

# Genetics of Equine Insect Bite Hypersensitivity and Genetic Diversity in Horses

Merina Shrestha

**ISSN:** 1652-6880

**ISBN (print version):** 978-91-576-8775-3

**ISBN (electronic version):** 978-91-576-8776-0

**ISBN:** 978-94-6343-016-6

**DOI:** <http://dx.doi.org/10.18174/396810>

### **Thesis committee**

#### **Promotors**

Prof. Dr H.J. Bovenhuis

Professor of Animal Breeding and Genomics Centre  
Wageningen University and Research, The Netherlands

Prof. Dr D.J. de Koning

Professor of Animal Breeding and Genetics  
Swedish University of Agricultural Sciences, Sweden

#### **Co-promotors**

Dr. B.J. Ducro

Associate professor, Animal Breeding and Genomics Centre  
Wageningen University and Research, The Netherlands

Dr. A. M. Johansson

Associate professor, Department of Animal Breeding and Genetics  
Swedish University of Agricultural Sciences, Sweden

#### **Other members (assessment committee)**

Dr K. F. Stock, Vereinigte Informationssysteme Tierhaltung W.V., Germany

Prof. Dr H. Savelkoul, Wageningen University and Research, The Netherlands

Dr T. Slotte, Stockholm University, Sweden

Prof. Dr N. Lundenheim, Swedish University of Agricultural Sciences, Sweden

The research presented in this doctoral thesis was conducted under the joint auspices of the Swedish University of Agricultural Sciences and the Graduate School Wageningen Institute of Animal Sciences of Wageningen University and is part of the Erasmus Mundus Joint Doctorate program "EGS-ABG".

# Genetics of Equine Insect Bite Hypersensitivity and Genetic Diversity in Horses

**Merina Shrestha**



ACTA UNIVERSITATIS AGRICULTURAE SUECIAE  
DOCTORAL THESIS No 2017:1

## **Thesis**

submitted in fulfillment of the requirements for the degree of doctor from

**Swedish University of Agricultural Sciences**

by the authority of the Board of the Faculty of Veterinary Medicine and Animal  
Science and from

**Wageningen University**

by the authority of the Rector Magnificus, Prof. Dr. A.P.J. Mol,

in the presence of the

Thesis Committee appointed by the Academic Board of Wageningen University  
and

the Board of the Faculty of Veterinary Medicine and Animal Science at the Swedish  
University of Agricultural Sciences

to be defended in public

on Friday 27 January, 2017

at 9.15 a.m. at the Swedish University of Agricultural Sciences

Acta Universitatis agriculturae Sueciae

2017:1

Cover: Merina Shrestha (design and drawing)  
(Suggestions: Cano Merkan and Juan Cordero)

ISSN 1652-6880

ISBN (print version) 978-91-576-8775-3

ISBN (electronic version) 978-91-576-8776-0

© 2017 Merina Shrestha, Uppsala

Print: SLU Service/Repro, Uppsala 2017

ISBN: 978-94-6343-016-6

DOI: <http://dx.doi.org/10.18174/396810>

# Genetics of Equine Insect Bite Hypersensitivity and Genetic Diversity in Horses

## Abstract

Genetic variation contributing to the phenotypic variation was utilized in this thesis to understand the genetic background of a complex trait IBH, and to understand genetic diversity and relationships between various horse populations.

IBH is the most common skin allergic disorder in horses, caused by bites of midges, predominantly *Culicoides* species. It affects various horse breeds worldwide. With no effective treatment, IBH degrades horse health and causes economic loss. In this thesis, we used genome-wide SNPs to identify regions contributing to genetic variance of IBH susceptibility. We also investigated influence of increased number of horses and dense SNPs on identification of genomic regions associated to IBH susceptibility. Multiple genomic regions with small effects were observed in Studies I-III. Interesting genomic regions in the Icelandic horse population across the studies I and II, was observed on chromosomes 1, 7, 10, 15 and 17. The percentage of the genetic variance explained by top ten windows increased from 3.07% (Study I) to 6.56% (Study II). Novel genomic regions were identified when number of Icelandic horses was increased in Study II. Using dense SNPs on the Exmoor pony population we identified novel genomic regions, on chr 8, associated to IBH susceptibility, though with borderline significance.

In Study IV, pre-conceived understanding about evolutionary history of horse populations matched obtained results from investigation of genetic relationships within Dutch warmblood populations (pairwise mean  $F_{ST} \leq 0.070$ ), and within pony-like populations (pairwise mean  $F_{ST} \leq 0.078$ ). Horse populations with similar genetic background might share similar genetic components for IBH susceptibility. The Friesian horse population had lowest diversity (mean inbreeding coefficients:  $f_i$ : 30.4%,  $f_{ROH}$ = 22.2%) in Study IV and was genetically distinct ( $F_{ST}$  ranged from 0.13 to 0.17). This might be a result of a history of several population bottlenecks and selection on a closed breeding scheme. Low diversity in immunity related genes, observed in the Friesian horse population, might have led to increased prevalence of IBH. Similarly, low susceptibility of IBH in a warmblood population, KWPN sport horse population might be due to high genetic diversity ( $f_i$  =-6.9%). High genetic diversity in KWPN sport horse population might be a result of an open breeding scheme and interbreeding with other warmblood populations.

**Keywords:** horse, genome-wide association study, insect bite hypersensitivity, *Culicoides*, case-control design, genetic diversity, inbreeding, genetic relationships, homozygosity, signatures of selection

*Author's address:* Merina Shrestha, SLU, Department of Animal Breeding and Genetics; P.O. Box 7023, 750 07 Uppsala, Sweden; *E-mail:* merina.shrestha@slu.se

# Dedication

To my family ...





# Contents

<b>List of Publications</b>	<b>11</b>
<b>Abbreviations</b>	<b>12</b>
<b>1 Introduction</b>	<b>13</b>
1.1 Genetic variation	13
1.2 Genetic variants	15
1.2.1 Single nucleotide polymorphism	15
1.3 Methods to estimate genetic variation	16
1.3.1 Genome-wide association study	16
1.3.2 Allergy	18
1.3.2.1 Hypersensitivity reactions	18
1.3.2.2 Pathogenesis of allergy	19
1.3.2.3 Allergens	19
1.3.2.4 Allergy in humans	19
1.3.2.5 Allergy and Major histocompatibility complex	20
1.3.2.6 Allergy in horses	22
1.3.2.7 Insect bite hypersensitivity	22
<b>2 Aims of the thesis</b>	<b>27</b>
<b>3 Summary of the studies</b>	<b>29</b>
3.1 Materials and Methods (Study I-III)	29
3.1.1 Composition of data	29
3.1.2 Phenotype assessment	30
3.1.3 Genotyping	31
3.1.4 Quality control and Population stratification analysis	31
3.1.5 Genome-wide association study	32
3.1.6 Validation study	33
3.2 Main findings	34
3.2.1 Population stratification	34
3.2.2 Genome-wide association study	36
3.3 Materials and Methods (Study IV)	39
3.3.1 Horse data and quality control	39
3.3.2 Inbreeding coefficient	39
3.3.3 Genetic relationships within and among horse populations	39
3.3.4 Runs of homozygosity	40

3.4	Main findings	40
3.4.1	Estimates of inbreeding coefficient	40
3.4.2	Genetic relationships within and among horse populations	40
3.4.3	Runs of homozygosity	41
<b>4</b>	<b>General discussion</b>	<b>43</b>
4.1	Genotype	43
4.1.1	IBH: complex allergic disorder	43
4.1.2	Increased sample size and number of genetic variants	48
4.1.3	Heritability	49
4.1.4	Copy number variants and Epigenetics	51
4.1.5	Study design	52
4.1.6	Major histocompatibility complex	54
4.2	Phenotype	56
4.2.1	Diagnosis of IBH	56
4.2.1.1	Clinical symptoms	56
4.2.1.2	<i>In-vivo</i> and <i>in-vitro</i> diagnostic methods	57
4.2.1.3	Estimated breeding values	60
4.3	Environment	61
4.4	Genetic diversity and IBH susceptibility	62
4.4.1	Genetic relationships and IBH susceptibility	64
<b>5</b>	<b>Conclusions</b>	<b>65</b>
<b>6</b>	<b>Sammanfattning (in Swedish)</b>	<b>67</b>
<b>7</b>	<b>Samenvatting (in Dutch)</b>	<b>69</b>
	<b>References</b>	<b>71</b>
	<b>Acknowledgements</b>	<b>83</b>
	<b>Training and Education</b>	<b>85</b>

## List of Publications

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

- I Shrestha, M., Eriksson, S., Schurink, A., Andersson, L.S., Sundquist, M., Frey, R., Broström, H., Bergström, T., Ducro, B., Lindgren, G. (2015). Genome-wide association study of insect bite hypersensitivity in Swedish-born Icelandic horses. *Journal of Heredity* 106, 366-74.
- II Shrestha, M., Eriksson, S., Ducro, B.J., Sundquist, M., Thomas, R., Schurink, A., Lindgren, G. (2016). Genetic risk factors for equine insect bite hypersensitivity confirmed in Icelandic horses and Exmoor ponies. (Manuscript).
- III Velie, B.D.\*, Shrestha, M\*, François, L., Schurink, A., Tesfayonas, Y.G., Stinckens, A., Blott, S., Ducro, B.J., Mikko, S., Thomas, R., Swinburne, J.E., Sundqvist, M., Eriksson, S., Buys, N., Lindgren, G. (2016). Using an inbred horse breed in a high density genome-wide scan for genetic risk factors of insect bite hypersensitivity (IBH). *PLoS ONE* 11(4): e0152966. doi:10.1371/journal.pone.0152966.
- IV Shrestha, M., Ducro, B.J., Eriksson, S., Bosse, M., Back, W., Johansson, A.M., Schurink, A. (2016). Understanding the evolutionary history and genetic relationships between the horse populations sampled in the Netherlands. (Manuscript).

\* These authors contributed equally

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

## Abbreviations

AD	Atopic dermatitis
CNVs	Copy number variants
DNA	Deoxyribonucleic acid
DNAm	DNA methylation
EBV	Estimated breeding value
ELA	Equine leukocyte antigen
ELISA	Enzyme-linked immunosorbent assay
GWA	Genome-wide association
HRT	Histamine release test
IBH	Insect bite hypersensitivity
IDT	Intradermal test
Ig	Immunoglobulin
LD	Linkage disequilibrium
mb	Mega bases
MDS	Multidimensional scaling
MHC	Major histocompatibility complex
mRNA	Messenger ribonucleic acid
OR	Odds ratio
QTL	Quantitative trait locus
RAO	Recurrent airway obstruction
RFLPs	Restriction fragment polymorphisms
ROH	Runs of homozygosity
SNP	Single nucleotide polymorphism
TDT	Transmission disequilibrium test

# 1 Introduction

## 1.1 Genetic variation

Mankind has improved phenotypes of animals and plants via artificial selection and breeding. In the past selection was performed taking into account only visible traits for example coat color in horses, patterns in flowers etc. or traits that can be measured such as volume of milk produced by a cow, body weight of chicken, performance in horses etc. This variation in phenotype (P) is a sum of both genetic factors (G) and environmental factors (E). For example, a large production of milk by a cow depends on the cow's own genes and the environment of the cow such as quality of feed provided to the cow. A contribution of both G and E is observed in complex traits like the skin allergic disorder studied in this thesis, insect bite hypersensitivity (IBH) in horses.

The genetic component influencing a phenotype of an individual could be due to a single gene (monogenic) or multiple genes (polygenic). A complex trait is influenced by multiple genes and environment. A complex trait is usually modelled as a product of multiple genes, each with small effects. It could also be that it is a result of multiple genes with a few genes having major effect on the phenotype. An additional source of genetic variation affecting complex traits is caused by interaction of multiple genes (epistasis) present at different loci in the genome. The proportion of phenotypic variance caused by genetic variation is known as heritability (Goddard & Hayes, 2009). In the field of animal breeding and genetics, the narrow sense heritability ( $h^2$ ) is much of interest.  $h^2$  refers to the proportion of phenotypic variation caused by additive genetic variation (Goddard & Hayes, 2009). The additive genetic variation can most effectively be selected on. The additive genetic variance is usually denoted as  $\sigma_A^2$ , and is the variance of the breeding values of the individuals in the population.

Breeding denotes programs, and procedures aimed at modifying phenotypes in populations to improve production or performance and reduce prevalence of diseases. Breeding programs rely on selection of the best individuals to produce the next generation. Traditionally, the breeding values were estimated from records of own and relatives' phenotypes, pedigree information and the heritability concept of phenotypes. Information on traits of interest from progenies, parents and other relatives will increase the accuracy of estimated breeding values (EBVs). Selection of an individual as a breeding animal by estimating EBVs does not require understanding of the biological mechanisms behind the trait of interest. However, this traditional method of estimating breeding values is not suitable for all kinds of traits. For a trait such as IBH in horses, with low heritability and a smaller number of progeny per stallion, the EBVs have low accuracy.

Nowadays with arrival of molecular technologies, genetic variation is utilized by using different forms of genetic evaluations. Through the concept of genomic selection, breeding values can be estimated based on genome-wide genetic variants without prior knowledge of their influence on the traits. Advancements in molecular technologies have also given us opportunity to understand the biological mechanisms behind the traits. By associating genetic variants with a trait of interest, genomic regions and subsequently genes influencing the outcome of the trait can be identified. The knowledge of genes influencing a trait can be utilized to select individuals in order to change the trait of interest.

Genetic variation has also been successfully utilized to understand the evolutionary history of populations by studying the population genetic structure, linkage and recombination rates. In case of domestic animals; selection and breeding has played an important role in shaping their genome. Knowledge regarding genetic diversity of populations and relationships between various populations was traditionally estimated using pedigree information and now by using genetic variants.

Estimation of the extent of genetic variation in the populations has been based on identification of genetic variants in visible traits (color, shape, pattern or other morphological aspects) and measurements of performance traits, allozymes (enzymes and proteins), nucleotide and amino acid sequences, and mutants (Hedrick, 2011).

## 1.2 Genetic variants

Different forms of the genetic variants at nucleotide sequence level have been identified in the genome such as restriction fragment length polymorphisms (RFLPs), microsatellites, single nucleotide polymorphisms (SNPs). The genetic variants such as structural variants (copy number variants: insertion, deletion, duplications) change quantity of a sequence. Other types of genetic variants do not necessarily change the quantity of a sequence, for example: inversions and translocations. RFLPs were the first genetic variants to be developed and were used to map disease in humans (Gusella et al. 1983). Then microsatellites were developed. Microsatellites are short stretches of oligonucleotides repeats of variable lengths. In a population, a microsatellite locus can have many alleles. But microsatellites are present in lower density in the genome than SNPs.

### 1.2.1 Single nucleotide polymorphism

The most common form of genetic variant investigated is SNP. A SNP is a specific locus in the genome where a single nucleotide (base pairs) differs between individuals in a population. SNPs have low polymorphic (predominantly bi-allelic) nature but it is compensated by their abundance in the genome, ease of genotyping or automation (Kruglyak, 1997) and low mutation rates. In principle, a SNP can contain any of the four nucleotides. The mutation rate is estimated at  $10^{-8}$  changes per nucleotide per generation (Li *et al.*, 1996). It corresponds to ~100 new SNPs per individual (Kondrashov, 1995). Hence, the chance of a second independent mutation (introducing a different nucleotide) occurring in the same locus is very low.

Previously, a horse could be genotyped for around 50,000 SNPs using a commercial genotyping array. A horse can now be commercially genotyped for around 670,000 SNPs. Similarly; in other species such as cattle recent commercial genotyping array contains around 777,000 SNPs. In humans, larger number of SNPs is used compared to domestic animals.

The SNPs have been used in different fields such as marker-assisted selection, genomic selection, mapping of quantitative trait loci, and genome-wide association (GWA) studies. We can also utilize SNPs to study evolutionary history by investigating allele frequency, length of homozygosity, patterns and distribution of homozygosity.

### 1.3 Methods to estimate genetic variation

Estimation of genetic variation includes prediction of outcome of the genotype-phenotype relationships (Goddard & Hayes, 2009). The estimation of genetic variation also includes discovering or mapping of genes and pathways for humans and domesticated animals (Goddard & Hayes, 2009). The traditional methods for mapping genes and pathways used linkage analysis, a family-based method to search for the chromosomal location associated with a trait locus by demonstrating co-segregation of the disease, or another trait, with genetic markers of known chromosomal location (Ott *et al.*, 2011). Linkage analysis has been successful for mapping Mendelian traits in humans such as Huntington's disease (Gusella *et al.*, 1983), and cystic fibrosis (Tsui *et al.*, 1985). It has also been successfully used to identify genes regarding susceptibility to the diseases like breast cancer (Walsh & King, 2007) and diabetes (Julier *et al.*, 1991), and also traits in domestic animals such as milk production (Georges *et al.*, 1995). In the past, it was suggested that linkage analysis had a limited power to detect weak effects of genes in complex human diseases (Risch & Merikangas, 1996) and thus linkage disequilibrium (LD) association mapping abbreviated to "association mapping", was used. LD between two polymorphic loci can be quantified by measuring the correlation of allele frequencies at the loci.

#### 1.3.1 Genome-wide association study

Association studies are mainly carried out on a genome-wide basis, referred to as genome-wide association (GWA) studies. GWA studies do not require previous knowledge about the trait under investigation unlike candidate gene studies. GWA studies have the capacity to identify novel loci influencing the trait of interest. In GWA studies genetic variation is assayed in the entire genome by relating genotypes of large number of genetic variants, typically SNPs, to phenotypes without prior hypothesis about trait mechanisms (Bonnelykke *et al.*, 2015; Ball, 2013). The idea behind GWA studies is described in Figure 1. GWA studies are designed to detect associations with causal variants (causal variants in LD with SNPs) that are relatively common in the population. The SNPs in the genotyping arrays are chosen to be common (with minor allele frequency higher than 5%). GWA studies assume common causal variants for disease. Hence, the SNPs linked to the genomic regions with stronger effects on the expression of disease can be identified in GWA studies (Stock *et al.*, 2016).

The genomic regions identified through GWA studies can help to understand the biological mechanisms underlying the trait of interest, and can improve genomic prediction models. If the identified genomic regions are



supported by identification of causal variants, then it can enable marker-based test or direct gene tests.

In humans, GWA studies have contributed in identification of different pathways in autoimmune diseases as reviewed by Visscher *et al.* (2012), identification of multiple susceptibility loci for breast cancer as reviewed by Skol *et al.* (2016) to mention some examples. In livestock and companion animals, GWA studies have identified multiple genomic regions for health and disease related traits (Bermingham *et al.*, 2014; Karlsson *et al.*, 2013), fertility related traits (Nayeri *et al.*, 2016; Schneider *et al.*, 2015; Hoglund *et al.*, 2014), production traits (Nayeri *et al.*, 2016; Wolc *et al.*, 2012) among other examples. GWA studies have also led to discovery of causative mutations for several traits in dogs and cattle as reviewed in Goddard and Hayes (2009). Similarly, in horses, GWA studies have identified genomic regions for various traits and diseases such as body size, diseases such as equine uveitis, osteochondritis dissecans, performance traits (jumping ability) (as reviewed in (Stock *et al.*, 2016; Finno & Bannasch, 2014)). Identification of SNPs followed by sequencing identified causal mutations in the genes associated with gaitedness (Andersson *et al.*, 2012a), diseases such as Lavender foal syndrome (Brooks *et al.*, 2010), and Foal immunodeficiency syndrome (Fox-Clipsham *et al.*, 2011).

In this thesis, GWA studies were performed to increase our understanding about a skin allergic disorder in horses, equine insect bite hypersensitivity (IBH). Before moving towards IBH, the general idea regarding pathogenesis of allergy, allergens, allergy in humans and horses will be presented below. This will be followed by a short introduction to IBH in horses.

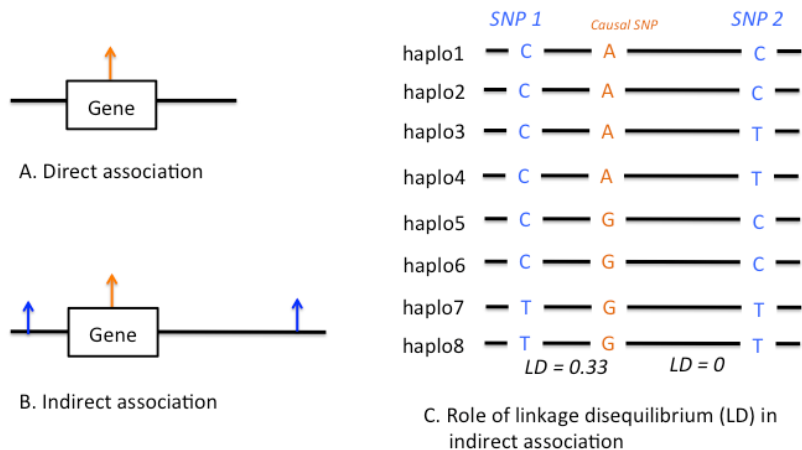


Figure 1. Orange and blue arrows represent SNPs. The orange SNP influences our trait of interest. A) Orange SNP is genotyped. Association analysis will identify orange SNP leading to direct association. B) Only blue SNPs are genotyped. Association analysis will not identify causal SNP but will identify blue SNP depending on LD between each blue SNPs and orange SNP. C) SNP1 and SNP2 are genotyped for eight observations. Only SNP1 will be associated with trait of interest as it is in LD (0.03) with causal SNP. The concept for this figure has been derived from Hirschhorn and Daly (2005) and Jungerius (2004).

### 1.3.2 Allergy

#### 1.3.2.1 Hypersensitivity reactions

Exposure to harmless antigens (allergens) can lead to abnormal immune responses, both innate and adaptive, in susceptible individuals. These allergens manifest signs and symptoms of over reactivity or ‘hypersensitivity reactions’. The hypersensitivity reactions are categorized based on different immune mediators and duration taken to observe symptoms due to reactions. Allergic disorders are IgE mediated hypersensitivity reactions. A type I hypersensitivity reaction is mediated by Immunoglobulin E (IgE) (Galli et al. 2008). An IgE mediated hypersensitive reaction can occur within minutes of allergen exposure (type I). The IgE mediated hypersensitive reactions can also occur after 2 to 4 hours and peaks 6 to 9 hours after allergen exposure. Such IgE

mediated hypersensitivity reactions are known as type IV hypersensitivity reactions (Klug *et al.*, 2009). Allergic reactions in humans include allergic asthma, allergic rhinitis (hay fever), eczema and food allergies.

#### *1.3.2.2 Pathogenesis of allergy*

When an individual is exposed to an allergen, depending on the susceptibility of an individual, either a healthy or an allergic immune response towards the allergen occurs (Traidl-Hoffmann *et al.*, 2009). The susceptibility of an individual depends in part on the genetics of an individual.

After exposure, allergen epitopes are presented to allergen-specific helper T cells ( $T_H0$ ) by Major histocompatibility (MHC) class II molecules on the surface of antigen presenting cells (APC). In IBH, immature dendritic cells present allergens to  $T_H0$  cells (Meulenbroeks, 2015b). In susceptible individuals,  $T_H0$  cells develop into effector T cells known as type 2 helper T cells ( $T_H2$ ). A  $T_H2$  immune response leads to the development of B cells that will produce IgE (IgE switched B cells). The IgE switched B cells transform to IgE producing plasma cells and produce allergen-specific IgE. The allergen-specific IgE binds to effector cells like mast cells and basophils. This binding after exposure to the allergen, results in cellular activation and release of inflammatory mediators (histamines, leukotrienes and cytokines) responsible for the symptoms of allergic disorders.

#### *1.3.2.3 Allergens*

Allergens like food, plant pollen, drugs, insect products, mold, animal hair, latex and others can give rise to different allergic reactions. The route of exposure, dose and function of the allergens is important for allergenicity (Traidl-Hoffman *et al.* 2009). Allergens that elicit type I hypersensitivity reactions are usually described as harmless environmental substances. Most of the allergens are proteins and glycoproteins (Traidl-Hoffmann *et al.*, 2009). However, some of the allergens do have biological activities influencing their allergenicity for example a major allergen of house dust mite, Der p I. Der p I causes allergic reactions like eczema, asthma and rhinitis. The allergen regulates IgE production (Hewitt *et al.*, 1995) and has also been shown to bias the immune response in favour of  $T_H2$  with the proteolytic activity (Ghaemmaghami *et al.*, 2002).

#### *1.3.2.4 Allergy in humans*

Allergic disorders are a common problem in humans and animals. The symptoms caused by allergies can range from mild discomfort to fatal anaphylactic shock. Roughly 25% of people in the developed world suffer

from allergic disorders (Galli *et al.*, 2008). Moreover, the prevalence of allergic disorders is increasing in the industrialized countries.

Allergic disorders are complex as many genes and environmental factors influence development of allergic disorders. In humans, high genetic predisposition regarding various allergic disorders has been suggested. Estimated heritability of asthma and hay fever in 3,808 pairs of Australian twins population ranged from 60 to 70% (Duffy *et al.*, 1990). The heritability estimates of childhood asthma, hay fever, eczema and food allergies in 25,306 Swedish twins ranged from 60 to 82% (Ullemar *et al.*, 2016).

The polygenic nature of allergic disorders has been confirmed in several studies. For example three candidate genes: *IL1R1*, *IL18R1* and *IL18RAP* were significantly associated with susceptibility of allergic asthma in a family-based association study in humans (Reijmerink *et al.*, 2008). Similarly, in a GWA study of atopic dermatitis (AD) in children, a locus on *EDC* gene (chr 1), five loci spanning between *RAD50* locus to *IL13* locus (chr 5), seven loci on MHC region (chr 6), and two loci close to *LRRC32* gene (chr 11) showed significant association (Weidinger *et al.*, 2013). Odds ratios of AD for the associated loci ranged between 0.64 and 1.68.

Atopic dermatitis in humans and IBH in horses have similar features (Schaffartzik *et al.*, 2012). Polygenic influence in development of IBH in horses has been observed in our study along with other studies. Along with polygenic nature, development of allergic disorders is influenced by several environmental factors. In humans, a reduced risk of allergic disorders was observed with exposure to environmental factors such as farm animal and products (Douwes *et al.*, 2008), and to the domestic dogs (Thorsteinsdottir *et al.*, 2016). Also a higher intake of a diet rich in antioxidants (Devereux & Seaton, 2005) and higher number of siblings (Kinra *et al.*, 2006) reduced the risk of allergic disorders. Gender also seemed to influence the risk of allergic disorder (Uekert *et al.*, 2006) where boys showed high susceptibility to allergy.

#### *1.3.2.5 Allergy and Major histocompatibility complex*

Major histocompatibility complex (MHC) is a general term used for the genomic region encoding MHC molecules. In humans, the MHC region is denoted as Human leukocyte antigen (HLA) and located on the short arm of chromosome (chr) 6 while in the horse, MHC region is located on chr 20. The MHC region contains three classes of genes: class I, class II and class III genes. The arrangement of three regions in the horse genome can be viewed in Figure 2. The MHC region contains genes encoding MHC molecules and other genes related with immunity. MHC class I genes encodes glycoproteins expressed on the surface of nearly all nucleated cells (Owen *et al.*, 2009).

MHC class I molecules bind peptides and present them to cytotoxic (CD8+) T cells to initiate immune response. MHC class II molecules encodes glycoproteins expressed predominantly on antigen presenting cells (APCs) such as macrophages, dendritic cells and B cells (Owen *et al.*, 2009). The MHC class II glycoproteins present antigens to the helper (CD4+) T cells and initiate immune response. The MHC class II region spans more than one mega bases (mb) and harbours two *Eqca-DQA* loci (*DQA1* and *DQA2*), two *Eqca-DQB* loci (*DQB1* and *DQB2*), three *Eqca-DRB* loci (*DRB1*, *DRB2* and *DRB3*), and one *Eqca-DRA* locus which is polymorphic in horses (Gustafson *et al.*, 2003). The map and annotation of MHC class II region is still on progress. For example, three loci has been reported for *Eqca-DQB* gene in MHC class II region (Immuno Polymorphism Database).

In humans, HLA region has been associated with allergic disorders as reviewed in Tamari *et al.* (2013). In horses, the MHC region has been associated with allergic disorders such as recurrent airway obstruction (RAO) susceptibility (Curik *et al.*, 2003) and IBH susceptibility. IBH susceptibility has been associated to MHC class II region in different serological and candidate gene studies (Klumplerova *et al.*, 2013; Andersson *et al.*, 2012b; Marti *et al.*, 1992). A same microsatellite marker was significantly associated with IBH susceptibility in Swedish-born Icelandic horses and Exmoor ponies (Andersson *et al.*, 2012b). Also, *Eqca-DRB3* gene in MHC class II region was associated with IBH susceptibility in Swedish-born Icelandic horses (Andersson *et al.*, 2012b). Similarly, *Eqca-DRA* gene in MHC class II region showed significant association with IBH in Icelandic horses in Switzerland that were imported from Iceland (Klumplerova *et al.*, 2013).

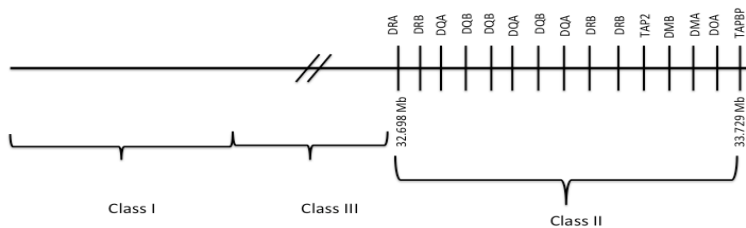


Figure 2. Position of MHC class I, II and III regions in horse genome. MHC class II region spans around 1mega bases (tentative position: 32.698 Mb – 33.729 Mb). This figure is an illustration from the study by Klumplerova *et al.* (2013).

#### 1.3.2.6 Allergy in horses

In horses, allergic disorders such as insect bite hypersensitivity (IBH), recurrent airway obstruction (RAO) or heaves, allergic rhino-conjunctivitis and recurrent urticaria or hives are IgE mediated hypersensitivity reactions (Fadok, 2013). RAO and IBH are commonly occurring allergic disorders in horses. These are the most studied allergic disorders in horses.

RAO, or heaves, is a hypersensitivity reaction to inhaled stable dust, including mold spores (Swinburne *et al.*, 2009). It is one of the most common disorders of the lower respiratory tract of mature horses and is influenced by multiple genes. Two loci on chr 13 and chr 15 have been associated with RAO susceptibility in Warmblood horses (Swinburne *et al.*, 2009). Another gene *TSLP* was associated with development of RAO in horses (Klukowska-Rotzler *et al.*, 2012). mRNA expressions of *TSLP* cytokine in clinical RAO cases were significantly higher ( $p < 0.05$ ) than clinical RAO controls. The mRNA expression of *TSLP* cytokine also increased significantly after 30-day exposure to moldy hay in experimental RAO cases compared to no exposure to moldy hay. While significant increase in mRNA expression of *TSLP* was not observed for experimental RAO controls. *TSLP* cytokine produced by damaged epithelial cells induce  $T_H2$  immune response leading to hypersensitive reactions (Fadok, 2013). *TSLP* gene has also been significantly associated with IBH susceptibility in Icelandic horses (Klumplerova *et al.*, 2013). Hence, some common genetic components might be shared between IBH and RAO manifesting allergic immune response.

#### 1.3.2.7 Insect Bite Hypersensitivity

##### *Background:*

IBH is a skin allergic reaction in horses. Such skin allergic disorders have been observed in many different horse breeds and in other species for instance sheep, cattle, donkey (Correa *et al.*, 2007; Yeruham *et al.*, 1993), dogs (Pucheu-Haston, 2016; Chamberlain, 1978) and hippopotamus (Spriggs & Reeder, 2012). IBH is known by different names in different parts of the world: “Queensland itch” in Australia, “Kasen disease” in Japan, “Sweet itch” in Great Britain, “Sommareksem” in Nordic countries, and the similar “Sommerekzem” in Germany (Schaffartzik *et al.*, 2012; Brostrom *et al.*, 1987).

IBH is caused by bites of female midges of *Culicoides* species, and potentially also female black flies of *Simulium* species (Hellberg *et al.*, 2009; Schaffartzik *et al.*, 2009). It is a reaction towards allergens present in the saliva of these biting midges. IBH has been described as a type I hypersensitivity reaction with release of histamine and other inflammatory mediators from basophils and mast cells (Jonsdottir *et al.*, 2015). However the involvement of

a type IV hypersensitivity reaction has also been suggested (Jonsdottir *et al.*, 2015; Schaffartzik *et al.*, 2012).

After exposure to the allergens some horses show symptoms of itchiness and scratch to get relief from itch. That leads to clinical symptoms such as scratches on the skin surface, hair loss, lesions, and possibly secondary infections. The clinical symptoms appear mainly around mane and tail, but also in other areas of body (Peeters *et al.*, 2014; Eriksson *et al.*, 2008). Most horses typically show clinical symptoms around spring and summer (Schurink *et al.*, 2013a) when *Culicoides* species are active. The symptoms disappear during autumn. Such occurrence of the clinical symptoms in a particular season is one of the characteristics that facilitate diagnosis of IBH. In severe cases, clinical symptoms persist. The horses develop clinical symptoms at various age, however average age of onset is observed to be 2 – 4 years (Schurink, 2012).

The worldwide prevalence of IBH in different horse populations ranges from 3% in Great Britain to 60% in Australia as reviewed in Schaffartzik *et al.* (2012). *Culicoides* species have not been observed in Iceland. Icelandic horses exported from Iceland to the European continent had IBH prevalence of more than 50% as reviewed in Schaffartzik *et al.* (2012). The estimated prevalence of IBH in Swedish-born Icelandic horses ranged from 6.7% (441 horses used, (Brostrom *et al.*, 1987)) to 8% (1,250 horses used, (Eriksson *et al.*, 2008)). Knowledge regarding IBH prevalence in Exmoor pony population is limited. However, IBH has been observed as a significant problem in this pony population (Andersson *et al.*, 2012b; ExmoorPonySociety).

Some horses are genetically predisposed to get IBH susceptibility. Multiple genes seem to influence the genetic predisposition. Heritability of IBH on the continuous scale was estimated in Dutch Friesian breeding mares (0.16, SE = 0.06) (Schurink *et al.*, 2011), Dutch Shetland breeding mares (0.24, SE = 0.06) (Schurink *et al.*, 2009), Swedish-born Icelandic horses (0.30, SE < 0.20) (Eriksson *et al.*, 2008), and Belgian warmblood horses (0.65 for clinical status) (Peeters *et al.*, 2015).

Environmental factors play essential role in development of IBH. It is important to investigate an effect of the environmental factors while estimating true genetic component influencing IBH susceptibility in horses. The environmental factors related to exposure to allergens like region of habitat of horses (exposure to allergens) (Eriksson *et al.*, 2008; Brostrom *et al.*, 1987), time period of collection of data (Schurink *et al.*, 2013b), protective rugs (Olsen *et al.*, 2011) have been observed to influence IBH prevalence. The environmental factors related to mares such as age, coat colour, wither height, body condition (amount of fat stored in the body) (Peeters *et al.*, 2014; Schurink *et al.*, 2013b) had significant effect on IBH prevalence. The effect of

sex in development of IBH susceptibility is inconclusive (Peeters *et al.*, 2011; Eriksson *et al.*, 2008; Brostrom *et al.*, 1987).

#### *Diagnosis:*

Diagnosis of IBH is performed by observing the clinical symptoms and examining the medical history of horses. The diagnosis is also performed using different *in-vivo* and *in-vitro* diagnostic tests such as intradermal tests (IDT) (Langner *et al.*, 2008; Fadok & Greiner, 1990), allergen-specific IgE and IgG enzyme-linked immunosorbent assay (ELISA) (Meide *et al.*, 2014; van der Meide *et al.*, 2014; Langner *et al.*, 2008), basophil degranulation test/histamine release test (HRT) (Langner *et al.*, 2008) and the cellular antigen stimulation tests/ sulfidoleukotriene test (Klumplerova *et al.*, 2013; Marti *et al.*, 1999). The diagnostic tests helps to identify sensitized horses (horses that show IgE reactivity but without clinical symptoms) from unsensitized horses. Observing presence of the clinical symptoms along with medical history of the horses is considered as a gold standard method for diagnosing IBH. However, it is suggested to supplement this method by other *in-vitro* or *in-vivo* diagnostic methods.

#### *Prevention and Treatment:*

IBH causes extreme discomfort; decrease quality of life for horses, and results in poor horse welfare. Horses are used for various purposes like sports, recreation, agriculture etc. Horses suffering from IBH are unfit for riding, and have a reduced value (horse's commercial value on average ranges from €1700 to €10,000) (Horsegene) in the market creating economic loss for the horse owner. In some cases, horses suffering from IBH are euthanized.

The best method to reduce IBH prevalence is to avoid an exposure to the biting midges totally. Total avoidance of exposure might not be possible, however exposure can be decreased by use of blankets on horses, overnight stabling (avoiding the time when the biting midges are active for example sunrise and sunset) (Olsen *et al.*, 2011). Other methods include relocating horses to low-risk areas and use of insect repellents as reviewed in Schaffartzik *et al.* (2012).

#### *Research regarding genetics of IBH:*

With various researches we are able to understand that IBH susceptibility is indeed a complex disorder where both environmental and genetic factors influence its development. Environmental factors have been identified in various horse populations (Peeters *et al.*, 2014; Schurink *et al.*, 2013b; Eriksson *et al.*, 2008). The genetic component influencing the susceptibility



has also been estimated. Further, immunity related genes located at MHC region and other regions have also been associated with IBH susceptibility in various candidate gene studies (Klumplerova *et al.*, 2013; Vychodilova *et al.*, 2013; Andersson *et al.*, 2012b). In candidate gene studies, the genes to be studies are included based on the previous knowledge of association of IBH like trait in other species for example humans. GWA studies (Schurink *et al.*, 2013a; Schurink *et al.*, 2012) have been performed in different breeds to identify and quantify the genomic regions associated with IBH susceptibility. However, GWA studies are limited.



## 2 Aims of the thesis

The overall aims of this thesis were to utilize genetic variation in horse populations to understand the genetics behind IBH in order to be able to reduce its prevalence, and to understand the evolutionary history, diversity, and genetic differences between various horse populations.

Studies I-III investigated if any genome-wide genetic variants in an Icelandic horse population and an Exmoor pony population were associated with IBH. Identifying genetic variants can subsequently lead to identifying causal genes that influence IBH susceptibility in the studied horse populations. If causal genes can be identified and biological pathways behind IBH development are better understood, this could be used to develop therapeutic measures to reduce the prevalence of IBH. The identified genetic variants can also be used to formulate genetic tests to predict the risk of IBH susceptibility of an individual horse. The horses with a higher risk of IBH susceptibility could be removed from breeding to reduce the prevalence of IBH.

Study IV investigated inbreeding level, distribution and patterns of runs of homozygosity segments, and differences between various horse populations utilizing the genome-wide genetic variants. This information in the genome was investigated if they reflect assumptions based on our understanding of evolution of these breeds throughout the history.

Mentioned below are the specific objectives of the thesis:

### Study I - III

- To determine the genetic variants associated with IBH.
- To determine the genetic variants associated with IBH utilizing a combined dataset and validating associated genetic variants.
- To determine the genetic variants associated with IBH utilizing increased SNPs density in an inbred horse population.

#### Study IV

- To investigate level of inbreeding and evolutionary history of various horse populations.
- To quantify differences in the genome between these populations.

## 3 Summary of the studies

This thesis comprises three studies (I-III) related to insect bite hypersensitivity (IBH), and one study related to genetic diversity and genetic differences between various horse populations (Study IV). The Studies I-III were performed to identify genetic variants (genomic regions) associated with IBH. Study IV was performed to understand qualitative and quantitative genetic relationships along with evolutionary history between various horse populations.

The summary of the first three Studies I – III will be presented first which will be followed by summary of Study IV.

### 3.1 Materials and methods (Study I-III)

#### 3.1.1 Composition of data

Information on Studies I – III regarding horse breed, number of horses in different categories (affected, unaffected, female, male, type of genotyping array used etc.) are presented in Table 1.

Study I consisted of 209 Icelandic horses. The majority (97%) of Icelandic horses were born in Sweden and the remaining horses were imported from Iceland, the Netherlands and Norway. These horses were paternal half-sibs.

Study II consisted of 355 Icelandic horses and 280 Exmoor ponies. The 355 Icelandic horses consisted of the 209 Icelandic horses included in study I, and 146 Icelandic horses sampled in the Netherlands. The data consisted of Icelandic horses born in Sweden (57.2%), in the Netherlands (34.6%), in Iceland (5.6%) or in other European countries (2.5%). 76.9% of the Icelandic horses were in paternal half-sib groups. The remaining Icelandic horses were maternal half-sibs and full-sibs. For validation purpose in the Study II, the 280 Exmoor ponies that were also included in Study III were used.

Study III consisted of 280 Exmoor ponies genotyped using latest genotyping array. Out of 336 available Exmoor ponies, 26 ponies were removed due to unclear status of IBH, incomplete pedigree information (less than 4 generations) or not recorded as a purebred. Average relatedness was estimated for the remaining 310 ponies using the genetic software Contribution, Inbreeding, Coancestry (CFC) (Sargolzaei *et al.*, 2006). We aimed for including ponies with low average relatedness to the group. Furthermore, 280 Exmoor ponies were selected from 310 ponies based on several steps of quality control to ensure gender balance (mostly females were removed) and good quality of DNA (blood samples from young ponies were given priority).

### 3.1.2 Phenotype assessment

In Study I, data collection began already in 2005. The data was collected by a questionnaire, that we distributed through websites, horse magazines, and the Icelandic horse association. Information about the horse (name, registration number, birth year, sex, country of origin, current location), IBH (status, season and age of onset of IBH, severity of IBH, affected areas on body), other allergies, and IBH status of the dam were collected through the questionnaire. The horses were classified as affected or unaffected based on the clinical symptoms of IBH observed and reported by the horse owners. Veterinarians later confirmed the symptoms while collecting blood samples. The distinct seasonal character of IBH facilitates diagnosis. Horses that showed seasonal symptoms of IBH for at least two grazing seasons were considered affected horses. The unaffected horses had never developed clinical symptoms of IBH and were at least 4 years old. The age of all horses ranged from 4 till 27.

In study II, besides the Icelandic horses sampled in Sweden as described earlier, additional Icelandic horses were sampled in 2010 in the Netherlands through publications on various horse related websites. The unaffected and affected horses were matched on different characteristics such as gender, age, sire and coat colour. A veterinarian and a researcher visited the horse owners. They conducted an IBH related questionnaire to the horse owners and scored IBH phenotypes. The affected horses showed clinical symptoms of IBH, and seasonality was taken into account. To increase reliability of diagnosis of IBH, the following criteria were considered: the unaffected horses must have had an exposure to *Culicoides spp.* and should have never developed clinical symptoms of IBH, the unaffected horses should be at least 4 years of age, the unaffected horses must have been exposed to *Culicoides spp.* for at least a year, and had never showed clinical symptoms of IBH. The age of horses ranged from 4 till 35 years.

In study III, data was collected from 2008 till 2011 through an open call to owners via the Exmoor pony society and online postings. The questionnaire contained information about the pony (name, gender, registration number, birth year, location, history of any disease), the pony's owner (name, location), environmental conditions (stabling, grazing, humidity, shelter, preventive measures: blankets), clinical symptoms of IBH (body location of crust, swelling, wounds, hair loss, skin thickening; age of onset of symptoms), and IBH status of sire and dam. The affected ponies showed clinical symptoms of IBH and were categorized by their owners as mildly, moderately or severely affected by IBH based on clinical symptoms and preventive measures implemented. The unaffected ponies did not show any clinical symptoms of IBH. Not all answers to the questions regarding environmental conditions were complete. For 24 Exmoor ponies age information was not available and around 30 unaffected Exmoor ponies out of 280 Exmoor ponies were 1 – 3 years old. Similar age distribution was observed in affected (mean = 10.2, median = 8) and unaffected group (mean = 10.9, median = 9).

### 3.1.3 Genotyping

Genome-wide evenly spaced SNPs are being increasingly utilized in association studies to identify SNPs close in LD with causal variants associated with traits of interest. To identify genetic variants, in our studies I-III we collected SNPs information using the most up-to-date genotyping array available at that time. In Study I, the first genotyping array developed for the horse, the Illumina Equine50KSNP BeadChip, was used to genotype 54,602 SNPs in 209 Swedish-born Icelandic horses. In Study II, the Illumina Equine70KSNP BeadChip was used to genotype 65,157 SNPs in 146 Icelandic horses sampled in the Netherlands. In Study III, the 670K Affymetrix Axiom array, was used to genotype 670,796 SNPs in 280 Exmoor ponies. With the development of new genotyping arrays in time, we could increase the number of SNPs being genotyped in populations. A large number of evenly spaced SNPs better represents the horse genome. With increase in density of SNPs in the genome, there is increase in chance of high LD among SNPs. If there is high LD among SNPs, chances of these SNPs being in LD with causal variants also increases.

### 3.1.4 Quality control and Population stratification analysis

Quality control and population stratification analyses were performed using GenABEL package (Aulchenko *et al.*, 2007) in the statistical software R. The quality control thresholds for the studies I-III are presented in Table 1. A stratified population can lead to false positive associations especially in case-

control GWA studies (Balding 2006). Hence, the potential presence of subpopulations was investigated using post-quality control dataset (Table 1) in Studies I – III. Multidimensional scaling (MDS) plots were constructed to visualize the distribution of affected and unaffected horses. In Study I and II, the result from a preliminary mixed model approach with single SNPs using the GenABEL package in R was compared with the expected distribution of P-values to obtain the genomic inflation factor ( $\lambda$ ), which was used to detect if there was stratification. In the study III, the genomic inflation factors were calculated using different models.

### 3.1.5 Genome-wide association study

In studies I and II, a case-control association test was performed using the Bayesian variable selection method Bayes C implemented in version 4.0 of GenSel software available online (<http://biggs.ansci.iastate.edu/>). In this method, all SNPs were fitted simultaneously as random effects in the model with only few SNPs assigned to having an effect and the remaining SNPs assigned to having no effect. No covariates were used in the model. The prior for marker variance was derived from the additive genetic variance based on the heritability of 0.27 for IBH as a binary trait (Eriksson *et al.*, 2008). Two hundred thousand Markov Chain Monte Carlo cycles were performed. On each cycle, 99.9% of all SNPs were assumed to have no effect in the model. The first 50% of the cycles were discarded as a burn-in period and estimates from every 100<sup>th</sup> cycle were saved.

The horse genome build (EquCab 2.0) was divided into 2,371 non-overlapping 1-mb windows in the study I, and 2,373 non-overlapping 1-mb windows in the study II. The windows explaining more than the expected percentage (for example in the study I:  $100\%/2,371 = 0.042\%$ ) of genetic variance were considered to be non-randomly associated with IBH. Within these 1-mb windows, the SNP with highest model frequency was considered to be associated with IBH. Model frequency of the SNP refers to the proportion of the post burn-in cycles that included the SNP in the model.

In the study III, the association tests were performed using the GenABEL package (Aulchenko *et al.*, 2007) in the statistical software R. Case-control association tests were performed using structured association approach and a principal component analysis (PCA) method. Association tests were also performed using the severity of IBH. A mixed-model structured association approach and PCA method was applied. A single SNP was fitted in all the models without any covariates.



### 3.1.6 Validation study

In study II, the SNPs with highest model frequency were selected for validation. The selected SNPs, located on chromosome 1, 3 and X, were used for validation in 280 Exmoor ponies. The genotype information for the selected SNPs was obtained via Taqman SNP Genotyping Assay. The association between IBH in Exmoor ponies and the selected SNPs were investigated using a chi square test with 1 degree of freedom in the online statistical software ([www.vassarstats.net](http://www.vassarstats.net)).

Table 1. Horse breeds, genotyping platforms, number of horses in total and in different categories, number of SNPs pre and post quality control, and quality control thresholds utilized in the studies I, II and III

	<b>Study I</b>	<b>Study II</b>	<b>Study III</b>
Horse breed	Icelandic horses	Icelandic horses Exmoor ponies	Exmoor ponies
Genotyping platform	Illumina Equine SNP50 Genotyping BeadChip	Icelandic horses: Illumina SNP50 and Illumina SNP70 Genotyping BeadChip	670K Axiom Equine Genotyping Array
Number of SNPs	54,602	45,986	670,796
<b>Number of horses</b>	209	355	280
Affected	104	177	110
Unaffected	105	178	170
<b>Number of sires and dams</b>	42 sires and 207 dams	136 sires and 333 dams	107 sires and 226 dams
<b>Gender</b>			
Female (Affected; Unaffected)	104 (51; 53)	198 (102; 96)	186 (68; 118)
Male (Affected; Unaffected)	105 (53; 52)	157 (75; 82)	94 (42; 52)
<b>Quality control thresholds</b>			
Genotyping frequency per horse for inclusion	$\geq 95\%$	$\geq 95\%$	$> 90\%$
Genotyping frequency per SNP for inclusion	$< 95\%$	$< 95\%$	$< 90\%$
Minor allele frequency per SNP for exclusion	$< 5\%$	$< 5\%$	$< 0.05\%$
<b>Post quality control</b>			
Number of horses	209	355	268
Number of SNPs	38,486	33,523	452,457

## 3.2 Main findings

### 3.2.1 Population stratification:

In Study I two separate populations were observed in MDS plot (Figure 3). The value of the genomic inflation factor was 0.81 indicating no stratification for IBH status however. In Study II, two separate populations were observed in MDS plot (Figure 4). The value of the genomic inflation factor was 0.92 indicating no stratification. In Study III, separate populations were not observed in MDS plot (Figure 5). The genomic inflation factor values differed for four different models applied in the study. The values ranged from 1.02 (mixed-model structured association approach) to 1.36 (structured association approach).

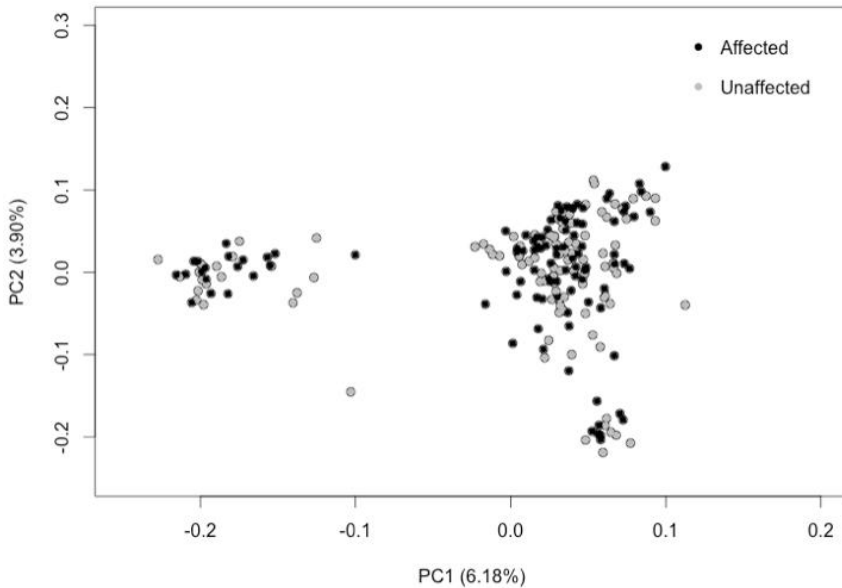


Figure 3. Multidimensional scaling plot for 209 Icelandic horses in Study I. Principal component 1 (PC1) explained 6.18% of the variance and Principal component 2 (PC2) explained 3.90%. Each dot represents an individual.

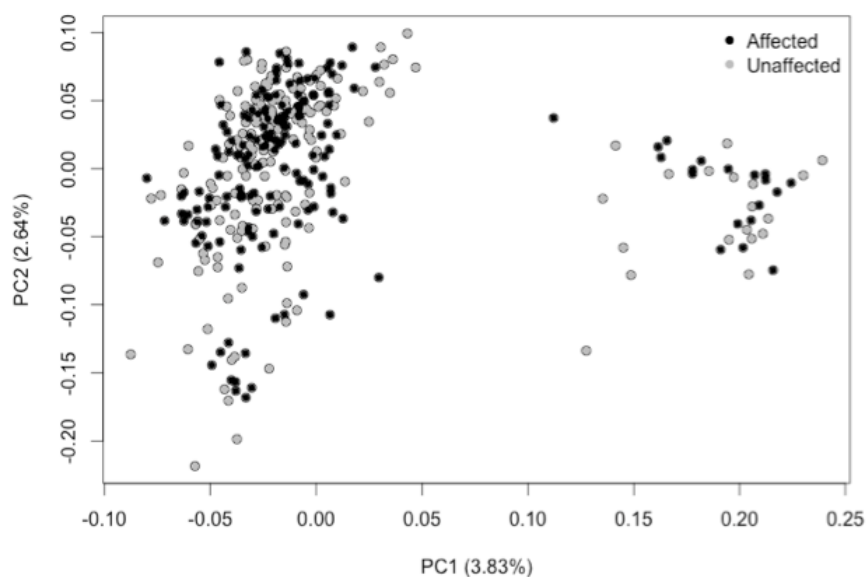


Figure 4. Multidimensional scaling plot for 355 Icelandic horses in Study II. Principal component 1 (PC1) explained 3.83% of the variance and Principal component 2 (PC2) explained 2.64%. Each dot represents an individual.

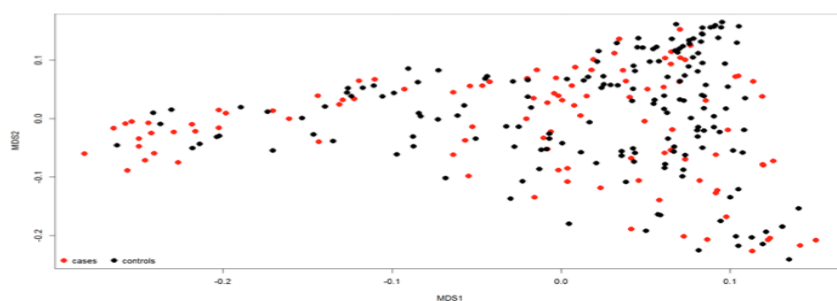


Figure 5. Multidimensional scaling plot for 268 Exmoor ponies in Study III. Each dot represents an individual.

### 3.2.2 Genome-wide association study:

Top 10 1-mb windows explaining highest percentage of genetic variance of IBH, in Study I and Study II are presented in Table 2. In Study I, 29 1-mb windows explained  $\geq 0.14\%$  of the genetic variance of IBH (data not shown), and these windows were investigated for immunity-related genes, allergy-related genes, and genes directly or indirectly involved in skin function. Immunity-related genes were observed in the genomic regions on chromosomes (chr) 7, 10, and 17.

A 1-mb window at 57 mb on chr 7 contained a gene cathepsin C (*CTSC*). *CTSC* was located 600 kb downstream of the associated SNP BIEC2-1002972 within this window. *CTSC* been shown to play a role in MHC class II antigen presentation in humans (Rebhan *et al.*, 1998).

A 1-mb window located at 35 mb on chr 10 contained the gene inhibitor of bruton agammaglobulinemia tyrosine kinase (*IBTK*) 159 kb downstream of the SNP BIEC2-115349 associated with IBH. *IBTK* regulates the activity of Bruton's tyrosine kinase (*BTK*), which is required for B-cell development in humans and mice (Spatuzza *et al.*, 2008). Progenitor B cells transform into mature B cells and lead to the production of different immunoglobulins: IgG, IgM, IgA, IgD, and IgE. B cells play a role in the adaptive immune system, and defects in B-cell development, selection, and function can lead to the development of immunodeficiency, allergies, and autoimmunity (Pieper *et al.*, 2013).

A 1-mb window located at 77 mb on chr 17 contained a gene, insulin receptor substrate 2 (*IRS2*) located 105 kb upstream of the associated SNP BIEC2-385377. *IRS2* is associated with the total level of IgE in asthmatic patients and plays a role in the modulation of antibody levels (Acevedo *et al.*, 2009). In humans, *IRS2* interacts with the immune response interleukin 9 (*IL-9*) signaling pathway and is one of the related super-pathways (Rebhan *et al.*, 1998). In mice *IL-9* increases production of IgE (Petit-Frere *et al.*, 1993) and showed that IgE mediates type I hypersensitivity reactions.

In Study II, for Icelandic horses, top ten 1-mb non-randomly associated windows (Table 2) were investigated for immunity-related genes, allergy-related genes, and genes directly or indirectly involved in skin function. Immunity-related genes were observed close to the associated windows on chr 1 and chr 15.

On chr 1, RAS guanyl releasing protein 1 (calcium and DAG20 regulated) (*RASGRP1*) was located upstream [149,706,774-149,775,059 bp] of the associated 1-mb window 152 mb. *RASGRP1* plays an essential role in IgE mediated allergic response in mice. In humans, *RASGRP1* regulates T and B cell development and is associated with the autoimmune disease systemic

lupus erythematosus (GeneCards). In horses, *RASGRP1* gene has 64 orthologues (Ensembl (Equ Cab 2)) and is 92.7% similar to human *RASGRP1* implying similar role of *RASGRP1* in horses and humans. Any disruption in the regulation of T and B cell development may influence the production of IgE and likely in the development of IBH, as IBH is an IgE mediated hypersensitivity reaction (Klumperova *et al.*, 2013).

On chr 15, a gene Interleukin 1 receptor antagonist (*IL1RN*) located on chr 15 [16,048,492-16,054,21210 bp], ~4 mb upstream of the window at 19 mb in the study I. *IL1RN* was one of the candidate genes investigated for an association with IBH susceptibility in Old Kladruby horses (Vychodilova *et al.*, 2013). One can speculate that associated region could be a regulatory region for *IL1RN*.

In the Exmoor ponies, two SNPs on chr 1 were associated with IBH. SNP BIEC2\_65455 (OR = 0.65, 95% confidence interval: 0.46 – 0.92) was successfully genotyped in 277 ponies. The risk allele frequency for this SNP was 0.50 in the unaffected group and 0.60 in the affected group. SNP BIEC2\_65299 (OR = 1.52, 95% confidence interval: 1.08 – 2.15) was successfully genotyped in 278 ponies. The risk allele frequency was 0.49 in the unaffected group and 0.60 in the affected group.

In Study III, no single SNP demonstrated genome-wide significance in any of the four GWA analyses performed. However, two regions on chr 8 were studied closer. First region chr.8: 70,269,986– 71,065,803, contained a SNP AX-104130346 that got the lowest p-value ( $P_{\text{unadjusted}}$ ) in 3 out of the 4 analyses and bordered on genome-wide significance ( $P_{\text{genome-wide}}$ ) in two analyses. The risk allele frequency was 0.46 in the unaffected group and 0.54 in affected group. This SNP was located in DCC netrin 1 receptor gene (*DCC*). In humans, *DCC* has been associated with apoptosis and functions as a tumor suppressor (GeneCards). Second region chr 8: 78,377,554– 78,880,555, consisted of 4 SNPs. Within this region a SNP AX-104330407 was focused on because this SNP resulted in the lowest p-value in one of the GWA analyses and was in the top 10 results of the other three GWA analyses. Immunity related gene *TNFRSF11A* was observed closest to the SNP AX-104330407. *TNFRSF11A* plays a role in autoinflammatory disorder in humans (Jeru *et al.*, 2014). *TNFRSF11A* might also play a role in hypersensitive reactions because autoinflammatory disorder and allergic disorders are caused by dysfunction of the immune system.

Table 2. Top 10, 1-mb non-overlapping windows in Study I and Study II ranked according to explained highest percentage of genetic variance of IBH, along with risk allele frequency of SNPs associated with IBH within a window

	1-mb windows explaining highest genetic variance (%)					Associated SNP within 1-mb window				
	Chr <sup>1</sup>	mb <sup>2</sup>	No. SNP in window	Gen. var (%) <sup>3</sup>	p>AVG <sup>4</sup>	SNP name	MF <sup>5</sup>	Position of SNP (bp)	Risk allele frequency	
									Unaff <sup>6</sup>	Aff <sup>7</sup>
Study I	32	73 <sup>a</sup>	20	0.51	0.04	BIEC2-1132313	0.009	73,129,947	0.34	0.66
	32	74 <sup>b</sup>	16	0.36	0.04	BIEC2-1133113	0.004	74,332,689	0.45	0.55
	17	77 <sup>c</sup>	17	0.34	0.03	BIEC2-385377	0.016	77,343,575	0.43	0.57
	18	26 <sup>c</sup>	15	0.34	0.03	BIEC2-409581	0.011	26,766,516	0.43	0.57
	10	41	17	0.33	0.04	BIEC2-119028	0.009	41,699,776	0.35	0.65
	32	76	17	0.27	0.02	BIEC2-1133893	0.009	76,174,026	0.40	0.60
	10	35	10	0.24	0.02	BIEC2-115349	0.007	35,942,977	0.42	0.58
	10	49	14	0.23	0.02	BIEC2-121644	0.004	49,192,501	0.47	0.53
	32	78	15	0.23	0.03	BIEC2-1134793	0.006	78,794,348	0.42	0.58
	5	71	17	0.22	0.03	BIEC2-917852	0.005	71,366,629	0.45	0.55
Study II	32	82 <sup>a</sup>	12	1.71	0.10	BIEC2_1138724	0.085	82,842,940	0.42	0.58
	1	153 <sup>b</sup>	14	1.02	0.07	BIEC2_65455	0.054	153,075,829	0.46	0.54
	22	25 <sup>c</sup>	12	0.62	0.05	BIEC2_589497	0.038	25,515,664	0.46	0.54
	3	35	17	0.54	0.07	BIEC2_776764	0.017	35,408,432	0.44	0.56
	18	32	21	0.52	0.05	BIEC2_410288	0.012	32,615,336	0.43	0.57
	32	76	21	0.45	0.05	BIEC2_1133893	0.017	76,174,026	0.41	0.59
	4	100	18	0.44	0.05	BIEC2_878810	0.031	100,879,322	0.44	0.56
	11	0	13	0.43	0.05	BIEC2_133565	0.015	300,347	0.39	0.61
	1	152	19	0.42	0.05	BIEC2_65299	0.013	152,871,014	0.46	0.54
	15	19	17	0.41	0.04	BIEC2_293386	0.013	19,652,562	0.42	0.58

<sup>1</sup> Chr: Chromosome number.

<sup>2</sup>mb: Mega base pair, 24 mb specifies a region from 24 mb to 25 mb (1-mb).

<sup>3</sup>Gen. var (%): Percentage of genetic variance explained by the window calculated using every 100<sup>th</sup> cycle after burn-in cycles.

<sup>4</sup>p>AVG: Proportion of cycles where the window explained more than the expected genetic variance 0.042%.

<sup>5</sup>MF: Model frequency is a proportion of post burn-in cycles where the SNP included in the model had an effect.

<sup>6</sup>Unaff: Unaffected group of horses

<sup>7</sup>Aff: Affected group of horses

a, b, c: top three windows that explained the highest, second highest and third highest percentage of the genetic variance.

### 3.3 Materials and methods (Study IV)

#### 3.3.1 Horse data and quality control

Study IV consisted of 203 horses representing nine populations (Table 3). The SNPs ( $n = 45,986$ ) overlapping between two genotyping platform: Illumina® EquineSNP50 Genotyping BeadChip and Illumina® EquineSNP70 Genotyping BeadChip were used. Quality control was performed using PLINK software (Purcell *et al.*, 2007). SNPs present on the X chromosome were removed followed by the SNPs with a genotyping rate  $< 90\%$ , and a minor allele frequency (MAF)  $< 1\%$ . Horses with genotyping rate  $> 90\%$  were accepted for further analysis. A dataset pruned based on LD was also created apart from a dataset without LD based pruning.

All analyses were performed using PLINK software (version 1.07) (Purcell *et al.*, 2007) unless specified.

#### 3.3.2 Inbreeding coefficient

Inbreeding coefficients were estimated to study the genetic diversity status within the populations. SNPs in LD can reduce the information-content of a dataset (Lopes *et al.*, 2013) and may lead to a less accurate estimation of inbreeding coefficients. Inbreeding coefficients ( $f_i$ ) were estimated using data obtained after LD-based SNP pruning.

Inbreeding results in homozygosity. Hence, inbreeding coefficient was also determined by calculating the proportion of the autosomal genome that was homozygous. Runs of homozygosity (ROHs), long stretches of homozygous SNPs, were identified to determine the homozygous segments in the genome. Inbreeding coefficients ( $f_{i,ROH}$ ) was estimated as defined by (McQuillan *et al.*, 2008).

#### 3.3.3 Genetic relationships within and among horse populations

The genome-wide genetic relationships within and between the 203 horses from nine horse populations were visualized using MDS plot. The MDS plot also served as a quality control, as it exposes potential breed-misclassification of a horse.

To get a quantitative measure of genetic differentiation between the horse populations, fixation index  $F_{ST}$  was estimated in PLINK (version 1.9) (Purcell *et al.*, 2007).  $F_{ST}$  values per SNP were estimated based on the method described by (Weir & Cockerham, 1984).  $F_{ST}$  values per SNP between two populations were averaged to establish the genetic differentiation between these populations.

#### 3.3.4 Runs of homozygosity

Runs of homozygosity (ROHs), in the genome were identified to detect genomic regions potentially under selection. To discover ROHs, a window of 50 SNPs was slid across the genome, one SNP at a time. We considered genomic regions in more detail when an at least partially overlapping ROH was present in at least 25% of all horses across populations, or in at least 80% of the horses within a population. The genomic regions where no or hardly any ROH was detected were also considered in detail.

### 3.4 Main Findings

#### 3.4.1 Estimates of Inbreeding coefficient

In general, inbreeding in populations that allow introgression was lower compared to inbreeding in closed populations. The highest mean inbreeding coefficients were observed in the Friesian horse ( $f_i = 30.4\%$ ,  $f_{i,ROH} = 22.2\%$ ) and Shetland pony populations ( $f_i = 24.0\%$ ,  $f_{i,ROH} = 14.3\%$ ). The lowest mean inbreeding coefficients were observed in the KWPN sport horse ( $f_i = 6.9\%$ ) and Icelandic horse ( $f_{i,ROH} = 4.3\%$ ). Lowest inbreeding was observed in the KWPN sport horses as a result of introgression from other European warmblood populations to enhance performance traits.

#### 3.4.2 Genetic relationships within and among horse populations

The quantified relationships among populations were as expected based on the history and classifications of the populations. The relationships quantified using the MDS plot (Figure 6) and  $F_{ST}$  values were similar. In MDS plot, warmblood populations, coldblood populations and pony- like populations clustered together (Figure 6). The four Dutch warmblood horse populations: the Gelder horses, harness horses, KWPN sport horses and Groninger horses showed moderate genetic differentiation (pairwise mean  $F_{ST} \leq 0.07$ ). This moderate differentiation between the Dutch warmblood populations confirms the origin of the populations. Exchange of genetic materials occurred between Gelder horses (light draft horse) and Groninger horses (heavy draft horse) long time ago to enhance each other's trait (Hendricks, 1995). Gelder horses, harness horses and KWPN sport horses constitute the KWPN horse population, whereas Groninger horses are preserved and registered by the Groninger horse association as a separate warmblood population.

The pony-like populations clustering together in the MDS plot, showed moderate genetic differentiation ( $F_{ST} = 0.078$ ). The pony-like populations, Shetland ponies and Icelandic horses, likely shared ancestors more recently than with other populations (Hendricks, 1995).



Estimated genetic differentiation was greatest for the Friesian horse population ( $F_{ST}$  ranged from 0.13 to 0.17).

### 3.4.3 Runs of homozygosity

Genomic regions with SNPs present in a homozygous segment in many horses might indicate signatures of selection, or, conserved regions throughout evolutionary history. Across populations,  $\geq 25\%$  of all horses studied had SNPs in a homozygous segment located on chr 7, chr 10 and chr 11. Within populations, genomic regions potentially indicating signatures of selection were located on chr 1 for Shetland ponies and on chr 8 and chr 24 for Friesian horses. Many genes were located within the genomic regions identified across and within populations.

Genomic regions that contained very few or no SNPs that were present in a homozygous segment were identified on chr 5, chr 8, chr 18 and chr 20. The genomic region on chr 5 (14.1-15.0Mb) contained the gene, *LXN* (OMIM: 609305). The region surrounding *LXN* did not contain SNPs that were present in a homozygous segment. The change in expression of *LXN* decreased the number of hematopoietic stem cells in laboratory mice (Liang *et al.*, 2007). The hematopoietic stem cells are considered a model for endogenous stem cell populations. Natural variation in the size of endogenous stem cell populations is important for several biological processes like homeostatic tissue regeneration and potentially organismal longevity. Based on these findings, *LXN* seems to be a potential candidate gene for which, from an evolutionary perspective, variation is preserved. The genomic region on chr 8 (0.1-0.7 mb) contained a gene *MED15* (OMIM: 607372), essential for signal transduction of for instance *TFGB* and Activin (Yang *et al.*, 2006). Additional research is required to understand if these genes are conserved based on evolutionary perspective.

Table 3. Data description including the number of horses (N), country of origin, type of studbook, and a description of the 9 populations sampled in the Netherlands in Study IV

Population	N	Country of origin	Type	Studbook	Description
Belgian draft horse	23	Belgium	Coldblood	Closed	Heavy draft horse
Friesian horse	20	the Netherlands	Coldblood	Closed	Harness and riding horse
Gelder horse	20	the Netherlands	Warmblood	Open, introgression of warmblood horses	Light draft and riding horse
Groninger horse	20	the Netherlands	Warmblood	Closed	Heavy draft and riding horse
Harness horse	20	the Netherlands	Warmblood	Open, introgression of Hackney and American Saddlebred horses	Harness horse
Icelandic horse	39	Iceland	Coldblood	Closed	Gaited, riding horse
KWPN sport horse	18	the Netherlands	Warmblood	Open, introgression of Thoroughbred and warmblood sport horses	Sport horse, jumping or dressage
Lipizzaner horse	23	Lipica, modern-day Slovenia	Warmblood	Closed, use of several sire and dam lines	Riding horse, Spanish riding school
Shetland pony	20	Shetland Isles, Scotland	Coldblood	Closed, four categories based on withers height	Harness and riding pony

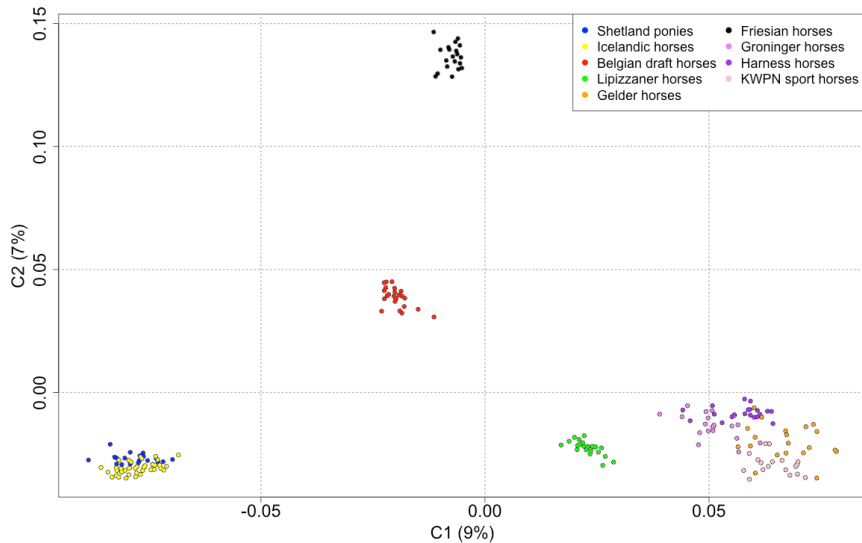


Figure 6. Multidimensional scaling plot of 203 horses from 9 populations in Study IV. The X-axis represents the first component (C1) and Y-axis the second component (C2).

## 4 General discussion

In this thesis, we utilized genetic variation in two horse populations to understand the genetics behind IBH susceptibility. We also utilized the genetic variation within and between horse populations to understand their genetic diversity status, genomic relationships, and evolutionary history of the populations.

The first section of this discussion will focus on genotypic, phenotypic and environmental aspects of IBH which is required to better estimate the genetic variation of IBH susceptibility. This section will discuss findings regarding IBH susceptibility in Studies I-III, and discuss alternatives to better capture the genetic variants associated with IBH.

The second section of this discussion will address how knowledge of genetic relationships between various horse populations might contribute to understand IBH susceptibility across populations.

### 4.1 Genotype

To better understand the genetic variation of IBH susceptibility we analyzed different datasets that included different number of samples and different number of SNPs. We will address the findings regarding IBH from the Studies I-III and discuss consequences of increasing sample size, number of SNPs on the outcome of the different studies.

#### 4.1.1 IBH: complex allergic disorder

IBH is considered a polygenic disorder, which is also influenced by environmental factors. This concept has been supported by candidate gene studies and by GWA studies (Table 4). Studies I-III performed in this thesis also support this concept. In Study I and II, we estimated an effect of multiple windows defined along the genome, on IBH susceptibility. In Study I, 29 non-overlapping windows, 1-mb in length, explained 0.14 – 0.51% of the genetic

variance (data not shown). Similarly in Study II, 30 non-overlapping windows explained 0.19 – 1.71% of the genetic variance of IBH (data not shown). The study by Schurink *et al.* (2012) on IBH susceptibility also showed 20 windows explaining 0.14 – 0.62% of the genetic variance of IBH in Shetland pony mares and 0.14 – 0.65% of the genetic variance in Icelandic horses. The identified windows in the Study I and II influenced IBH with small effects. By using a similar method of identifying genomic regions, in layer chickens, Wolc *et al.* (2012) estimated the influence of multiple 1-mb non-overlapping windows on the complex traits, egg uniformity and egg weight. Top 10 1-mb windows associated with egg weight (eggs laid between 26 and 28 weeks of age), explained 1.1 – 28.0% of the genetic variance (Wolc *et al.*, 2012). With large number of samples (29,000 individuals) genomic regions for the complex traits were identified (Wolc *et al.*, 2012). Also for IBH, a larger number of horses would increase the chances of finding regions with effects on IBH.

To compare the effect sizes of the different associated SNPs with IBH within the Studies I, II and III, and also with previous studies (Table 4) we calculated odds ratio (OR) for the SNPs (Table 5). In general, OR represents the odds that an outcome will occur given a particular exposure, compared to the odds of the outcome occurring in the absence of that exposure (Szumilas, 2010). OR also provides a measure of association (maybe not statistically significant) for example between a SNP (or allele of SNP) and trait of interest where OR = 1 indicates lack of association; OR > 1 and OR < 1 corresponds with increasing and decreasing odds respectively, with presence of that allele of SNP. In case of association studies, OR is an exponential function of the allele substitution effect ( $e^{b1}$ , where  $b1$  is an allele substitution effect) associated with a one-unit increase in the exposure (Szumilas, 2010). In this thesis, OR was determined as  $e^{b1}$ . In Study I and II, we estimated OR based on an allele substitution effects of a minor allele of the top SNPs within the 10 windows that had highest association with IBH (Table 4). In the studies I and II, the effects of the SNPs (allele substitution effects) were estimated based on study by Kizilkaya *et al.* (2010) implemented in GenSel software. In Study III, we used allele substitution effects for a minor allele of the top 10 SNPs that were obtained using mixed model structured association (Table 5) because this model gave the lowest value for genomic inflation factor in this study.

The OR values of the SNPs in Study III (Table 5) ranged from 1.6 to 8.4, and were similar to OR values in the study by Schurink *et al.* (2013a). In our case, OR of 1.6 denotes that a heterozygote horse carrying a minor allele (risk allele) will have probability of developing IBH by 1.6 times compared to a horse homozygous for the major allele (protective allele). However, these OR

values obtained in the Study III by fitting a single SNP in a model might be overestimated because of small number of horses in the study (Xu, 2003).

The OR values in the Studies I and II (Table 5) were lower compared to in Study III (Table 5) and other studies. This might be because allele substitution effects of the SNPs associated with IBH in Study I and II are partial effects of the SNPs (Garrick & Fernando, 2013). In the Study I and II, multiple SNPs were simultaneously fitted in the model. When multiple SNPs fitted simultaneously are in high LD the partial effects of the SNPs tend to get smaller (Garrick & Fernando, 2013). Multiple SNPs being in high LD is likely present in the studies I and II, resulting in small effect sizes for the associated SNPs compared to the effects as estimated in the study III.

Table 4. Odds Ratio of associated genetic variants with IBH susceptibility in previous candidate gene studies and GWA studies

Reference	Study type	Breeds	N <sup>a</sup>	Phenotype	Genetic variants used	Associated genetic variants	OR <sup>b</sup>
Andersson et al. 2012	Candidate gene study	Icelandic horse	187	Clinical symptoms	6 Microsatellites markers on MHC region; Exon2 sequence from <i>Eqca-DRA</i> gene and <i>Eqca-DRB3</i> gene on MHC class II region	COR112 (microsatellite); <i>Eqca-DRB3</i>	4.19; 3.60
Andersson et al. 2012	Candidate gene study	Exmoor pony	240	Clinical symptoms	6 Microsatellites markers on MHC region; Exon2 sequence from <i>Eqca-DRA</i> gene and <i>Eqca-DRB3</i> gene on MHC class II region	COR112	1.48
Klumplerova et al. 2012	Candidate gene study	Icelandic horse	89	Clinical symptoms	5 Microsatellite markers on MHC region; Exon2 sequence from <i>Eqca-DRA</i> gene on MHC class II region	<i>DRA</i> gene	0.25
					20 SNPs in 17 candidate allergy-related genes	Combined genotypes: <i>CD14</i> , <i>IL23R</i> , <i>TSLP</i> and <i>TGFB3</i>	3.76-5.16
Vychodilova	Candidate	Kladruby	61	Clinical	40 SNPs in 29 candidate genes	Genotypes containing 2 to 5	0.06 -

Reference	Study type	Breeds	N <sup>a</sup>	Phenotype	Genetic variants used	Associated genetic variants	OR <sup>b</sup>
et al. 2012	gene study	horse	mares	symptoms	(function in immune response and allergy); MHC region: Exon 2 sequence of <i>Eqca-DRA</i> gene and <i>Eqca-DQA</i> gene	genes: <i>IFNG</i> , <i>TGFB1</i> , <i>JAK2</i> , <i>TSLP</i> and <i>IVL</i>	12.00
Vychodilova et al. 2012	Candidate gene study	Kladruby horse	58 mares	IgE level: binary and continuous	40 SNPs in 29 candidate genes (function in immune response and allergy); MHC region: Exon 2 sequence of <i>Eqca-DRA</i> gene and <i>Eqca-DQA</i> gene	Binary: <i>IGHE</i> , <i>ELA-DRA</i> , and <i>IL10/b</i> Continuous: <i>IGHE</i> , <i>FCERIA</i> , <i>IL4</i> , <i>IL4R</i> , <i>IL10</i> , <i>IL1RA</i> , and <i>JAK2</i>	0.02 - 13.62
Vychodilova et al. 2012	Candidate gene study	Kladruby horse	30 mares	Gene expression in skin biopsies and peripheral blood leukocytes	SNPs located on genes: <i>IFNG</i> , <i>TGFB1</i> , <i>JAK2</i> , <i>TSLP</i> and <i>IVL</i>	<i>IFNG</i> and <i>TGFB1</i>	
Schurink et al. 2012	Genome-wide association study	Icelandic horse	146	Clinical symptoms	65,175 SNPs (Illumina EquineSNP70 Genotyping BeadChip)	20 1-mb windows on 9 chromosomes explaining >0.14% of genetic variance of IBH: 1, 3, 4, 5, 6, 7, 9, 11, 15, 18, 19, 20, 23, and X	
Schurink et al. 2012	Genome-wide association study	Shetland pony	200	Clinical symptoms	65,175 SNPs (Illumina EquineSNP70 Genotyping BeadChip)	20 1-mb windows on 14 chromosomes explaining >0.14% of genetic variance of IBH: 3, 7, 8, 11, 17, 20, 23, 27, and 28	
Schurink et al. 2013	Genome-wide association study	Shetland pony	180	Clinical symptoms	54,602 SNPs (Illumina EquineSNP50 Genotyping BeadChip)	24 SNPs on 12 chromosomes: 1, 3, 4, 6, 7, 8, 9, 11, 17, 20, 27, 30, and X	1.94 - 5.95

<sup>a</sup>N: Number of horses; <sup>b</sup>OR: Odds Ratio

Table 5. Odds ratio for top 10 SNPs for Study I- III

Chromosome	SNP name	SNP position	Allele substitution effect	Odds Ratio
Study I				
5	BIEC2-917852	71,366,629	-0.001	0.999
10	BIEC2-119028	41,699,776	0.002	1.002
10	BIEC2-115349	35,942,977	-0.001	0.999
10	BIEC2-121644	49,192,501	-0.001	0.999
17	BIEC2-385377	77,343,575	-0.004	0.996
18	BIEC2-409581	26,766,516	-0.003	0.997
32	BIEC2-1132313	73,129,947	0.002	1.002
32	BIEC2-1133113	74,332,689	-0.001	0.999
32	BIEC2-1133893	76,174,026	0.002	1.002
32	BIEC2-1134793	78,794,348	-0.001	0.999
Study II				
1	BIEC2_65299	152,871,014	-0.003	0.997
1	BIEC2_65455	153,075,829	-0.016	0.984
3	BIEC2_776764	35,408,432	-0.004	0.996
4	BIEC2_878810	100,879,322	-0.008	0.992
11	BIEC2_133565	300,347	0.003	1.003
15	BIEC2_293386	19,652,562	0.003	1.003
18	BIEC2_410288	32,615,336	0.003	1.003
22	BIEC2_589497	25,515,664	-0.010	0.990
32	BIEC2_1138724	82,842,940	-0.021	0.979
32	BIEC2_1133893	76,174,026	0.003	1.003
Study III				
2	AX-103372605	116,550,518	2.134	8.445
4	AX-103676516	16,228,799	0.874	2.395
6	AX-102955701	55,403,540	1.652	5.217
7	AX-103802842	36,418,118	1.779	5.924
8	AX-104843731	62,711,228	1.332	3.788
8	AX-104130346	71,065,803	0.517	1.677
9	AX-104273278	9,674,307	2.006	7.434
16	AX-104129425	79,101,000	1.502	4.490
20	AX-103533051	29,990,169	1.679	5.362
26	AX-104585909	14,984,244	1.843	6.319

#### 4.1.2 Increased sample size and number of genetic variants

Effect sizes of the genetic variants are likely underestimated due to incomplete LD between the genetic variants and causal variants. However, adding samples in the dataset might also lead to stratified population and one should be cautious while performing such procedure. To increase sample size in Study II, we combined data from 209 Icelandic horses sampled in Sweden and 146 Icelandic horses sampled in the Netherlands. Two populations were genotyped using different genotyping platforms (Table 1). The minor allele frequencies of overlapped SNPs between two genotyping platforms were investigated in these two Icelandic horse populations. The correlation of 91% between minor allele frequencies of the SNPs indicated no stratification in the combined data.

By increasing sample size from 209 horses (Study I) to 355 horses (Study II), we observed increment in the percentage of the genetic variance explained by non-randomly associated windows with IBH. The top 10 windows non-randomly associated with IBH in Study I explained 3.07% of the genetic variance and in Study II explained 6.56% of the genetic variance of IBH. We compared total number of windows explaining  $\geq 0.14\%$  of the genetic variance of IBH between the Study I (29 1-mb windows) and the Study II (73 1-mb windows). We observed an increase in the percentage of the genetic variance explained by 1-mb window on chr X (82 mb, 76 mb, 4 mb) and on chr 9 (50 mb) in the Study II compared to the Study I, at the range 1.49% to 0.03%. In Study II, 146 Icelandic horses were added to form a combined data. The 146 additional Icelandic horses in Study II had previously been studied for IBH susceptibility (Schurink *et al.*, 2012) by using similar approach as in Study II. Hence, we compared non-randomly associated windows obtained from a combined data in the Study II with the top 20 1-mb windows presented in the study by Schurink *et al.* (2012). The percentage of the genetic variance explained by five 1-mb windows on chr 1 (7 mb), chr 3 (35mb), chr 7 (55mb), chr 15 (19mb) and chr 18 (32mb) in the Study II compared to study by Schurink *et al.* (2012), increased at the range 0.34% to 0.08%.

In Study I -III, we used the latest equine genotyping array available at that time (Table 1). In Study III, borderline association of SNPs located at chr 8 with IBH susceptibility was observed. Association of chr 8 with IBH susceptibility has not been observed before in any previous GWA studies. This might be most likely due to use of less dense SNP genotyping arrays in the Study I and II, and previous GWA studies. In Study III, 425 SNPs covered genomic region 70-71 mb on chr 8, and genomic region 78 mb was covered by 234 SNPs. While in the Studies I and II, less than 50 SNPs covered genomic region 70-71mb, and less than 25 SNPs covered genomic region 78 mb on chr 8 (Figure 7). The percentage of the genetic variance explained by 1-mb



windows on chr 8 [71 mb, 72 mb and 78 mb] in the Studies I and II was  $\leq 0.07\%$ .

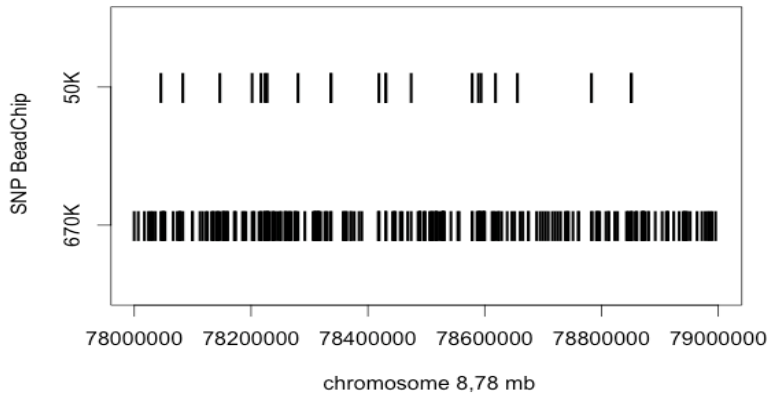


Figure 7. SNPs density at 78mb on chromosome 8 in 50K and 670K SNP chip. A line represents one SNP.

#### 4.1.3 Heritability

In the Study III, the heritability in Exmoor ponies was estimated to be 39.70% using IBH severity as a trait in a mixed model structured association approach. This estimated heritability is comparable to pedigree based heritability estimate (33%) of IBH (severity as trait) for Swedish-born Icelandic horse population (Eriksson *et al.*, 2008). Estimated heritability in the Study III could be inflated because environmental factors were not included in the model as answers obtained from participants via questionnaire, regarding the environmental factors, were not consistent in the Study III.

The genomic heritability, proportion of total variance explained by all SNPs, estimated in the Study I (11.4%) and the Study II (12.7%) was lower compared to pedigree based heritability of 27%, which was used to calculate prior for the genetic variance of IBH in Studies I and II. The heritability for IBH as a binary trait was estimated to be 27% (Eriksson *et al.*, 2008). Such scenario, referred to as missing heritability, has been encountered in GWA studies of complex traits for example human height, for disease such as schizophrenia, and others as reviewed in Maher (2008). In the study by Schurink *et al.* (2012), all SNPs explained 13% of variance as compared to

24% obtained by using pedigree information in Dutch Shetland pony mares. Estimates of heritability using genome-wide SNPs was 30% lower compared to the pedigree-based heritability for egg weight (Wolc *et al.*, 2012). Use of higher value of the probability,  $\pi = 99.5\%$  was suggested as one of the reason for low heritability estimate (Wolc *et al.*, 2012). The probability “ $\pi$ ,” indicates the percentage of SNPs assigned to having no effect on the trait, on each iteration. In the Studies I and II, we assigned 99.9% of the SNPs to have no effect on the trait IBH which could be one reason for low heritability estimates. Use of paternal half-sibs to design a matched case-control structure in the Study I might have influenced estimate of heritability.

To fill the gap of missing heritability, several methods has been proposed (Maher, 2008) such as increasing SNPs density, increasing sample size, and investigating alternative forms of genetic variants. In case of IBH, increment in number of Icelandic horse population gave similar values for SNP based heritability estimates in the Studies I and II. It might be case that rare variants are involved in influencing IBH susceptibility. To capture such rare variants more horses are required (Robinson *et al.*, 2014). Using dense SNPs in Exmoor pony population we estimated higher SNP based heritability as mentioned before. To better capture the genetic variation alternative forms of genetic variants could also be investigated. These genetic variants might influence IBH susceptibility in horses, however following section will be focused on CNVs and epigenetics. The potential of CNVs and epigenetics to better identify genetic variants in IBH and explain larger portion of genetic variance will be discussed.

The heritability estimates of allergies reported in human are higher compared to IBH heritability estimated (as binary trait) in different horse breeds, 16% to 27% (Schurink *et al.*, 2011; Schurink *et al.*, 2009; Eriksson *et al.*, 2008). Heritability estimates in humans utilized Monozygotic (MZ) and Dizygotic (DZ) twin pairs where only three essential statistics (MZ resemblance, DZ resemblance and total phenotypic variation) can be determined. However, many other genetic and non-genetic causal components might be contributing to the total phenotypic variation in the twin study designs. If phenotypic variation is not partitioned properly into causal sources, confounding will lead to bias in the estimate of heritability. For full sibs or DZ twins it is difficult to separate the dominance component from the common environmental component that might lead to upward bias in heritability estimates (Visscher *et al.*, 2008).

#### 4.1.4 Copy number variants and Epigenetics

##### *Copy number variants*

Copy number variants (CNVs) are segments of DNA ~1kb in length or larger in size present at different copy number in comparison with reference genome (Feuk *et al.*, 2006). Fifteen percentage of heritability estimate of schizophrenia was explained by 4 CNVs (Nagao, 2015). Use of CNVs for heritability estimation of IBH in horses is limited.

Analysis of CNVs related to allergy is highly relevant because studies on CNVs have revealed that the majority of genes containing CNVs are involved in immune response in humans (Redon *et al.*, 2006), in cattle (Bickhart *et al.*, 2012), and in horses (Ghosh *et al.*, 2014; Metzger *et al.*, 2013). Changes occurring in CNVs may cause changes in expression of the genes containing CNVs by either dysregulating RNA levels (increasing or decreasing), or by modifying the structure of the gene.

In horses, studies regarding association between IBH and CNVs are limited. However in humans, CNVs have been associated to allergic disorder such as atopic dermatitis (AD) (Brown *et al.*, 2012). AD in humans and IBH in horses have shown similar features (Schaffartzik *et al.*, 2012). Similar to IBH, AD is chronically relapsing, non-contagious, pruritic, IgE mediated hypersensitivity reaction, and share similar histopathological (microscopic examination of tissues) findings with IBH. In humans, CNVs within Filaggrin (*FLG*) gene was observed to be associated with AD susceptibility (Brown *et al.*, 2012). *FLG* gene alleles contained 10, 11 or 12 tandem repeats of ~972 bp in 3<sup>rd</sup> exon of *FLG*. Individuals with different alleles expressed different amount of *FLG* in the epidermis. Each addition of copy number of *FLG* reduced OR of AD susceptibility by a factor of 0.88 (CI 0.78 – 0.98) which suggest CNVs in *FLG* as a dose-dependent contribution to AD susceptibility. In horses also, role of *FLG* in IBH susceptibility was investigated. Expression of *FLG* in skin biopsies between IBH affected and unaffected Belgian warmblood horses did not show any significant difference. However, incorrect annotation of *FLG* complementary DNA (cDNA) sequence in the horse genome was suggested (Peeters, 2013). Therefore, a further investigation of CNVs in *FLG* in horses and association of *FLG* with IBH susceptibility seems to be required.

Direct investigation of CNVs is suggested as only biallelic CNVs (deletions and duplications) can be tagged by SNPs (Redon *et al.*, 2006). It is challenging to tag complex CNVs or multiallelic CNVs (Beckmann *et al.*, 2007) using SNPs. Hence, investigation of CNVs might contribute to identify genetic variants influencing IBH susceptibility, and also contribute to explain missing heritability.

## Epigenetics

Epigenetic variations occur due to environmental factors and lead to changes in gene expressions without altering nucleotides on the sequences. These variations occur through DNA methylation (DNAm), histone modification and micro RNAs (Baye *et al.*, 2010).

In humans, epigenetic modifications have been associated with susceptibility to different allergies in candidate gene approach and genome-wide approach. Associations of the epigenetic modifications have been investigated in a specific region of a candidate gene, forkhead transcription factor (*FOXP3*) gene (Hinz *et al.*, 2012). Children with low number of demethylated Treg cells (low *FOXP3* gene expression) showed a higher risk of developing AD (OR: 1.55, 95% CI = 1.00-2.41) and food allergies (OR: 1.5, 95% CI = 1.06-2.25) in the first year of life. Increase in susceptibility towards allergy in children with low number of demethylated Treg cells suggests a role of DNAm patterns in *FOXP3* gene in susceptibility towards allergy. In horses also effector T cells are suppressed in a similar way by Treg cells as in humans (Meulenbroeks, 2015a). Any changes in the stability of Treg cells may lead to abnormal immune response resulting in allergic response in horses. Hence, investigation of DNAm patterns of Treg cells regarding IBH is also of interest in horses. In horses, increased level of DNAm in *PAX5* gene in bone marrow, further supported by decreased mRNA expression of *PAX5* gene was observed in common variable immune deficient (CVID) horses (Tallmadge *et al.*, 2015). However, epigenetic modifications have not been analyzed in relation to IBH susceptibility in horses. With significance of epigenetic modifications to maintain balance in different levels of the immune system (for example T-cell differentiation), epigenetics is an important aspect that should be considered. Such a study should be considered to understand biological mechanisms underlying IBH susceptibility such that specific, efficient therapy can be introduced to reduce the prevalence of IBH and thus contributing to horse-welfare.

### 4.1.5 Study design

An ideal design of a GWA study depends on the genetic architecture of the trait of interest. In Studies I, II and III, we performed GWA analysis to find genomic regions associated with IBH, a common complex trait. In Studies I and II, a case-control design was used. In Study III, GWA analyses were performed using IBH as binary trait as well as a continuous trait, taking severity of the clinical symptoms into account. A case-control design is sensitive to population stratification, which might lead to false positive results if not taken into account. In the studies I, II and III, we visualized if any

subpopulations were present via MDS plot. Although two subpopulations were present in the MDS plot in the study I (Figure 3) and the study II (Figure 4), the genomic inflation factor was less than 1, which confirmed no effect of stratification. In the studies I and II, we fitted multiple SNPs simultaneously in a model. In such case, some of the SNPs will simultaneously account for known or unknown population structure (Fernando & Garrick, 2013).

In the Study III, although the MDS plot (Figure 5) did not show a structure with subpopulations in Exmoor pony population, the values for genomic inflation factor obtained after fitting models such as structured association and principal component analysis-based method were greater than 1 showing effect of stratification. The significance values for associated SNPs were then corrected using inflation factor obtained in respective models. Among the investigated models to investigate association, only mixed model structure association showed a genomic inflation factor close to one in the Study III.

A case-control design has also been used in the candidate gene association studies (Table 4). The candidate gene studies using case-control designs might also suffer from population stratification (Cardon & Palmer, 2003), that can increase the risk of inaccurate estimation of genetic variation if not taken into account. Also, the candidate gene studies will exclude vast information of the genome that may lead to inaccurate estimation of genetic variation. A matched case-control design was used in the Studies I – III. To correct for the population stratification due to pedigree, paternal-half sibs were used in the Studies I and II, and full-sibs were used in Study III (Schurink *et al.*, 2012).

We observed association of multiple SNPs, with small effects, with IBH susceptibility in the studies I, II, and III. In Studies I and II, we observed that top SNPs associated with IBH had small effects. We increased SNPs density in the Study III, none of the SNPs was significantly associated with IBH susceptibility after correcting for multiple testing. This hints that to tag genetic variants for a complex trait such as IBH we need much larger sample size compared to sample size used in the studies I-III. In GWA studies for complex traits, use of large sample size and dense SNP for mapping traits in domestic species has been suggested (Goddard & Hayes, 2009). However, data collection from large number of horses is challenging compared to other domestic species as cattle, pigs because horses are owned by private owners and are distributed in different geographical areas. It is also challenging to apply a strict identical protocol to diagnose IBH and gather consistent information about environmental factors in large number of horses gathered from different areas. In such case, large number of horses for IBH can be gathered by forming large collaborative groups of researchers with the goals of sharing material, data, and standardizing diagnosis and experimental methods

for IBH. The genetic diversity study performed in 814 horses from 30 different breeds (correcting for breeds in case of IBH association analysis) could be one example of a large international collaboration (Petersen *et al.*, 2013) in a field of horse genetics. As an alternative to collecting large samples for GWA studies we can also look towards including other types of study designs such as family-based designs along with GWA studies as a next step.

#### *Family-based design*

A robust way to identify and characterize variants influencing disease susceptibility including IBH could be to use a test that combine both linkage and association mapping, transmission disequilibrium test (TDT) (Ott *et al.*, 2011; Spielman *et al.*, 1993). TDT estimates frequency of allele transmission from parents to offspring on heterozygous loci. This design is not susceptible to population stratification as allele or genotype frequencies are not considered (Ott *et al.*, 2011). Usually TDT is performed to identify linkage after association has been established between a genetic variant and trait of interest, and act like a validating step for association studies (Spielman *et al.*, 1993). In humans, linkage and GWA studies have identified variants associated with diseases such as Crohn's disease as reviewed in Ott *et al.* (2011) and AD (Enomoto *et al.*, 2007). TDT has been used to confirm association results when many associations has been identified as reviewed in Ott *et al.* (2011). In the current case of IBH, where multiple genetic variants has been identified, with significant or suggestive associations, in different studies, performing TDT is an interesting option.

#### 4.1.6 Major histocompatibility complex

The MHC region has been significantly associated with IBH susceptibility in candidate gene studies (Klumplerova *et al.*, 2013; Andersson *et al.*, 2012b) but causal variant has not been identified yet. GWA studies for IBH susceptibility can help to confirm previously associated genomic region and also to identify novel genomic regions associated with IBH susceptibility. In the Studies I and II, approximately five 1-Mb windows covered the MHC region and surroundings on chr 20. In the Study I, each of the five windows explained 0.04-0.05% of the total genetic variance of IBH susceptibility. One of the five 1-Mb windows at a genomic region 33mb on chr 20 in Study II explained 0.15% of the genetic variance compared to 0.05% for the same corresponding window in Study I. This is most likely due to the increment in number of horses in Study II (355 horses) compared to in Study I (209 horses). A window at 35mb on chr 20 explained highest percentage (0.62%) of the genetic variance for IBH susceptibility in Dutch Shetland pony mares (Schurink *et al.*,

2012). A window at 30mb on chr 20 explained 0.17% of the genetic variance in Icelandic horses (Schurink *et al.*, 2012).

Previous GWA studies including the Studies I and II used SNP chip with less dense SNPs. We investigated the number of SNPs around MHC region and surroundings. The number of SNPs spanning MHC region and surroundings (~30-34 Mb) was around 100 in the Studies I and II. This number increased considerably to more than 6.3 thousand in Study III where we used the 670K Genotyping Array. We also observed that SNPs in MHC region are not evenly distributed (Figure 8) based on the current available genetic map of MHC region for horse genome (Equcab 2.0).

In Study III, a SNP 33.32 mb downstream the *DRB* gene in the MHC region was observed within top 10 SNPs associated with IBH susceptibility. After correcting for multiple tests none of the SNPs showed genome-wide significance in the Study III. Using large number of horses might increase the significance of the genetic variants.

In humans, MHC region contains high gene density, gene duplication, diversity and has low recombination events (Traherne, 2008; Gustafson *et al.*, 2003). The genes in MHC region are suggested to be extremely polymorphic for example in humans, more than 2,000 alleles are observed in the *DRB* loci of MHC class II region (Immuno Polymorphism Database). In humans, the structural and copy number variants are regarded as an important source of diversity in MHC region (Ghosh *et al.*, 2014; Traherne, 2008). Also, other sources of variation such as epistasis, epigenetics might play role in diversity among genes on MHC region and influence disease susceptibility (Traherne, 2008). Hence, in case of IBH susceptibility, different sources of variants should also be considered in addition to SNPs.

Mapping of MHC region is still on progress. Any conclusion regarding analysis of the MHC region is preliminary as fine mapping and annotation needs to be performed to further improve the possibilities to study this region. Improvements might be observed via studies focused on the MHC region alone in horses, supplemented by work to generate third assembly of horse genome with improved maps and annotation of genes. The third assembly of horse genome will be available soon (Kalbfleisch *et al.*, 2014). Hence, to investigate influence of MHC region on IBH susceptibility, sequencing MHC region might be a good strategy. Advancement in sequencing technologies has been successful to produce long reads of sequences with less error which can give correct map and annotation of regions with segmental duplication like MHC to construct fine map of regions containing segmental duplications less complex.

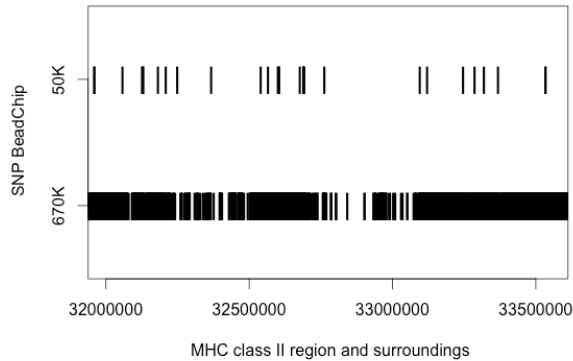


Figure 8. SNP density on MHC class II region and surroundings, from 50K SNP chip (on top) and 670K SNP chip (on bottom). A gap on 670K SNP chip is seen around 3.3 mb.

## 4.2 Phenotype

An accurate phenotyping of a trait is paramount in any studies. Accurate phenotyping can compensate in part for the requirement of large number of samples. To better estimate the genetic variations in case of IBH too, accurate phenotyping, which can be translated to accurate diagnosis of affected and unaffected horses, is of immense importance. Diagnosis of IBH in horses is a challenging procedure and should be performed with caution and considering essential criteria discussed earlier in the thesis. Here, the discussion will focus on other methods that can act as a supplement to the methods, to further improve reliability of IBH status of horses.

### 4.2.1 Diagnosis of IBH

#### 4.2.1.1 Clinical symptoms

The clinical symptoms of IBH include intense itching leading to scratching, hair loss, scaling, thickening of skin, open wounds and secondary infections (Schurink *et al.*, 2011; Bjornsdottir *et al.*, 2006). The clinical symptoms can in general be recognized by horse owners themselves. This enables the possibility to collect phenotypes on a large number of horses in a relatively easy way by consulting horse-owners, or inspectors. Itching might however also be caused by ecto-parasites such as mites or lice that can also enhance innate inflammatory responses, adaptive immune reactions and altered behavior



(Wilson, 2014), and additional information is required to exclude other causes of itch. This additional information can be collected through offering a questionnaire to owners with questions about location of the clinical symptoms on the horse (typically mane and tail), seasonality, and recurrence of symptoms in subsequent years, medical history of horse, type of preventive measures (stabling, rugs), and therapeutic measures against IBH. In particular the seasonality of IBH is a distinct characteristic of this allergic disorder, by which itching caused by ecto-parasites can be ruled out.

Collection of IBH data was performed through questionnaire in various studies (Peeters *et al.*, 2015; Andersson *et al.*, 2012b; Eriksson *et al.*, 2008), including the studies described in this thesis. Horse owners have additional knowledge about the health of the horse, over a longer period of time as opposed to a once only observation (Peeters *et al.*, 2015; Eriksson *et al.*, 2008). Practical limitations of gathering data using a questionnaire might be that mainly the horse owners who have affected horses will be eager to reply such that the collected data does not represent the population (Schurink, 2012). Non-responding should therefore be reduced as much as possible.

The clinical symptoms combined with well-documented medical history of the horse are still considered as the golden standard for diagnosis of IBH (Langner *et al.*, 2008; Baselgia *et al.*, 2006). Including falsely diagnosed horses in studies will reduce the power of investigating IBH susceptibility. Therefore horses included in the control group should certainly be not sensitive to IBH. Certainty about sensitivity is absent if horses are not exposed to *Culicoides* or if use of preventive measures such as keeping horses in the stables and use of protective rugs is limiting exposure to *Culicoides* spp. Certainty about the chosen unaffected horses was increased in this thesis by including Icelandic horses in Studies I and II that were at least 4 years of age, which is beyond the commonly assumed age of onset for IBH of 2 to 4 years. In Study I, unaffected Icelandic horses had an exposure to *Culicoides* spp. and had never developed clinical symptoms. Unaffected Icelandic horses sampled in the Netherlands, in Study II, were at least one year at risk of developing symptoms of IBH.

#### 4.2.1.2 *In-vivo and in-vitro diagnostic methods*

Different *in-vivo* and *in-vitro* diagnostic tests such as intradermal tests (IDT) (Langner *et al.*, 2008; Fadok & Greiner, 1990), allergen-specific IgE and IgG enzyme-linked immunosorbent assay (ELISA) (van der Meide *et al.*, 2014; Langner *et al.*, 2008), basophil degranulation test/histamine release test (HRT) (Langner *et al.*, 2008) and the cellular antigen stimulation tests/sulfoxidoleukotriene test (Klumplerova *et al.*, 2013; Marti *et al.*, 1999) can

supplement the clinical diagnosis of IBH (Meulenbroeks, 2015a), as a more detailed description about the allergic reaction can be given.

IBH affected horses showed increased IgE reactivity against allergens from *Culicoides spp.* compared to IBH unaffected horses (van der Meide *et al.*, 2014; Hellberg *et al.*, 2009). Diagnostic tests for IBH do not always match results from clinical inspection for IBH as e.g. shown in the study by van der Meide *et al.* (2014) by using recombinant allergens. In an ELISA test, 76 to 85% of IBH affected horses showed IgE reactivity towards recombinant allergens developed from *Culicoides obsoletus* complex (van der Meide *et al.*, 2014). The horses that did not show IgE reactivity, but had clinical symptoms of IBH, were likely showing symptoms from cross-reactions. Use of individual recombinant allergens in ELISA can distinguish genuine sensitization from sensitization against cross-reactive allergens (van der Meide *et al.*, 2014). The horses developing genuine sensitization might have different genetic background compared to the horses developing cross-reactive allergens, which should be accounted for in genetic analyses.

Different biological mechanisms in the horses might be acting towards different allergens. Hence, it is important to study a group of horses sensitized to allergens from the same *Culicoides spp.*. In the Netherlands, *Culicoides obsoletus* are found to be the major cause of IBH (Raaij *et al.*, 2008; Rijdt *et al.*, 2008). Icelandic horses and Shetland ponies sampled in the Netherlands showed increased IgE reactivity towards recombinant allergens from *Culicoides obsoletus* compared to a homolog from *Culicoides sonorensis* (van der Meide *et al.*, 2013). Hence, it is important to know which *Culicoides spp.* are causing IBH to avoid false diagnosis. *Culicoides spp.* causing IBH varies by country and region (Rijdt *et al.*, 2008). In Study I, Icelandic horses were sampled from five different regions in Sweden. The knowledge regarding type of biting midges causing IBH in Sweden is limited (Eriksson *et al.*, 2008), and type of *Culicoides spp.* might vary between regions. In Study III, most of the Exmoor ponies were sampled from England in the United Kingdom (UK). *Culicoides pulicaris* were potentially the cause of IBH in England (Mellor & McCaig, 1974). The recombinant allergens derived from *Culicoides obsoletus* complex (van der Meide *et al.*, 2014) might not be suitable for horses sampled in Sweden and the UK. Diagnostic tests can help to identify the *Culicoides spp.* that is responsible for IBH in a population. Additionally, allergen specific IgE levels, as derived from a diagnostic test can be used as a quantitative trait (Vychodilova *et al.*, 2013). The quantitative trait has a potential to improve reliability of individual phenotype and seems to be more suitable to perform GWA study (Schurink, 2012), as it allows for variation of degree of sensitization.

For diagnosis of IBH, thorough clinical examination and medical history of the horses as well as diagnostic tests should be included to increase reliability of diagnosis of IBH. The specificity and sensitivity of the type of diagnostic tests used in the study should also be taken into account. In a study by Klumplerova *et al.* (2013), the diagnostic tests served as a quality control and horses that had positive scores in diagnostic tests but had no clinical symptoms (sensitized horses) of IBH were removed from further analysis. The genetic background might be different between sensitized unaffected horses (positive for diagnostic tests but do not show clinical symptoms) and non-sensitized unaffected horses (negative for diagnostic test and do not show clinical symptoms). Classifying sensitized and non-sensitized unaffected horses in the control group in GWA studies can lead to reduction in power, assuming that the genetic background for IBH susceptibility differs between sensitized and non-sensitized unaffected horses. Hence, the horses showing both clinical symptoms of IBH and positive diagnostic tests can be classified as affected horses, and the horses with no clinical symptoms and negative diagnostic tests can be classified as unaffected horses. This step should be followed after careful evaluation of the medical history, exposure to the allergens, and other criteria discussed earlier.

Because of the polygenic inheritance of IBH, the GWA studies require large number of affected and unaffected individuals. An ideal situation would be to perform diagnostic tests for all the horses in the study. However, there are some limitations for using diagnostic tests such as requirement of fresh blood samples for example in CAST, HRT (Marti *et al.*, 1999), skilled manpower (for example veterinarian) for blood collection (Schurink, 2012), and economical issues. In addition, these diagnostic tests are invasive and requires ethical approval (Schurink, 2012). Performing diagnostic tests such as IDT is intensive than obtaining blood samples and performing ELISA tests. Also, IDT poses a risk of sensitizing non-sensitized horse with allergens introduced through the test. Using only clinical symptoms and medical history to identify unaffected horses for a control group poses a certain risk of selection of false negatives. This risk will be reduced when using diagnostic tests such as ELISA to confirm IBH status of horses categorized as unaffected horses from the clinical symptoms and medical history.

In humans, dose-response relationship for peanut allergy was observed (Hourihane *et al.*, 1997). Fourteen highly sensitive individuals, allergic to peanut proteins, were supplied with different dosage of peanut proteins. Few individuals showed allergic reactions to the lowest dosage of peanut protein supplied whereas other individuals did not show reactions to the highest dose of protein supplied. Such dose-response relationship may exist between

*Culicoides spp.* allergens and horses, and should be considered for IBH. Dose-response relationship may influence diagnosis of IBH in unaffected horses. Hence, it should be made sure that unaffected horses have sufficient exposure to *Culicoides spp.* allergens. Including sufficient number of affected and unaffected horses from one farm/location would be an ideal scenario because such scenario ensures that all the horses in the study are exposed to similar allergens/environment for similar period of time. It is quite challenging to find large number of horses in one farm/location because the private owners residing in different locations keep horses.

#### 4.2.1.3 Estimated breeding values

Determining IBH-sensitivity of a horse based on its own phenotypic expression might be considerably inaccurate because the IBH status is depending on exposure to *Culicoides spp.* (at a minimum density). In situations of strong dependence on environmental conditions, a predicted phenotype rather than the realized phenotype can better be used as trait under study in research and a prediction can be derived from the estimated breeding values (EBVs). Horses with highest and lowest EBVs for IBH can be selected for the association study, similar as was performed by Aguilar *et al.* (2016) in dairy cattle for somatic cell score. Cows with extreme high and low EBVs, that is 2 SD above and below the average EBVs, were selected for the association studies (Aguilar *et al.*, 2016). A similar procedure can be applied for IBH in horses. The selected horses can be assigned based on the breeding values to the affected and unaffected group in association studies or alternatively the breeding values can directly be used as a continuous trait.

Using EBVs directly as a phenotype in GWA studies resulted in large number of false positives in a study by Ekine *et al.* (2013). Hence de-regressed EBVs, also adjusted for parent average effects, can better be used for GWA studies.

Accuracy of EBVs depends on heritability of trait, type and amount of information sources provided. For example, the accuracy of an EBV based on phenotypes of the sire and its progeny will be higher than when based only on half-sibs of the sire. When EBV is based on fewer information sources, EBVs tend to regress towards the mean. The accuracy of EBVs will particularly benefit from inclusion of (large) progeny groups. Using phenotypes derived from EBVs is therefore less suitable for studies on mares and on small breeds, because accuracies of EBVs are low due to small progeny groups. In addition, if there is no system to keep routine record of IBH status, it will be challenging to gather large amount of data.

### 4.3 Environment

IBH is a multifactorial trait influenced by genetic and environmental factors. Environmental factors such as age, coat colour, body condition of horse; exposure of horse to allergens; preventive measures such as stabling of horses, and use of protective rugs have been suggested to influence development of IBH (Peeters *et al.*, 2014; Schurink *et al.*, 2013a; Olsen *et al.*, 2011; Eriksson *et al.*, 2008; Brostrom *et al.*, 1987). The effect of sex on IBH development is inconclusive (Peeters *et al.*, 2014; Eriksson *et al.*, 2008; Brostrom *et al.*, 1987). The environmental factors play an important role in development of IBH and thereby influence IBH prevalence.

In the Studies I-III, the effect of sex didn't show significance for IBH susceptibility for Icelandic horse populations or Exmoor pony population, and the sex was not included as a covariate in the analyses. Age has been found to be of importance for IBH, with the age of onset of IBH around 2-4 years (Schurink, 2012). Age was part of the selection criteria for horses to be included in the studies, and was not included as an effect in the models in Study I-II. In Study III age was unknown for a number of horses and not included as a covariate.

Exposure to *Culicoides spp.* is crucial for development of clinical symptoms of IBH. The *Culicoides spp.* are weak fliers, and they tend to remain close to their breeding sites (Schurink, 2012). Considering this behavior of the *Culicoides* species it is possible that even within a region, density of *Culicoides* species can differ. As the presence of *Culicoides spp* vary on different geographical area, exposure to allergens also differ based on geographical area of horses. In Northern Sweden *Culicoides* are not densely present and to avoid controls that had not been exposed to allergens, only cases were selected from this area in Study I

The majority of the horses (97%) in the Study I was born in Sweden. In the Study II, in a combined data, 57.2% of Icelandic horses were born in Sweden, 34.6% of Icelandic horses in the Netherlands, 5.6% of Icelandic horses in Iceland, and remaining were born in European countries. Immune response of horse population might differ if the allergens differ, e.g. if different species of insects cause IBH. *Culicoides obsoletus* has been identified as a cause of IBH in the Netherlands (Raat *et al.*, 2008; Rijt *et al.*, 2008), but knowledge is lacking on which species are of importance for IBH in Sweden and the UK. Hence, studies regarding *Culicoides spp* causing IBH are encouraged.

## 4.4 Genetic diversity and IBH susceptibility

Selection along with evolutionary mechanisms such as mutation, migration, recombination and genetic drift have led to the evolution of approximately 500 genetically diverse horse breed types and non-breed types. Genetic diversity within these horse populations is of interest as the genetic diversity is essential to maintain survival and adaptive potential of a population (Oliehoek *et al.*, 2009), and to avoid inbreeding depression. Events such as population bottlenecks, closed breeding practices along with selection have shown to reduce genetic diversity. Reduced genetic diversity have been associated with reduced production and survival traits (Leroy *et al.*, 2015; Thompson *et al.*, 2000). Reduced genetic diversity has also been associated with an increase in the level of infections and disease susceptibility (Cheng *et al.*, 2012; Acevedo-Whitehouse *et al.*, 2003). In a population of Tasmanian devil (*Sacrophilus harrisi*), found in Tasmania Island in Australia, reduced diversity of the MHC genes has been suggested as a risk for development of a contagious cancer, devil face tumour disease (DFTD) (Siddle *et al.*, 2007). It is suggested that DFTD is able to avoid the immune system due to lack of diversity in MHC genes.

Reduced genetic diversity might also increase risk of allergic disorders. Homozygosity across the entire MHC class II region was observed to increase risk of IBH susceptibility (Andersson *et al.*, 2012b).

In study IV, genetic diversity status of nine horse populations, sampled from the Netherlands, was assessed using SNPs. Inbreeding coefficients ( $f$ ) and runs of homozygosity ( $F_{ROH}$ ) were investigated to get status of the genetic diversity of the nine horse populations. The horse populations used in the study IV are Belgian draft horses, Friesian horses, Icelandic horses, Shetland ponies, Lipizzaner horses, Groninger horses, and Warmblood population which include Gelder horses, Harness horses and KWPN sport horses. Many horse populations worldwide are affected by IBH. IBH prevalence was observed to be 18.2% for Dutch Friesian mare population (Schurink *et al.*, 2011) and 8.8% for Dutch Shetland pony mare population (Schurink *et al.*, 2009). The prevalence of IBH for Warmblood population in the Netherlands was observed to be lower compared to Friesian horse population, Shetland pony population and Icelandic horse population (van den Boom *et al.*, 2008). Belgian warmblood horses showed IBH prevalence of 9.8% (Peeters *et al.*, 2011). The prevalence of IBH for Swedish Icelandic horse population was observed to be 8% (Eriksson *et al.*, 2008). The knowledge regarding IBH prevalence in other breeds (Gelder horses, Groninger horses, Lipizzaner horses, Belgian draft horses) is limited. Heritability of IBH was estimated in Dutch Friesian breeding mares (0.16, SE = 0.06) (Schurink *et al.*, 2011), Dutch Shetland

breeding mares (0.24, SE = 0.06) (Schurink *et al.*, 2009), and Swedish-born Icelandic horses (0.33, SE < 0.2) (Eriksson *et al.*, 2008).

The Friesian horse population was observed to be the most inbred horse population (mean inbreeding coefficients:  $f_i$ : 30.4%,  $f_{iROH}$ = 22.2%) in Study IV. High level of inbreeding in the Friesian horse population has also been observed in other studies (van de Goor *et al.* 2011). A history of several population bottlenecks, a closed breeding scheme along with selection has led to an increase in levels of inbreeding in the Friesian horse population. The Friesian horse population showed the highest number of long stretches of ROHs (homozygous regions) suggesting recent occurrence of inbreeding in the population. The Friesian horse population appeared genetically different than other horse populations in Study IV. Hence, the genomic regions where  $\geq 50\%$  of the horses within the Friesian horse populations showed ROHs was selected for further investigation (data not shown in Study IV). The genes, present in the selected genomic regions, were listed using Ensembl (EquCab 2.0). The biological functions of known genes on the list were classified using PANTHER classification system (Mi *et al.*, 2013). Using PANTHER statistical tool, we also performed statistical overrepresentation test for the list of genes. This test classifies the functions of the genes into different biological processes and performs binomial test on obtained processes to identify overrepresentation and underrepresentation of the biological processes (Mi *et al.*, 2013). In the Friesian horse population we observed significant overrepresentation of the immune response biological process. Majority of the genes present in the selected genomic regions of the Friesian horse population were involved in immunity implying low diversity in the immunity related genes. One can speculate that low genetic diversity in immunity related genes in the Friesian horse population might have led to an increase in IBH susceptibility.

In Study IV, KWPN sport horses showed lower level of inbreeding than other populations. This warmblood population had less and short ROHs. The diverse genome of the KWPN sport horses might be due to its open breeding scheme and interbreeding with other warmblood populations. Additionally, Thoroughbreds are introgressed in breeding of the KWPN sport horses to enhance performance. Introgression would result in a breaking down of ROHs, reducing in particular the length of ROHs and creating diversity. The warmblood populations were observed to have low IBH prevalence (van den Boom *et al.*, 2008). One can hypothesize that a diverse genetic background in the warmblood population as the KWPN sport horses might be associated to factor to low IBH prevalence.

#### 4.4.1 Genetic relationships and IBH susceptibility

IBH is a polygenic skin allergic reaction. Different horse populations might share similar genetic background for IBH susceptibility. In Study IV, we observed moderate genetic differentiation between the Icelandic horse population and the Shetland pony population ( $F_{ST} = 0.078$ ). Dutch Shetland ponies and Dutch Icelandic horses also showed IgE reactivity towards same recombinant allergens of *Culicoides obsoletus* complex (van der Meide *et al.*, 2014) which suggests presence of common genetic background for IBH susceptibility. In Study I, 1-mb window located at 66mb on chr 9 explained 0.15% of genetic variance for IBH susceptibility in Icelandic horse populations. A SNP, on chr 9 significantly associated with IBH susceptibility in Dutch Shetland ponies (Schurink *et al.*, 2013a), was located 452,819 bp upstream from the associated SNP in Study I. This suggests that some of the genes involved in IBH susceptibility can be found across different horse populations. Similarly, in Study I, a window located at 77 mb on chromosome 17 explained 0.34% of the genetic variance for IBH susceptibility. This window was one of the top 3 windows associated with IBH. In Shetland pony mares, two 1-mb windows located at 75 and 76 mb on chr 17 explained 0.30% and 0.16% of the genetic variance respectively (Schurink *et al.*, 2012). In Study I, a gene *IRS2* functioning as immunity-related gene was identified at 77 mb on chr 17. The identification of these neighboring windows on chromosome 17 in 2 separate breeds further strengthen the potential role of *IRS2* in IBH. With above-mentioned observations one can speculate that among multiple genes influencing IBH susceptibility, association of a few common genes can be identified in multiple populations.



## 5 Conclusions

With references to the aims of this thesis, the specific conclusions are:

Study I – III:

- I. Non-random association of multiple genetic variants in studies I – III showed that IBH susceptibility is influenced by many genes (most likely each with small effects).
- II. Increasing sample size (dataset) explained a higher proportion of the genetic variance of IBH.
- III. Utilizing increased number of genetic variants (density of genetic variants) in the genome, and inbred horse identified novel genomic regions.

Study IV:

- I. The highest and lowest level of mean inbreeding was observed in Friesian horse population and KWPN sport horse population.
- II. The Friesian horse population was genetically most different than other horse populations. Pairwise genetic differentiation between Dutch warmblood populations (Gelder horse, Groninger horse and KWPN sport horse) was lowest compared to other pairs.
- III. Potential regions of selection within populations were observed on chromosomes 1, 8 and 24; across populations were observed on chromosomes 7, 10 and 11.
- IV. The results in this study reflected expected knowledge based on the evolutionary history of the studied horse populations.



## 6 Sammanfattning

Genetisk variation som bidrar till den fenotypiska variationen användes i denna avhandling för att förstå den genetiska bakgrunden till en komplex egenskap IBH, och för att förstå den genetiska mångfalden och relationer mellan olika hästpopulationer.

IBH är den vanligaste allergiska hudsjukdomen hos hästar som orsakas av bett av knott, främst Culicoidesarter. Det påverkar olika hästraser i världen. Utan effektiv behandling så försämrar IBH hästars hälsa och orsakar ekonomiska förluster. I denna avhandling använde vi SNP-markörer spridda över hela genomet för att identifiera områden som bidrar till genetisk variation för IBH-känslighet. Vi undersökte också inverkan av ökat antal hästar och täta SNP på identifiering av genomiska regioner som är associerade till IBH-känslighet. Flera regioner med små effekter observerades i studie I-III. Intressanta regioner i islandshästpopulationen, i studierna I och II, observerades på kromosomerna 1, 7, 10, 15 och 17. Den procentuella andelen av den genetiska variansen som kan förklaras av tio regioner ökade från 3.07% (studie I) till 6.56% (studie II). Nya regioner identifierades när antalet islandshästar ökades i studie II. Användning av täta SNP i exmoorponnypopulationen ledde till att vi identifierade nya genomiska regioner associerade till IBH-känslighet på kromosom 8 men de låg på gränsen till att ha betydelse.

I studie IV visade det sig att förkunskaper om den evolutionära historien hos hästpopulationer matchade de erhållna resultaten från undersökningen av genetiska relationer inom populationer av holländskt varmblod (parvis medel  $F_{ST} \leq 0.070$ ), och inom ponnyliknande populationer (parvis medel  $F_{ST} \leq 0.078$ ). Hästpopulationer med liknande genetisk bakgrund kan dela liknande genetiska komponenter för IBH-känslighet. Populationen av Frieserhäst hade lägst variation (medelvärde inavelskoefficienterna:  $f_i = 30.4\%$ ,  $f_{iROH} = 22.2\%$ ) i studie IV och var genetiskt distinkta ( $F_{ST}$  till andra populationer varierade från 0.13 till 0.17). Detta kan vara ett resultat av en historia med flera

flaskhalseffekter och urval i en sluten population. Låg variation i immunitetsrelaterade gener observerades i Frieserhästuppopulationen och skulle ha kunnat leda till ökad förekomst av IBH. På liknande sätt kan låg känslighet för IBH i den varmblodiga populationen KWPN sporthäst bero på hög genetisk variation (= -6.9%). Hög genetisk variation i KWPN sporthäst kan vara ett resultat av ett öppet avelssystem och korsning med andra varmbloodspopulationer.

## 7 Samenvatting

In dit proefschrift is de genetische achtergrond van het complexe kenmerk Staart- en maneneczeem (SME) bij paarden bestudeerd, alsmede de genetische diversiteit bij en verwantschap tussen paardenrassen.

SME bij paarden is een allergische huidaandoening, die wordt veroorzaakt door beten van knutten (*Culicoides*). SME komt wereldwijd voor bij meerdere paardenrassen. Er bestaat geen effectieve behandeling tegen de aandoening en de aandoening is zeer belastend voor de gezondheid en welzijn van paarden. Bovendien leidt de aandoening tot waardevermindering van het paard.

In dit proefschrift hebben we gebruik gemaakt genomewide SNPs om genoomregio's te identificeren die bijdragen aan de genetische variatie van SME-gevoeligheid. We onderzochten tevens het effect van verhoging van het aantal paarden en van SNP-dichtheid op de identificatie van betrokken genomische regio's.

In de studies I-III werden meerdere regio's gevonden met kleine effecten op SME-gevoeligheid. Voor IJslandse paarden, zoals geanalyseerd in studies I en II werden betrokken regio's gevonden op de chromosomen 1, 7, 10, 15 en 17. Het percentage van de genetische variantie verklaard door de top tien windows steeg van 3.07% (Studie I) tot 6.56% (Studie II). Het groter aantal onderzochte IJslandse paarden in Studie II resulteerde in de identificatie van meer genomische regio's die betrokken zijn bij SME-gevoeligheid. In studie III werd een high density SNPs in de analyse van Exmoor ponies gebruikt en werden nieuwe genomische regio's, op chromosoom 8 geïdentificeerd, alhoewel de significantie beperkt was.

De resultaten van onderzoek IV ondersteunen de vanuit de historie veronderstelde genetische relaties binnen de Nederlandse warmbloed populaties (paarsgewijs gemiddelde  $F_{ST} \leq 0.070$ ), en binnen de pony-achtige populaties (paarsgewijs gemiddelde  $F_{ST} \leq 0.078$ ). Paardenpopulaties met dezelfde genetische achtergrond kunnen daarmee ook dezelfde genetische componenten voor SME-gevoeligheid delen. Het Friese Paard had de laagste diversiteit (gemiddelde inteeltcoëfficiënten:  $f_i$ : 30.4%,  $f_{iROH} = 22.2\%$ ) in Study

IV en was genetisch het meest verschillend van de andere populaties ( $F_{ST}$  varieerde 0.13-0.17). Dit kan een gevolg zijn van een geschiedenis met meerdere genetische bottlenecks en van selectie binnen een gesloten fokprogramma. Lage diversiteit in immuun-gerelateerde genen, zoals waargenomen in het Friese Paard, zou mogelijk hebben geleid tot een grotere prevalentie van SME. De vermeende lage prevalentie van SME in warmbloedpaarden zou het gevolg kunnen zijn van de hoge genetische diversiteit (= -6.9%). Hoge genetische diversiteit in de KWPN-sportpaard kan het gevolg zijn van een open fokkerij en kruisen met andere warmbloed populaties.

## References

- Acevedo, N., Mercado, D., Vergara, C., Sánchez, J., Kennedy, M., Jiménez, S., Fernández, A., Gutiérrez, M., Puerta, L. & Caraballo, L. (2009). Association between total immunoglobulin E and antibody responses to naturally acquired *Ascaris lumbricoides* infection and polymorphisms of immune system related *LIG4*, *TNFSF13B* and *IRS2* genes. *Clinical Experimental Immunology*, 157, pp. 282-290.
- Acevedo-Whitehouse, K., Gulland, F., Greig, D. & Amos, W. (2003). Inbreeding: Disease susceptibility in California sea lions. *Nature*, 422(6927), p. 35.
- Aguilar, M.D.a., Ponce, S.I.R.a., Lopez, F.J.R., Padilla, E.G.a., Pelaez, C.G.V.a., Bagnato, A. & Strillacci, M.G. (2016). Genome-wide association study for milk somatic cell score in holstein cattle using copy number variation as markers. *Journal of Animal Breeding and Genetics*, pp. 1-11.
- Andersson, L.S., Larhammar, M., Memic, F., Wootz, H., Schwochow, D., Rubin, C.J., Patra, K., Arnason, T., Wellbring, L., Hjalms, G., Imsland, F., Petersen, J.L., McCue, M.E., Mickelson, J.R., Cothran, G., Ahituv, N., Roepstorff, L., Mikko, S., Vallstedt, A., Lindgren, G., Andersson, L. & Kullander, K. (2012a). Mutations in *DMRT3* affect locomotion in horses and spinal circuit function in mice. *Nature*, 488(7413), pp. 642-6.
- Andersson, L.S., Swinburne, J.E., Meadows, J.R., Brostrom, H., Eriksson, S., Fikse, W.F., Frey, R., Sundquist, M., Tseng, C.T., Mikko, S. & Lindgren, G. (2012b). The same ELA class II risk factors confer equine insect bite hypersensitivity in two distinct populations. *Immunogenetics*, 64(3), pp. 201-8.
- Aulchenko, Y.S., Ripke, S., Isaacs, A. & van Duijn, C.M. (2007). GenABEL: an R library for genome-wide association analysis. *Bioinformatics*, 23(10), pp. 1294-6.
- Ball, R.D. (2013). Designing a GWAS: power, sample size, and data structure. *Methods Mol Biol*, 1019, pp. 37-98.
- Baselgia, S., Doherr, M.G., Mellor, P., Torsteinsdottir, S., Jermann, T., Zurbriggen, A., Jungi, T. & Marti, E. (2006). Evaluation of an in vitro

- sulphidoleukotriene release test for diagnosis of insect bite hypersensitivity in horses. *Equine Vet J*, 38(1), pp. 40-6.
- Baye, T.M., Martin, L.J. & Khurana Hershey, G.K. (2010). Application of genetic/genomic approaches to allergic disorders. *J Allergy Clin Immunol*, 126(3), pp. 425-36; quiz 437-8.
- Beckmann, J.S., Estivill, X. & Antonarakis, S.E. (2007). Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. *Nat Rev Genet*, 8(8), pp. 639-46.
- Bermingham, M.L., Bishop, S.C., Woolliams, J.A., Pong-Wong, R., Allen, A.R., McBride, S.H., Ryder, J.J., Wright, D.M., Skuce, R.A., McDowell, S.W. & Glass, E.J. (2014). Genome-wide association study identifies novel loci associated with resistance to bovine tuberculosis. *Heredity (Edinb)*, 112(5), pp. 543-51.
- Bickhart, D.M., Hou, Y., Schroeder, S.G., Alkan, C., Cardone, M.F., Matukumalli, L.K., Song, J., Schnabel, R.D., Ventura, M., Taylor, J.F., Garcia, J.F., Van Tassell, C.P., Sonstegard, T.S., Eichler, E.E. & Liu, G.E. (2012). Copy number variation of individual cattle genomes using next-generation sequencing. *Genome research*, 22(4), pp. 778-90.
- 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, p. 3.
- Bonnelykke, K., Sparks, R., Waage, J. & Milner, J.D. (2015). Genetics of allergy and allergic sensitization: common variants, rare mutations. *Curr Opin Immunol*, 36, pp. 115-126.
- Brooks, S.A., Gabreski, N., Miller, D., Brisbin, A., Brown, H.E., Streeter, C., Mezey, J., Cook, D. & Antczak, D.F. (2010). Whole-genome SNP association in the horse: identification of a deletion in myosin Va responsible for Lavender Foal Syndrome. *PLoS Genet*, 6(4), p. e1000909.
- Brostrom, H., Larsson, A. & Troedsson, M. (1987). Allergic dermatitis (sweet itch) of Icelandic horses in Sweden: an epidemiological study. *Equine Vet J*, 19(3), pp. 229-36.
- Brown, S.J., Kroboth, K., Sandilands, A., Campbell, L.E., Pohler, E., Kezic, S., Cordell, H.J., McLean, W.H.I. & Irvine, A.D. (2012). Intragenic copy number variation within filaggrin contributes to the risk of atopic dermatitis with a dose-dependent effect. *J Invest Dermatol*, 132(1), pp. 98-104.
- Cardon, L.R. & Palmer, L.J. (2003). Population stratification and spurious allelic association. *Lancet*, 361(9357), pp. 598-604.
- Chamberlain, K.W. (1978). Clinical signs and diagnosis of atopic disease in the dog. *J Small Anim Pract*, 19(9), pp. 493-505.
- Cheng, Y., Sanderson, C., Jones, M. & Belov, K. (2012). Low MHC class II diversity in the Tasmanian devil (*Sarcophilus harrisii*). *Immunogenetics*, 64(7), pp. 525-33.
- Correa, T.G., Ferreira, J.M., Riet-Correa, G., Ruas, J.L., Schild, A.L., Riet-Correa, F., Guimaraes, A. & Felipe-Bauer, M.L. (2007). Seasonal allergic dermatitis in sheep in southern Brazil caused by *Culicoides insignis* (Diptera: Ceratopogonidae). *Vet Parasitol*, 145(1-2), pp. 181-5.



- Curik, I., Fraser, D., Eder, C., Achmann, R., Swinburne, J., Cramer, R., Brem, G., Solkner, J. & Marti, E. (2003). Association between the MHC gene region and variation of serum IgE levels against specific mould allergens in the horse. *Genetics, selection, evolution : GSE*, 35 Suppl 1, pp. S177-90.
- Devereux, G. & Seaton, A. (2005). Diet as a risk factor for atopy and asthma. *The Journal of allergy and clinical immunology*, 115(6), pp. 1109-17; quiz 1118.
- Douwes, J., Cheng, S., Travier, N., Cohet, C., Niesink, A., McKenzie, J., Cunningham, C., Le Gros, G., von Mutius, E. & Pearce, N. (2008). Farm exposure in utero may protect against asthma, hay fever and eczema. *Eur Respir J*, 32(3), pp. 603-11.
- Duffy, D.L., Martin, N.G., Battistutta, D., Hopper, J.L. & Mathews, J.D. (1990). Genetics of Asthma and Hay-Fever in Australian Twins. *Am Rev Respir Dis*, 142(6), pp. 1351-1358.
- Ekine, C.C., Rowe, S.J., Bishop, S.C. & Koning, D.J.d. (2013). Why breeding values estimated using familial data 1 should not be used for Genome-Wide Association Studies. *G3: Genes/Genomes/Genetics*.
- Enomoto, H., Noguchi, E., Iijima, S., Takahashi, T., Hayakawa, K., Ito, M., Kano, T., Aoki, T., Suzuki, Y., Koga, M., Tamari, M., Shiohara, T., Otsuka, F. & Arinami, T. (2007). Single nucleotide polymorphism-based genome-wide linkage analysis in Japanese atopic dermatitis families. *BMC Dermatol*, 7, p. 5.
- Ensembl (Equ Cab 2) *Gene: RASGRP1*  
[http://www.ensembl.org/Equus\\_caballus/Gene/Compare\\_Ortholog?db=core;g=ENSECAG00000010114;r=1:149706774-149775059;t=ENSECAT00000010916](http://www.ensembl.org/Equus_caballus/Gene/Compare_Ortholog?db=core;g=ENSECAG00000010114;r=1:149706774-149775059;t=ENSECAT00000010916) [17th December].
- Eriksson, S., Grandinson, K., Fikse, W.F., Lindberg, L., Mikko, S., Brostrom, H., Frey, R., Sundquist, M. & Lindgren, G. (2008). Genetic analysis of insect bite hypersensitivity (summer eczema) in Icelandic horses. *Animal*, 2(3), pp. 360-5.
- ExmoorPonySociety *Welcome to the world of Exmoor Ponies*.  
<http://www.exmoorponysociety.org.uk/> [19th October].
- Fadok, V.A. (2013). Update on equine allergies. *Vet Clin North Am Equine Pract*, 29(3), pp. 541-50.
- Fadok, V.A. & Greiner, E.C. (1990). Equine insect hypersensitivity: skin test and biopsy results correlated with clinical data. *Equine Vet J*, 22(4), pp. 236-40.
- Fernando, R.L. & Garrick, D. (2013). Bayesian methods applied to GWAS. *Methods Mol Biol*, 1019, pp. 237-74.
- Feuk, L., Marshall, C.R., Wintle, R.F. & Scherer, S.W. (2006). Structural variants: changing the landscape of chromosomes and design of disease studies. *Hum Mol Genet*, 15 Spec No 1, pp. R57-66.
- Finno, C.J. & Bannasch, D.L. (2014). Applied equine genetics. *Equine Vet J*, 46(5), pp. 538-44.
- Fox-Clipsham, L.Y., Carter, S.D., Goodhead, I., Hall, N., Knottenbelt, D.C., May, P.D., Ollier, W.E. & Swinburne, J.E. (2011). Identification of a mutation

- associated with fatal Foal Immunodeficiency Syndrome in the Fell and Dales pony. *PLoS Genet*, 7(7), p. e1002133.
- Galli, S.J., Tsai, M. & Piliponsky, A.M. (2008). The development of allergic inflammation. *Nature*, 454(7203), pp. 445-454.
- Garrick, D.J. & Fernando, R.L. (2013). Implementing a QTL detection study (GWAS) using genomic prediction methodology. *Methods in molecular biology (Clifton, N J)*, 1019, pp. 275-98.
- GeneCards *GeneCards: The Human Database*. [http://www.genecards.org/cgi-bin/carddisp.pl?gene=DCC&keywords=DCC,netrin,1,receptor,gene-aliases\\_descriptions](http://www.genecards.org/cgi-bin/carddisp.pl?gene=DCC&keywords=DCC,netrin,1,receptor,gene-aliases_descriptions).
- GeneCards *GeneCards: The Human Gene Database*. <http://www.genecards.org> [17th December].
- Georges, M., Nielsen, D., Mackinnon, M., Mishra, A., Okimoto, R., Pasquino, A.T., Sargeant, L.S., Sorensen, A., Steele, M.R., Zhao, X.Y., Womack, J.E. & Hoeschele, I. (1995). Mapping Quantitative Trait Loci Controlling Milk-Production in Dairy-Cattle by Exploiting Progeny Testing. *Genetics*, 139(2), pp. 907-920.
- Ghaemmaghami, A.M., Gough, L., Sewell, H.F. & Shakib, F. (2002). The proteolytic activity of the major dust mite allergen Der p 1 conditions dendritic cells to produce less interleukin-12: allergen-induced Th2 bias determined at the dendritic cell level. *Clinical and Experimental Allergy*, 32, pp. 1468-1475.
- Ghosh, S., Qu, Z., Das, P.J., Fang, E., Juras, R., Cothran, E.G., McDonell, S., Kenney, D.G., Lear, T.L., Adelson, D.L., Chowdhary, B.P. & Raudsepp, T. (2014). Copy number variation in the horse genome. *PLoS Genet*, 10(10), p. e1004712.
- Goddard, M.E. & Hayes, B.J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics*, 10(6), pp. 381-391.
- Gusella, J.F., Wexler, N.S., Conneally, P.M., Naylor, S.L., Anderson, M.A., Tanzi, R.E., Watkins, P.C., Ottina, K., Wallace, M.R., Sakaguchi, A.Y. & et al. (1983). A polymorphic DNA marker genetically linked to Huntington's disease. *Nature*, 306(5940), pp. 234-8.
- Gustafson, A.L., Tallmadge, R.L., Ramlachan, N., Miller, D., Bird, H., Antczak, D.F., Raudsepp, T., Chowdhary, B.P. & Skow, L.C. (2003). An ordered BAC contig map of the equine major histocompatibility complex. *Cytogenet Genome Res*, 102(1-4), pp. 189-95.
- Hedrick, P.W. (2011). *Genetics of populations*. 4th. ed. Sudbury, Mass.: Jones and Bartlett Publishers.
- Hellberg, W., Mellor, P.S., Torsteinsdottir, S. & Marti, E. (2009). Insect bite hypersensitivity in the horse: comparison of IgE-binding proteins in salivary gland extracts from *Simulium vittatum* and *Culicoides nubeculosus*. *Vet Immunol Immunopathol*, 132(1), pp. 62-7.
- Hendricks, B.L. (1995). *International Encyclopedia of Horse Breeds*. University of Oklahoma press.
- Hewitt, C.R.A., Brown, A.P., Hart, B.J. & Pritchard, D.I. (1995). A Major House-Dust Mite Allergen Disrupts the Immunoglobulin-E Network by

- Selectively Cleaving Cd23 - Innate Protection by Antiproteases. *J Exp Med*, 182(5), pp. 1537-1544.
- Hinz, D., Bauer, M., Roder, S., Olek, S., Huehn, J., Sack, U., Borte, M., Simon, J.C., Lehmann, I., Herberth, G. & group, L.s. (2012). Cord blood Tregs with stable FOXP3 expression are influenced by prenatal environment and associated with atopic dermatitis at the age of one year. *Allergy*, 67(3), pp. 380-9.
- Hirschhorn, J.N. & Daly, M.J. (2005). Genome-wide association studies for common diseases and complex traits. *Nature reviews Genetics*, 6(2), pp. 95-108.
- Hoglund, J.K., Sahana, G., Guldbrandtsen, B. & Lund, M.S. (2014). Validation of associations for female fertility traits in Nordic Holstein, Nordic Red and Jersey dairy cattle. *BMC Genet*, 15, p. 8.
- Horsegene *Insect Bite Hypersensitivity*.  
[http://www.bi.w.kuleuven.be/GENLOG/horsegene/insect\\_bite\\_hypersensitivity.html](http://www.bi.w.kuleuven.be/GENLOG/horsegene/insect_bite_hypersensitivity.html) [19th October].
- Hourihane, J.B., Kilburn, S., Nordlee, J., Hefle, S., Taylor, S. & Warner, J. (1997). An evaluation of the sensitivity of subjects with peanut allergy to very low doses of peanut protein: a randomized, double-blind, placebo-controlled food challenge study. *Journal of allergy clinical immunology*, 100(5), pp. 596-600.
- Immuno Polymorphism Database. <https://www.ebi.ac.uk/ipd/imgt/hla/stats.html> [1st October].
- Jeru, I., Cochet, E., Duquesnoy, P., Hentgen, V., Copin, B., Mitjavila-Garcia, M.T., Sheykholeslami, S., Le Borgne, G., Dastot-Le Moal, F., Malan, V., Karabina, S., Mahevas, M., Chantot-Bastaraud, S., Lecron, J.C., Faivre, L. & Amselem, S. (2014). Brief Report: Involvement of TNFRSF11A molecular defects in autoinflammatory disorders. *Arthritis Rheumatol*, 66(9), pp. 2621-7.
- Jonsdottir, S., Hamza, E., Janda, J., Rhyner, C., Meinke, A., Marti, E., Svansson, V. & Torsteinsdottir, S. (2015). Developing a preventive immunization approach against insect bite hypersensitivity using recombinant allergens: A pilot study. *Vet Immunol Immunop*, 166(1-2), pp. 8-21.
- Julier, C., Hyer, R.N., Davies, J., Merlin, F., Soularue, P., Briant, L., Cathelineau, G., Deschamps, I., Rotter, J.I., Froguel, P., Boitard, C., Bell, J.I. & Lathrop, G.M. (1991). Insulin-Igf2 Region on Chromosome-11p Encodes a Gene Implicated in Hla-Dr4-Dependent Diabetes Susceptibility. *Nature*, 354(6349), pp. 155-159.
- Jungerius, B. (2004). *Implementation of SNPs in pig genetics: LD and QTL analysis*. Diss. The Netherlands: Wageningen University.
- Kalbfleisch, T., Rebolledo-Mendez, J., Orlando, L. & MacLeod, J.N. (2014). Resources, and Progress Towards a Fully Annotated EquCab3. *Plant & Animal Genome XXII*. San Diego, CA, USA, p. W279.
- Karlsson, E.K., Sigurdsson, S., Ivansson, E., Thomas, R., Elvers, I., Wright, J., Howald, C., Tonomura, N., Perloski, M., Swofford, R., Biagi, T., Fryc, S., Anderson, N., Courtay-Cahen, C., Youell, L., Ricketts, S.L., Mandlebaum, S., Rivera, P., von Euler, H., Kisseberth, W.C., London,

- C.A., Lander, E.S., Couto, G., Comstock, K., Starkey, M.P., Modiano, J.F., Breen, M. & Lindblad-Toh, K. (2013). Genome-wide analyses implicate 33 loci in heritable dog osteosarcoma, including regulatory variants near CDKN2A/B. *Genome Biol*, 14(12), p. R132.
- Kinra, S., Davey Smith, G., Jeffreys, M., Gunnell, D., Galobardes, B. & McCarron, P. (2006). Association between sibship size and allergic diseases in the Glasgow Alumni Study. *Thorax*, 61(1), pp. 48-53.
- Kizilkaya, K., Fernando, R.L. & Garrick, D.J. (2010). Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. *J Anim Sci*, 88(2), pp. 544-51.
- Klug, W.S., Cummings, M.R., Spencer, C.A. & Palladino, M.A. (2009). *Concepts of Genetics*: Pearson International Edition.
- Klukowska-Rotzler, J., Marti, E., Lavoie, J.P., Ainsworth, D.M., Gerber, V., Zurbriggen, A. & Janda, J. (2012). Expression of thymic stromal lymphopoietin in equine recurrent airway obstruction. *Vet Immunol Immunopathol*, 146(1), pp. 46-52.
- Klumplerova, M., Vychodilova, L., Bobrova, O., Cvanova, M., Futas, J., Janova, E., Vyskocil, M., Vrtkova, I., Putnova, L., Dusek, L., Marti, E. & Horin, P. (2013). Major histocompatibility complex and other allergy-related candidate genes associated with insect bite hypersensitivity in Icelandic horses. *Mol Biol Rep*, 40(4), pp. 3333-40.
- Kondrashov, A.S. (1995). Contamination of the Genome by Very Slightly Deleterious Mutations: Why Have We Not Died 100 Times Over? *J. theor. Biol*, 175, pp. 583-594.
- Kruglyak, L. (1997). The use of a genetic map of biallelic markers in linkage studies. *Nat Genet*, 17(1), pp. 21-24.
- Langner, K.F., Darpel, K.E., Drolet, B.S., Fischer, A., Hampel, S., Heselhous, J.E., Mellor, P.S., Mertens, P.P. & Leibold, W. (2008). Comparison of cellular and humoral immunoassays for the assessment of summer eczema in horses. *Vet Immunol Immunopathol*, 122(1-2), pp. 126-37.
- Leroy, G., Phocas, F., Hedan, B., Verrier, E. & Rognon, X. (2015). Inbreeding impact on litter size and survival in selected canine breeds. *Vet J*, 203(1), pp. 74-8.
- Li, W.H., Ellsworth, D.L., Krushkal, J., Chang, B.H. & Hewett-Emmett, D. (1996). Rates of nucleotide substitution in primates and rodents and the generation-time effect hypothesis. *Mol Phylogenet Evol*, 5(1), pp. 182-7.
- Liang, Y., Jansen, M., Aronow, B., Geiger, H. & Van Zant, G. (2007). The quantitative trait gene latexin influences the size of the hematopoietic stem cell population in mice. *Nat Genet*, 39(2), pp. 178-88.
- Lopes, M.S., Silva, F.F., Harlizius, B., Duijvesteijn, N., Lopes, P.S., Guimaraes, S.E.F. & Knol, E.F. (2013). Improved estimation of inbreeding and kinship in pigs using optimized SNP panels. *Bmc Genet*, 14.
- Maher, B. (2008). Personal genomes: The case of the missing heritability. *Nature*, 456(7218), pp. 18-21.

- 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), pp. 113-7.
- Marti, E., Urwyler, A., Neuenschwander, M., Eicher, R., Meier, D., de Weck, A.L., Gerber, H., Lazary, S. & Dahinden, C.A. (1999). Sulfidoleukotriene generation from peripheral blood leukocytes of horses affected with insect bite dermal hypersensitivity. *Vet Immunol Immunopathol*, 71(3-4), pp. 307-20.
- McQuillan, R., Leutenegger, A.L., Abdel-Rahman, R., Franklin, C.S., Pericic, M., Barac-Lauc, L., Smolej-Narancic, N., Janicijevic, B., Polasek, O., Tenesa, A., MacLeod, A.K., Farrington, S.M., Rudan, P., Hayward, C., Vitart, V., Rudan, I., Wild, S.H., Dunlop, M.G., Wright, A.F., Campbell, H. & Wilson, J.F. (2008). Runs of homozygosity in European populations. *Am J Hum Genet*, 83(3), pp. 359-372.
- Meide, N.M.A.v.d., Savelkoul, H.F.J., Meulenbroeks, C., Ducro, B.J. & Tijhaar, E. (2014). Evaluation of a diagnostic ELISA for insect bite hypersensitivity in horses using recombinant Obsoletus complex allergens. *The Veterinary Journal*, 200, pp. 31-37.
- Mellor, P.S. & McCaig, J. (1974). The probable cause of "sweet itch" in England. *Veterinary Record*, 95(18), pp. 411-415.
- Metzger, J., Philipp, U., Lopes, M.S., da Camara Machado, A., Felicetti, M., Silvestrelli, M. & Distl, O. (2013). Analysis of copy number variants by three detection algorithms and their association with body size in horses. *BMC genomics*, 14, p. 487.
- Meulenbroeks, C. (2015a). *Equine insect bite hypersensitivity*. Diss. Utrecht, The Netherlands: Utrecht University.
- Meulenbroeks, C. (2015b). *Equine insect bite hypersensitivity: Pathogenesis, diagnosis and immunomodulation*. Diss.: Utrecht University.
- Mi, H.Y., Muruganujan, A., Casagrande, J.T. & Thomas, P.D. (2013). Large-scale gene function analysis with the PANTHER classification system. *Nat Protoc*, 8(8), pp. 1551-1566.
- Nagao, Y. (2015). Copy number variations play important roles in heredity of common diseases: a novel method to calculate heritability of a polymorphism. *Sci Rep*, 5, p. 17156.
- Nayeri, S., Sargolzaei, M., Abo-Ismael, M.K., May, N., Miller, S.P., Schenkel, F., Moore, S.S. & Stothard, P. (2016). Genome-wide association for milk production and female fertility traits in Canadian dairy Holstein cattle. *BMC Genet*, 17(1), p. 75.
- Oliehoek, P.A., Bijma, P. & van der Meijden, A. (2009). History and structure of the closed pedigreed population of Icelandic Sheepdogs. *Genetics, selection, evolution : GSE*, 41, p. 39.
- Olsen, L., Bondesson, U., Brostrom, H., Olsson, U., Mazogi, B., Sundqvist, M., Tjalve, H. & Ingvast-Larsson, C. (2011). Pharmacokinetics and effects of cetirizine in horses with insect bite hypersensitivity. *Vet J*, 187(3), pp. 347-51.
- Ott, J., Kamatani, Y. & Lathrop, M. (2011). Family-based designs for genome-wide association studies. *Nature Reviews Genetics*, 12(7), pp. 465-474.

- Owen, J.A., Punt, J., Stranford, S.A. & Jones, P.P. (2009). *Kuby Immunology*. United States of America: Susan Winslow.
- Peeters, L.M. (2013). *Genetic and immunological analysis of insect bite hypersensitivity in belgian warmblood horses*. Diss. Belgium: KU Leuven.
- Peeters, L.M., Janssens, S., Brebels, M. & Buys, N. (2015). Genetic parameters and estimated breeding values of insect bite hypersensitivity in Belgian Warmblood horses. *Vet J*, 206(3), pp. 420-2.
- Peeters, L.M., Janssens, S., Coussé, A. & Buys, N. (2014). Insect bite hypersensitivity in Belgian warmblood horses: prevalence and risk factors. *Vlaams Diergeneeskundig Tijdschrift*, 83, pp. 240-249.
- Peeters, L.M., Verlinden, T., Brebels, M., Buys, N. & Janssens, S. (2011). Environmental factors affecting the prevalence of insect bite hypersensitivity in Belgian warmblood horses in Vlaanderen. *Commun Agric Appl Biol Sci*, 76(1), pp. 205-9.
- Petersen, J.L., Mickelson, J.R., Cothran, E.G., Andersson, L.S., Axelsson, J., Bailey, E., Bannasch, D., Binns, M.M., Borges, A.S., Brama, P., da Camara Machado, A., Distl, O., Felicetti, M., Fox-Clipsham, L., Graves, K.T., Guerin, G., Haase, B., Hasegawa, T., Hemmann, K., Hill, E.W., Leeb, T., Lindgren, G., Lohi, H., Lopes, M.S., McGivney, B.A., Mikko, S., Orr, N., Penedo, M.C., Piercy, R.J., Raekallio, M., Rieder, S., Roed, K.H., Silvestrelli, M., Swinburne, J., Tozaki, T., Vaudin, M., C, M.W. & McCue, M.E. (2013). Genetic diversity in the modern horse illustrated from genome-wide SNP data. *PLoS One*, 8(1), p. e54997.
- Petit-Frere, C., Dugas, B., Braquet, P. & Mencia-Huerta, J.M. (1993). Interleukin-9 potentiates the interleukin-4-induced IgE and IgG1 release from murine B lymphocytes. *Immunology*, 79(1), pp. 146-51.
- Pieper, K., Grimbacher, B. & Eibel, H. (2013). B-cell biology and development. *J Allergy Clin Immunol*, 131(4), pp. 959-71.
- Pucheu-Haston, C.M. (2016). Atopic dermatitis in the domestic dog. *Clin Dermatol*, 34(2), pp. 299-303.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J. & Sham, P.C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, 81(3), pp. 559-575.
- Raat, I.J.d., Boom, R.v.d., Poppel, M.v. & Oldruitenborgh-Oosterbaan, M.M.v. (2008). The effect of a topical insecticide containing permethrin on the number of Culicoides midges caught near horses with and without insect bite hypersensitivity in the Netherlands. *Tijdschr Diergeneeskd*, 133(20), pp. 838-842.
- Rebhan, M., Chalifa-Caspi, V., Prilusky, J. & Lancet, D. (1998). GeneCards: a novel functional genomics compendium with automated data mining and query reformulation support. *Bioinformatics*, 14(8), pp. 656-64.
- Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., Cho, E.K., Dallaire, S., Freeman, J.L., Gonzalez, J.R., Gratacos, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J.R., Marshall, C.R., Mei, R., Montgomery,

- L., Nishimura, K., Okamura, K., Shen, F., Somerville, M.J., Tchinda, J., Valsesia, A., Woodward, C., Yang, F., Zhang, J., Zerjal, T., Zhang, J., Armengol, L., Conrad, D.F., Estivill, X., Tyler-Smith, C., Carter, N.P., Aburatani, H., Lee, C., Jones, K.W., Scherer, S.W. & Hurles, M.E. (2006). Global variation in copy number in the human genome. *Nature*, 444(7118), pp. 444-54.
- Reijmerink, N.E., Postma, D.S., Bruinenberg, M., Nolte, I.M., Meyers, D.A., Bleecker, E.R. & Koppelman, G.H. (2008). Association of IL1RL1, IL18R1, and IL18RAP gene cluster polymorphisms with asthma and atopy. *J Allergy Clin Immunol*, 122(3), pp. 651-4 e8.
- Rijt, R.v.d., Boom, R.v.d., Jongema, Y. & Oldruitenborgh-Oosterbaan, M.M.S.v. (2008). Culicoides species attracted to horses with and without insect hypersensitivity. *The Veterinary Journal*, 178(1), pp. 91-97.
- Risch, N. & Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science*, 273(5281), pp. 1516-1517.
- Robinson, M.R., Wray, N.R. & Visscher, P.M. (2014). Explaining additional genetic variation in complex traits. *Trends Genet*, 30(4), pp. 124-32.
- Sargolzaei, M., Iwaisaki, H. & Colleau, J. (2006). CFC, A tool for monitoring genetic diversity. *Proceedings of the 8th World Congress on Genetics Applied to Livestock Production*. Belo Horizonte, Brazil, pp. 27-28.
- Schaffartzik, A., Hamza, E., Janda, J., Crameri, R., Marti, E. & Rhyner, C. (2012). Equine insect bite hypersensitivity: what do we know? *Vet Immunol Immunop*, 147(3-4), pp. 113-26.
- Schaffartzik, A., Weichel, M., Crameri, R., Bjornsdottir, T.S., Prisi, C., Rhyner, C., Torsteinsdottir, S. & Marti, E. (2009). Cloning of IgE-binding proteins from *Simulium vittatum* and their potential significance as allergens for equine insect bite hypersensitivity. *Vet Immunol Immunopathol*, 132(1), pp. 68-77.
- Schneider, J.F., Miles, J.R., Brown-Brandl, T.M., Nienaber, J.A., Rohrer, G.A. & Vallet, J.L. (2015). Genomewide association analysis for average birth interval and stillbirth in swine. *J Anim Sci*, 93(2), pp. 529-40.
- Schurink, A. (2012). *Insect bite hypersensitivity in horses: genetic and epidemiological analysis*. Diss. Wageningen, The Netherlands: Wageningen University.
- Schurink, A., Ducro, B.J., Bastiaansen, J.W.M., Frankena, K. & van Arendonk, J.A.M. (2013a). Genome-wide association study of insect bite hypersensitivity in Dutch Shetland pony mares. *Anim Genet*, 44(1), pp. 44-52.
- 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), pp. 1286-93.
- Schurink, A., Podesta, S.C., Ducro, B.J., van Arendonk, J.A. & Frankena, K. (2013b). Risk factors for insect bite hypersensitivity in Friesian horses and Shetland ponies in The Netherlands. *Vet J*, 195(3), pp. 382-4.
- Schurink, A., van Grevenhof, E.M., Ducro, B.J. & van Arendonk, J.A.M. (2009). Heritability and repeatability of insect bite hypersensitivity in Dutch Shetland breeding mares. *J Anim Sci*, 87(2), pp. 484-90.

- Schurink, A., Wolc, A., Ducro, B.J., Frankena, K., Garrick, D.J., Dekkers, J.C.M. & van Arendonk, J.A.M. (2012). Genome-wide association study of insect bite hypersensitivity in two horse populations in the Netherlands. *Genet Sel Evol*, 44.
- Siddle, H.V., Kreiss, A., Eldridge, M.D.B., Noonan, E., Clarke, C.J., Pyecroft, S., Woods, G.M. & Belov, K. (2007). Transmission of a fatal clonal tumor by biting occurs due to depleted MHC diversity in a threatened carnivorous marsupial. *Proc Natl Acad Sci U S A*, 104(41), pp. 16221-6.
- Skol, A.D., Sasaki, M.M. & Onel, K. (2016). The genetics of breast cancer risk in the post-genome era: thoughts on study design to move past BRCA and towards clinical relevance. *Breast Cancer Res*, 18(1), p. 99.
- Spatuzza, C., Schiavone, M., Di Salle, E., Janda, E., Sardiello, M., Fiume, G., Fierro, O., Simonetta, M., Argiriou, N., Faraonio, R., Capparelli, R., Quinto, I. & Scala, G. (2008). Physical and functional characterization of the genetic locus of IBtk, an inhibitor of Bruton's tyrosine kinase: evidence for three protein isoforms of IBtk. *Nucleic Acids Res*, 36(13), pp. 4402-16.
- Spielman, R.S., McGinnis, R.E. & Ewens, W.J. (1993). Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet*, 52(3), pp. 506-16.
- Spriggs, M. & Reeder, C. (2012). Treatment of vasculitis and dermatitis in a 59-yr-old Nile hippopotamus (*Hippopotamus amphibius*). *J Zoo Wildl Med*, 43(3), pp. 652-6.
- Stock, K.F., Ricard, A., Jönsson, L. & Mark, T. (2016). Genomic applications in horse breeding. *Animal Frontiers*, 6(1), pp. 45-52.
- Swinburne, J.E., Bogle, H., Klukowska-Rotzler, J., Drogemuller, M., Leeb, T., Temperton, E., Dolf, G. & Gerber, V. (2009). A whole-genome scan for recurrent airway obstruction in Warmblood sport horses indicates two positional candidate regions. *Mamm Genome*, 20(8), pp. 504-15.
- Szumilas, M. (2010). Explaining odds ratios. *J Can Acad Child Adolesc Psychiatry*, 19(3), pp. 227-9.
- Tallmadge, R.L., Shen, L., Tseng, C.T., Miller, S.C., Barry, J. & Felipe, M.J. (2015). Bone marrow transcriptome and epigenome profiles of equine common variable immunodeficiency patients unveil block of B lymphocyte differentiation. *Clin Immunol*, 160(2), pp. 261-76.
- Tamari, M., Tanaka, S. & Hirota, T. (2013). Genome-wide association studies of allergic diseases. *Allergol Int*, 62(1), pp. 21-8.
- Thompson, J.R., Everett, R.W. & Hammerschmidt, N.L. (2000). Effects of inbreeding on production and survival in Holsteins. *J Dairy Sci*, 83(8), pp. 1856-64.
- Thorsteinsdottir, S., Thyssen, J.P., Stokholm, J., Vissing, N.H., Waage, J. & Bisgaard, H. (2016). Domestic Dog Exposure at birth reduces the Incidence of Atopic Dermatitis. *Allergy*.
- Traherne, J.A. (2008). Human MHC architecture and evolution: implications for disease association studies. *Int J Immunogenet*, 35(3), pp. 179-92.
- Traidl-Hoffmann, C., Jakob, T. & Behrendt, H. (2009). Determinants of allergenicity. *J Allergy Clin Immunol*, 123(3), pp. 558-66.



- Tsui, L.C., Buchwald, M., Barker, D., Braman, J.C., Knowlton, R., Schumm, J.W., Eiberg, H., Mohr, J., Kennedy, D., Plavsic, N., Zsiga, M., Markiewicz, D., Akots, G., Brown, V., Helms, C., Gravius, T., Parker, C., Rediker, K. & Doniskeller, H. (1985). Cystic-Fibrosis Locus Defined by a Genetically Linked Polymorphic DNA Marker. *Science*, 230(4729), pp. 1054-1057.
- Uekert, S.J., Akan, G., Evans, M.D., Li, Z., Roberg, K., Tisler, C., Dasilva, D., Anderson, E., Gangnon, R., Allen, D.B., Gern, J.E. & Lemanske, R.F., Jr. (2006). Sex-related differences in immune development and the expression of atopy in early childhood. *The Journal of allergy and clinical immunology*, 118(6), pp. 1375-81.
- Ullemar, V., Magnusson, P.K.E., Lundholm, C., Zettergren, A., Melén, E., Lichtenstein, P. & Almqvist, C. (2016). Heritability and confirmation of genetic association studies for childhood asthma in twins. *Allergy*, 71, pp. 230-238.
- van den Boom, R., Ducro, B. & Sloet van Oldruitenborgh-Oosterbaan, M.M. (2008). Identification of factors associated with the development of insect bite hypersensitivity in horses in The Netherlands. *Tijdschr Diergeneeskd*, 133(13), pp. 554-9.
- van der Meide, N.M., Roders, N., Sloet van Oldruitenborgh-Oosterbaan, M.M., Schaap, P.J., van Oers, M.M., Leibold, W., Savelkoul, H.F. & Tijhaar, E. (2013). Cloning and expression of candidate allergens from *Culicoides obsoletus* for diagnosis of insect bite hypersensitivity in horses. *Vet Immunol Immunopathol*, 153(3-4), pp. 227-39.
- van der Meide, N.M., Savelkoul, H.F., Meulenbroeks, C., Ducro, B.J. & Tijhaar, E. (2014). Evaluation of a diagnostic ELISA for insect bite hypersensitivity in horses using recombinant *Obsoletus* complex allergens. *Vet J*, 200(1), pp. 31-7.
- Visscher, P.M., Brown, M.A., McCarthy, M.I. & Yang, J. (2012). Five years of GWAS discovery. *Am J Hum Genet*, 90(1), pp. 7-24.
- Visscher, P.M., Hill, W.G. & Wray, N.R. (2008). Heritability in the genomics era--concepts and misconceptions. *Nat Rev Genet*, 9(4), pp. 255-66.
- Vychodilova, L., Matiasovic, J., Bobrova, O., Futas, J., Klumplerova, M., Stejskalova, K., Cvanova, M., Janova, E., Osickova, J., Vyskocil, M., Sedlinska, M., Dusek, L., Marti, E. & Horin, P. (2013). Immunogenomic analysis of insect bite hypersensitivity in a model horse population. *Vet Immunol Immunop*, 152(3-4), pp. 260-268.
- Walsh, T. & King, M.C. (2007). Ten genes for inherited breast cancer. *Cancer Cell*, 11(2), pp. 103-105.
- Weidinger, S., Willis-Owen, S.A., Kamatani, Y., Baurecht, H., Morar, N., Liang, L., Edser, P., Street, T., Rodriguez, E., O'Regan, G.M., Beattie, P., Folster-Holst, R., Franke, A., Novak, N., Fahy, C.M., Winge, M.C., Kabesch, M., Illig, T., Heath, S., Soderhall, C., Melen, E., Pershagen, G., Kere, J., Bradley, M., Lieden, A., Nordenskjold, M., Harper, J.I., McLean, W.H., Brown, S.J., Cookson, W.O., Lathrop, G.M., Irvine, A.D. & Moffatt, M.F. (2013). A genome-wide association study of atopic dermatitis identifies loci with overlapping effects on asthma and psoriasis. *Hum Mol Genet*, 22(23), pp. 4841-56.

- Weir, B.S. & Cockerham, C.C. (1984). Estimating F-Statistics for the Analysis of Population-Structure. *Evolution*, 38(6), pp. 1358-1370.
- Wilson, A.D. (2014). Immune responses to ectoparasites of horses, with a focus on insect bite hypersensitivity. *Parasite Immunol*, 36(11), pp. 560-72.
- Wolc, A., Arango, J., Settar, P., Fulton, J.E., O'Sullivan, N.P., Preisinger, R., Habier, D., Fernando, R., Garrick, D.J., Hill, W.G. & Dekkers, J.C. (2012). Genome-wide association analysis and genetic architecture of egg weight and egg uniformity in layer chickens. *Anim Genet*, 43 Suppl 1, pp. 87-96.
- Xu, S. (2003). Theoretical basis of the Beavis effect. *Genetics*, 165(4), pp. 2259-68.
- Yang, F., Vought, B.W., Satterlee, J.S., Walker, A.K., Jim Sun, Z.Y., Watts, J.L., DeBeaumont, R., Saito, R.M., Hyberts, S.G., Yang, S., Macol, C., Iyer, L., Tjian, R., van den Heuvel, S., Hart, A.C., Wagner, G. & Naar, A.M. (2006). An ARC/Mediator subunit required for SREBP control of cholesterol and lipid homeostasis. *Nature*, 442(7103), pp. 700-4.
- Yeruham, I., Braverman, Y. & Orgad, U. (1993). Field observations in Israel on hypersensitivity in cattle, sheep and donkeys caused by Culicoides. *Aust Vet J*, 70(9), pp. 348-52.

## Acknowledgements

It's not an end result but the journey to reach to the end matters. My journey of more than four years has been filled with happiness, laughter, sad moments, moments where I felt lost, a mixture of all the emotions possible. I met many people throughout this journey. I would like to sincerely thank you all who shared this journey with me and made it memorable. I learnt a lot personally and professionally throughout this period.

Firstly, I would like to thank my past and current supervisors in Sweden and the Netherlands: Anna Maria Johansson, Anouk Schurink, Bart Ducro, Susanne Eriksson, Tomas Bergström, and Gabriella Lindgren. I appreciate all of your contribution so that I could reach to this point today. Without your expertise, suggestions, motivations, critical analyses, I doubt if I would have made this far. **Susanne**, you have been like a guardian to me. You supported me through troubling times and helped me to overcome many obstacles during my PhD study. I am really grateful. **Anouk**, thanks for being my friend in addition to being my supervisor. I really like your idea of sitting down with me, and discussing your comments and suggestions. It helped me a lot and it's something I would like to apply myself. **Anna**, I came to you as a person with a baggage and you sheltered me by accepting to supervise me. I want to thank you for your support especially during the last moment of the study. **Tomas**, I always enjoy speaking with you. You helped to hone my knowledge in laboratory skills during my masters. I enjoyed the time listening to music and working in lab with you. **Bart**, I am grateful towards your insightful comments and critical questions on the manuscripts. Your expertise in writing is the one that I hope I will be able to attain someday. **Gabriella**, I would like to thank you for giving me the opportunity to work as your PhD student. You introduced me towards the field of horse genetics and it's something that is valuable to me.

I would like to acknowledge **EGS-ABG program** for giving me opportunity to pursue my doctoral degree and experience academic environment of two universities, Swedish University of Agricultural Sciences (SLU) and Wageningen University (WU). I would like to thank all **PhD students and staff in Department of Animal Breeding and Genetics, SLU**. I have enjoyed taking FIKA with you all. Celebrating special occasions with sweet treats makes it even more special. **Cano**, thanks for all the laughter. I thing your charming personality brightens everyone's day. Thanks to my friends in Sweden: **Agnese, Andre, Berihu, Bingjie, Chrissy, Durga, Josh, Juan, Kim, Lin, Nancy, Sandrine, Sangeet, Suzana and Thu**. You all made me feel like home even though I was far away from home. Karaoke nights, I will always remember.

**Agnese**, we made a long journey from our first day of master studies till here now as PhD students. I still remember our first moment in Norway when Tormod wrote an equation of Henderson's mixed model on board. We looked at each other's face with no clue. You made me adventurous (a bit), and took me out of my comfort zone. **Chrissy**, thanks a lot for tasty cookies, you know how much I love them. **Sandrine**, thanks for your support during tough writing period, I will remember to relax my little body. **Thu**, thanks for being there as a friend, girl. Things happen when you expect it the least. **Juan**, process of understanding you has made me understand myself better. Thanks for your support during tough times and teaching me to express myself in a healthy way.

I would also like to thank all **PhD students and staff in Animal Breeding and Genomics Centre, WU**. Dear friends at wageningen **Amabel, Coralia, Claudia, Mathijs, Zih-Hua**, and **Pamela**, your company made my stay memorable. I didn't know what to expect on my first arrival to Wageningen but it became smooth ride with you guys around.

Last but not the least, my family back at home, Maa and Dad, without your guidance, love and support I don't think I will be who I am. Thanks ☺

Merina Shrestha  
Uppsala, December 2016

## *Training and Education*



<b>The Basic Package (8 ECTS)</b>	<b>Year</b>
Welcome to EGS-ABG	2012
Fall Research school “Animal Breeding and Society” EGS-ABG	2013
Research Ethics (POG0054)	2014
Fall Research School "Animal breeding and society Sustainability concept in animal breeding" EGS-ABG	2015
<b>Scientific exposure (10.4 ECTS)</b>	
<b><i>International conferences (4.8 ECTS)</i></b>	
10 <sup>th</sup> Dorothy Russel Havemeyer Horse Genome Workshop, Azores, Portugal	2013
EAAP-European association of animal production, Nantes, France	2013
The Biology of Genomes, New York, USA	2014
11 <sup>th</sup> Dorothy Russel Havemeyer Horse Genome Workshop, Hanover, Germany	2015
Plant and Animal Genome (PAG) Conference XXIV, San Diego, USA	2016
<b><i>Seminars and workshop (0.6 ECTS)</i></b>	
One-year seminar at SLU, Uppsala, Sweden	2014
Half-time seminar at SLU, Uppsala, Sweden	2015
<b><i>Presentations (5 ECTS)</i></b>	
Poster Presentation in 10 <sup>th</sup> Havemeyer Conference, Azores, Portugal	2013
Oral Presentation in EAAP, Nantes, France	2013
Poster Presentation in The Biology of Genomes, New York, USA	2014
Poster Presentation in 11 <sup>th</sup> Havemeyer Conference, Hanover, Germany	2015
Poster Presentation in PAG Conference XXIV, San Diego, USA	2016

**In-Depth Studies (22 ECTS)**

---

***Disciplinary and interdisciplinary courses (22 ECTS)***

Domestic Animal Genetic Resources, Uppsala, Sweden	2012
Statistical methods in Animal Breeding, Uppsala, Sweden	2012
Basic Perl Programming for the Biological Sciences, Uppsala, Sweden	2012
Modern Statistics for the Life Sciences (ABG-30806), Wageningen, The Netherlands	2013
Statistical Methods in Genetic Mapping- R based course, Uppsala, Sweden	2013
Introduction to Genome Browsers, Stockholm, Sweden	2014
Linear Models in Animal Breeding (NOVA course), Lofoten, Norway	2015

---

**Professional Skills Support Courses (13.5 ECTS)**

Introduction to SAS software for statistical analysis (SLU-P0044), Uppsala, Sweden	2012
Seminar series at SLU, Uppsala, Sweden	2012-2016
How to write and publish a scientific paper (PVG0027), Uppsala, Sweden	2014
Genetic analysis using ASReml4.0, Wageningen, The Netherlands	2014
Introduction to Bioinformatics using Next Generation Sequencing data, Uppsala, Sweden	2014

---

**Didactic Skills Training (7.2 ECTS)*****Lecture***

Lectures on course VM0048 (Basic Biomedicine- “Biomedicinsk Baskurs”)	2012-2013
Lab supervisions and group discussions on “Genome Analysis” course	2013

***Supervising projects***

Supervising 2 short projects, Uppsala, Sweden	2013- 2014
Supervising 1 MSc thesis	2015

<b>Total</b>	<b>61.1 ECTS</b>
--------------	------------------