	6	25
Unplaced						
Population	0	5	12	15	20	85
Stallions	2	10	19	22	28	56
Mares	0	4	9	11	15	37
Geldings	0	4	12	15	22	85
Earnings						
Population	0	62 540	155 800	329400	348 400	2 997 000
Stallions	32 510	281 100	541 600	928000	1 480 000	2 997 000
Mares	5 000	52 710	121 800	191000	214 600	1 826 000
Geldings	0	62 360	132 500	222900	309 900	1 630 000
Earnings/start						
Population	0	2 360	3 830	6207	6 112	57 070
Stallions	2270	5 352	6907	12200	16770	53540
Mares	1161	2 381	3 451	6030	5 270	57 070
Geldings	0	2 068	3 632	3969	5 133	15 000
Race time volt (seconds)						
Population	81.40	86.90	89.00	89.32	91.20	104.00
Stallions	81.40	83.72	85.35	85.87	87.17	94.20
Mares	85.10	88.50	89.50	90.17	91.40	102.10
Geldings	83.40	87.40	89.30	89.95	91.70	104.00
Race time auto (seconds)						
Population	80.10	85.60	87.70	88.24	89.50	177.40
Stallions	80.10	82.80	83.70	84.43	86.45	91.70
Mares	84.60	87.82	88.75	89.41	90.40	98.20
Geldings	81.80	86.05	87.20	89.05	89.25	177.40
Blup						
Population	101.00	112.00	116.00	115.5	119.00	126.00
Stallions	109.00	113.00	118.00	118,00	122.2	126.00
Mares	107.00	113.00	117.00	116.7	119.2	126.00
Geldings	101.00	110.00	113.00	113.5	117.00	123.00
Inbreeding						
Population	0,00	5.34	6.13	6.53	7.53	14.33
Stallions	3.7	5.35	6.03	6.59	7.33	12.10
Mares	3.43	5.30	6.27	6.65	7.77	14.33
Geldings	0.00	5.36	6.03	6.38	7.31	12.11

17. Supplementary Table 6

Min, max, mean and median for all the traits analysed stratified by genotype in the three SNPs

SNP3737	TT (n=459)				CT (n=23)			
Total	**Min**	**Mean**	**Median**	**Max**	**Min**	**Mean**	**Median**	**Max**
No of starts	1	38.14	28.00	222	4	36.22	24.00	85
No of victories	0	3.95	2.00	53	0	5.26	3.00	29
Victories (freq.)	0	0.09	0.07	1	0	0.14	0.09	0.75
No of placings 1-3	0	11.22	7.00	101	0	13.04	8.00	56
Placings 1-3 (freq.)	0	0.26	0.25	1	0	0.32	0.32	0.75
Unplaced	0	18.95	14.00	133	0	16.65	14.00	38
Unplaced (freq)	0	0.51	0.50	1	0	0.47	0.45	0.76
Earnings (SEK)	0	230600	97850	3189000	0	244097	90000	1261000
Earnings/Start	0	47055	3500	57070	0	6292	3750	33910
Blup	95.0	111.7	112.0	126.0	97.0	112.8	114.0	124.0
Race time auto (sec)	80.1	89.1	86.50	177.40	83.3	87.5	85.1	92.9
Race time volt (sec)	81.4	90.6	90.0	106.9	83.9	90.6	89.3	104.4

SNP8604	TT (n=425)				TG (n=57)			
Total	**Min**	**Mean**	**Median**	**Max**	**Min**	**Mean**	**Median**	**Max**
No of starts	1	37.41	28.00	210	1	42.77	30.00	222
No of victories	0	3.91	2.00	53	0	4.72	2.00	24
Victories (freq.)	0	0.09	0.06	1	0	0.12	0.10	0.75
No of placings 1-3	0	11.09	7.00	101	0	12.89	8.00	53
Placings 1-3 (freq.)	0	0.26	0.25	1	0	0.28	0.27	0.75
Unplaced	0	18.55	14.00	121	0	21.00	13.00	133
Unplaced (freq.)	0	0.57	0.50	1	0	0.48	0.50	1
Earnings (SEK)	0	224892	97617	3189000	0	278612	99800	2678000
Earnings/Start	0	4670	3436	57070	0	5612	4257	28980
Blup	95.0	111.9	112.00	126.0	95.0	110.4	110.0	123.0
Race time auto (sec)	80.1	89.1	86.5	177.40	80.2	87.9	84.8	95.4
Race time volt (sec)	81.4	90.6	90.0	106.9	82.4	90.3	89.6	105.7

SNP5826	AA (n=390)				AG (n=88)				GG (n=4)			
Total	**Min**	**Mean**	**Median**	**Max**	**Min**	**Mean**	**Median**	**Max**	**Min**	**Mean**	**Median**	**Max**
No of starts	1	38.02	29.00	222	1	38.67	25.50	186	7	26.75	24.00	52
No of victories	0	3.83	2.00	40	0	4.72	2.00	53	1	5.50	5.00	11
Victories (freq.)	0	0.09	0.07	1	0	0.09	0.06	0.31	0.13	0.21	0.14	0.42
No of placings 1-3	0	11.24	7.00	72	0	11.75	6.00	101	1	8.00	9.00	13
Placings 1-3 (freq.)	0	0.26	0.25	1	0	0.25	0.23	1	0.14	0.28	0.24	0.50
Unplaced	0	19	14.00	133	0	19.24	13.00	86	4	12.25	10.00	25
Unplaced (freq.)	0	0.51	0.50	1	0	0.49	0.50	1	0.31	0.48	0.51	0.57
Earnings (SEK)	0	227800	101900	3189000	0	254500	76720	2997000	26000	163900	171700	286200
Earnings/Start	0	4850	3544	57070	0	4510	3264	22300	3714	5938	4515	11010
Blup	95.0	111.7	112.0	126.0	95.0	111.8	111.5	125.0	110.0	113.2	114.0	115.0
Race time auto (sec)	80.2	89.0	88.6	177.4	80.10	88.7	88.3	99.3	88.3	90.1	88.4	93.7
Race time volt (sec)	81.4	90.6	90.1	106.9	82.7	90.5	90.6	102.8	88.8	92.0	90.8	97.60