Equine Trait Mapping

From Disease Loci to the Discovery of a Major Gene Controlling Vertebrate Locomotion

Lisa S Andersson
Faculty of Veterinary Medicine and Animal Science
Department of Animal Breeding and Genetics
Uppsala

Doctoral Thesis
Swedish University of Agricultural Sciences
Uppsala 2012

Acta Universitatis agriculturae Sueciae
2012:65

ISSN 1652-6880
ISBN 978-91-576-7712-9

Print: SLU Service/Repro, Uppsala 2012

Equine Trait Mapping. From Disease Loci to the Discovery of a Major Gene Controlling Vertebrate Locomotion.

Abstract

Assigning function to genes is essential for a better understanding of biological systems. To date, approximately half of the genes in the vertebrate genome have known function. Domestic animals are a rich source for trait mapping and in this thesis we have mapped three distinct equine phenotypes. The result provides increased knowledge regarding gene function and importantly, practical implications for horse welfare. In paper I and IV, we confirm that Equine Multiple Congenital Ocular Anomalies (MCOA) syndrome is inherited as an incompletely dominant trait ($p= 2.2x10^{-16}$). By first identifying a 208 kb identity-by decent (IBD) region and subsequently excluding polymorphic sites identified through Illumina sequencing, we conclude that the gene *PMEL* causes these defects in horse. Our findings, together with functional analyses recently published, support that the cause of MCOA syndrome is a missense mutation (Arg625Cys) near the transmembrane region of PMEL that results in altered biochemical properties. In paper II we show that variants in the MHC-II region influence the susceptibility to equine Insect Bite Hypersensitivity with the same marker risk allele identified in two distinct populations, OR 4.19 ($p= 2.3x10^{-5}$) and 1.48 ($p= 0.04$) for Icelandic horses and Exmoor ponies respectively. In addition, homozygosity across the MHC-II region confers a higher risk of developing disease, OR= 2.67 ($p= 1.3x10^{-3}$). Finally, in paper III we utilize the EquineSNP50 BeadChip to identify the first *Gait* locus in horse. A highly significant SNP ($EMP2= 2.0x10^{-4}$) was identified to be consistent with a recessive mode of inheritance for the lateral gait pace in Icelandic horses, and confirmed in an independent sample set ($p= 2.4x10^{-14}$). Illumina sequencing of an established IBD region identified a nonsense mutation in the gene *DMRT3*. A clearly dichotomous distribution in a panel of gaited and non-gaited breeds revealed that the *DMRT3* mutation is permissive for a variety of alternate gaits. The mutation also has a favorable effect in harness racing horses. Functional characterization of the truncated protein demonstrated correct localization and an intact DNA binding profile. mRNA expression in a small population of commissural neurons from the spinal cord was confirmed in mutant and wild type horses. Further, a *DMRT3* null mouse displayed a change in spinal cord circuit signaling and locomotion. These findings reveal a new molecule involved in the regulation of limb movement.

Keywords: horse, association mapping, GWAS, MCOA, Multiple Congenital Ocular Anomalies, IBH, Insect Bite Hypersensitivity, pace, locomotion, athletic performance

Author's address: Lisa Andersson, SLU, Department of Animal Breeding and Genetics, P.O. Box 597, 751 24 Uppsala, Sweden. *E-mail:* Lisa.Andersson@slu.se

Contents

List of Publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I **Andersson LS**, Lyberg K, Cothran G, Ramsey DT, Juras R, Mikko S, Ekesten B, Ewart S, Lindgren G (2011). Targeted analysis of four breeds narrows equine Multiple Congenital Ocular Anomalies locus to 208 kilobases. *Mammalian genome* 22 (5-6):353-360.

II **Andersson LS**, Swinburne JE, Meadows JR, Brostrom H, Eriksson S, Fikse WF, Frey R, Sundquist M, Tseng CT, Mikko S, Lindgren G (2011). The same ELA class II risk factors confer equine insect bite hypersensitivity in two distinct populations. *Immunogenetics 64(3), 201-208*.

III **Andersson LS**, Larhammar M, Memic F, Wootz H, Schwochow D, Rubin CJ, Patra K, Arnason T, Wellbring L, Hjälm G, Imsland F, Petersen JL, McCue ME, Mickelson JR, Cothran G, Ahituv N, Roepstorff L, Mikko S, Vallstedt A, Lindgren G, Andersson L, Kullander K (2012). Mutations in *DMRT3* affect locomotion in horses and spinal circuit function in mice. *Accepted in Nature.*

IV **Lisa S Andersson**, Maria Wilbe, Agnese Viluma, Gus Cothran, Björn Ekesten, Susan Ewart, Gabriella Lindgren. Next Generation Sequencing Establishes Equine Multiple Congenital Ocular Anomalies (MCOA) is Caused by Mutant *PMEL* (*manuscript*).

Papers I-III are reproduced with the permission of the publishers.

Related Work by the Author

(Not included in the thesis)

Andersson LS, Juras R, Ramsey DT, Eason-Butler J, Ewart S, Cothran G, Lindgren G (2008) Equine Multiple Congenital Ocular Anomalies maps to a 4.9 megabase interval on horse chromosome 6. *BMC genetics* 9:88.

Andersson LS, Hogstrom C, Mikko S, Eriksson S, Grandinson K, Brostrom H, Frey R, Sundquist M, Lindgren G (2009) Polymorphisms in *SPINK5* do not associate with insect bite hypersensitivity in Icelandic horses born in Sweden. *Animal genetics* 40 (5):790-791.

Heimann M, Janda J, Sigurdardottir OG, Svansson V, Klukowska J, von Tscharner C, Doherr M, Brostrom H, **Andersson LS**, Einarsson S, Marti E, Torsteinsdottir S (2011) Skin-infiltrating T cells and cytokine expression in Icelandic horses affected with insect bite hypersensitivity: a possible role for regulatory T cells. *Veterinary immunology and immunopathology* 140 (1-2):63-74.

Andersson LS, Axelsson J, Dubielzig RR, Lindgren G, Ekesten B (2011) Multiple congenital ocular anomalies in Icelandic horses. *BMC veterinary research* 7:21.

Abbreviations

CI	confidence interval
CPG	central pattern generator
DMRT3	doublesex and mab-3 related transcription factor 3
DNA	deoxyribonucleic acid
ELA	equine leukocyte antigen
EMSA	electrophoretic mobility shift assay
GFP	green fluorescent protein
GWAS	genome-wide association study
IBD	identity by decent
IBH	Insect Bite Hypersensitivity
IGF2	insulin-like growth factor 2
kb	kilo bases
LD	linkage disequilibrium
Mb	mega bases
MCOA	Multiple Congenital Ocular Anomalies
MHC	major histocompatibility complex
mRNA	messenger ribonucleic acid
OR	odds ratio
PCR	polymerase chain reaction
PMEL	premelanosome protein
qPCR	quantitative polymerase chain reaction
QTL	quantitative trait locus
QTN	quantitative trait nucleotide

RPE	retinal pigment epithelium
SNP	single nucleotide polymorphism
TMD	transmembrane domain
wt	wild type

Introduction

Genomic studies in horses have, in many regards, been facilitated by similar studies in other species such as in human and mouse. The genome projects of these species, and later that of the horse, were initiated with the construction of whole genome linkage maps, mainly consisting of microsatellite marker data. Following the development of Sanger sequencing technology, a draft version of the human genome was available in 2000 (Lander *et al.*, 2001; Venter *et al.*, 2001). The horse was actually one of the first placental animals with a fully sequenced genome and a draft version was completed by 2007 (Wade *et al.*, 2009). As part of these sequencing efforts, a large number of single nucleotide polymorphisms (SNPs) were identified and commercial SNP chips developed. These technologies revolutionized genetic trait mapping, in part by speeding up the previously very time consuming and labor intense genotyping process. Today, extremely high throughput sequencing, which produce an increasing amount of data at a rapidly declining cost, have become an attractive alternative to SNP chips. For instance, the human genome sequencing funded by Celera Genomics took three years to complete and had a budget of approximately US$300,000,000. In 2011, we sequenced the whole genome of two horses in just a few days and at a cost of approximately $18,000. The past 20 years has truly been a golden age for molecular genetics. The remarkable developments in technology have been followed by a burst in the number of genes identified to be associated with phenotypic traits and improved our understanding of the genome in terms of regulation, recombination, mutation, transposable elements and evolution.

However, further research is required before we fully understand the human genome. To date, we only have functional information, vague for some, on approximately half of the ~21,000 genes in the mammalian genome (Alberts, 2011). There is an interesting two-way relationship joining the research of the

human genome and that of domestic animals such as the horse. The human genome was instrumental in the annotation of genes within the horse genome; over 16,000 are orthologous (Wade *et al.*, 2009). However, genetic research in domestic animals and in other model organisms can be integral to the functional annotation of the human genome i.e. what is the purpose of each gene product. Functional annotation is vital in order to increase our knowledge about the human body and assist in disease prevention, diagnostics and treatment. Forward genetics, such as knocking out genes in mouse and zebrafish, is one way of characterizing gene functions. An alternative approach is reverse genetics, starting with the recognition of an interesting heritable trait followed by the identification of the causative gene and mutation. The identification of causal variants sometimes takes advantage of comparative information from similar disorders or traits in other species. All mapped traits will add to this knowledge bank and will help scientists working with different species worldwide.

Molecular genetic research in domestic animals is also important from an animal welfare and/ or breeding scheme perspective. Once genomic mutations or even genomic regions harboring quantitative trait loci (QTLs) have been identified, marker assist selection can be utilized to select for desirable traits and against disease. This is also important from an ethical perspective, since it is our responsibility to breed healthy animals suited to their purpose. The aim of the studies included in this thesis was thus two-fold. First, to use the power of domestic animal genetics to assign function to mammalian genes and secondly, to use the results to address equine breeding schemes and animal welfare concerns.

The use of domestic animals for assigning function to genes

Selection has caused an accumulation of mutations

Domestic animals are a fantastic resource for gene mapping. For thousands of years, since the dawn of domestication, humans have selected spontaneous mutations in farm and companion animals. These mutations include those with an obvious phenotypic effect, such as coat color, but also those with a more subtle effect on traits such as behavior, disease resistance and production. Horses have similarly been selected for athletic performance traits. As a result, domesticated animals are much more phenotypically diverse than their wild ancestors. Artificial selection has caused an accumulation of both positive and negative mutations and has thus created a pool that can be exploited in gene

mapping studies. The results from such experiments may serve to elucidate the function of different genes in the mammalian genome, including humans (Andersson & Georges, 2004; Andersson, 2001). Bottlenecks, genetic drift and natural adaptations further add to the phenotypic diversity seen in domestic animals.

A fundamental reason for using domestic animals in disease trait mapping lies in the fact that loci with large effects are more easily mapped than loci with small effects. Because of small population sizes and strong human artificial selection, numerous highly penetrant mutations (causing different traits and diseases) are present in relatively high frequency in domestic animals compared to the human race. Genetic hitchhiking or the extensive spread of recessive alleles through the use of popular sires, will also affect a population's gene frequency. A recessive mutation has to reach a sufficiently high allele frequency before breeders become aware of the phenotypic problem, however once such disease is identified, it can now be quite straightforward to map the causative loci using only a few diseased and healthy animals (Charlier *et al.*, 2008). In yet other cases, highly penetrant, deleterious mutations can actually be selected for unintentionally. This occurs when a mutation that confers a valuable trait also has negative pleiotropic effects. Genetically this can be seen as heterozygous individuals displaying the valuable trait whilst homozygous individuals display an additional disease phenotype. The Rhodesian ridgeback dog breed provides one such example. The desired dorsal hair ridge is caused by a dominant 133 kilo base (kb) duplication, however dogs that are homozygous for the duplication have a much higher risk of developing the neural-tube defect, dermoid sinus (Salmon Hillbertz *et al.*, 2007). Multiple Congenital Ocular Anomalies (MCOA) syndrome characterized in paper I is another example. It has a high frequency in several horse breeds as the same mutation also causes the valued Silver coat color.

Why are domestic animals a good model?

There are several reasons why domestic animals are of significant value for trait mapping compared to humans (Andersson, 2009; Goddard & Hayes, 2009; Georges, 2007):

- Easy access to large pedigrees, either through the creation of an experimental cross or by taking advantage of existing farm animals and their corresponding pedigree records. Pedigrees are particularly valuable for linkage mapping, but they also facilitate the phasing of alleles.

II Easier to collect tissues samples compared to in humans. However, for companion animals such as dogs and horses, some tissues are obviously problematic to collect.

III Higher linkage disequilibrium (LD), especially within breeds. Between breeds the LD is lower. This makes mapping in domestic animals favorable since one can take advantage of the high LD within breeds for genome scans and the low LD between breeds for fine mapping, a so called two-stage approach (Lindblad-Toh *et al.*, 2005).

IV Often reduced environmental noise because of comparable housing across the whole population.

V An enrichment of mutations with large effects.

VI The same trait is often present in several breeds, which facilitates both mapping and the confirmation of genetic results. This phenomena can be due to gene flow or because breeds share a common ancestor that carried the trait-causing mutation. In horses, one or a few individuals (usually stallions), with a particular desired trait are commonly intercrossed between breeds to improve it. In fact, there is not a single known coat color unique to any one breed (Bowling & Ruvinsky, 2000).

VII Less genetic heterogeneity including (i) locus heterogeneity, which is the case when different genes are responsible for similar phenotypes and (ii) allelic heterogeneity, which is the presence of different alleles within a single gene causing similar phenotypes. In humans genetic heterogeneity is common. For example, over 1,000 different mutations in the gene *cystic fibrosis transmembrane conductance regulator* (*CFTR*) are associated with cystic fibrosis (Andersson, 2009). In contrast, many breeds of domestic animals have small effective population sizes and thus a high degree of homogeneity.

Most of the facts presented above have been taken advantage of in order to complete either the research presented in this thesis, or the preliminary experiments upon which this thesis work is based. For example, pedigrees were used in the linkage analysis of MCOA syndrome (not included here) and in the phasing of major histocompatibility complex (MHC) class II haplotypes (paper II), tissues from both live and slaughtered animals were collected for all

projects and the relatively high LD in horse was explored in the pace genome-wide association study (GWAS) (paper III). In addition, in the study of MCOA syndrome (paper I) we used haplotype analysis across four different breeds to narrow the genetic interval and in the pace study (paper IV) we genotyped additional horse breeds and found the same causative mutation in a large subset of these.

The clear benefits of domestic animals compared to human populations for trait mapping have been highlighted, however, when performing this task in any species there are several aspects that must be taken into consideration. For instance, when setting up the experimental design the predicted inheritance pattern, genome structure of the particular animals and resources at hand must all be weighed.

Mapping traits genes

Complex versus monogenic traits

A complex trait, also called a multifactorial or quantitative trait, depends on the cumulative action of many genes as well as environmental factors. In contrast, a monogenic trait, also called a Mendelian trait, is not influenced by the environment and is affected by only one gene. The study of monogenic traits began with the Austrian scientist Gregor Mendel and his research on peas and flowers at the beginning of the 19th century. Monogenic traits have been essential in revealing the mechanisms of heredity and represent the majority of mapped mutations. Thousands of QTL have also been mapped (Andersson, 2009) but in only a small fraction of these has the causative mutation, the quantitative trait nucleotide (QTN), been identified. Three commonly cited studies where the QTN has been determined are the muscle growth influencing *insulin-like growth factor 2* (*IGF2)* loci in pigs and the *myostatin* (*MSTN*) gene in sheep (Clop *et al.*, 2006; Van Laere *et al.*, 2003) and the milk fat content effecting *acyl-coenzyme A:diacylglycerol Acyltransferase* (*DGAT1*) gene in cattle (Grisart *et al.*, 2002).

Additional QTNs have been identified through their effect on monogenic disorders. An example of such is the missense mutation in the *ryanodine receptor 1* (*RYR1*) gene causing a monogenic disorder called malignant hyperthermia; later proved to influence lean muscle mass in pigs (Fujii *et al.*, 1991). Another example is illustrated by the mutations in the gene *serine peptidase inhibitor, kazal type 5* (*SPINK5*). They cause a rare recessive

disorder called Netherton syndrome (Chavanas *et al.*, 2000), and were later also associated with atopic dermatitis (Walley *et al.*, 2001). We investigated if polymorphisms in this gene were associated with equine Insect Bite Hypersensitivity (IBH), but no such correlation could be established (Andersson *et al.*, 2009). The *doublesex and mab-3 related transcription factor 3* (*DMRT3*) nonsense mutation described in paper III of this thesis, is another example that can be added to the list of known QTNs. It was first identified when we studied the recessive trait "pace" in Icelandic horses. We have now shown that the *DMRT3* QTN has a large effect on athletic performance in harness racing horses. Similarly, the *myostatin* gene has been shown to be involved in athletic performance traits in horse with different genotypes correlating with optimum race distance in Thoroughbreds (Binns *et al.*, 2010; Hill *et al.*, 2010).

The lack of significant association in many GWAS of complex traits may support Fisher's traditional infinitesimal model which assumes a very large (infinite) number of loci each have extremely small effects. However, the fact that QTNs with a measurable effect on quantitative traits have been identified supports the alternate so-called quasi- infinitesimal model i.e. the segregation of genes with large effects for complex traits. The relative weight between these two theories has long been debated. Our finding, that the *DMRT3* nonsense mutation has a large effect on racing performance, supports the quasi- infinitesimal model and may give assurance that there are still genes with large effect on complex traits left to be identified.

Unfortunately, it is hard to know beforehand if there is a major locus segregating for any given trait. The heritability of a trait is a poor predictor of effect sizes. One example is height in humans. Even though it has a heritability estimate of ~80%, the hundreds of genetic variants identified to date only explain ~10% of the phenotypic variation (Lango Allen *et al.*, 2010). In contrast, the results from our study of pace (paper III) provides a very interesting example of a phenotype that was long considered to be a quantitative trait with a heritability estimate of 60% (Albertsdottir *et al.*, 2011), but by changing the phenotypic classifications we were able to identify a monogenic component. The quality of the pace is still a quantitative character, but the *DMRT3* nonsense mutation is permissive for ability to pace and has a simple recessive mode of inheritance in Icelandic horses.

Mapping strategies

There are two different major strategies most commonly adopted to map trait genes; (i) the candidate gene approach and (ii) an unprejudiced approach where one searches the entire genome in order to find a chromosomal region, and ultimately the gene, causing the phenotype of interest. The later can be performed either by genotyping evenly distributed markers or by whole genome sequencing. In the candidate gene approach, a gene is selected based on knowledge about gene function. If one manages to predict the correct gene, this approach can be extremely fast and cheap. However, it often fails due to our limited knowledge about the function of many genes. In the studies presented here, we have utilized both strategies. In paper I, we used the candidate gene approach and genotyped markers in the vicinity of the *premelanosome protein* (*PMEL*) gene in order to find the mutation causing MCOA syndrome. The *PMEL* gene had previously been selected as a candidate gene for Silver coat color because of its known association to hypopigmented phenotypes in other species. Likewise, the MHC-II region investigated in paper II, had been linked to various allergic phenotypes in other species, making it a obvious candidate region for equine IBH. In contrast, there weren't any convincing candidate genes for the ability to pace and we therefore performed a genome-wide scan using the Illumina equine50kSNP chip.

Two common methodologies for establishing connections between phenotype and genotype are linkage mapping and association mapping. Historically, when genotyping was a very laborious and time-consuming process, linkage mapping was readily used because this approach requires fewer markers. In linkage mapping one quantifies the co-segregation of marker alleles and a trait within a pedigree. When DNA is replicated and inherited from parent to child it is passed on in large chromosome blocks. This is caused by the fact that the number of recombinations in a single meiosis is relatively low. Each marker can therefore tag a large region and so fewer markers are necessary to cover the entire genome. As a consequence, the identified chromosomal region(s) are usually very large, which is a drawback in most linkage studies. In contrast, association analysis investigates subjects from a whole population and so takes advantage of historical recombination. This generally results in much shorter chromosomal regions, but requires more markers. With today's technological advances, such as SNP chips and whole genome re-sequencing, this is no longer a major problem. All studies presented in this thesis are based on association mapping.

Yet another mapping strategy is F_{st}- and selective sweep scans. These are particularly useful for finding regions with large differences in allele frequencies between populations or when one expects that there has been strong selection for a particular trait, in turn leading to long homozygous stretches. These methodologies can be utilized when a breed is thought to be fixed for a mutation; an instance where neither linkage nor association mapping can be applied because of their dependence of segregation. Sweep or F_{st} scans can be performed on both whole genome sequencing data (Rubin *et al.*, 2010) or SNP chip data (Vaysse *et al.*, 2011). One major drawback with these scans however, is that it may be difficult to establish the actual phenotypic characteristics causing significant signals.

Factors that influence GWAS

Many issues such as mutation effect size, density of informative markers, the degree of LD and the number of genotyped subjects, influence the success of GWAS. Several of these entities can, to some extent, be compensated for by each other e.g., one needs to increase the number of subjects if the mutation effect size is small or if the LD between marker and mutation is low. A major advantage of using domestic animals over humans is in fact the higher within breed LD that directly affects the number of markers needed for a whole genome scan. For example, the newest dog SNP chip contains 170,000 markers and the horse chip 74,000. In contrast, significant LD in human usually only extends for tens of kb (Gabriel *et al.*, 2002) and the newest human SNP chip therefore contains over one million markers. However, LD varies significantly between different populations even within species.

A very important aspect to keep in mind when performing GWAS is the risk of false positives due to population stratification. This phenomenon occurs when a population is divided into sub-groups displaying both different marker allele and disease frequency. This is common in domestic animals and is often caused by the use of popular sires. Population stratification can be detected by comparing the observed with the expected distribution of p-values. Obviously, false positives can also be caused when an inappropriate method is used to correct for multiple testing. This is a central aspect of GWAS when considering that many thousands of statistical tests are executed. Bonferroni correction assumes independency between tests. This is usually not the case for GWAS because of LD between adjacent markers and this correction is thus usually considered to be too stringent. In the GWAS presented in this thesis we have used permutations to correct for multiple testing.

A downside to association studies lies in the fact that it is impossible to establish association if the haplotype on which the mutation first arose, the ancestral haplotype, is common within the population. We have performed two separate GWAS for monogenic traits in horse (data not shown) and one reason for the lack of significant results may be that the disease mutation is not tagged by any unique SNP. Another problematic scenario is if there are multiple causative alleles at a single locus. If these are tagged by opposite marker alleles the association signals might cancel each other out.

Fine mapping and the identification of causal variants

The size of the trait identity-by-decent (IBD) region depends both on the number of recombination events that have occurred since the mutation appeared and on the selective pressure attached to it. If the mutation is recent, there will be a large IBD region common to all cases. This may facilitate the identification of the region but makes identification of causal variants much more difficult. Today, a helpful strategy is to sequence the whole region using next generation technology, which identifies all polymorphic sites where cases and controls differ. In paper IV we performed long range PCRs prior to sequencing, but other methodologies such as sequence capture can also be applied. With the continuing drop in sequencing cost, we decided to sequence the entire genome of two horses in paper III. Reads aligning to the established IBD region were subsequently extracted and analyzed. When sequencing large regions, one usually ends up with a very large number of potential causative mutations. A way to overcome this, especially useful for monogenic traits, is to choose controls with an ancestral haplotype or haplotypes very similar to the disease haplotype. We utilized this approach in paper IV, and the result was just two candidate mutations. In fact, unless the mutation happened on a very uncommon haplotype chances are quite high that the ancestral haplotype still resides in the current population of domestic animals (Andersson, 2009).

If the ancestral haplotype could not be identified, and/or if the trait is complex and thus without a dichotomous phenotypic classification, numerous polymorphic sites are usually identified as potentially causative. In these cases, one can first focus on the sites that are either in coding regions or in conserved elements such as those identified through the 29 mammals project (Lindblad-Toh *et al.*, 2011). In the pace project presented in paper III we identified 65 polymorphic sites within the sequenced 438 kb IBD region. One of these was a very good candidate for functional validation since it introduced a stop codon

into a transcription factor. There is an obvious pitfall to this methodology, it will not detect a gain of function mutation generated within an "unconstrained" region, for instance the generation of a transcription factor binding site. Many times no obvious candidate mutation can be identified. The usual methodology is then to genotype all the polymorphic sites in a large number of subjects, here even individuals that do not carry the trait are of value since they can be used to exclude causality. Further, quantitative PCR (qPCR) can be used to test for differential gene expression within the associated region, which could aid in the identification of the most likely candidate.

Obviously fine mapping monogenic traits, which have a simple one-to-one relationship between genotype and phenotype, is much easier than fine mapping QTLs. Identification of the actual QTNs is hampered by the difficulty in assigning an exact phenotypic class to each sample. Here, domestic animals potentially have a very convenient advantage over human studies. Although the quantitative trait is a metric value, for each locus individuals can be defined categorically (except for multi-allelic mutations such as multiple duplications). Sometimes the genotype of an individual at any given QTL can be determined by the phenotypic distribution of the trait among their progeny, at least if the effect size is high (Van Laere *et al.*, 2003).

Functional validation

Depending on the nature of the causative mutation, different functional validation strategies can be adopted. For example, if the mutation is in a predicted promoter or enhancer element, the amount of mRNA in mutant versus wild type (wt) subjects can be measured using qPCR. Mutations in the 3' and 5' end of genes can also influence mRNA stability. Stop mutations commonly subject the mRNA to nonsense-mediated decay. If the mutation alters amino acid sequence, one can test if the protein retains its biochemical properties through enzymatic assays, signaling assays, or investigate its ability to form correct structures etc. The binding properties of mutated transcription factors can be analyzed with electrophoretic mobility shift assay (EMSA) and their capacity to regulate expression with luciferase assays. Mutations that cause changes in the signal peptide sequence are a special case, as these can direct the protein to the wrong locality. This error may be elucidated by green fluoresces protein (GFP) tagging. A fantastic resource for the functional validation of mutations is to either disrupt the gene in mouse, a so-called knockout mouse, or to induce the exact same mutation, a knock-in mouse. This

section by no means covers all functional validations that can be performed, but gives a short outline of some the most commonly used methods.

The horse

History of the horse

The horse (*Equus caballus*) belongs to the order Perissodactyla (odd-toed hoofed mammals). There are three existing families in this order: the Equidae (horses, asses and zebras), the Rhinocerotidae (rhinos) and the Tapiridae (tapirs). The horse was domesticated on the Eurasian steppe 5,000-6,000 years ago (Ludwig *et al.*, 2009). They do not seem to have undergone a tight domestication bottleneck, rather both genetic and archaeological evidence suggests multiple domestication events (Cieslak *et al.*, 2010; Jansen *et al.*, 2002; Vila *et al.*, 2001) and the high diversity seen in maternal linages suggests continued gene flow between domestic and wild stock (Lippold *et al.*, 2011b). In contrast, sequence analysis of the Y-chromosome suggests a limited number of patrilines (Lippold *et al.*, 2011a), indicating a strong sex bias in the domestication process. Horses throughout history have played an important roll in human civilization through transportation, use in agriculture, service in war, food production and as sport- and companion animals. Horses have primarily been bred for performance and health and the major selective focus is on athletic traits i.e. mutations that influence riding or driving. This is probably the main reason why horses are not as morphologically diverse, or extreme, as many dog breeds. Still, there has been some selection for fancy phenotypes such as coat color and miniature stature. There are approximately 500 recognized horse breeds and the worldwide population is greater than 58 million (FAOSTAT 2007). There is much long-range haplotype sharing between breeds, at least in comparison to dogs, suggesting a close interbreed relationship and gene flow between many horse populations (Wade *et al.*, 2009).

Historic background to the field of horse genetics

The horse has 32 pairs of chromosomes (2n= 64) including the sex chromosomes as revealed in 1959 (Rothfels *et al.*, 1959). Investigations into the mode of inheritance for coat colors marked the start of horse genetic research in the beginning of the 20th century (Hurst, 1906). The first markers to be used with the aim of constructing genetics maps were genes, in which alleles could be distinguished by visual examination. A new method to

distinguish alleles for different proteins was possible after the development of biochemical techniques such as gel electrophoresis. These included the scoring of blood groups (ABO-series), variants in blood serum proteins and different kinds of immunological proteins, such as the equine leukocyte antigen (ELA). In the beginning of the 1980's, 14 marker genes belonging to four different linkage groups had been characterized in horse and an additional three markers had been assigned to the X chromosome. In 1978, tobiano was the first coat color gene to be assigned to a linkage group (Trommershausen-Smith, 1978). Four years later Andersson and Sandberg mapped both the chestnut and roan loci using 15 blood group and protein loci (Andersson & Sandberg, 1982). They were able to map these traits even with this sparse dataset thanks to a very high number of horses, over 30,000. In fact, that horse material was collected from the same biobank used in the present investigations, a repository that has now increased to include more than 270,000 horse samples.

The development of dense genome maps was, for many years, hampered by the lack of available genetic markers. This problem was solved with the emergence of different cloning techniques. In 1992, the first equine microsatellites were characterized (Ellegren *et al.*, 1992) and in the same decade, one of the first large maps to incorporate microsatellites in horse was produced in Uppsala, with over 100 markers assigned to different linkage groups (Lindgren *et al.*, 1998). Because of the discovery of microsatellites, the 1990's experienced a rapid increase in the number of markers available for linkage studies. Several physical maps, such as cytogenic and radiation hybrid maps, were developed as a complement to the genetic maps. These provided higher resolution than linkage maps, as they did not rely on recombination. Further, they were also invaluable for assigning the orientation and position of linkage groups to different chromosomes. In combination, the years of effort from scientists world wide, resulted in a high-resolution map of almost 5,000 markers, assigned to chromosomes, correctly orientated and supplemented by comparative information from mouse and human. In fact the horse genome became one of best characterized among domesticated animals, as reviewed in (Chowdhary & Raudsepp, 2008). The constructed maps were utilized in genome scans performed to search for genes underlying various traits and served as a bank of markers that could be utilized in candidate gene studies, phylogenetic studies and parental testing. The maps were also instrumental to the whole genome sequencing of horse by providing a framework for read mapping and to ensure scaffolds were assigned in the correct location and orientation.

Sequencing of the horse genome

In February 2007 the first draft of the horse genome was completed and a second version (EqaCab2) was released a few months later. The genome was sequenced to approximately 7X coverage using traditional Sanger technologies and a whole genome shotgun approach. The horse sequenced was a Thoroughbred mare called Twilight (Wade *et al.*, 2009), selected because of her low level of heterozygosity. Approximately 96% of the sequence was anchored to chromosomes with a predicted total genome length of 2.43 giga bases and many chromosomes showed a high degree of conserved synteny to human chromosomes. Almost half of the assembly constitutes repetitive sequence, mostly long interspersed nuclear elements. The predicted number of protein-coding genes, 20,322, is similar to predictions in other eutherian mammals. Of these annotated genes, 16,617 have predicted human orthologous - 15,027 predicted one-to-one orthologous. The gene predictions were verified by transcriptome analysis of eight equine tissues and 88% of the 169,073 predicted exons were expressed (Wade *et al.*, 2009). Most notable among the gene families that have undergone paralogous expansion compared to human, are keratin genes, which may affect hoof formation and opsin genes, important for photoreception. The latter might be important for predator detection.

Approximately 1.2 million SNPs were identified and used to construct a horse specific SNP chip. The SNPs are a combination of the heterozygous positions identified in Twilight, and those found from a comparison of approximately 100,000 random reads from each of seven different horse breeds. Two SNP chips have now been released to facilitate trait gene mapping in horses. The first, EquineSNP50 BeadChip compromised just over 54,000 SNPs and was used in paper III of this thesis. An improved version was recently released which contains approximately 70,000 SNPs. The degree of within breed LD in horse is on average intermediate to dogs and humans, with LD approximately five times lower than in dogs and five times higher than in humans. Except for in Thoroughbred, LD declines rapidly over the first 50-100 kb in most breeds and reaches background levels within 1-2 mega bases (Mb) (McCue *et al.*, 2012; Wade *et al.*, 2009). The available horse genome sequence, and the development of the genomic tools that it facilitated, have revolutionized molecular genetics research in this species. Both the sequence and the new tools have been instrumental in the studies presented in this thesis and will continue to be so in the future - for all scientists working with horse or comparative molecular genetics.

Aims of this Thesis

The overall goal of this thesis was to identify genes associated with three distinct phenotypes in horse. Identification of these genes would have important implications for horse welfare and could inform breeding decisions. Furthermore, mapping trait genes in domestic animals aids in the functional annotation of the human genome, as well as the genomes of other economic and/or culturally important species.

The specific aims were:

I To determine the genetic cause of equine Multiple Congenital Ocular Anomalies (MCOA) syndrome, to elucidate the debated mode of inheritance of the disease as well as to estimate penetrance.

II To investigate if genetic variants in the equine MHC-II region influence susceptibility to Insect Bite Hypersensitivity (IBH).

III To identify loci responsible for the lateral gait "pace" in Icelandic horses through whole genome association mapping.

Present Studies

Mutant *PMEL* causes Multiple Congenital Ocular Anomalies syndrome (Papers I and IV)

Background

Multiple Congenital Ocular Anomalies (MCOA) syndrome is a hereditary condition in horse consisting of a wide range of ocular defects (Ramsey *et al.*, 1999). Most affected horses have a coat color called Silver (Figure 1). In 2006, this coat color was shown to be associated with a dominant missense mutation in *PMEL* (Brunberg *et al.*, 2006). Silver horses display a dilution of eumelanin (black and brown pigment), into gray or white. This is most noticeable in the mane and tail, while the body of the horse becomes slightly diluted in color and often displays a dappled pattern. The color is especially desired in Rocky Mountain horses and Silver horses are thus more frequently used for breeding. The consequence is an unusually high frequency of both Silver (approximately 55% of registered individuals), and MCOA syndrome horses in this breed. The documented association between *PMEL* and Silver gave us the opportunity to quantify the degree of linkage between the coat color and MCOA syndrome. In a paper published by us in 2008 we used four families of Rocky Mountain horses to map the *MCOA* locus to a 4.9 Mb region on chromosome 6, containing ~190 genes (Andersson *et al.*, 2008).

Figure 1. A Silver colored Rocky Mountain Horse. The typical shiny white mane and tail, as well as a slightly diluted body color with dapples, is seen in this genetically black Silver colored horse. The horse has also been diagnosed with MCOA.

It has been argued that MCOA syndrome was caused by a recent mutation and was restricted to horses that are closely related to the Rocky Mountain Horse, however it is now recognized that the mutation is old, at least ~3,000 years (Ludwig *et al.*, 2009) and present in a diverse set of breeds including the Icelandic Horse, Shetland Pony, Exmoor Pony, American Miniature Horse, Belgian Draft and Morgan Horse, Kentucky Mountain Saddle Horse and Mountain Pleasure Horse (Andersson *et al.*, 2011; Komaromy *et al.*, 2011; Pinard & Basrur, 2011; Plummer & Ramsey, 2011).

Horses with MCOA syndrome can be categorized into two separate groups depending on disease severity and pedigree examination has suggested a incomplete dominant mode of inheritance (Ewart *et al.*, 2000). The genetic study performed by us was in agreement with this hypothesis (Andersson *et al.*, 2008). Horses that are heterozygous for the disease-causing allele have less severe ocular anomalies, mainly displaying peripheral cysts of the iris, ciliary body or retina, called the Cyst phenotype. A smaller number of these horses also have moderate retinal dysplasia, or retinal detachment, which appears to be an extension of these cysts. Homozygous mutant horses present with multiple anomalies and have the so-called MCOA phenotype (Figure 2). They encompass all clinical signs included in the Cyst phenotype concurrent with iris hypoplasia, retina dysplasia, iridocorneal angle abnormalities, congenital cataracts, cornea globosa, iridocorneal adhesions and opacification, as well as pupils with a decreased or absent light response and that do not dilate when administered mydriatic drugs. A single horse does not necessarily possess all of these clinical signs and there are additional defects, which are detected less frequently (Andersson *et al.*, 2011; Grahn *et al.*, 2008; Ramsey *et al.*, 1999).

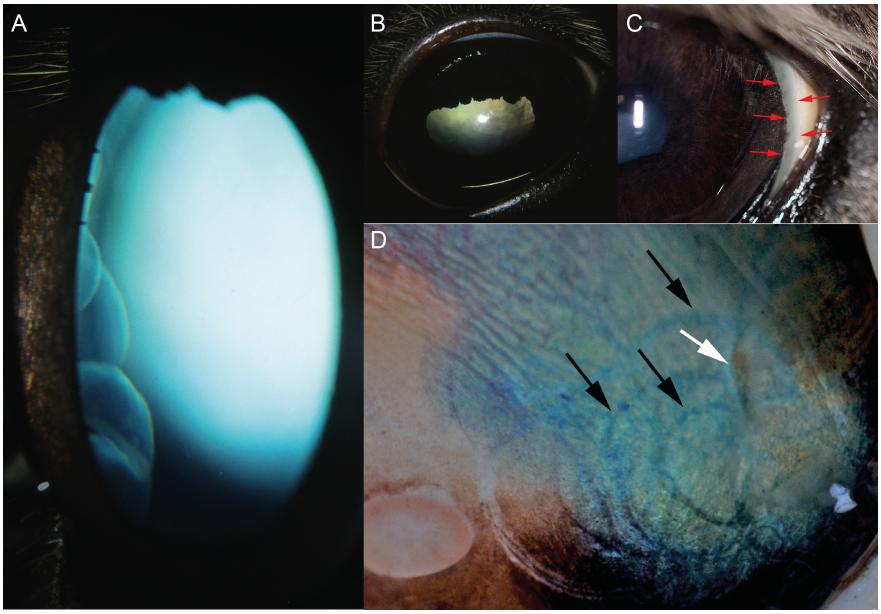

Figure 2. Clinical features of MCOA in Icelandic horses homozygous for the *PMEL* missense mutation. **A**) Multiple iridociliary cysts in the temporal quadrant of the eye. **B**) A pupil unable to dilate normally and cataractous changes in the lens contributing to visual impairment. **C**) A regional abnormality of the pectinate ligament heralded by a widened, white perilimbal zone temporally (red arrows). **D**) Post-mortem specimen showing multiple demarcation lines in the fundus (black arrows) extending towards the optic nerve head and an area where the retina is still detached (white arrow). Photos: Björn Ekesten.

Mutations in *PMEL* have previously been shown to regulate hypopigmented phenotypes, not only in horses, but also in mouse, chicken, zebrafish, dog and cattle (Kuehn & Weikard, 2007; Kerje *et al.*, 2004; Kwon *et al.*, 1995). Only the zebrafish and dog have reported concurrent ocular anomalies, although in different forms than those observed in the horse (Schonthaler *et al.*, 2005; Gelatt *et al.*, 1981; Dausch *et al.*, 1978). The zebrafish mutant, *fading vision*, is characterized by hypopigmented retinal pigment epithelium (RPE) and most likely because of RPE alteration, the outer segments of photoreceptors are strongly reduced in length or completely absent. Interestingly, the retina of adults shows a partial recovery to normal morphology (Schonthaler *et al.*, 2005). No abnormalities in the other segments of the eye were reported. *PMEL* mutant dogs are called “merles” and typically have blue eyes (Clark *et al.*, 2006). Double merle dogs have an elevated frequency of ocular anomalies including microthalmia, microcoria, colobomas and increased ocular pressure (Gelatt *et al.*, 1981; Dausch *et al.*, 1978). However, not all double merles are affected by this pleiotropic effect and surprisingly, the proposed causative SINE element has also been found in a few non-merle dogs (Clark *et al.*,

2006). In order to study complete loss of function of PMEL, a mouse line with inactivated *Pmel* has recently been created (Hellstrom *et al.*, 2011). This study showed that *Pmel* inactivated mice have a minor visible dilution of the coat color but no detectable ocular anomalies. In yet another recent study of the biochemical properties of PMEL it was shown that mutations in or near the transmembrane domain (TMD) of PMEL, like the *Silver* mutation in horse, alter the capacity of this region to self-associate (Watt *et al.*, 2011). The mutations alters fibril formation into a more compact structure and this in turn leads to severe pigment loss. In chicken, the effect of the mutation causing the Dominant white phenotype was shown to be similar to that in horse, however ocular examination of these birds could not identify any structural changes or other eye anomalies (Karlsson *et al.*, 2009). The aim of this project was to identify the causative mutation for MCOA syndrome and to establish if *PMEL* is the gene responsible; a debated theory given that no ocular anomalies are seen in most of the *PMEL* mutant species, including the null mouse.

Results and discussion

In paper I, we extended our previous MCOA syndrome study to include 465 horses (Cyst phenotype, n= 246; MCOA phenotype, n= 83) from four different breeds including Rocky Mountain horses, American Miniature horses, Kentucky Mountain Saddle horses, and Icelandic horses. IBD mapping pinpointed a 208 kb genomic interval harboring the MCOA mutation (Table 1). The identified interval is gene-dense, harboring 15 genes. Several of these are known, or proposed to be, involved in ocular development or disease. In addition, the study verified that the disease is inherited as an incompletely dominant trait and that the two distinct phenotypic categories, Cyst and MCOA, are caused by one and two copies of the mutant allele, respectively ($p= 2.20x10^{-16}$). We were unable to detect clinical signs of MCOA syndrome in six percent of the genetically characterized heterozygous horses. This might be due to modifier loci or due to our inability to detect small cysts if positioned just posterior of the iris. Importantly, many of these non-penetrant horses have produced affected offspring when crossed with non-carriers.

Table 1. Phased data of 13 markers defines a 208 kb IBD haplotype. Six distinct haplotypes were identified in the four analyzed breeds. The minimal IBD region is marked in light grey and the haplotypes crucial for excluding genes marked in dark grey.

Haplotype	No of chr	*MS1*	*N1*	*MS14*	*PMEL -ex11*	*MS3*	*8345*	*TKY284*	*8377*	*MS13*	*SMARCC -int24*	*SMARCC -int19*	*8453*	*8475*	Breed
A	355	269	T	247	T	233	C	177 or 175	C	222	G	G	A	T	RH, KY
B	12	263	G	247	T	233	C	177	C	222	G	G	G	A	RH, KY
C.1	10	263	G	247	T	233	C	177	C	222	G	G	A	A	MINI
D.1	21	263	G	247	T	231	C	177	C	222	G	G	G	A	IS
D.2	1	263	G	247	T	231	C	177	C	222	G	T	G	A	IS
C.2	10	263	G	247	T	233	C	177	C	222	A	T	G	A	MINI
Position		0	33	50	58	115	118	161	180	227	240	242	297	360	

No of chr= The number of chromosomes containing of each haplotype. RH= Rocky Mountain Horse, KY= Kentucky Mountain Horse, MINI= American Miniature Horse, IS= Icelandic Horse. Position= The chromosome position of marker *MS1* is 73,607,795 and is here set as the reference point zero (0). The distances to the following markers are given in kb.

In paper IV we used Illumina TruSeq technology to sequence the entire IBD region in five homozygous MCOA horses from three different breeds, one horse with the heterozygous Cyst phenotype and four healthy controls. In an attempt to minimize the number of candidate mutations, we carefully selected healthy horses with haplotypes as similar as possible to the MCOA haplotype in conjunction with cases representing different breeds. After filtering out all SNPs that did not follow the correct mode of inheritance, i.e. horses with the MCOA phenotype homozygous for one allele, Cyst phenotype horse heterozygous, unaffected horses homozygous for the opposite allele, only two SNPs remained. Both of these SNPs are positioned within the *PMEL* gene, demonstrating that this gene is in fact important for ocular development or at least has the ability to change eye structure into a pathologic form when mutated. One of the identified SNPs is intronic and in an unconstrained region according to 29 mammals conservation scores. The other SNP however introduces an amino acid change in the cytoplasmic religion of PMEL (Arg625Cys).

The molecular effects of this missense mutation have recently been studied in detail (Watt *et al.*, 2011). Most importantly, the mutation is positioned very close to the TMD and whereas the wild type TMD is not able to oligomerize, human PMEL with an introduced mutation analogues to the horse mutation ($hPMEL^{R625C}$) gains this function and can readily self oligomerize. The horse mutation causes a replacement of the middle of three consecutive arginines to a cysteine. The new oligomerization ability of the TMD was therefore suggested to be a consequence of decreased electrostatic repulsion between the adjacent basic RRR motifs through the removal of a positive charge. Upon maturation of profibrils into elongated fibrillar sheets, $hPMEL^{R625C}$ show marked differences to wild type. In mouse melanocyte cell lines expressing the latter, most stage III and IV melanosomes were densely pigmented. In cells expressing mutant form, there was less space between the fibrils and the sheets appeared abnormally tightly packed. Importantly, and probably because of this arrangement, the organelles seemed to lack pigmentation. These findings suggest that either the fibrils were no longer capable of binding to melanins or that melanin production was somehow inhibited. Electron microscopy analysis revealed that cells expressing mutant PMEL were hypopigmented, harbored fewer pigmented melanosomes and were enriched in early stage melanosomes (Watt *et al.*, 2011). This data reveals a clear functional effect of the coding SNP and together with our genetic finding, demonstrates that the missense mutation in exon 11 of *PMEL* causes MCOA syndrome in horse.

Why do changes in PMEL fibrilliar sheet formation cause such dramatic morphological alternations to the horse eye? Does PMEL have other functions besides those of melanin processing? The anomalies seen in eyes of horses may suggest alterations in cellular differentiation and/or migration. *PMEL* is known to be expressed in two different kinds of pigment cells in the eye, the neural ectoderm derived retina- iris- and ciliary body pigmented epithelia and in neural crest derived uveal melanocytes distributed in the middle layer of the eye (choroid, iris and ciliary body) (Hu *et al.*, 2008). The purpose of these cells are numerous, so if the abnormal sheet formation severely encumbers cell capacities, severe morphologic and functional changes to the eyes would not be surprising. RPE provides metabolic support for the retina, is involved in the phagocytosis of the constantly shedding photoreceptor outer segment, converts and stores retinoid, absorbs scattered light, facilitates transportation of ions, fluid and nutrition and secrets several growth factors (Strauss, 2005). Less is known about the other two pigmented epithelia but the ciliary pigment epithelium is important for the production of aqueous humor whilst the iris pigment epithelium is important for dilation of the pupil. The uveal melanocytes produce extracellular matrix degrading enzymes as well as growth factors, such as vascular endothelial growth factor, which may control the blood circulation of the eye (Hu *et al.*, 2002). A first step in the continued analysis of the horse mutation may be to use *in situ* mRNA hybridization and PMEL protein immunohistochemistry in mutant and normal eye tissue to compare if either the expression levels and/or localizations are altered in homozygous mutants. Timing might then be of great importance as it is possible that *PMEL* expression influences development during a short time window and therefore it is necessary to analyze tissue from the exact developmental stage in order to see differences in expression.

Another interesting question is why some species present ocular alterations and some do not? This is likely because of the genetic background, for example; we know that *PMEL* expression is *microphthalmia-associated transcription factor* (*Mitf*) dependent (Baxter & Pavan, 2003). An interesting study would be to quantify the ocular anomalies, since they vary significantly between affected individuals, and to then investigate if this has a correlation to genotypes at other coat color loci, such as the *Extension* and *Agouti* loci. The same would apply to the genetically characterized heterozygous horses that were not diagnosed with any ocular anomalies. Perhaps the most powerful approach to address the function of the mutation is by the generation of a knock-in mouse. This would provide an opportunity to monitor *PMEL* expression during eye

embryological development and further give us the opportunity to investigate the effect of mutant *PMEL* on different color backgrounds.

The risk of MCOA syndrome and its link to the silver mutation should be taken into consideration in breeding decisions. Horses with the MCOA phenotype are at particular risk of having impaired vision, and difficulties in adapting to changing light conditions are probably a common phenomenon in these horses. Some individuals have more severe impairment of their vision, causing abnormal behavior and an inability to perform. Breeding *PMEL* mutation carriers only to known non-carriers would practically eliminate the risk of producing horses with the vision threatening abnormalities caused by this syndrome. The *PMEL* mutation does not have an obvious effect on phenomelanin (red and yellow pigment), making it virtually impossible to detect phenotypically in horses lacking dark pigment, for example chestnuts, white grays, palominos, pearl- and champagne-colored horses. These horses must be genotyped in order to determine their carrier status.

Variation in the MHC-II region increases the susceptibility to Insect Bite Hypersensitivity (Paper II)

Background

Insect Bite Hypersensitivity (IBH) is the most common allergic disease in horses and affects numerous breeds worldwide. It is generally caused by an allergic IgE mediated type I hypersensitivity reaction against proteins in the saliva of various species of the biting midges, *Culicoides ssp* (Schaffartzik *et al.*, 2012; Mellor & McCraig, 1974; Ishihara, 1957). The prevalence of IBH among Icelandic horses born in Sweden is approximately 8% (Eriksson *et al.*, 2008). In contrast, the prevalence is sometimes as high as 26–35% in horses exported from Iceland, most likely because *Culicoides* does not exist in Iceland and thus these horses have not encountered the antigen in early life (Schaffartzik *et al.*, 2012; Bjornsdottir *et al.*, 2006; Brostrom *et al.*, 1987). This theory was strengthen by a recent study performed on horses exported as weanlings which revealed that if a horse is exposed to allergens at the age of 7-10 months, they do not develop IBH at a higher frequency than horses born in Europe.

IBH is clearly seasonal, determined by the activity of insects and is classed as a chronic disease. Symptoms include intensely pruritic lesions, urticaria, oedema and papules (Figure 3). These are usually most prominent along the dorsal

midline, in particular the tail base, mane, head, withers and rump (Schaffartzik *et al.*, 2012). Horses suffering from IBH that have been left unattended can easily be recognized by the almost complete loss of hair from the mane and tail and continuous scratching may cause open wounds. Secondary symptoms include lichenification, crusts and scaling (Brostrom *et al.*, 1987). The ability of affected horses to graze might be impaired due to constant itching and rubbing.

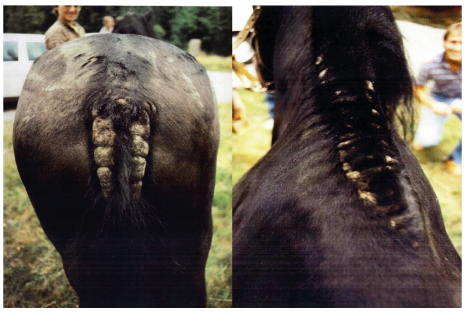

Figure 3. Icelandic horse affected with IBH. This 10 year-old gelding, imported to Sweden from Iceland, has scratched away most of its mane and tail as a consequence of IBH. These photos, taken at a late, chronic, stage of the disease, show the formation of transverse ridges and folds produced by the prolonged self-induced skin trauma. Secondary thickening of the skin gives this animal a leathery bark-like appearance (lichenification). Source: Hans Broström.

We have estimated the heritability of IBH for Icelandic horses born in Sweden at 0.08 on the visible scale and 0.33 on an underlying, continuous scale (Eriksson *et al.*, 2008). Similar estimates have been reported for Shetland ponies and Friesian horses (Schurink *et al.*, 2011; Schurink *et al.*, 2009). In the study presented in paper II, we decided to investigate the major histocompatibility complexes (MHC) II because: i) certain human leukocyte antigen (HLA) class II variants are associated with an increased risk of atopic dermatitis in man, a disease that partially resembles equine IBH (Kiyohara *et al.*, 2008; Saeki *et al.*, 1995); ii) histopathological analysis of IBH affected skin reveals an increased number of MHC-II positive cells, presumably Langerhans cells and iii) two studies have shown serologically that certain equine leukocyte antigen (ELA) class II specificities are linked to IBH susceptibility (Lazary *et al.*, 1994; Marti *et al.*, 1992; Halldorsdottir *et al.*, 1991). We wanted to confirm these associations because the larger of these studies did not seem to consider the relationship among horses sampled and therefore the result might be influenced by population stratification. Furthermore, recent investigations on the MHC class II region, showed that multiple microsatellite haplotypes could be associated with a single serotype (Tseng *et al.*, 2010) and so DNA genotyping therefore provides better resolution than the earlier methodologies. The ELA class II region in horse is positioned on chromosome 20 and harbors three *DQA* loci, two *DQB* loci and three *DRB* loci (Tseng *et al.*, 2010). Only one copy of the *DRA* locus is present, however contrary to many other species,

this gene is polymorphic in the domestic horse, with at least four different alleles identified (Brown *et al.*, 2004; Albright-Fraser *et al.*, 1996; Bailey, 1994).

Results and discussion

In an effort to establish if specific ELA-II alleles are associated with IBH status we tested 94 IBH affected and 93 unaffected Icelandic horses. These horses were paternal half-sibs sired by 42 different stallions. The average number of offspring per stallion was 4.5. In order to standardize the environmental effect we only included horses born in Sweden. Also, these horses had previously been genotyped on the Illumina Equine SNP50 BeadChip, making it possible to ensure that our study did not suffer from the effects of stratification. The result was subsequently followed up in a second, independent population consisting of 106 unaffected and 80 IBH-affected Exmoor ponies. All horses were genotyped at four microsatellite markers positioned within the MHC class II region and were sequenced for the highly polymorphic exons two from *DRA* and *DRB3* respectively.

Marker *COR112* was clearly associated with IBH status (Table 2) with allele *274* enriched in cases (f= 0.22) compared to controls (f= 0.06), odds ratio 4.19 (95% CI= 2.00-9.44, p= 2.34×10^{-5}). In the Exmoor ponies, *COR112* allele *274* was likewise associated (p_{raw}= 0.043) but was much more common, with a frequency of 0.60 in affected horses compared to 0.50 in healthy ponies. Thirty SNPs and two indels, which after phasing constitute eight different alleles, were identified in the *DRB3* gene. The frequency of *DRB3* allele *1.2* was three times higher in affected Icelandic horses compared to controls and was significantly associated with IBH status (p_{raw}= 0.004, OR= 3.60). Unexpectedly, we also found that *DRB3* allele *1* was possibly protective (p_{raw}= 0.015) in Icelandic horses. It was present at a frequency of 0.51 in healthy horses but only 0.37 in affected horses, giving an odds ratio of 0.58 (95% CI= 0.38-0.91). However, these two *DRB3* alleles differ only by one intronic SNP and therefore likely produce the same peptide. Hence, the associations that these alleles display are probably not due to their phenotypic effect but rather due to LD with causal variants. Further, this marker was not associated to IBH status in Exmoor ponies.

Table 2. Association results between ELA markers and IBH in Icelandic horses.

					IBH		Healthy				
Marker	**Position (bp)**	**p-value (2xN)**	**df**	**A_d**	**n**	**p(A_d)**	**n**	**p(A_d)**	**p-values (2x2)**	**OR**	**95% OR**
DRA	32690939	0.025	2	*0201*	94	0.17	91	0.08	0.012	2.28	1.15-4.71
COR112	33282436	0.001	5	*274*	86	0.22	87	0.06	$2.34x10^{-5}$	4.19	2.00-9.44
DRB3	33364480	0.008	3	*1.2*	91	0.13	91	0.04	0.004	3.60	1.45-10.2
COR113	33480825	0.271	3	*282*	89	0.14	86	0.07	0.037	2.17	1.01-4.93
UM011	33510120	0.151	4	*180*	90	0.13	86	0.06	0.033	2.25	1.02-5.26
COR114	33516304	0.354	4	*259*	89	0.13	89	0.08	0.229	1.61	0.77-3.45

df= degrees of freedom, A_d= allele most strongly associated with disease, p(A_d)= frequency of A_d and OR= odds ratios

Interesting, when we combined microsatellite and sequencing data in order to investigate the pattern of homozygosity throughout the MHC class II region we could conclude that affected and healthy horses had different degrees of homozygosity. In Icelandic horses, 0.13 of affected individuals were homozygous, compared to only 0.04 of the control horses. Homozygous horses were more commonly observed among the Exmoor Pony samples, with frequencies of 0.30 and 0.15 respectively. A Cochran-Mantel-Haenszel test, which combines data from both breeds, showed that homozygosity in the region confers a higher risk of developing disease, with an odds ratio of 2.67 (p= 0.0013, 95% CI= 1.22-4.66). The data presented in this study confirms the association of ELA alleles to IBH reported earlier and presents a novel link between homozygosity in the MHC-II region and disease frequency. It is still unclear what the causal variants are or how homozygosity influences disease susceptibility. Each requires further research.

Future prospects and practical implications

It would be of great interest to pinpoint the causal ELA variant(s) and to understand their role(s) in disease manifestation. Unfortunately fine mapping is made somewhat difficult by two major aspects: i) the horse MHC-II region is not yet fully characterized and, like humans, many large allele specific duplications are likely to exist and ii) to prove causality is often difficult in complex diseases, where a simple one-to-one relationship between genotype and phenotype does not exist. Potential causative variants may also be present in healthy individuals that have not developed the disease due to other genetic or environmental factors. However, further fine mapping through the examination of allele frequency differences in a larger horse material and using a more dense marker set should certainly be possible considering i) the

associated region presented here is not yet well defined as a clear 3' border has not been established and ii) the examined region is large, with the two most distant markers, *DRA* and *COR114,* positioned 825 kb apart. Subsequent re-sequencing of selected individuals can be exploited to catalogue all potential causative mutants and subsequent characterization, including examination of variant conservation scores and qPCRs, may isolate those mutations most likely to play a role in IBH susceptibility.

Genes outside the MHC-II region most likely also contribute to disease susceptibility. We have performed one GWAS using 104 healthy and 105 affected Icelandic horses, however could not detect any SNPs reaching genome wide significance. This was likely due to a lack of power because of small effect sizes relative to the number of genotyped subjects. The marker density was probably also a limiting factor. We utilized the Illumina Equine SNP50 BeadChip but only approximately 38,400 SNPs remained both after quality control filtering and excluding SNPs with a minor allele frequency of 0.05. Recently an IBH susceptibility GWAS was performed in Shetland ponies, and whilst some promising chromosomal regions were shown to be associated to disease, it was not clear if these reached genome wide significance (Schurink *et al.*, 2012). As a collaborative effort we are now planning a meta-analysis in an effort to increase our chances of find genes with smaller effects. The best way of overcoming limitations in marker densities is to do whole genome re-sequencing. This methodology has the advantage of providing data for mapping, fine mapping and the identification of potential causative polymorphic sites in one single experiment.

To date, our advice to horse breeders is to limit the extent of inbreeding as this might influence IBH susceptibility by increasing the levels of homozygosity in the MHC-II region. However, we do not recommend the direct testing of these markers with the aim of minimizing the frequency of the *214*-allele and there are many reasons for this. First and foremost, we do not yet know the causative variant and the degree of LD between this and the *COR112* marker. This means that inherently, a number of horses will be incorrectly diagnosed either as carriers or non-carriers. Second, we do not yet know the effect size of the mutation, and so do not know how a *COR112* breeding strategy would affect the long-term health of the breed. Each of these test factors needs to be explained to horse owners carefully, so as to avoid an "over-belief" in the MCH-II results. These results, if used prematurely in breeding could potentially have undesirable side effects.

Identification of the first *Gait* locus in horse (Paper III)

Background

Most horse breeds can only perform the three basic gaits; "walk", "trot" and "canter". However, some breeds are called "gaited" as they are able to perform the lateral two-beat gait "pace" (Figure 4a), and/or different varieties of ambling gaits. The ambling gait may vary slightly between breeds but is always characterized by a four-beat footfall, which makes it particularly smooth and comfortable for the rider. The variable parameter is primarily if there is an exactly even timing between footfall patterns, as in the Icelandic horse gait "tölt", or if there is a slight delay, leading to either a lateral or diagonal coupling of the legs. There are over 60 gaited breeds worldwide (Hendricks, 1995) and these are more common in America and Asia than in Europe. Gaitedness has been under strong selection in many breeds and has also been the basis for the formation of new breeds. In principle, all Icelandic horses can tölt but not all of them have the ability to pace. Thus horses of this breed are categorized as being either four-gaited (walk, trot, canter, tölt) or five-gaited (walk, trot, canter, tölt and pace). The segregation of this trait made it possible to investigate the underlying genetic cause, with the potential of revealing interesting biological insights into the regulation of locomotion. The heritability for pace has been estimated at 0.60 in Icelandic horses (Albertsdottir *et al.*, 2011), and a recessive mode of inheritance had previously been suggested in Standardbred horses (Cothran *et al.*, 1987).

Most quadrupeds trot (Figure 4b) and even biped humans swing their arms in simultaneous motion with their diagonal leg. However some species (such as camels and llama) move in pace instead of trot, whilst yet others perform an ambling gait (such as giraffe and elephant). Neuronal circuits in the spinal cord generate the rhythmic and coordinated neural output required for limb movement (Goulding, 2009; Nishimaru & Kakizaki, 2009). These neuronal circuits are collectively referred to as central pattern generators (CPGs) and have been shown to function even without external input (sensory feedback and input from the brain). CPGs coordinate the spatial and temporal activity of the joint muscles from all four limbs and also regulate the speed and stability of gait rhythm. Different neurons in the CPG can be characterized based on their location, projections, morphology, neurotransmitters and more recently, on their marker gene expression pattern profile, such as cell type specific transcription factors (Kullander, 2005). Although several specific populations of neurons have been identified to date, substantial research is required to elucidate their function and relative importance in the pattern generating

mechanism. Here, inhibitory neurons are just as important as their excitatory counterparts for the generation of alternating movements, as flexion can only occur in combination with a relaxation of the antagonistic muscles. While over one thousand genes have been linked to uncoordinated movements in *C. elegans*, and more than two hundred mutant lines with walking defects have been identified in *Drosophila* (Kullander, 2005) many fewer genes have been identified in mouse. Two mouse lines with spectacular locomotive defects are the *EPH receptor A4* null mouse and its ligand pheno-copy, the null *ephrin-B3* mouse; both of which display a rabbit like parallel-coupled gait (Kullander et al. 2003; Dottori et al. 1998). A complete switch from alternating to synchronous drug induced fictive locomotion can also be identified in mice with mutant *netrin-1* (Rabe et al. 2009). All three of these genes are expressed in commissural neurons, establishing an important role for this cell type in left and right synchronization.

Results and discussion

In paper III we further exploited the available Illumina Equine SNP50chip data from 209 Icelandic horses by correlating this to the phenotypic description of each horses' gait. A questionnaire was sent to all owners of genotyped animals and 30 replied that their horse was four-gaited and 40 reported to have a five-gaited (pacing) horse. A GWAS revealed a single highly significant SNP positioned on chromosome 23. The association was strongest when a recessive mode of inheritance for pace was considered (p_{raw}= 1.7 x 10^{-9}, EMP2= 2.0 x 10^{-4}, Figure 4c). The result was verified in an independent sample set of 190 horses (p= 2.4x10^{-14}, Table 3). It is worth noting that of the horses classified as four-gaited, as many as 31% are homozygous mutant and should, according to the model, be able to pace. A proportion of this discrepancy is likely due to the use of phenotypic categorization from breeding field tests scores. Pace is a very demanding gait, and there are many factors influencing why a five-gaited horse presents as four-gaited on the day of evaluation. These include the level of experience for both horse and rider and the mental and physical condition of the animal at the time. Even so, the presence of modifying loci cannot be excluded. Another interesting observation is the low number of wild type homozygots (Table 3). This is in concordance with the GWAS and reflects the fact that Icelandic horse breeders less often cross two four-gaited horses to each other.

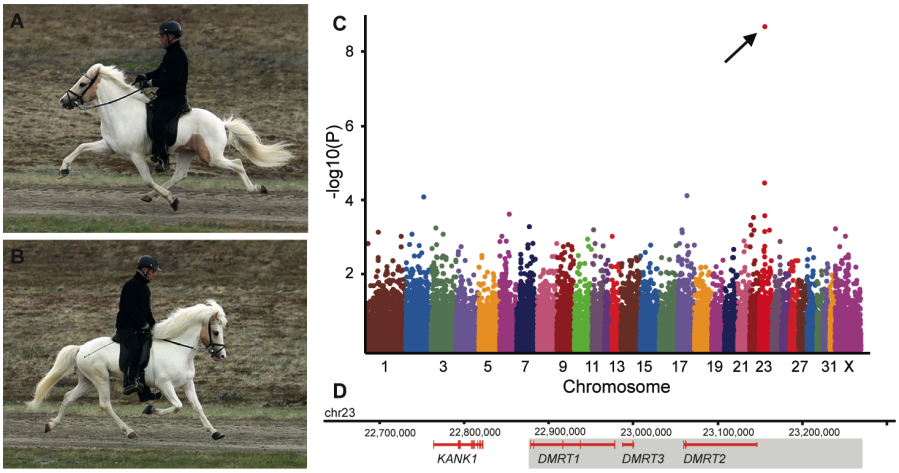

Figure 4. Identification of a DMRT3 mutation in horses. **A**) A pacing Icelandic horse, fore and hind legs on the same side of the body are synchronized. **B)** A trotting Icelandic horse, the fore and hind legs on the opposite side of the body are synchronized diagonally. **C**) Genome-wide association analysis revealed a highly significant association between the ability to pace and SNP BIEC2_620109 on horse chromosome 23 (p_{raw}=1.7 x 10^{-9}, EMP2 = 2.0 x 10^{-4}). **D**) The 684 kb genomic interval associated with the *Gait* locus. The shaded region defines the minimum IBD region (438 kb).

We re-sequenced the whole genome of one four-gaited and one five-gaited horse and extracted reads mapping to an established 438 kb IBD region (Figure 4d). We identified 65 variants unique to the *Gait* haplotype, including a very good candidate mutation; a cysteine to adenine transversion in exon two of the transcription factor *DMRT3* creating a premature stop codon. The mutation is predicted to cause a truncated protein lacking the last 174 amino acids, which is ~40% of the entire sequence and would result in the loss of many highly conserved residues. The mutation was genotyped in a panel of gaited and non-gaited horse breeds and the results showed remarkable concordance to the alternative gait phenotype (Table 3). Nearly all individuals from gaited breeds were homozygous mutant, regardless of whether their alternate gait is characterized by the lateral pace, regular beat ambling or lateral/diagonal couplets ambling. Thus, when considering the data across breeds, the *DMRT3* mutation is permissive for the ability to perform these alternate gaits. Training is probably one important factor affecting the gait phenotype but other loci that regulate and fine-tune the pattern of locomotion are likely to exist. Interestingly, the mutation was also found at high frequencies in horses used for harness racing and we were able to connect this to racing performance. First, we performed a blind study on 61 Standardbred horses in training at a race-camp near Uppsala. Two horses had difficulties in sustaining trot at high

speed. Genotyping revealed that these two were heterozygous, while the remaining 59 horses were homozygous mutant (p= 0.0005). Thus the mutation enables horses to trot at high speed without breaking into gallop, the natural gait for horses at high speed. Further, in a population-based investigation we could conclude that homozygous mutant horses have on average superior breeding values (best linear unbiased prediction, BLUP- values) and higher earnings.

Table 3. DMRT3 genotype frequencies among horse populations.

Breed	Number	CC[1]	CA	AA
A. Gaited horses				
Icelandic horses[2]				
Four-gaited	124	0.02	0.67	0.31
Five-gaited	66	0.00	0.02	0.98
Random sample	162	0.00	0.22	0.78
Other gaited horses				
Kentucky Mountain Saddle Horse	22	0.00	0.09	0.91
Missouri Fox Trotter	40	0.00	0.00	1.00
Paso Fino	45	0.00	0.00	1.00
Peruvian Paso	19	0.00	0.00	1.00
Rocky Mountain Horse	17	0.00	0.00	1.00
Tennessee Walking Horse	33	0.00	0.03	0.97
B. Non-gaited horses				
Arabian Horse	18	1.00	0.00	0.00
Gotland Pony	28	1.00	0.00	0.00
North-Swedish Draft Horse	31	1.00	0.00	0.00
Przewalski's Horse	6	1.00	0.00	0.00
Shetland Pony	20	1.00	0.00	0.00
Swedish Ardennes	22	1.00	0.00	0.00
Swedish Warmblood	64	1.00	0.00	0.00
Thoroughbred	29	1.00	0.00	0.00
C. Horses bred for harness racing				
Standardbred, trotter (Sweden)	270	0.00	0.06	0.94
Standardbred, trotter (USA)	57	0.00	0.00	1.00
Standardbred, pacer (USA)	40	0.00	0.00	1.00
French Trotter (France)	47	0.06	0.34	0.60

[1]Wild type. [2]These do not include the horses included in the initial GWAS and therefore provides a replication of the original highly significant association.

We studied the functional effects of the mutation (Figure 5). First, the truncated *DMRT3* mRNA is not subjected to nonsense-mediated decay. This is in agreement with studies showing that this process primarily effects mRNA with mutations prior to the last intron (Culbertson & Leeds, 2003). Next, we used EMSA and confirmed that the protein still retains its DNA binding capacity. It may therefore be a dominant negative mutation with normal DNA-binding properties but with defective interaction with other proteins. We tagged the protein with GPF and expressed the construct in human glioma U251 cells. Both mutant and wild type protein localized to the nucleus. Furthermore, *in situ* mRNA hybridization in horse spinal cord tissue sections established that both mutant and wild type *DMRT3* is expressed in a small population of neurons located in the ventral horn and around the central canal.

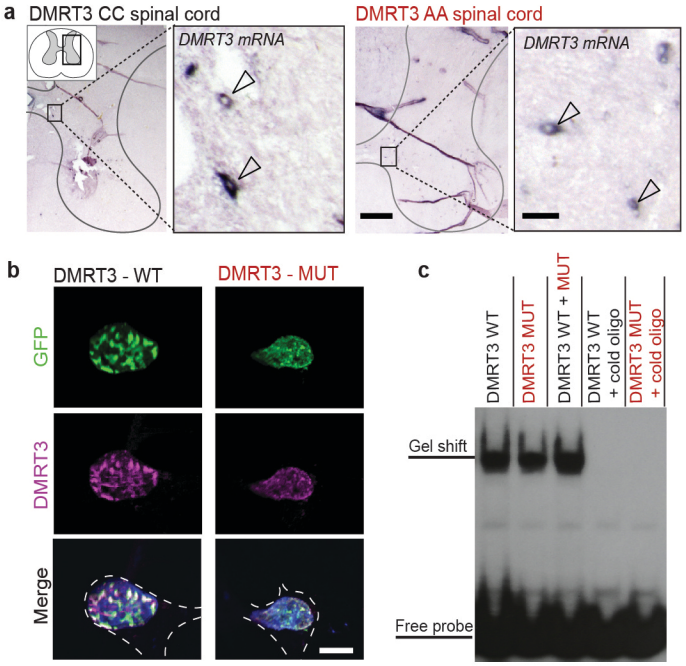

Figure 5. Functional characterization of the horse *DMRT3* mutation. **A**) mRNA *in situ* hybridization of tissue sections from the lumbar spinal cord of wild type (*CC*, left) and mutant (*AA*, right) horse. Arrows mark neurons with *DMRT3* expression. **B**) Transfection experiments in human glioma U251 cells. eGFP-DMRT3 proteins (green) corresponding to the wild type and mutant forms coincides with anti-DMRT3 antibody labeling (purple), both showing nuclear localization (co-localization with DAPI, blue). **C**) EMSA using an oligonucleotide containing a DMRT3-binding motif and in vitro-translated myc-tagged DMRT3 wild type and mutant proteins. Scale bars: 1 mm (a), 100 µm (a, close-up), 10 µm (b).

A *Dmrt3* knockout mouse model had previously been generated which displayed shorter lifespan compared to their wild type counterpart and occasionally male sexual development abnormalities (Ahituv *et al.*, 2007). We further characterized this mouse model with a focus on locomotion and spinal cord neuron development and also investigated *Dmrt3* expression during different developmental stages in wild type mice. The result led us to conclude an important role for *Dmrt3* in gait control. The major outcomes of the investigations were:

I Already in E12.5, the gene is expressed in a small subset of commissural neurons i.e. those that cross the spinal cord midline. These neurons:
   i. belong to the dI6 interneurons
   ii. are inhibitory
   iii. connect to motor neurons

II The dI6 population is of similar size in wild type and null mice. However, there is a 58% increase in Wilms Tumor 1 (WT1) positive neurons, which also belong to dI6, and fewer commissural neurons in null mice.

III Central pattern generated output, as analyzed in isolated neonatal spinal cords using drug induced fictive locomotion, revealed strikingly uncoordinated and irregular rhythm signaling in null mice.

IV In air stepping experiments at postnatal day (P) 1 and P4, alternating hind limb movements were almost completely absent in *Dmrt3* null mice. Alternating movements were regularly seen in wild type mice as well as in the front legs of null mice.

V When challenged in water, knockout mice displayed decreased swim duration and twitching movements rarely observed in wild type.

VI Null mice had difficulties in sustaining the same high speed as wild type mice when placed on a treadmill.

VII On the treadmill, there were several significant deviations in gait parameters suggesting that at slower velocities null mice take longer strides.

In paper III, we give abundant evidence that the *DMRT3* nonsense mutation causes the gait phenotype in horses and that we have identified a new molecule important for the control of locomotion. The loss of Dmrt3 in mouse might

result in a more permissive and flexible state of the locomotive circuit. Likewise, horses with the ability to perform alternate gaits should benefit from a loosely hardwired neuronal circuit and indeed, all such tested horses carry the *DMRT3* nonsense mutation. Moreover, the *DMRT3* mutation has accumulated in harness race horses for which the transition from trot to gallop leads to race disqualification. The clear dichotomous distribution of the mutation in different breeds is likely due to the fact that the mutation has unfavorable effects for some purposes such as gallop racing, dressage, show jumping and heavy drafting. Supporting this, we show that Icelandic horses carrying at least one wild type allele have superior breeding field test scores for trot and canter. Further, Standardbred trotters, most of which are homozygous mutant, often have difficulties in canter, hindering their performance in dressage and show jumping.

Future prospects and practical implications

The discovered connection between the *DMRT3* gene and locomotion opens a whole new field of research opportunities. First, which gene activities are dependent on this transcription factor? ChipSeq (chromatin immunoprecipitation in combination with massive parallel DNA sequencing) from wt as well as transcriptome comparisons between wt and knockout mice would give interesting information about target genes and how the *DMRT3* mutation affects transcription levels. Also, we know that the truncated DMRT3 retains its DNA binding capacities, but we do not know if it can still activate/repress transcription. We are currently conducting luciferase experiments in order to investigate this. Further, the horse mutation is interesting since it has an effect in heterozygotes. The generation of a knock-in mouse could give us the opportunity to further characterize the effect of the protein truncation, especially its dominant negative effect. In the null mouse, it is quite interesting that we observe such a dramatic phenotype in spinal cord signaling during development, but a milder phenotype in adults. There seems to be a compensatory mechanism, something that is also suggested by the observed expansion of the WT positive cells and an increase in *Dmrt1* expression. The generation of a conditional knockout would answer important questions regarding the $DMRT3^{+}$ subset of cells without prior establishment of compensatory effects during embryonic development.

In a comparative genetic effort, we have started to investigate if the *DMRT3* gene is inactivated in other species not moving in a regular trot. Furthermore, mutations in *DMRT3* are likely to exist in humans and might lead to defects in

synchronization of limb movement or other diseases connected to locomotion. Novel *DMRT3* mutations might also be found in horses bred for other purposes and sequencing of coding regions and conserved elements could be performed in a number of highly specialized breeds.

When did the horse *DMRT3* nonsense mutation first arise and how has it spread to the worldwide distribution seen today? The mutation may have occurred for the first time in a foal living among the big horse herds of Mongolia and China and then spread to Europe and finally reached North and South America with the settlers. Extremely few gaited horses are found in Africa and these originate from Mongolian and American Saddle bred horses (Hendricks, 1995). We are currently testing this hypothesis by first deep sequencing the established IBD region in 16 diverse breeds. The distribution of mutations may serve to tell us something about the relatedness of the haplotypes but will also identify polymorphic sites. These variants can be used in a larger worldwide haplotype analysis of gaited and none-gaited breeds. We are also genotyping the mutation in ancient samples.

Important modifier loci are likely to exist. As a first step toward their identification, we are planning whole genome re-sequencing of homozygous mutant harness racing horses which compete in pace and trot respectively. Furthermore, collecting phenotypic information about the pattern of locomotion in gaited horse breeds segregating for the *DMRT3* mutation can give us the opportunity to investigate the effect of the mutation in each specific breed. It is likely that the homozygous and heterozygous horses will differ phenotypically, as they do in the Icelandic Horse breed. It is worth noting that in many horse breeds pace is not a valued gait and riders train their ambling horses to avoid this lateral and sometimes quite uncomfortable gait. In Icelandic horses, pace is only ridden over short distances at a very high speed.

Obviously, the identification of this major *Gait* locus in horses will have practical implications. The large effects seen in racing performance make the mutation particularly valuable for diagnostic testing in these horses. Whilst this may lead to faster genetic gains, saving money and time for owners and trainers, it is of equal importance to a horse welfare perspective, since only horses with the right pre-conditions may be set in training. *CA* and *CC* horses likely need more training and many never reach the level required for entering a race, which might have led to a decision to slaughter the animal in the past. However, *CA* and *CC* Standardbred horses are likely to become better riding horses compared to *AA* (because of a superior ability to canter), which might

increase their value for this market instead. There is also an increased risk of injury in *CA* and *CC* horses when used for harness racing as they have difficulties in synchronizing trot at high speed. DNA testing may thus aid in matching the right horse to the right purpose, which is positive for both trainers and horses. Genetic testing can also be applied to Icelandic horses. This application would trace if the wild type allele from a four-gaited sire and/or dam has been transmitted to offspring. The identification of this major locus can also aid in the creation of new gaited horse breeds. For example, there is a group within the Morgan breed called “gaited Morgans” and genetic testing of these horses might aid in the creation of a distinct breed. The opportunities that are emerging are fascinating and it will be interesting to see how testing will be used and how this shapes the future development of different breeds.

Conclusions

The studies presented in this thesis have assisted in the functional annotation of the mammalian genome and also have practical implications for horse welfare. Reconnecting to the aims of this thesis, the specific conclusions are:

I Mutant *PMEL* causes Multiple Congenital Ocular Anomalies syndrome. The disease is undoubtedly inherited as an incompletely dominant trait. We were unable to detect anomalies in 6% of carriers, potentially because of incomplete penetrance.

II Variants in the MHC-II region influence susceptibility to Insect Bite Hypersensitivity. The same marker risk allele was present in two distinct populations, OR 4.19 ($p= 2.34x10^{-5}$) and 1.48 ($p= 0.043$) for Icelandic horses and Exmoor ponies respectively. In addition, homozygousity across the MHC-II region confers a higher risk of developing disease, OR= 2.67 ($p= 0.0013$).

III We have identified the first *Gait* locus in horses. A nonsense mutation in the transcription factor *DMRT3* is permissive for alternative gaits across multiple breeds. The lateral gait pace show recessive inheritance with incomplete penetrance in Icelandic horses. The *DMRT3* mutation also has a favorable effect on racing performance. Further analysis in mice shows that gene disruption changes spinal cord signaling, neuron development and ultimately locomotion.

References

Ahituv, N., Zhu, Y., Visel, A., Holt, A., Afzal, V., Pennacchio, L.A. & Rubin, E.M. (2007). Deletion of ultraconserved elements yields viable mice. *PLoS Biol* 5(9), e234.

Alberts, B. (2011). A grand challenge in biology. *Science* 333(6047), 1200.

Albertsdottir, E., Eriksson, S., Sigurdsson, A. & Arnason, T. (2011). Genetic analysis of 'breeding field test status' in Icelandic horses. *J Anim Breed Genet* 128(2), 124-32.

Albright-Fraser, D.G., Reid, R., Gerber, V. & Bailey, E. (1996). Polymorphism of DRA among equids. *Immunogenetics* 43(5), 315-7.

Andersson, L. (2001). Genetic dissection of phenotypic diversity in farm animals. *Nat Rev Genet* 2(2), 130-8.

Andersson, L. (2009). Genome-wide association analysis in domestic animals: a powerful approach for genetic dissection of trait loci. *Genetica* 136(2), 341-9.

Andersson, L. & Georges, M. (2004). Domestic-animal genomics: deciphering the genetics of complex traits. *Nat Rev Genet* 5(3), 202-12.

Andersson, L. & Sandberg, K. (1982). A linkage group composed of three coat color genes and three serum protein loci in horses. *J Hered* 73(2), 91-4.

Andersson, L.S., Axelsson, J., Dubielzig, R.R., Lindgren, G. & Ekesten, B. (2011). Multiple congenital ocular anomalies in Icelandic horses. *BMC Vet Res* 7, 21.

Andersson, L.S., Hogstrom, C., Mikko, S., Eriksson, S., Grandinson, K., Brostrom, H., Frey, R., Sundquist, M. & Lindgren, G. (2009). Polymorphisms in SPINK5 do not associate with insect bite hypersensitivity in Icelandic horses born in Sweden. *Anim Genet* 40(5), 790-1.

Andersson, L.S., Juras, R., Ramsey, D.T., Eason-Butler, J., Ewart, S., Cothran, G. & Lindgren, G. (2008). Equine Multiple Congenital Ocular Anomalies maps to a 4.9 megabase interval on horse chromosome 6. *BMC Genet* 9, 88.

Bailey, E. (1994). Variation within the antigen binding site of the major histocompatibility complex DRA gene of domestic horses. *Equine. Infect. Dis*. VII, 123–126.

Baxter, L.L. & Pavan, W.J. (2003). Pmel17 expression is Mitf-dependent and reveals cranial melanoblast migration during murine development. *Gene Expr Patterns* 3(6), 703-7.

Binns, M.M., Boehler, D.A. & 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. *Anim Genet* 41 Suppl 2, 154-8.

Bjornsdottir, S., Sigvaldadottir, J., Brostrom, H., Langvad, B. & Sigurdsson, A. (2006). Summer eczema in exported Icelandic horses: influence of environmental and genetic factors. *Acta Vet Scand* 48, 3.

Bowling, A.T. & Ruvinsky, A. (2000). *The Genetics of the Horse*: CAB International.

Brostrom, H., Larsson, A. & Troedsson, M. (1987). Allergic dermatitis (sweet itch) of Icelandic horses in Sweden: an epidemiological study. *Equine Vet J* 19(3), 229-36.

Brown, J.J., Thomson, W., Clegg, P., Eyre, S., Kennedy, L.J., Matthews, J., Carter, S. & Ollier, W.E. (2004). Polymorphisms of the equine major histocompatibility complex class II DRA locus. *Tissue Antigens* 64(2), 173-9.

Brunberg, E., Andersson, L., Cothran, G., Sandberg, K., Mikko, S. & Lindgren, G. (2006). A missense mutation in PMEL17 is associated with the Silver coat color in the horse. *BMC Genet* 7, 46.

Charlier, C., Coppieters, W., Rollin, F., Desmecht, D., Agerholm, J.S., Cambisano, N., Carta, E., Dardano, S., Dive, M., Fasquelle, C., Frennet, J.C., Hanset, R., Hubin, X., Jorgensen, C., Karim, L., Kent, M., Harvey, K., Pearce, B.R., Simon, P., Tama, N., Nie, H., Vandeputte, S., Lien, S., Longeri, M., Fredholm, M., Harvey, R.J. & Georges, M. (2008). Highly effective SNP-based association mapping and management of recessive defects in livestock. *Nat Genet* 40(4), 449-54.

Chavanas, S., Bodemer, C., Rochat, A., Hamel-Teillac, D., Ali, M., Irvine, A.D., Bonafe, J.L., Wilkinson, J., Taieb, A., Barrandon, Y., Harper, J.I., de Prost, Y. & Hovnanian, A. (2000). Mutations in SPINK5, encoding a serine protease inhibitor, cause Netherton syndrome. *Nat Genet* 25(2), 141-2.

Chowdhary, B.P. & Raudsepp, T. (2008). The horse genome derby: racing from map to whole genome sequence. *Chromosome Res* 16(1), 109-27.

Cieslak, M., Pruvost, M., Benecke, N., Hofreiter, M., Morales, A., Reissmann, M. & Ludwig, A. (2010). Origin and history of mitochondrial DNA lineages in domestic horses. *PLoS One* 5(12), e15311.

Clark, L.A., Wahl, J.M., Rees, C.A. & Murphy, K.E. (2006). Retrotransposon insertion in SILV is responsible for merle patterning of the domestic dog. *Proc Natl Acad Sci U S A* 103(5), 1376-81.

Clop, A., Marcq, F., Takeda, H., Pirottin, D., Tordoir, X., Bibe, B., Bouix, J., Caiment, F., Elsen, J.M., Eychenne, F., Larzul, C., Laville, E., Meish, F., Milenkovic, D., Tobin, J., Charlier, C. & Georges, M. (2006). A mutation

creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nat Genet* 38(7), 813-8.

Cothran, E.G., MacCluer, J.W., Weitkamp, L.R. & Bailey, E. (1987). Genetic differentiation associated with gait within American standardbred horses. *Anim Genet* 18(4), 285-96.

Culbertson, M.R. & Leeds, P.F. (2003). Looking at mRNA decay pathways through the window of molecular evolution. *Curr Opin Genet Dev* 13(2), 207-14.

Dausch, D., Wegner, W., Michaelis, W. & Reetz, I. (1978). [Eye changes in the merle syndrome in the dog (author's transl)]. *Albrecht Von Graefes Arch Klin Exp Ophthalmol* 206(2), 135-50.

Ellegren, H., Johansson, M., Sandberg, K. & Andersson, L. (1992). Cloning of highly polymorphic microsatellites in the horse. *Anim Genet* 23(2), 133-42.

Eriksson, S., Grandinson, K., Fikse, W.F., Lindberg, L., Mikko, S., Broström, H., Frey, R., Sundquist, M. & Lindgren, G. (2008). Genetic analysis of insect bite hypersensitivity (summer eczema) in Icelandic horses. *Animal* 2(3), pp 360–365.

Ewart, S.L., Ramsey, D.T., Xu, J. & Meyers, D. (2000). The horse homolog of congenital aniridia conforms to codominant inheritance. *J Hered* 91(2), 93-8.

Fujii, J., Otsu, K., Zorzato, F., de Leon, S., Khanna, V.K., Weiler, J.E., O'Brien, P.J. & MacLennan, D.H. (1991). Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. *Science* 253(5018), 448-51.

Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., Liu-Cordero, S.N., Rotimi, C., Adeyemo, A., Cooper, R., Ward, R., Lander, E.S., Daly, M.J. & Altshuler, D. (2002). The structure of haplotype blocks in the human genome. *Science* 296(5576), 2225-9.

Gelatt, K.N., Powell, N.G. & Huston, K. (1981). Inheritance of microphthalmia with coloboma in the Australian shepherd dog. *Am J Vet Res* 42(10), 1686-90.

Georges, M. (2007). Mapping, fine mapping, and molecular dissection of quantitative trait Loci in domestic animals. *Annu Rev Genomics Hum Genet* 8, 131-62.

Goddard, M.E. & Hayes, B.J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat Rev Genet* 10(6), 381-91.

Goulding, M. (2009). Circuits controlling vertebrate locomotion: moving in a new direction. *Nat Rev Neurosci* 10(7), 507-18.

Grahn, B.H., Pinard, C., Archer, S., Bellone, R., Forsyth, G. & Sandmeyer, L.S. (2008). Congenital ocular anomalies in purebred and crossbred Rocky and Kentucky Mountain horses in Canada. *Can Vet J* 49(7), 675-81.

Grisart, B., Coppieters, W., Farnir, F., Karim, L., Ford, C., Berzi, P., Cambisano, N., Mni, M., Reid, S., Simon, P., Spelman, R., Georges, M. & Snell, R. (2002). Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. *Genome Res* 12(2), 222-31.

Halldorsdottir, S., Lazary, S., Gunnarsson, E. & Larsen, H.J. (1991). Distribution of leucocyte antigens in Icelandic horses affected with summer eczema compared to non-affected horses. *Equine Vet J* 23(4), 300-2.

Hellstrom, A.R., Watt, B., Fard, S.S., Tenza, D., Mannstrom, P., Narfstrom, K., Ekesten, B., Ito, S., Wakamatsu, K., Larsson, J., Ulfendahl, M., Kullander, K., Raposo, G., Kerje, S., Hallbook, F., Marks, M.S. & Andersson, L. (2011). Inactivation of Pmel alters melanosome shape but has only a subtle effect on visible pigmentation. *PLoS Genet* 7(9), e1002285.

Hendricks, B. (1995). *International encyclopedia of horse breeds*. 1. ed: Univeristy 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.E. (2010). A sequence polymorphism in MSTN predicts sprinting ability and racing stamina in thoroughbred horses. *PLoS One* 5(1), e8645.

Hu, D.N., Savage, H.E. & Roberts, J.E. (2002). Uveal melanocytes, ocular pigment epithelium, and Muller cells in culture: in vitro toxicology. *Int J Toxicol* 21(6), 465-72.

Hu, D.N., Simon, J.D. & Sarna, T. (2008). Role of ocular melanin in ophthalmic physiology and pathology. *Photochem Photobiol* 84(3), 639-44.

Hurst, C. (1906). On the inheritance of coat colour in horses. *Proceedings of the Royal Society* (B77), 388-394.

Ishihara, U.H. (1957). Studies on summer mange ("Kasen" disease) of the horse in Japan. 111. Skin sensitivity tests with insect allergens. *Bull Nat Inst Anim Health Tokyo* 32, 217-233.

Jansen, T., Forster, P., Levine, M.A., Oelke, H., Hurles, M., Renfrew, C., Weber, J. & Olek, K. (2002). Mitochondrial DNA and the origins of the domestic horse. *Proc Natl Acad Sci U S A* 99(16), 10905-10.

Karlsson, A.C., Kerje, S., Hallbook, F. & Jensen, P. (2009). The Dominant white mutation in the PMEL17 gene does not cause visual impairment in chickens. *Vet Ophthalmol* 12(5), 292-8.

Kerje, S., Sharma, P., Gunnarsson, U., Kim, H., Bagchi, S., Fredriksson, R., Schutz, K., Jensen, P., von Heijne, G., Okimoto, R. & Andersson, L. (2004). The Dominant white, Dun and Smoky color variants in chicken are associated with insertion/deletion polymorphisms in the PMEL17 gene. *Genetics* 168(3), 1507-18.

Kiyohara, C., Tanaka, K. & Miyake, Y. (2008). Genetic susceptibility to atopic dermatitis. *Allergol Int* 57(1), 39-56.

Komaromy, A.M., Rowlan, J.S., La Croix, N.C. & Mangan, B.G. (2011). Equine Multiple Congenital Ocular Anomalies (MCOA) syndrome in PMEL17 (Silver) mutant ponies: five cases. *Vet Ophthalmol* 14(5), 313-20.

Kuehn, C. & Weikard, R. (2007). Multiple splice variants within the bovine silver homologue (SILV) gene affecting coat color in cattle indicate a function additional to fibril formation in melanophores. *BMC Genomics* 8, 335.

Kullander, K. (2005). Genetics moving to neuronal networks. *Trends Neurosci* 28(5), 239-47.

Kwon, B.S., Halaban, R., Ponnazhagan, S., Kim, K., Chintamaneni, C., Bennett, D. & Pickard, R.T. (1995). Mouse silver mutation is caused by a single base insertion in the putative cytoplasmic domain of Pmel 17. *Nucleic Acids Res* 23(1), 154-8.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M.V., Kaul, R., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., de la Bastide, M., Dedhia,

N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E.V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J.V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.P., Yeh, R.F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Patrinos, A., Morgan, M.J., de Jong, P., Catanese, J.J., Osoegawa, K., Shizuya, H., Choi, S. & Chen, Y.J. (2001). Initial sequencing and analysis of the human genome. *Nature* 409(6822), 860-921.

Lango Allen, H., Estrada, K., Lettre, G., Berndt, S.I., Weedon, M.N., Rivadeneira, F., Willer, C.J., Jackson, A.U., Vedantam, S., Raychaudhuri, S., Ferreira, T., Wood, A.R., Weyant, R.J., Segre, A.V., Speliotes, E.K., Wheeler, E., Soranzo, N., Park, J.H., Yang, J., Gudbjartsson, D., Heard-Costa, N.L., Randall, J.C., Qi, L., Vernon Smith, A., Magi, R., Pastinen, T., Liang, L., Heid, I.M., Luan, J., Thorleifsson, G., Winkler, T.W., Goddard, M.E., Sin Lo, K., Palmer, C., Workalemahu, T., Aulchenko, Y.S., Johansson, A., Zillikens, M.C., Feitosa, M.F., Esko, T., Johnson, T., Ketkar, S., Kraft, P., Mangino, M., Prokopenko, I., Absher, D., Albrecht, E., Ernst, F., Glazer, N.L., Hayward, C., Hottenga, J.J., Jacobs, K.B., Knowles, J.W., Kutalik, Z., Monda, K.L., Polasek, O., Preuss, M., Rayner, N.W., Robertson, N.R., Steinthorsdottir, V., Tyrer, J.P., Voight, B.F., Wiklund, F., Xu, J., Zhao, J.H., Nyholt, D.R., Pellikka, N., Perola, M., Perry, J.R., Surakka, I., Tammesoo, M.L., Altmaier, E.L., Amin, N., Aspelund, T., Bhangale, T., Boucher, G., Chasman, D.I., Chen, C., Coin, L., Cooper, M.N., Dixon, A.L., Gibson, Q., Grundberg, E., Hao, K., Juhani Junttila, M., Kaplan, L.M., Kettunen, J., Konig, I.R., Kwan, T., Lawrence, R.W., Levinson, D.F., Lorentzon, M., McKnight, B., Morris, A.P., Muller, M., Suh Ngwa, J., Purcell, S., Rafelt, S., Salem, R.M., Salvi, E., Sanna, S., Shi, J., Sovio, U., Thompson, J.R., Turchin, M.C., Vandenput, L., Verlaan, D.J., Vitart, V., White, C.C., Ziegler, A., Almgren, P., Balmforth, A.J., Campbell, H., Citterio, L., De Grandi, A., Dominiczak, A., Duan, J., Elliott, P., Elosua, R., Eriksson, J.G., Freimer, N.B., Geus, E.J., Glorioso, N., Haiqing, S., Hartikainen, A.L., Havulinna, A.S., Hicks, A.A., Hui, J., Igl, W., Illig, T., Jula, A., Kajantie, E., Kilpelainen, T.O., Koiranen, M., Kolcic, I., Koskinen, S., Kovacs, P., Laitinen, J., Liu, J., Lokki, M.L., Marusic, A., Maschio, A., Meitinger, T., Mulas, A., Pare, G., Parker, A.N., Peden, J.F., Petersmann, A., Pichler, I., Pietilainen, K.H., Pouta, A., Ridderstrale, M., Rotter, J.I., Sambrook, J.G., Sanders, A.R., Schmidt, C.O., Sinisalo, J.,

Smit, J.H., Stringham, H.M., Bragi Walters, G., Widen, E., Wild, S.H., Willemsen, G., Zagato, L., Zgaga, L., Zitting, P., Alavere, H., Farrall, M., McArdle, W.L., Nelis, M., Peters, M.J., Ripatti, S., van Meurs, J.B., Aben, K.K., Ardlie, K.G., Beckmann, J.S., Beilby, J.P., Bergman, R.N., Bergmann, S., Collins, F.S., Cusi, D., den Heijer, M., Eiriksdottir, G., Gejman, P.V., Hall, A.S., Hamsten, A., Huikuri, H.V., Iribarren, C., Kahonen, M., Kaprio, J., Kathiresan, S., Kiemeney, L., Kocher, T., Launer, L.J., Lehtimaki, T., Melander, O., Mosley, T.H., Jr., Musk, A.W., Nieminen, M.S., O'Donnell, C.J., Ohlsson, C., Oostra, B., Palmer, L.J., Raitakari, O., Ridker, P.M., Rioux, J.D., Rissanen, A., Rivolta, C., Schunkert, H., Shuldiner, A.R., Siscovick, D.S., Stumvoll, M., Tonjes, A., Tuomilehto, J., van Ommen, G.J., Viikari, J., Heath, A.C., Martin, N.G., Montgomery, G.W., Province, M.A., Kayser, M., Arnold, A.M., Atwood, L.D., Boerwinkle, E., Chanock, S.J., Deloukas, P., Gieger, C., Gronberg, H., Hall, P., Hattersley, A.T., Hengstenberg, C., Hoffman, W., Lathrop, G.M., Salomaa, V., Schreiber, S., Uda, M., Waterworth, D., Wright, A.F., Assimes, T.L., Barroso, I., Hofman, A., Mohlke, K.L., Boomsma, D.I., Caulfield, M.J., Cupples, L.A., Erdmann, J., Fox, C.S., Gudnason, V., Gyllensten, U., Harris, T.B., Hayes, R.B., Jarvelin, M.R., Mooser, V., Munroe, P.B., Ouwehand, W.H., Penninx, B.W., Pramstaller, P.P., Quertermous, T., Rudan, I., Samani, N.J., Spector, T.D., Volzke, H., Watkins, H., Wilson, J.F., Groop, L.C., Haritunians, T., Hu, F.B., Kaplan, R.C., Metspalu, A., North, K.E., Schlessinger, D., Wareham, N.J., Hunter, D.J., O'Connell, J.R., Strachan, D.P., Wichmann, H.E., Borecki, I.B., van Duijn, C.M., Schadt, E.E., Thorsteinsdottir, U., Peltonen, L., Uitterlinden, A.G., Visscher, P.M., Chatterjee, N., Loos, R.J., Boehnke, M., McCarthy, M.I., Ingelsson, E., Lindgren, C.M., Abecasis, G.R., Stefansson, K., Frayling, T.M. & Hirschhorn, J.N. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 467(7317), 832-8.

Lazary, S., Marti, E., Szalai, G., Gaillard, C. & Gerber, H. (1994). Studies on the frequency and associations of equine leucocyte antigens in sarcoid and summer dermatitis. *Anim Genet* 25 Suppl 1, 75-80.

Lindblad-Toh, K., Garber, M., Zuk, O., Lin, M.F., Parker, B.J., Washietl, S., Kheradpour, P., Ernst, J., Jordan, G., Mauceli, E., Ward, L.D., Lowe, C.B., Holloway, A.K., Clamp, M., Gnerre, S., Alfoldi, J., Beal, K., Chang, J., Clawson, H., Cuff, J., Di Palma, F., Fitzgerald, S., Flicek, P., Guttman, M., Hubisz, M.J., Jaffe, D.B., Jungreis, I., Kent, W.J., Kostka, D., Lara, M., Martins, A.L., Massingham, T., Moltke, I., Raney, B.J., Rasmussen, M.D., Robinson, J., Stark, A., Vilella, A.J., Wen, J., Xie, X., Zody, M.C., Baldwin, J., Bloom, T., Chin, C.W., Heiman, D., Nicol, R., Nusbaum, C., Young, S., Wilkinson, J., Worley, K.C., Kovar, C.L., Muzny, D.M., Gibbs, R.A., Cree, A., Dihn, H.H., Fowler, G., Jhangiani, S., Joshi, V., Lee, S., Lewis, L.R., Nazareth, L.V., Okwuonu, G., Santibanez, J., Warren, W.C., Mardis, E.R., Weinstock, G.M., Wilson, R.K., Delehaunty,

K., Dooling, D., Fronik, C., Fulton, L., Fulton, B., Graves, T., Minx, P., Sodergren, E., Birney, E., Margulies, E.H., Herrero, J., Green, E.D., Haussler, D., Siepel, A., Goldman, N., Pollard, K.S., Pedersen, J.S., Lander, E.S. & Kellis, M. (2011). A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* 478(7370), 476-82.

Lindblad-Toh, K., Wade, C.M., Mikkelsen, T.S., Karlsson, E.K., Jaffe, D.B., Kamal, M., Clamp, M., Chang, J.L., Kulbokas, E.J., 3rd, Zody, M.C., Mauceli, E., Xie, X., Breen, M., Wayne, R.K., Ostrander, E.A., Ponting, C.P., Galibert, F., Smith, D.R., DeJong, P.J., Kirkness, E., Alvarez, P., Biagi, T., Brockman, W., Butler, J., Chin, C.W., Cook, A., Cuff, J., Daly, M.J., DeCaprio, D., Gnerre, S., Grabherr, M., Kellis, M., Kleber, M., Bardeleben, C., Goodstadt, L., Heger, A., Hitte, C., Kim, L., Koepfli, K.P., Parker, H.G., Pollinger, J.P., Searle, S.M., Sutter, N.B., Thomas, R., Webber, C., Baldwin, J., Abebe, A., Abouelleil, A., Aftuck, L., Ait-Zahra, M., Aldredge, T., Allen, N., An, P., Anderson, S., Antoine, C., Arachchi, H., Aslam, A., Ayotte, L., Bachantsang, P., Barry, A., Bayul, T., Benamara, M., Berlin, A., Bessette, D., Blitshteyn, B., Bloom, T., Blye, J., Boguslavskiy, L., Bonnet, C., Boukhgalter, B., Brown, A., Cahill, P., Calixte, N., Camarata, J., Cheshatsang, Y., Chu, J., Citroen, M., Collymore, A., Cooke, P., Dawoe, T., Daza, R., Decktor, K., DeGray, S., Dhargay, N., Dooley, K., Dorje, P., Dorjee, K., Dorris, L., Duffey, N., Dupes, A., Egbiremolen, O., Elong, R., Falk, J., Farina, A., Faro, S., Ferguson, D., Ferreira, P., Fisher, S., FitzGerald, M., Foley, K., Foley, C., Franke, A., Friedrich, D., Gage, D., Garber, M., Gearin, G., Giannoukos, G., Goode, T., Goyette, A., Graham, J., Grandbois, E., Gyaltsen, K., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Healy, C., Hegarty, R., Honan, T., Horn, A., Houde, N., Hughes, L., Hunnicutt, L., Husby, M., Jester, B., Jones, C., Kamat, A., Kanga, B., Kells, C., Khazanovich, D., Kieu, A.C., Kisner, P., Kumar, M., Lance, K., Landers, T., Lara, M., Lee, W., Leger, J.P., Lennon, N., Leuper, L., LeVine, S., Liu, J., Liu, X., Lokyitsang, Y., Lokyitsang, T., Lui, A., Macdonald, J., Major, J., Marabella, R., Maru, K., Matthews, C., McDonough, S., Mehta, T., Meldrim, J., Melnikov, A., Meneus, L., Mihalev, A., Mihova, T., Miller, K., Mittelman, R., Mlenga, V., Mulrain, L., Munson, G., Navidi, A., Naylor, J., Nguyen, T., Nguyen, N., Nguyen, C., Nicol, R., Norbu, N., Norbu, C., Novod, N., Nyima, T., Olandt, P., O'Neill, B., O'Neill, K., Osman, S., Oyono, L., Patti, C., Perrin, D., Phunkhang, P., Pierre, F., Priest, M., Rachupka, A., Raghuraman, S., Rameau, R., Ray, V., Raymond, C., Rege, F., Rise, C., Rogers, J., Rogov, P., Sahalie, J., Settipalli, S., Sharpe, T., Shea, T., Sheehan, M., Sherpa, N., Shi, J., Shih, D., Sloan, J., Smith, C., Sparrow, T., Stalker, J., Stange-Thomann, N., Stavropoulos, S., Stone, C., Stone, S., Sykes, S., Tchuinga, P., Tenzing, P., Tesfaye, S., Thoulutsang, D., Thoulutsang, Y., Topham, K., Topping, I., Tsamla, T., Vassiliev, H., Venkataraman, V., Vo, A., Wangchuk, T., Wangdi, T., Weiand, M., Wilkinson, J., Wilson, A., Yadav, S., Yang, S.,

Yang, X., Young, G., Yu, Q., Zainoun, J., Zembek, L., Zimmer, A. & Lander, E.S. (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438(7069), 803-19.

Lindgren, G., Sandberg, K., Persson, H., Marklund, S., Breen, M., Sandgren, B., Carlsten, J. & Ellegren, H. (1998). A primary male autosomal linkage map of the horse genome. *Genome Res* 8(9), 951-66.

Lippold, S., Knapp, M., Kuznetsova, T., Leonard, J.A., Benecke, N., Ludwig, A., Rasmussen, M., Cooper, A., Weinstock, J., Willerslev, E., Shapiro, B. & Hofreiter, M. (2011a). Discovery of lost diversity of paternal horse lineages using ancient DNA. *Nat Commun* 2, 450.

Lippold, S., Matzke, N.J., Reissmann, M. & Hofreiter, M. (2011b). Whole mitochondrial genome sequencing of domestic horses reveals incorporation of extensive wild horse diversity during domestication. *BMC Evol Biol* 11, 328.

Ludwig, A., Pruvost, M., Reissmann, M., Benecke, N., Brockmann, G.A., Castanos, P., Cieslak, M., Lippold, S., Llorente, L., Malaspinas, A.S., Slatkin, M. & Hofreiter, M. (2009). Coat color variation at the beginning of horse domestication. *Science* 324(5926), 485.

Marti, E., Gerber, H. & Lazary, S. (1992). On the genetic basis of equine allergic diseases: II. Insect bite dermal hypersensitivity. *Equine Vet J* 24(2), 113-7.

McCue, M.E., Bannasch, D.L., Petersen, J.L., Gurr, J., Bailey, E., Binns, M.M., Distl, O., Guerin, G., Hasegawa, T., Hill, E.W., Leeb, T., Lindgren, G., Penedo, M.C., Roed, K.H., Ryder, O.A., Swinburne, J.E., Tozaki, T., Valberg, S.J., Vaudin, M., Lindblad-Toh, K., Wade, C.M. & Mickelson, J.R. (2012). A high density SNP array for the domestic horse and extant Perissodactyla: utility for association mapping, genetic diversity, and phylogeny studies. *PLoS Genet* 8(1), e1002451.

Mellor, P.S. & McCraig, J. (1974). The probable cause of "sweet itch" in England. *Vet Rec* 95(18), 411-5.

Nishimaru, H. & Kakizaki, M. (2009). The role of inhibitory neurotransmission in locomotor circuits of the developing mammalian spinal cord. *Acta Physiol (Oxf)* 197(2), 83-97.

Pinard, C.L. & Basrur, P.K. (2011). Ocular anomalies in a herd of Exmoor ponies in Canada. *Vet Ophthalmol* 14(2), 100-8.

Plummer, C.E. & Ramsey, D.T. (2011). A survey of ocular abnormalities in miniature horses. *Vet Ophthalmol* 14(4), 239-43.

Ramsey, D.T., Ewart, S.L., Render, J.A., Cook, C.S. & Latimer, C.A. (1999). Congenital ocular abnormalities of Rocky Mountain Horses. *Vet Ophthalmol* 2(1), 47-59.

Rothfels, K., Alexrad, A., Siminovitch, L. & McCulloch Parker, R. (1959). The origin of altered cell lines from mouse, monkey and man, as indicated by chromosome and transplantation studies. *Proceeding of Canadian Cancer Research Conference* (3), 189-214.

Rubin, C.J., Zody, M.C., Eriksson, J., Meadows, J.R., Sherwood, E., Webster, M.T., Jiang, L., Ingman, M., Sharpe, T., Ka, S., Hallbook, F., Besnier, F., Carlborg, O., Bed'hom, B., Tixier-Boichard, M., Jensen, P., Siegel, P., Lindblad-Toh, K. & Andersson, L. (2010). Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* 464(7288), 587-91.

Saeki, H., Kuwata, S., Nakagawa, H., Etoh, T., Yanagisawa, M., Miyamoto, M., Tokunaga, K., Juji, T. & Shibata, Y. (1995). Analysis of disease-associated amino acid epitopes on HLA class II molecules in atopic dermatitis. *J Allergy Clin Immunol* 96(6 Pt 2), 1061-8.

Salmon Hillbertz, N.H., Isaksson, M., Karlsson, E.K., Hellmen, E., Pielberg, G.R., Savolainen, P., Wade, C.M., von Euler, H., Gustafson, U., Hedhammar, A., Nilsson, M., Lindblad-Toh, K., Andersson, L. & Andersson, G. (2007). Duplication of FGF3, FGF4, FGF19 and ORAOV1 causes hair ridge and predisposition to dermoid sinus in Ridgeback dogs. *Nat Genet* 39(11), 1318-20.

Schaffartzik, A., Hamza, E., Janda, J., Crameri, R., Marti, E. & Rhyner, C. (2012). Equine insect bite hypersensitivity: What do we know? *Vet Immunol Immunopathol* 147(3-4), 113-26.

Schonthaler, H.B., Lampert, J.M., von Lintig, J., Schwarz, H., Geisler, R. & Neuhauss, S.C. (2005). A mutation in the silver gene leads to defects in melanosome biogenesis and alterations in the visual system in the zebrafish mutant fading vision. *Dev Biol* 284(2), 421-36.

Schurink, A., Ducro, B.J., Bastiaansen, J.W., Frankena, K. & van Arendonk, J.A. (2012). Genome-wide association study of insect bite hypersensitivity in Dutch Shetland pony mares. *Anim Genet*.

Schurink, A., Ducro, B.J., Heuven, H.C. & van Arendonk, J.A. (2011). Genetic parameters of insect bite hypersensitivity in Dutch Friesian broodmares. *J Anim Sci* 89(5), 1286-93.

Schurink, A., van Grevenhof, E.M., Ducro, B.J. & van Arendonk, J.A. (2009). Heritability and repeatability of insect bite hypersensitivity in Dutch Shetland breeding mares. *J Anim Sci* 87(2), 484-90.

Strauss, O. (2005). The retinal pigment epithelium in visual function. *Physiol Rev* 85(3), 845-81.

Trommershausen-Smith, A. (1978). Linkage of tobiano coat spotting and albumin markers in a pony family. *J Hered* 69(4), 214-6.

Tseng, C.T., Miller, D., Cassano, J., Bailey, E. & Antczak, D.F. (2010). Identification of equine major histocompatibility complex haplotypes using polymorphic microsatellites. *Anim Genet* 41 Suppl 2, 150-3.

Van Laere, A.S., Nguyen, M., Braunschweig, M., Nezer, C., Collette, C., Moreau, L., Archibald, A.L., Haley, C.S., Buys, N., Tally, M., Andersson, G., Georges, M. & Andersson, L. (2003). A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature* 425(6960), 832-6.

Vaysse, A., Ratnakumar, A., Derrien, T., Axelsson, E., Rosengren Pielberg, G., Sigurdsson, S., Fall, T., Seppala, E.H., Hansen, M.S., Lawley, C.T., Karlsson, E.K., Bannasch, D., Vila, C., Lohi, H., Galibert, F., Fredholm, M., Haggstrom, J., Hedhammar, A., Andre, C., Lindblad-Toh, K., Hitte, C. & Webster, M.T. (2011). Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. *PLoS Genet* 7(10), e1002316.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G.V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M.L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N.N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J.F., Guigo, R., Campbell, M.J., Sjolander, K.V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang,

Y.H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A. & Zhu, X. (2001). The sequence of the human genome. *Science* 291(5507), 1304-51.

Vila, C., Leonard, J.A., Gotherstrom, A., Marklund, S., Sandberg, K., Liden, K., Wayne, R.K. & Ellegren, H. (2001). Widespread origins of domestic horse lineages. *Science* 291(5503), 474-7.

Wade, C.M., Giulotto, E., Sigurdsson, S., Zoli, M., Gnerre, S., Imsland, F., Lear, T.L., Adelson, D.L., Bailey, E., Bellone, R.R., Blocker, H., Distl, O., Edgar, R.C., Garber, M., Leeb, T., Mauceli, E., MacLeod, J.N., Penedo, M.C., Raison, J.M., Sharpe, T., Vogel, J., Andersson, L., Antczak, D.F., Biagi, T., Binns, M.M., Chowdhary, B.P., Coleman, S.J., Della Valle, G., Fryc, S., Guerin, G., Hasegawa, T., Hill, E.W., Jurka, J., Kiialainen, A., Lindgren, G., Liu, J., Magnani, E., Mickelson, J.R., Murray, J., Nergadze, S.G., Onofrio, R., Pedroni, S., Piras, M.F., Raudsepp, T., Rocchi, M., Roed, K.H., Ryder, O.A., Searle, S., Skow, L., Swinburne, J.E., Syvanen, A.C., Tozaki, T., Valberg, S.J., Vaudin, M., White, J.R., Zody, M.C., Lander, E.S. & Lindblad-Toh, K. (2009). Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science* 326(5954), 865-7.

Walley, A.J., Chavanas, S., Moffatt, M.F., Esnouf, R.M., Ubhi, B., Lawrence, R., Wong, K., Abecasis, G.R., Jones, E.Y., Harper, J.I., Hovnanian, A. & Cookson, W.O. (2001). Gene polymorphism in Netherton and common atopic disease. *Nat Genet* 29(2), 175-8.

Watt, B., Tenza, D., Lemmon, M.A., Kerje, S., Raposo, G., Andersson, L. & Marks, M.S. (2011). Mutations in or near the transmembrane domain alter PMEL amyloid formation from functional to pathogenic. *PLoS Genet* 7(9), e1002286.

Acknowledgements

Först och främst vill jag tacka min huvudhandledare **Gabriella Lindgren**. Utan dig hade inget av detta varit möjligt. Tack för att du sökte upp mig med ett erbjudande att göra examensarbete och senare även en doktorandtjänst inom molekylärgenetik på häst. Jag hade inte kunnat drömma om att få jobba med en så lyckad kombination. Tack för alla kloka råd, uppskattande ord och förnuftig vägledning! Tack även till mina biträdande handledare **Hans Broström** och **Sofia Mikko**. Vi har haft många trevliga stunder tillsammans då vi diskuterat hästforskning.

Jag vill också säga stort tack till alla personer som har bidragit till de studier som ligger till grund för denna avhandling, tack **Björn**, **Calle**, **Fatima**, **Freddy**, **Freyja**, **Göran H**, **Hanna W**, **Jennifer**, **Katarina L**, **Klas**, **Lars**, **Lisbeth**, **Maria**, **Marie**, **Martin**, **Rebecka**, **Susanne** och **Thorvaldur**. Er hjälp har varit ovärderlig och ni har gjort forskningen till ett nöje! Tack **Leif** för all hjälp i *DMRT3* projektet (”Det Är För Bra För Att Vara Sant-projektet”), alla intressanta vetenskapliga samtal och att du visat att det lönar sig att kämpa in i det sista.

Special thanks to **Doreen** and **Marta** for valued assistance in the *DMRT3* project. I’m glad to have such competent people involved in this exciting research project! Thanks **Agnese** for help with genotyping in the MCOA project. Thanks also to all people who have contributed with knowledge and horse samples including **Gus Cothran, Susan Ewart**, **June Swinburne** and the Minnesota group **Jessica, Molly** and **Jim**. I have enjoyed collaborating with all of you!

Tack även **Göran A** för vetenskaplig input inom det funktionella området och för att du är en så bra avdelningschef. **Ulla**, tack för handledning i labbet och att du har hjälp till med laboratorier när det har krisat! Du är en klippa. Tack

Gudrun och senare **Helena** för hjälp i administrativa frågor. **Jennifer**, tack för all korrekturläsning – av abstracts, postrar, presentationer, artiklar och nu även avhandlingen. Din hjälp har varit helt ovärderlig. Jag skriver detta på svenska då jag vet att du förstår ☺. Tack också **Martin** för hjälp med figurer till denna avhandling och till artikel III.

Thanks to **all past and present members** of the lab for creating such a nice working environment! I whish you the best of luck with your own research projects.

Speciellt tack till mina förra rumskamrater **Maria, Jonas, Anders** och senare även **Abhi**. Ni har gjort tiden som doktorand jätterolig och det hade inte varit detsamma utan Er!! Jag vill också tacka ”tjejerna” för en otroligt kul och minnesvärd tid med mycket samtal, fest och dans, speciellt **Jeanette, Jennifer, Maria** och **Mia**, även om den senare inte gick ut med oss speciellt ofta ☺.

Jag vill även passa på att tacka alla bidragande veterinärer och hästägare. Utan er hjälp med prover, fenotypiska beskrivningar och tillåtelse till undersökningar hade inte dessa studier varit möjliga. Jag minns med extra glädje de dagar jag har fått lämna labbet och kontoret för åka ut ”på fältet” tack **Hasse, Björn, Jeanette, Doreen,** och **Martin** för att ni varit med på dessa resor ☺. Tack även nuvarande och dåvarande personal på husdjursgenetiska laboratoriet för hjälp med prover.

Sist men inte minst vill jag tacka min familj för att ni har ställt upp och stöttat! Speciellt tack till **mamma** och **pappa** för att ni mutade mig så att jag skulle få ok betyg på gymnasiet och som har gjort att jag aldrig någonsin behövt oroa mig för något annat än studierna under min tid i Uppsala. Tack också min underbara sambo **Marcus** för ovärderligt stöd och uppmuntran!