UPPSALA
UNIVERSITET

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 736

Evolution of MHC Genes and MHC Gene Expression

KARIN BERGGREN BREMDAL

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2010

ISSN 1651-6214
ISBN 978-91-554-7792-9
urn:nbn:se:uu:diva-122011

Dissertation presented at Uppsala University to be publicly examined in Evolutionsbiologiskt centrum, Zootissalen, Villavägen 9, Uppsala, Friday, May 21, 2010 at 09:00 for the degree of Doctor of Philosophy. The examination will be conducted in English.

Abstract
Berggren Bremdal, K. 2010. Evolution of MHC Genes and MHC Gene Expression. Acta Universitatis Upsaliensis. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology* 736. 69 pp. Uppsala. ISBN 978-91-554-7792-9.

Polymorphism in coding regions and regions controlling gene expression is the major determinant of adaptive differences in natural populations. Genes of the major histocompatibility complex (MHC) possess a high level of genetic variation, which is maintained by selection over long coalescence times. MHC genes encode antigen-presenting molecules in the adaptive immune system, which protects the host from infectious diseases. However, MHC molecules may also present self-peptides and for most autoimmune diseases there is a genetic factor associated with the MHC.

MHC genes have been used to learn about the interplay of selection and historical population events. In domestic dogs and their progenitor, the wolf, I explored factors associated with domestication and breed formation and their influence not only on MHC coding regions but also on the haplotypic structure of the class II region. Polymorphism and strong selection was demonstrated in the proximal promoters of MHC genes in dogs and wolves. Hence, genetic variation associated with MHC gene expression may have at least equal importance for a well functioning immune system. Associations between promoter sequences and particular coding alleles suggested allele-specific expression patterns. SNP haplotypes of the MHC class II region revealed ancestral as well as convergent haplotypes, in which combinations of alleles are kept by selection. Interestingly, weaker allelic associations were found between different genes and between coding regions and promoters in dogs compared to wolves. Potentially, this could cause insufficient defense against infections and predispose dogs to autoimmune diseases. For example, I identified a site in the promoter region that showed a consistent difference between haplotypes conferring susceptibility and protection to diabetes in dogs, which should be investigated further.

Furthermore, I investigated how selection and demographic changes associated with glacial and inter-glacial periods have affected MHC variation in European hedgehogs and extended the prevailing knowledge concerning their population history.

Keywords: major histocompatibilty complex, dog leukocyte antigen, balancing selection, linkage disequilibrium, promoter, diabetes mellitus, *Canis familiaris*, *Canis lupus*, *Erinaceus europaeus*, *Erinaceus concolor*

Karin Berggren Bremdal, Department of Evolution, Genomics and Systematics, Evolutionary Biology, Norbyv. 18C, Uppsala University, SE-75236 Uppsala, Sweden

ISSN 1651-6214
ISBN 978-91-554-7792-9
urn:nbn:se:uu:diva-122011 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-122011)

List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I Berggren, K.T., Ellegren, H., Hewitt, G.M., Seddon, J.M. (2005) Understanding the phylogeographic patterns of European hedgehogs, *Erinaceus concolor* and *E. europaeus* using the MHC. *Heredity*, 95(1):84–90

II Berggren, K.T., Seddon, J.M. (2005) MHC promoter polymorphism in grey wolves and domestic dogs. *Immunogenetics*, 57(3-4):267-272

III Berggren, K.T., Seddon, J.M. (2008) Allelic combinations of promoter and exon 2 in *DQB1* in dogs and wolves. *Journal of Molecular Evolution*, 67(1):76-84

IV Seddon, J.M., Berggren, K.T., Fleeman, L.M. (2010) Evolutionary history of DLA class II haplotypes in canine diabetes mellitus through single nucleotide polymorphism genotyping. *Tissue Antigens*, 75 (3): 218-226

V Berggren, K.T., Seddon, J.M. (2010) Linkage disequilibrium and haplotype patterns of the MHC class II region - a comparison between wolves and dogs. Manuscript

Contents

Abbreviations

A	Adenine
Aids	Acquired immune deficiency syndrome
Ag	Antigen
APC	Antigen presenting cell
B cell	B lymphocyte
BoLA	Bovine lymphocyte antigen
bp	Base pair(s)
C	Cytosine
CIITA	MHC class II transactivator
CREB	cAMP response element binding
D'	Measure of LD
d_N	Nonsynonymous substitution rate
DLA	Dog leukocyte antigen
DNA	Deoxyribonucleic acid
d_S	Synonymous substitution rate
EHH	Extended haplotype homozygosity
G	Guanine
HIV	Human immunodeficiency virus
HLA	Human leukocyte antigen
kb	Kilo base pair(s) (10^3 base pairs)
LD	Linkage disequilibrium
MHC	Major histocompatibility complex

Mb	Mega base pair(s) (10^6 base pairs)
NF-Y	Nuclear transcription factor Y
mtDNA	Mitochondrial DNA
Ne	Effective population size
r^2	Measure of LD
RFX	Regulatory factor X gene family
SNP	Single nucleotide polymorphism
SSCP	Single-strand conformation polymorphism
T	Thymine
T cell	T lymphocyte
TCR	T cell receptor
TH1	T helper cell type 1
TH2	T helper cell type 2
PBR	Peptide binding region
PCR	Polymerase chain reaction

Introduction

MHC structure and function

In the early days of immunological science, the major histocompatibility complex (MHC) molecule was recognized (and named) because of its role in the rejection of organs and tissues. It was discovered that transplantation of skin between genetically similar mice was successful while transplantation of skin between divergent mice strains resulted in rejection of the transplanted tissue (Snell 1948). Not until much later was the role of MHC as an intrinsic component of the immune system elucidated.

The role of MHC in the immune system

There is a constant battle, a kind of arms race, between the immune system and the foreign substances that regularly invade our bodies. The foreign substances may be bacteria, viruses or larger parasites, all referred to below as pathogens. The pathogens use an array of strategies to avoid the immune system of the host. The immune system tries to respond with a well-adjusted and often specific response with a mission to destroy the invader. Genetic variation is crucial for the immune system's ability to recognize and respond to the array of pathogens and their avoidance strategies (Frank 2002).

The immune system consists of a large number of components, which integrate in a complex system to protect the host. The full complexity of the immune system is explained in e.g. Abbas et al (2000) and the parts of special interest for this thesis are also reviewed in e.g. Meyer and Thomson (2001) and Piertney and Oliver (2006) but are briefly described here. Vertebrates depend on two types of immune defense. Firstly, the innate immune response, in which white blood cells, such as macrophages and neutrophils, secrete proteins that will destroy pathogens by phagocytosis, affords non-specific defense. Secondly, the adaptive immune system has the capacity to recognize specific elements and thereby respond with a defense that is adapted for that particular pathogen. If the adaptive immune system has recognized and responded against a pathogen, it will create memory cells with the capability to 'remember' the pathogen. If the host is once again exposed to the same or a very similar pathogen, the immune response will be even more efficient. The adaptive immune system is derived and not something with which we are born.

Central to the adaptive immune system are white blood cells called T lymphocytes (T cells) and B lymphocytes (B cells). Cytotoxic (CD8+) T cells conduct battles against intra-cellular pathogens, such as viruses. However cytotoxic T cells rely on co-stimulatory signals from another kind of T cell, CD4+ T helper cells, upon activation. This is called a cellular response or TH1 response (as T helper cell type 1 conduct the defense). B cells produce antibodies that act to combat extra-cellular pathogens. They also rely on co-stimulatory signals from T helper cells upon activation in the humoral response or TH2 (T helper cell type 2).

All T cells have T cell receptors (TCRs), which bind fragments of pathogens on the cell surface. Proteins that have been brought into the cell are broken into peptides and transported back to the cell surface where they are displayed by MHC molecules and are bound by TCR (Zinkernagel and Doherty 1974). Peptides that are presented by MHC molecules, recognized by the receptors and which trigger an immune response are often referred to as antigens and the part of the antigen, which physically binds to the receptor, is referred to as an epitope.

There are two types of MHC molecule, class I and class II. Cytotoxic T cells bind to epitopes if presented by MHC class I and T helper cells respond to antigen presentation by MHC class II molecules (*Figure 1*). MHC molecules therefore constitute a central role in the specific immune system. As my research has been conducted entirely on MHC class II, the focus from now on will be on that molecule and the genes encoding it.

The MHC molecule

The MHC class II molecule consists of two glycoprotein chains, the α-chain and the β-chain, which form a heterodimeric structure. Each chain consists of two domains, α1 + α2 and β1 + β2. The two outer domains, α1 and β1, interact to form a cleft in which peptides are bound for presentation to T helper cells. The cleft consists of a β–sheet surrounded by two α–helix structures. The α2 and β2 domains connect to segments with transmembrane residues followed by a cytoplasmic tail. The segment with transmembrane residues anchors the MHC molecule to the cell membrane of the antigen presenting cell. The molecular structure of MHC molecules is described in more detail in e.g. Madden (1995) and Hughes and Yeager (1998). *Figure 2* shows a simplified picture of the MHC class II molecule. While MHC class I molecules are expressed on all nucleated cells, MHC class II molecules are constantly expressed only on certain antigen presenting cells (APCs), such as macrophages, dendritic cells and B cells. However, MHC class II can also be expressed elsewhere through stimulation with certain cytokines such as interferon-γ and interleukin-4 (Glimcher and Kara 1992; Ting and Trowsdale 2002).

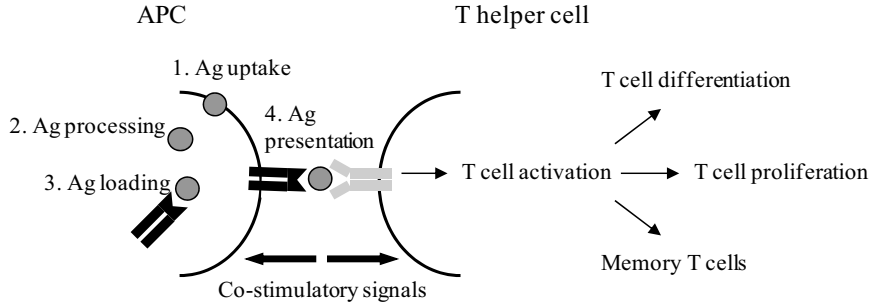


Figure 1. Antigen (Ag) presentation by MHC class II (black receptor) on the cell surface of antigen presenting cell (APC). The T-cell receptor (gray receptor) of a CD4+ T helper cell recognizes the Ag when presented by MHC class II and hence is activated.

Genetic structure of MHC

The MHC region comprises a large segment of DNA, in humans extending approximately 4 megabases (Mb). Within the MHC region we find genes of MHC class I and MHC class II but also genes with other immunological functions and non-immunological functions (Beck et al. 1999). Many of the sequences are pseudogenes. The arrangement of MHC genes seems conserved between eutherian mammals but also with other vertebrates although there are some organization differences such as intron length and MHC loci copy number (Kelley et al. 2005; Trowsdale 1995). The terminology also differs. The MHC region in humans is often referred to as the human leukocyte antigen (HLA), in mice it is called the H2 complex, in cattle bovine lymphocyte antigen (BoLA) and in dogs dog leukocyte antigen (DLA) to give some examples.

Not all genes within the MHC class I and class II region are encoding antigen presenting MHC molecules. Those that do are often referred to as the classical MHC genes. In most mammals the classical genes of MHC class II are encoded at the *DP*, *DQ* and *DR* loci. The α-chain and the β-chain of MHC class II molecules are encoded by separate genes; the A gene encodes the α-chain and the B gene encodes the β-chain. Hence there is an A gene and a B gene for each locus. Within each gene there are five to six different exons with interspersed intron sequences. Exon 2 encodes the main part of the α1 and β1 domains. As described above, these domains are involved in peptide binding. In many species there are also several copies of each gene, i.e. *DRB1* and *DRB2* etc, and, although most are expressed, some copies are pseudogenes. For example in humans multiple expressed as well as unexpressed copies of the *DRB* locus have been identified and for *DQA* and *DQB* single expressed genes are found together with pseudogene copies

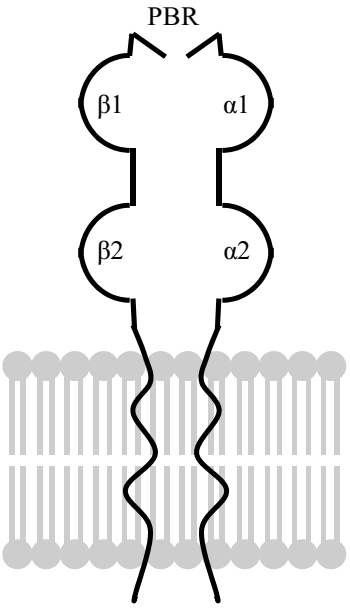


Figure 2. Schematic drawing of a MHC class II molecule. The α-chain and the β-chain make up a heterodimeric molecule. The two outer domains, α1 and β1, form a cleft in which peptides are bound. A transmembrane region anchors the molecule to the cell membrane of an antigen presenting cell.

(Beck et al. 1999). In dogs there are single functional genes for *DRB*, *DQA* and *DQB* (Wagner 2003) but an incomplete copy of the *DRB* gene has been reported in some dogs (Wagner 2003; Wagner et al. 1996b) as well as an incomplete *DQB* gene copy (Wagner et al. 1998). A simplified picture of the MHC class II region is shown in *Figure 3*. MHC genomic organization is reviewed in e.g. Meyer and Thomson (2001) and Piertney and Oliver (2006).

A single MHC molecule can bind only a limited number of different peptides, determined by the amino acid residues in certain regions of the peptide binding cleft. These regions of the molecule are often referred to as the peptide binding regions (PBR). The nucleotides encoding these regions have been found to be very variable in some genes, particularly in comparison with surrounding nucleotides (e.g. Hedrick et al. 1991; Hughes and Nei 1988; Hughes and Nei 1989; Parham et al. 1988). Nucleotide sites involved in peptide binding also have higher heterozygosity than surrounding nucleotides (Hedrick et al. 1991). The genes *DQA*, *DQB* and *DRB* show high levels of polymorphism in most mammals, *DRB* generally being the most variable and *DQA* generally the least variable. For example, in humans 878 different *DRB* alleles, 108 different *DQB* alleles and 35 different *DQA* alleles have been reported (IMGT, the international ImMunoGeneTics information system, April 2010, http://hla.alleles.org/nomenclature/stats.html). This makes

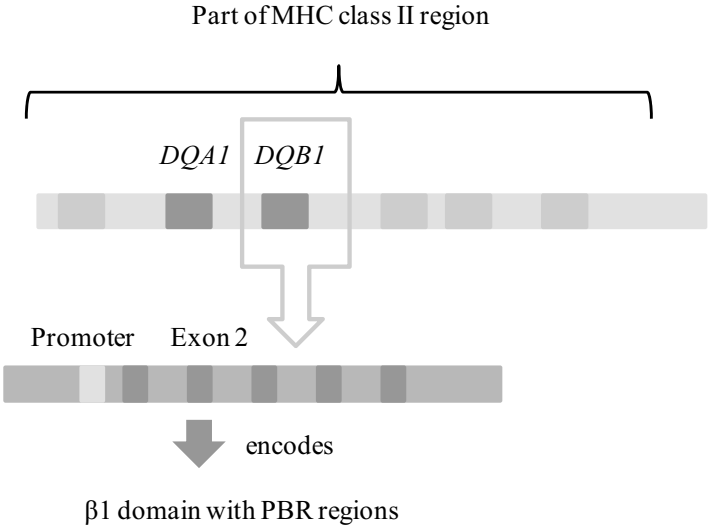


Figure 3. Schematic picture of a small part of the MHC class II region that covers the expressed *DQA1*and *DQB1* genes. Each gene consists of several exons with interspersed intron sequences. Exon 2 encodes the part of the MHC molecule that binds and presents antigens to T cells. The promoter region is a cis-acting regulatory element controlling the expression of the gene.

MHC genes among the most variable in the whole genome (Meyer and Thomson 2001; Piertney and Oliver 2006).

Certain MHC alleles at different loci are inherited together (as MHC haplotypes) more frequently than what would be expected at random. This phenomenon is called linkage disequilibrium (LD). Among diverged MHC class II haplotypes in humans, recombination events have been rare compared to expectations from the genome wide recombination rate (Raymond et al. 2005). Raymond et al. (2005) concluded that the most divergent class II haplotypes in humans have been evolving independently for approximately 40 million years. It is believed that *DQA* and *DQB* alleles could encode incompatible components of the *DQ* receptor and hence be strongly selected against but there are probably also preferred combinations of all *DRB*, *DQA* and *DQB* alleles. Strong LD and low recombination rate prevents negative (purifying) selection from removing recessive deleterious mutations on otherwise favorable haplotypes and these mutations may hence become fixed in some haplotypes. van Oosterhout (2009) suggested that this process may lead to heterozygote advantage for which the deleterious mutations will not be expressed.

Regulation of MHC gene expression

During the last decade many researchers in evolutionary biology have focused their interest in regions controlling the expression of genes. This has occurred because the level of polymorphism identified within coding genes is often insufficient to explain adaptive differences (Streelman and Kocher 2000).

The mechanism controlling expression of MHC genes have been well characterized (Benoist and Mathis 1990; Glimcher and Kara 1992; Guardiola et al. 1996; Ting and Trowsdale 2002). Constitutive, as well as cytokine-induced, expression of MHC class II is primarily controlled at the transcriptional level through the use of gene specific proximal promoters (as well as other cis-acting elements) and several transcription factors. Several binding sites for transcription factors have been identified within the proximal promoters. The binding sites are between seven and 14 base pairs (bp) in length and are located within a region 40-160 bp upstream from the transcriptional start site (Glimcher and Kara 1992). Furthest away from the transcriptional start site we find the W box (including the S box), followed by the X box (including the overlapping X1 and X2 boxes) and then closest to the transcriptional start site, the Y box. There are also some locus specific regions such as the T box of the *DQA1* promoter (Guardiola et al. 1996). Transcription factors such as CREB and various RFX and NF-Y proteins, anchor to the boxes to initiate transcription. The MHC class II transactivator (CIITA) (Radosevich and Ono 2003; Ting and Trowsdale 2002; van den Elsen et al. 2004) functions as a co-activator, which interconnects the transcription factors with each other.

The transcription factor binding sites show similarity across species. However, within-species polymorphism has frequently been observed in humans and mice (Andersen et al. 1991; Cowell et al. 1998; Janitz et al. 1997; Mitchison and Roes 2002; Perfetto et al. 1993; Singal and Qiu 1995). Transient transfection assays, in which MHC promoter elements are fused with a reporter gene to identify their functional level, show that some of these promoter polymorphisms can affect expression of *DR* and *DQ* genes (Janitz et al. 1997; Louis et al. 1994; Singal and Qiu 1995; Woolfrey and Nepom 1995).

As described above, T helper cells can either favor a cellular response (TH1) or a humoral response (TH2). The TH1/TH2 balance is regulated by the strength of signal released through the presentation of an antigen by MHC class II to T helper cells. The signal strength may depend on the epitope concentration, the affinity to which the MHC class II molecule can bind to the epitope and the concentration of MHC class II molecules (Guardiola et al. 1996). Modulation of MHC class II expression may hence be important for a well-functional immune system and polymorphism within the promoter region may be of evolutionary importance, in addition to polymorphism

commonly described within the PBR sites of the coding genes. Little is known about this in general, and analyses of natural populations have been absent.

Selection on MHC genes

The maximum number of MHC molecules that can be found within a single individual is limited in comparison to, for example, the total possible number of TCRs. Although MHC molecules are less specific in their epitope binding capacity they may still restrict the flexibility of the immune system and its ability to respond to the pathogens' avoidance strategies. This limitation makes them the target of strong selection; selection for high levels of genetic polymorphism and selection for particular alleles to be maintained in the population. Selection that acts in favor of polymorphism is called balancing selection.

Although most MHC researchers agree that MHC polymorphism is maintained by balancing selection, in many cases it has been difficult to obtain convincing evidence, as outlined in several reviews (Bernatchez and Landry 2003; Garrigan and Hedrick 2003; Hughes and Yeager 1998; Meyer and Thomson 2001; Piertney and Oliver 2006). Different approaches to test for selection are used depending if one wishes to infer selection in the contemporary population, that is, in the current generation or to infer selection over the history of populations or selection over the history of species (Garrigan and Hedrick 2003; Piertney and Oliver 2006). A common problem faced is that selection acts most efficiently in large populations. In smaller and fluctuating populations stochastic events such as random genetic drift may often override any effects of selection. Furthermore, inferences of selection should consider over what period of time does selection need to act to leave behind a detectable signal and over what period of time reduced selection is needed for a signal to be erased (Garrigan and Hedrick 2003). These factors are particularly important for inferences of selection acting on the current generation from effects on genotypic frequencies and fitness of heterozygotes (Piertney and Oliver 2006). Below, I have described some observations that have often been used to infer the presence of balancing selection. These observations may also help us understand the source of polymorphism at MHC genes. Correlating the observations with the function of the MHC helps to support the theory of positive natural selection as the driving force in shaping the genetic patterns of this specific genomic region.

1. Substitution rates

For the majority of protein coding genes, any mutation within the coding sequence that affects the properties of the gene product will result in reduced

fitness and will be lost due to purifying selection. However, at PBR sites of MHC genes, selection works to favor such property changes.

For a nucleotide substitution to be subjected to any kind of selection it has to change the amino acid residue and hence the protein structure. Such mutations are called non-synonymous substitutions. In contrast, synonymous substitutions maintain the same amino acid and hence are selectively neutral, and represent the underlying mutation rate. The rate of non-synonymous substitutions (d_N) (number of non-synonymous substitutions/non-synonymous site) is often compared to the synonymous substitution rate (d_S) (number of synonymous substitutions/synonymous site) to test for the direction of selection (Hill and Hastie 1987; Hughes and Nei 1988). Under neutrality, in which no selection is acting, d_N is expected to be equal to d_S. If the changes are disadvantageous and purifying selection operates, d_N is smaller than d_S ($d_N/d_S < 1$). In the case of MHC where selection acts in favor of amino acid changes, d_N is higher than d_S ($d_N/d_S > 1$), which indicates positive selection. There are different kinds of positive selection, such as directional selection when a single variant is selected and eventually becomes fixed in the population. However, the type of selection acting on PBR sites of MHC molecules is balancing selection, which acts to establish the variant at an equilibrium frequency and hence maintain nucleotide polymorphism within the population (Hughes and Nei 1988; Hughes and Nei 1989) (*Figure 4*).

The high degree of polymorphism in MHC genes could be explained by a high mutation rate. However, Hughes and Nei (1988; 1989) showed that d_N only exceeds d_S at the gene region coding for the PBR of the molecule. They also showed that the d_S value did not differ from other genes, thus reflecting a normal mutation rate. Hence, MHC polymorphism is specifically related to peptide binding and not to the genes in general.

A significant d_N/d_S ratio requires a considerably long period of time of mutation and selection. Garrigan and Hedrick (2003) showed with computer simulations that it is possible to achieve d_N/d_S ratio significance in the range of 10 000 generations if the population size is large and selection strong. However in most cases it would take hundreds of thousands of generations. Furthermore, Garrigan and Hedrick's (2003) simulations showed that it takes even longer for the d_N/d_S ratio to lose its significance if selection is lost. Therefore, using the d_N/d_S ratio as an estimator of balancing selection says little about the current state in the population. However it provides information that a MHC gene of a certain species has been under balancing selection during the history of that species or even pre-dating that species.

2. Trans-species polymorphism

Under neutral expectations, the number of generations for all alleles within a species to be traced back to a common ancestor is four times the effective

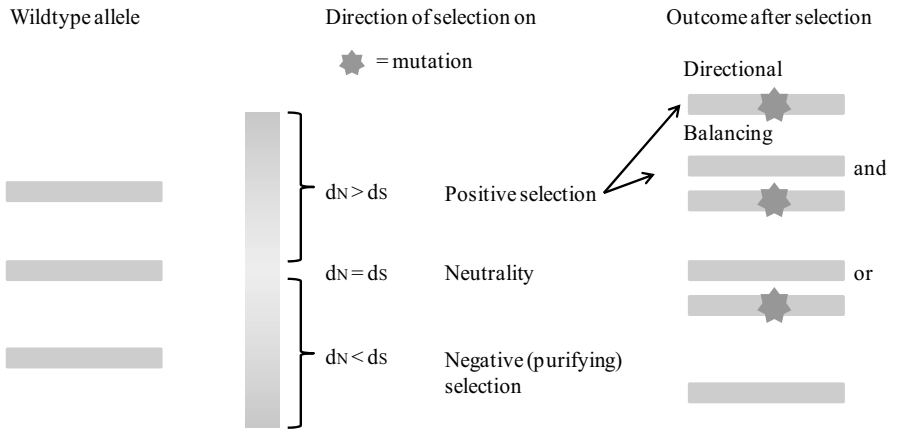


Figure 4. Outcome of selection acting on a new mutant (star) allele in a large population. d_N non-synonymous substitution rate, d_S synonymous substitution rate

population size (N_e), on average. This implies that if two species have been separated for more than $4N_e$ generations, all alleles in one species are more related to each other than to any allele in the other species. However, in the presence of balancing selection, as for MHC, alleles may persist for much longer periods of time, enabling alleles from different species to be more related than alleles from a single species (Takahata and Nei 1990). Consequently, in a phylogenetic tree, MHC alleles will cluster according to allelic lineages and not according to species. This has been shown, for example, in several studies on humans and chimpanzees (e.g. Gyllensten and Erlich 1989; Mayer et al. 1992), mice (McConnell et al. 1988), fish (Garrigan and Hedrick 2001; Graser et al. 1996) and birds (Richardson and Westerdahl 2003). This is referred to as trans-species polymorphism or trans-species evolution and implies that a lot of the variation that we see in MHC genes today is derived from ancestral species and not generated in each species following speciation (Figueroa et al. 1988; Klein 1987). The trans-species evolution theory is supported by unexpectedly large genetic distances between alleles, that is the number of nucleotide differences between them is often greater than expected for alleles within a species (Klein et al. 1998). Polymorphism has accumulated over long time and this is thought to be the most important way through which polymorphism has arisen. Like d_N/d_S ratios, trans-species polymorphism also indicates balancing selection over long history of time and is not useful at a short time scale.

There are other theories that have been proposed to explain the origin of diversity at MHC genes. Exchange of sequences, such as interlocus recombination and gene conversion, can shape variation at MHC alleles. It can result in alleles of a more recent origin showing divergence that otherwise would result from substitutions accumulating over long periods of time, as

indicated by trans-species polymorphism (Martinsohn et al. 1999). Although this mechanism is likely not the most important mechanism in which MHC polymorphism is derived, it may explain occasional patterns in localized regions (Bergstrom et al. 1998).

3. Allele frequency distribution and heterozygosity

For neutral genes, the normal pattern is to find a very common allele and some very rare variants. However in the presence of balancing selection, most alleles are instead found at intermediate and similar allele frequencies and we see few common alleles as well as few rare alleles. This has been demonstrated for MHC genes of many human populations (Hedrick and Thomson 1983).

Several alleles occurring at intermediate allele frequencies will result in an excess of heterozygotes and a lower level of homozygotes compared to that expected under neutrality. Under neutrality, when mutation and genetic drift are the only forces acting, there is an equilibrium distribution for heterozygosity, which will depend on the number of alleles at a given sample size (Watterson 1978). Higher than expected levels of heterozygosity has been interpreted as an indication of balancing selection. This can be assessed by tests of neutrality, such as the Ewens-Watterson test of neutrality, which is a statistical test in which homozygosity statistics from the data are compared to expected values under the null hypothesis of neutrality (Watterson 1978). Compared to the methods described above, this approach may allow for identification of balancing selection over a somewhat shorter time frame, such as over the history of populations of a given species.

It is important however to consider factors other than selection, which may affect allele frequencies and heterozygosity levels. Gene flow between populations may increase the number of rare alleles and hence result in an underestimation of increased heterozygosity as a result of balancing selection (Meyer and Thomson 2001). Population bottlenecks may also alter allele frequencies, with alleles lost more rapidly than heterozygosity is reduced and hence there would be an overestimate of the true signal of selection (Garrigan and Hedrick 2003). The Ewens-Watterson test of neutrality assumes that population size has remained constant over time, which is rarely true in natural populations (Garrigan and Hedrick 2003; Piertney and Oliver 2006). A common way to separate the effects of selection from demographic effects is to compare the patterns observed for MHC genes with those from a neutrally evolving marker such as microsatellites or mitochondrial DNA (mtDNA). Such neutral markers are only expected to be influenced by demographic factors and deviations between such markers and MHC can thus be attributed to selection (Garrigan and Hedrick 2003; Piertney and Oliver 2006).

Another way to show recent positive selection is to assess the conservation of MHC haplotype backgrounds. Single nucleotide polymorphisms (SNPs), which capture surrounding genetic variation, can be used to construct extended haplotypes and extended haplotype homozygosity (EHH) can be calculated to estimate the level of conservation. The extended haplotypes are sorted according to the sequence of a specified core region, which can be anything from a single SNP site to a large genomic region. EHH is the probability that two randomly chosen chromosomes (with identical core regions) are identical by descent and hence shows the transmission of an extended haplotype through time without recombination events (Sabeti et al. 2002). EHH is measured for each SNP site along the extended haplotype and the decay of EHH can be plotted against the distance from the selected core region. Unusually high EHH and a high frequency for a core haplotype, for example a specified MHC allele or the sites of a PBR region, indicates the presence of a mutation(s) that spread in the gene pool faster than expected under neutrality (Sabeti et al. 2002). In the absence of selection, high frequency alleles are expected to have been retained in the population sufficiently long for EHH to decay through recombination events. This method has been used to identify several alleles within the MHC region of humans, which show evidence of recent selective sweeps (de Bakker et al. 2006).

What drives balancing selection?

It is well recognized that MHC variation is maintained by balancing selection. However, there have been several theories postulated about the driving force behind this selection. The theories can be largely summarized into two types of mechanisms, disease-based and reproductive mechanisms (Bernatchez and Landry 2003; Meyer and Thomson 2001; Piertney and Oliver 2006). Disease-based theories are based on the assumption that a specific MHC allele is favored because of its ability to better recognize and bind a pathogen to trigger the immune system and hence provide protection from that pathogen.

An early disease-based theory, which attempted to explain the high MHC polymorphism, was presented by Doherty and Zinkernagel (1975). They claimed that, in a population exposed to a varied repertoire of pathogens, it would be advantageous for an individual to be heterozygous at MHC loci because each MHC molecule can recognize only a limited number of pathogens. Having a more varied array of MHC molecules would make an individual less vulnerable to various infections. This heterozygote advantage is often referred to as overdominant selection. In its simplest form it is assumed that all heterozygotes have equal and high fitness while all homozygotes have equal and low fitness. Takahata and Nei (1990) showed that this model could very well explain the persistence of alleles over time as indicated by trans-species polymorphism and the accumulation of mutations, which cause

large number of differences between alleles. However, the assumption of equal fitness for all heterozygotes (and all homozygotes) is not realistic and the overdominance model has been criticized as a non-realistic explanation of the selection-driving force (De Boer et al. 2004).

It has also been argued that, for the individual, high fitness is correlated with an optimal number of MHC alleles rather than maximum number. The reason for this is explained in terms of a trade off between having immunological flexibility provided by high genetic diversity at MHC and the higher risk of reaction against self-peptides that diversity brings (Kalbe et al. 2009; Nowak et al. 1992). Nevertheless, empirical studies have in some cases shown heterozygote advantage. An example, which will be discussed later, is HIV infected patients and their progression into aids (Carrington et al. 1999). Another is the association between the number of MHC alleles and the prevalence of avian malaria infection (Westerdahl et al. 2005). In a population of water voles with two MHC alleles, heterozygote individuals were more resistant to parasite infection than either of the two homozygotes, giving strong support to the overdominant theory in a straightforward case (Oliver et al. 2009).

Another disease-based theory, which has received extensive attention, is the negative frequency dependent selection theory. In this model, genotype fitness values are not fixed but change in proportion to allele frequencies. If a new pathogen is introduced or if a rare pathogen increases in frequency or if an established pathogen changes through mutation, a previously rare MHC allele which can recognize the pathogen would increase in fitness and hence in frequency. There will be a cyclic process of fitness values for both host genotypes and pathogen genotypes or pathogen types, where MHC allele frequencies fluctuate in time as pathogens adapt to them or are replaced by others (Clarke and Kirby 1966; Takahata and Nei 1990). Theoretically, it has been shown that these cycling processes could maintain many alleles in the population over time (Borghans et al. 2004).

A third disease based theory is the fluctuating selection hypothesis, which suggests spatial and temporal heterogeneity among pathogens so that the selective advantage of different MHC genes fluctuates in time or space over the life time of individuals or over the geographical range of a population (Hill et al. 1991). Compared to the negative frequency dependent theory, this model does not assume co-evolution between host and pathogen as the determining factor for pathogen fluctuations but allows for external factors to decide the distribution of pathogens (Spurgin and Richardson 2010).

The disease-based mechanisms have often been difficult to prove in spatial populations. The relative contribution of overdominance, negative frequency dependent selection and fluctuating selection has been widely discussed and in most cases all three mechanisms could explain observed patterns of genetic variability at MHC loci in empirical studies (Spurgin and Richardson 2010). There is also disagreement regarding how large the selec-

tion coefficient needs to be to explain the long coalescence times of alleles (Meyer and Thomson 2001; Piertney and Oliver 2006). Evidence for associations between particular MHC alleles and resistance to infectious diseases have, however, been shown in several natural populations (e.g. Langefors et al. 2001; Lohm et al. 2002; Meyer-Lucht and Sommer 2005; Paterson et al. 1998; Schwaiger et al. 1995).

As an alternative to pathogen-based models, various reproductive mechanisms have also been proposed to account for the polymorphism found at MHC genes. Mate choice based on MHC has been demonstrated in, for example, laboratory and wild mice (Egid and Brown 1989; Yamazaki et al. 1983), reviewed in Jordan and Bruford (1998). By disassortative mating, there is a reduction in the number of homozygotes and it is less likely that alleles are lost because of drift and can therefore be maintained in the population for a longer time (Hedrick 1992). MHC-based selective mating could be a mechanism to avoid inbreeding more generally, as MHC loci are highly polymorphic and individuals that share MHC alleles are very likely to be related (Potts et al. 1994).

It has also been suggested that interactions between a mother and her fetus can play a role in the maintenance of MHC polymorphism (Clarke and Kirby 1966). If a fetus has increased fitness when it has a MHC type that differ from its mother, the number of homozygous births will be reduced and the effect could explain a long coalescence time for alleles (Hedrick and Thomson 1988). In humans, it has been shown that spontaneous abortions are more common among couples that share the same MHC alleles (Thomas et al. 1985). It could be that a homozygous fetus is negatively selected due to direct effects of MHC but an alternative explanation is that it is more likely to be homozygous at nearby recessive deleterious loci and therefore have lower survival.

Reproductive mechanisms have been criticized due to the weak connection between them and the function of MHC in the immune system, and pathogen-based models are currently accepted to be the primary driving force of balancing selection. However, reproductive mechanisms may very well still contribute to shaping the patterns of polymorphism at MHC.

Factors affecting MHC polymorphism

As described previously, the signal of balancing selection is not always easily detected for MHC genes. Indeed, for some species (or populations), variation at the MHC is actually quite limited compared to related species (populations).

By far the most common reason to why selection has been insufficient to shape patterns at MHC genes is that in small populations, demographic processes may be a much greater force than selection in influencing the level of MHC polymorphism. This does not imply that balancing selection is re-

duced but signifies that the power of genetic drift has been stronger than the power of selection (Edwards and Potts 1996; Hedrick et al. 2001b). In such populations, reduced MHC polymorphism is correlated with low genetic variation across the whole genome. Endangered species with a very low population size have been shown to possess low levels of MHC polymorphism. For example, cheetahs (*Aconyx jubatus*) show low MHC diversity, which correlates with a genome-wide loss of diversity (O´Brien et al. 1985). A similar process has been shown for small, isolated populations of the Australian bush rat (*Rattus fuscipes greyii*) (Seddon and Baverstock 1999). Also, historical events such as bottlenecks can be reflected in the level of MHC polymorphism (Ellegren et al. 1993; Mikko and Andersson 1995). In many European species of plants and animals, the level of genetic variation has been affected by repeated bottlenecks associated with glacial periods (Hewitt 1999; Taberlet et al. 1998), although the effect on their MHC diversity is unknown.

Given that balancing selection is the mechanism maintaining MHC polymorphism, it is reasonable to expect that factors affecting the strength of selection would also affect the level of polymorphism at MHC genes. Assuming disease-based mechanisms as the driving force behind balancing selection, the load of different pathogens could affect the level of variation (Edwards and Potts 1996). Low levels of MHC variation have been found in several marine mammals (Murray et al. 1995; Slade 1992; Trowsdale et al. 1989) and it was hypothesized that low exposure to parasites in marine environments compared to terrestrial environments would reduce the selective pressure for maintaining high MHC polymorphism in marine mammals. However not all studies on marine mammals support this view (Hoelzel et al. 1999; Murray and White 1998). A study by Wegner et al. (2003) investigates the relationship between MHC diversity in three-spined sticklebacks and parasite diversity in different habitats. They confirm an association between MHC polymorphism and parasite diversity and show that there are only small differences in microsatellite polymorphism between the different habitats. This study supports the hypotheses that parasite load can influence MHC variation.

Why study MHC variation

Researchers have been interested in learning about MHC polymorphism for a number of different reasons.

1. MHC genes have often been stated as an excellent candidate gene for learning about natural selection and its influence on local adaptation in natural populations (Hedrick 1994). Studies from MHC may help us understand the factors that affect the strength of selection and how selection interacts with other forces such as drift. Further, the power of statistical tests to detect

selection can be evaluated by using the MHC (Garrigan and Hedrick 2003) and later applied when testing for selection in other parts of the genome where the signal of selection may not be as strong.

2. One may take advantage of the large genetic variability and the old allelic age offered by MHC genes. By using distribution patterns of a marker at which alleles persist over long time, conclusions may be drawn about historical events for which traces have been erased by drift through time in other genes. MHC alleles have, for example, been used to estimate effective population size in humans (Klein et al. 1990). Another example is Vilà et al. (2005) who used MHC diversity to estimate the number of founders in the domestication of dogs from wolves.

3. In conservation genetics, MHC genes are considered as a possible candidate gene that has the potential to directly affect disease resistance and reproductive success. However, MHC polymorphism may also provide valuable information about genetic variability in general (Edwards and Potts 1996). Hughes (1991) proposed that work concerning protection of endangered species and captive breeding programs should put MHC diversity as a central factor. Although, most conservation geneticists would agree that MHC polymorphism is relevant to consider, most would probably be cautious in placing too much emphasis on MHC variability considering the complexities of spatial fitness effects at the MHC (Edwards and Potts 1996). Nevertheless, parasites constitute a threat to endangered species, especially if a pathogen introduced to the environment is a novel threat for the endangered species. For example, Hedrick et al. (2001a) showed that the endangered fish *Gila topminnow* suffered from infections spread through occasional contact with guppies. Inbred strains and individuals homozygous for MHC had lower survival than outbred controls and heterozygotes, which could be attributed to lack of important MHC genes or to low genetic variation in general. Similar results were obtained for an endangered salmon species (Arkush et al. 2002).

4. From a medical point of view, MHC allele or MHC haplotype associations are central to many diseases. For example, there is a clear association between MHC type and the progression of HIV infection into aids. Carrington et al. (1999) found that HIV patients homozygous for MHC class I loci and/or with two specific class I alleles progressed to aids faster than HIV patients who were heterozygous for MHC class I loci or who lacked the two aids-associated alleles. Similarly, both MHC class I and class II alleles have been associated with protection from severe malaria infection (Hill et al. 1991).

There are also many examples of associations between MHC type and non-infectious diseases including many autoimmune diseases (Caillat-Zucman 2009; Shiina et al. 2004). One example is the association between *DR* and *DQ* alleles and insulin-dependent diabetes (diabetes type 1). *DR/DQ* haplotypes have been associated with increased susceptibility to, as well as

protection from, diabetes type 1 in humans (Erlich et al. 2008) and also in dogs (Kennedy et al. 2006). In some cases single amino acid polymorphisms have been shown to influence diabetes type 1 susceptibility patterns (Erlich et al. 2008). The association between MHC types and diseases are however rarely straightforward; individuals carrying an allele/haplotype associated with protection from a disease may still develop the disease and vice versa. Environmental factors, apart from genetic factors, often contribute to the susceptibility to autoimmune diseases (Caillat-Zucman 2009). Furthermore, strong LD makes it problematic to determine whether the cause of association to a disease depends on changes in the sequence of the protein binding regions (encoded by exon 2) of these genes or in nearby regions. Erlich et al. (2008) highlight the possibility that polymorphism in MHC regions other than the exon 2 alleles of *DR* and *DQ* could be important for diabetes type 1 susceptibility.

Abnormal expression of MHC class II has also been suggested to be associated with autoimmune diseases (Guardiola et al. 1996). An up-regulation of MHC class II and hence increased signal strength may direct the immune response towards a cellular (TH1) response (Baumgart et al. 1998). In turn, a bias towards a cellular response could predispose those individuals to autoimmune diseases (Mueller-Hilke and Mitchison 2006). An example is the correlation between MHC class II expression patterns and the susceptibility to, as well as the progression of, rheumatoid arthritis (Heldt et al. 2003). It has been hypothesized that balancing selection acts on the promoter regions to obtain an appropriate TH1/TH2 balance and hence maintain promoter polymorphism within the population (Mitchison et al. 1999; Mueller-Hilke and Mitchison 2006). Deviations from normal MHC class II expression patterns have also been associated with diseases such as severe immunodeficiency resulting from defects in the regulatory mechanisms of MHC class II expression (Mach et al. 1996; Mach et al. 1994).

Studied species

Below I have described relevant background information for the species studied in this thesis.

European hedgehogs

In Europe, there are two parapatric species of hedgehogs, the brown-breasted *Erinaceus europaeus* and the white-breasted *E. concolor* (Reeve 1994). *E. europaeus* is usually found in western Europe while *E. concolor* is found in eastern Europe. In hedgehogs, as in many other European species, the level and distribution of genetic variation has been strongly affected by repeated glacial periods (Hewitt 1999). Seddon et al. (2001) and Santucci et al. (1998)

showed a deep split between the two species using mtDNA haplotypes. Further subdivisions within the species were also identified, reflecting post-glacial colonization patterns. As for other animals, it is assumed that genetic diversity is highest in the regions where the hedgehogs survived during ice ages and from where they expanded during inter glacial periods (Hewitt 1999; Taberlet et al. 1998). Iberia, Italy and the Balkans constituted the most important refugia for European hedgehogs (Santucci et al. 1998; Seddon et al. 2001).

Both species are hibernating animals. During hibernation body temperature and many physiological systems such as the immune system, are greatly affected (Boyer and Barnes 1999; Burton and Reichman 1999).

Samples from both European species of hedgehogs have been used in this study.

Dogs

Dogs (*Canis familiaris*) were domesticated from grey wolves probably at some point between 15 000 and 100 000 years ago. The precise point in time has been widely discussed (Lindblad-Toh et al. 2005; Savolainen et al. 2002; Vila et al. 1997). Nonetheless, the domestication of dogs predates that of other domesticated animals, such as cattle, pigs, horses and chickens (Bruford et al. 2003). The first dog domestication event probably took place in East Asia as suggested by the distribution of genetic diversity (Savolainen et al. 2002). However, the number of founder events and the number of founders involved in the events has been a question of debate. Studies in which mtDNA has been used as a genetic marker suggest between four and six founder events (Savolainen et al. 2002; Vila et al. 1997). However, genetic diversity of this marker may have been lost through drift over time. Vila et al. (2005) used MHC diversity to estimate the number of founders and suggested an absolute minimum of 19-32, probably up to hundreds, founders of the dog population. The relatively large number of founders may be explained by a large original domestication event with many wolves contributing with genes to the dog gene pool or more likely by continuing hybridization between dogs and wolves (Vila et al. 2005). It is commonly known that dogs and wolves hybridize to produce fertile offspring (Verardi et al. 2006; Vila et al. 2003) and continued backcrossing was likely.

Dogs constitute the morphologically most diverse mammal species and we recognize hundreds of dog breeds, most of which are less than 150 years old (Ostrander and Wayne 2005; Parker et al. 2004). In evolutionary terms this is a very short time. Nevertheless, the formation of breeds has had a major effect on the distribution of genetic variability among dogs. For a puppy to be registered to a certain breed, it is required that both parents belong to that same breed (Ostrander and Wayne 2005; Parker et al. 2004). This requirement has resulted in limited or no gene flow between breeds and,

as a consequence, reduced genetic variability within breeds and genetic differentiation among breeds (Lindblad-Toh et al. 2005; Ostrander and Wayne 2005; Parker et al. 2004; Sutter et al. 2004).

The dog genome has been subjected to strong evolutionary forces as there has been strong artificial selection for traits associated with morphology and behavior (Saetre et al. 2004; Svartberg 2006). Such strong artificial selection on morphological and behavioral traits may have caused reduced selection on other traits. Bjornerfeldt et al. (2006) suggested that relaxation of selective constraint may have allowed accumulation of slightly deleterious mutations in the mitochondrial genome of dogs. The results were supported by Cruz et al. (2008) who used whole-genome SNP data to show that the d_N/d_S ratio is 50% higher in dogs than in wolves. Strong selection is expected to affect not only the selected locus but also nearby regions through genetic hitchhiking. Through a selective sweep, regions surrounding a selected region may lose their genetic variability and high LD with the selected region results (Kim and Nielsen 2004). However, it has been suggested that for artificial selection associated with domestication, the target region for selection may have been selectively neutral prior to domestication and the reduction of surrounding genetic variability may be less than from sweeps caused by natural selection (Innan and Kim 2004).

The dog genome has also been strongly influenced by drift as a result of strong bottlenecks in the early domestication of dogs and in the later formation of breeds. Gray et al. (2009) modeled the demographic patterns associated with the bottlenecks of domestication and breed formation. The authors reached the conclusion that the contraction as a result of the domestication resulted in only a modest reduction in nucleotide diversity compared with the contraction associated with breed formation. This may be explained by the potentially large degree of back-crossing between wolves and dogs as suggested from MHC data (Vila et al. 2005). A consequence of the bottlenecks is high LD within breeds, often 10-100 times more extensive than that found in humans (Gray et al. 2009; Lindblad-Toh et al. 2005; Sutter et al. 2004). The number of breed founders as well as the current and past popularity of a breed is reflected in the extent of LD (Gray et al. 2009; Lindblad-Toh et al. 2005; Sutter et al. 2004). Gray et al. (2009) concluded that the patterns of LD reflect population history in a similar way to nucleotide diversity levels.

One may take advantage of high LD when it comes to disease association mapping. In combination with haplotype sharing between breeds, high LD allows for disease association mapping with much fewer markers compared to the numbers of markers required in association studies of humans (Lindblad-Toh et al. 2005; Ostrander and Wayne 2005; Parker et al. 2004; Sutter et al. 2004). Dogs share many diseases, such as cancer, heart diseases and immune related diseases, with humans (Ostrander and Giniger 1997). For example, human autoimmune diseases such as type 1 diabetes, rheuma-

toid arthritis, haemolytic anaemia and Hashimoto's disease all have equivalents that are common in many dog breeds (Kennedy et al. 2007b). The release of the dog genome sequence (Kirkness et al. 2003; Lindblad-Toh et al. 2005) has enabled extensive evolutionary research on the dog genome and has further facilitated the dog as an excellent model organism for learning about selection processes as well as disease associations.

The samples used in this thesis come from various breeds from Scandinavia (paper II, III and V) and from Australia (paper IV).

Grey wolves

The wild progenitor of dogs, the grey wolf (*Canis lupus*), was once distributed across most of the Northern Hemisphere. However, during the last centuries the distribution area has been drastically reduced, leaving geographically and genetically isolated populations (Wayne et al. 1992). As for other species, wolves have been affected by glacial periods. However, because of high mobility in between glacial periods, some adaptation to living in glacial regions and changes of habitat distribution there is a lack of historical phylogeographical structure (Vila et al. 1999).

Based on a study of mtDNA, the total effective population size of female wolves was estimated to be 173 000 individuals, giving a total estimated population size of approximately a million wolves throughout the world. However, this is likely to be an overestimate because wolves are well surveyed and because the population declines are recent, resulting in high worldwide genetic variability that has not yet been affected by drift (Vila et al. 1999). The true worldwide population size is probably not more than 300 000 individuals (Vila et al. 1999).

In North America there is a large and continuous population distribution reaching throughout Alaska and Canada (Roy et al. 1994). In Europe the population is much more fragmented with recent bottlenecks (Ellegren et al. 1996; Wayne et al. 1992). By using samples from locations throughout Eurasia and North America, Vila et al. (1999) showed little genetic differentiation between wolves on both a regional scale as well as over continents. However they found indications of a local genetic structure as a consequence of recent restricted gene flow. In a similar way as for dogs, demographic history is reflected in the level of LD. Gray et al. (2009) showed that LD (measured as $r^2 = 0.2$) reached less than 10 kilo base pairs (kb) among wolves from Alaska, Canada and Yellowstone, compared with wolves from Spain and Sweden among which LD reached approximately 1 Mb, that is, 100 x longer.

The wolf samples used in this study come from the Finnish/Russian wolf population. Aspi et al. (2006) showed that wolves from Finland have kept their genetic variation in spite of rapid population decline during the last 150 years. Furthermore, there was no strong evidence of inbreeding. The Finnish

wolf population is connected with the Russian wolf population, which has also been affected by population declines (Pulliainen 1980). Aspi et al. (2006) also showed that, although a conservation management program has allowed for an increase in total population size during the last decade, the effective population size has not increased. This could be explained by increased isolation. The dispersal distances of Finnish wolves seem to have decreased over recent times (Aspi et al. 2006; Kojola et al. 2006). Shorter dispersal distances may lead to a more rapid loss of genetic diversity (Leonard et al. 2005). Nonetheless, the Finish/Russian wolf population still constitutes a genetically variably wolf population, with expected heterozygosity levels exceeding those of other European populations and most North American populations (Aspi et al. 2006).

Research aims

In my research I have tried to understand how evolutionary forces act to shape the genetic patterns of MHC class II genes. I have a combined research interest in evolutionary biology, conservation management and immunogenetics. For me, MHC is a natural choice of genetic system to use for a variety of purposes and the core of this thesis is to increase the understanding of MHC class II genes, how they evolve and how they can be used in various fields of research.

A specific interest has been to understand more about the evolution of MHC gene expression as I believe that changes in regulatory features of genes may be of even greater adaptive importance than changes of protein encoding genes themselves. Obviously, genetic regions involved in gene expression regulation are also subjected to evolutionary forces such as drift and selection. Although the regulatory features controlling MHC class II gene expression have been well characterized in human and laboratory mice, little is known about how evolutionary forces act on these elements in natural populations. To take it another step further, I was also interested to survey if and how protein encoding MHC genes and regulatory elements of MHC genes co-evolve. I believe that the integration and association of gene evolution and gene regulatory element evolution is central and significant in both biological and medical science.

Specific research aims:

Paper I

- Take advantage of the high level of MHC gene inter-species conservation to derive MHC class II exon 2 gene sequence from a previously unsurveyed species, in this case from two species of European hedgehogs.
- Use the old allelic age of MHC genes to improve the knowledge of phylogeographical patterns associated with glacial and inter glacial periods.
- Study how evolutionary forces such as selection and demography have contributed to shaping genetic patterns at exon 2 of MHC genes in hedgehogs.

Paper II

- Characterize MHC class II promoter sequences in Swedish dogs and Russian/Finnish wolves to detail the level and the location of polymorphism within these important gene regulatory elements.

Paper III

- Evaluate the signature of selection on promoter and exon 2 sequences in dogs and wolves.
- Survey haplotypic association patterns between *DQB1* exon 2 alleles and promoter variants and test how evolutionary forces associated with domestication and breed formation have influenced these patterns.
- Analyze phylogenetic relationships among promoter variants and exon 2 alleles as well as relationships among promoter/exon 2 haplotypes.

Paper IV

- Use SNP markers to determine the evolutionary history of MHC class II haplotypes to distinguish between conserved and convergent extended haplotypes in dogs.
- Apply the above information to analyze disease associations using diabetes mellitus as an example.
- Sequence the *DQB1* promoter region of dogs with diabetes mellitus and Australian control dogs to identify disease associated/geographically associated promoter variants.

Paper V

- Analyze how evolutionary forces, such as drift and selection associated with dog domestication, have affected MHC class II haplotypes, the association of exon 2 alleles and their genetic background, defined by extended SNP haplotypes.
- Further explore the processes involved in the generation and maintenance of MHC class II haplotypes.

Research investigations

The research investigations included in this thesis are presented as publications/manuscripts later on in this thesis. However, below I have described and discussed the background, the methodology and the results, putting my personal research aims in focus. Some details, especially concerning the methods, have been left out here while other aspects are presented or discussed more thoroughly.

Paper I - Understanding the phylogeographic patterns of European hedgehogs, *Erinaceus concolor* and *E. europaeus* using the MHC

In this study we derived the first MHC sequences from hedgehogs. We had access to information from mtDNA and also preliminary results from a nuclear intron sequence from the same samples. Results from mtDNA suggested a deep split between *E. concolor* and *E. europaeus* and further subdivisions within the species that corresponded to postulated glacial refugial regions (Santucci et al. 1998; Seddon et al. 2001). However, the preliminary results from nuclear sequence data failed to divide *E. europaeus* into different monophyletic subgroups. We intended to analyse how information from an additional nuclear marker with old allelic age could complement previous knowledge about hedgehog phylogeographical patterns associated with glacial periods.

We also wanted to explore the evolutionary forces that may have acted on MHC genes in hedgehogs. In particular, we wanted to investigate whether hibernation can affect the selection pressure on MHC genes. During hibernation, the immune system is severely suppressed (Burton and Reichman 1999). Hibernating animals seem to have evolved a system making it possible to defeat pathogens in spite of immunosuppression through the manipulation of body temperature. By relying on an alternative way of pathogen control, the selection pressure on MHC genes could be affected.

Material and Methods

In total, 22 samples from *E. concolor* and 62 samples from *E. europaeus*, were used in this study. Primers were designed to amplify *DRB1*, *DQA1* and *DQB1* by aligning published sequences from several mammals. Samples were sequenced and alleles were assigned in heterozygous samples using single-strand conformation polymorphism (SSCP) (Orita et al. 1989) and/or cloning.

PBR sites were assigned according to the predictions from human MHC genes (Brown et al. 1988; Brown et al. 1993). Phylogenetic relationships among alleles were analyzed in a neighbor-joining tree and the strength of selection was tested using d_N/d_S estimates. The number of *DQB1* alleles compared to the number of mtDNA types (MHC/mtDNA ratio) was calculated from a resampled unbiased data set (Table 1).

Deriving MHC sequences from a new species

Amplification of *DQA1* and *DQB1* was successful. However, the amplification of *DRB1* failed, most likely due to the lack of inter-species conservation. *DRB1* is the most variable locus and finding suitable regions for primers was indeed difficult. Because of this, primers may match the hedgehog sequence poorly and so the polymerase chain reaction (PCR) products will be non-specific. An alternative, but less likely explanation is that hedgehogs might lack the *DRB1* locus. It is known that some species lack certain loci. For example, the *DQ* locus is absent in cats (Yuhki et al. 2003).

Using MHC genes to understand phylogeographical patterns

The level of genetic diversity at *DQA1* was limited, with two alleles identified in each species. For *E. concolor* the geographical distribution of the two alleles was clearly subdivided by the Bosphorus. For *DQB1* ten alleles were found in *E. concolor* and six in *E. europaeus*. As for *DQA1*, there were no shared alleles across the Bosphorus in *E. concolor*. However, there was not a clear structure in relation to geographical barriers in *E. europaeus.* This was in concordance with previous preliminary results from nuclear intron sequences but contradicted those of mtDNA. At the *DQB1* locus, *E. concolor* showed greater variation than *E. europaeus* in number of alleles, nucleotide diversity, heterozygosity and also the level of *DQB1* variation in comparison to the level of mtDNA variation (Table 1). This difference was also observed for the nuclear intron sequence data (Seddon et al. 2001). However, the intra-species divergence, measured as number of nucleotide differences between alleles, was similar between the two species.

A possible explanation for the inconsistent pattern in mtDNA and nuclear loci is that the sorting of alleles was a result of repeated refugial separation

Table 1. Comparison between MHC (*DQB1*) and mitochondrial variation in hedgehogs

Species by mtDNA group	N	No. of *DQB1* alleles	No. of mtDNA types	MHC/mtDNA	Unbiased MHC/mtDNA
E. concolor	22	10	22	0.455	0.608 ± 0.0008
C1	17	8	17	0.471	
C2	5	2	5	0.400	
E. europaeus	62	7	55	0.127	0.138 ± 0.0002
E1	39	4	38	0.105	
E2	23	3	17	0.176	

Unbiased MHC/mtDNA ratios were estimated by creating 10 000 data sets of 22 *E. concolor* and 62 *E. europaeus* individuals by resampling with replacement. MtDNA groupings (E1, E2, C1, C2) were based on Seddon et al (2001).

and associated population bottlenecks. Recent bottlenecks associated with contractions to the refugia may have occurred at the same time in both species, accounting for the subdivision of mtDNA. An additional older separation in refugia for *E. concolor* allowed sorting of nuclear alleles in that species. A second explanation is a difference in past expansion patterns between the species. Rapid expansion events result in a stronger founder effect and thus more severe loss of genetic variation (Hewitt 1996; Ibrahim et al. 1996). Repeated expansion events for *E. europaeus* would have resulted in a reduction in both MHC and mtDNA variation. In contrast, slower expansion patterns for *E. concolor* in the past could have maintained MHC variation. Recent bottlenecks must then have been mild, affecting only mtDNA, which has an effective population size one-quarter that of nuclear DNA.

The use of MHC data, in addition to previous mtDNA and nuclear intron sequence data, made it possible to draw more detailed conclusions about the demographic changes associated with the glacial refugia and postglacial expansion events. This study showed the importance of using more than one marker in phylogeographic studies. As data from MHC and nuclear intron sequence followed the same pattern, the difference observed between MHC data and mtDNA data could be attributed to the difference between nuclear DNA and mtDNA in general and not to the special characteristics of MHC.

Evolutionary forces acting on hedgehog MHC genes

As described above, the distribution of MHC variation in European hedgehogs has been affected by repeated glacial and inter-glacial events and demographic events. Despite the dissimilarities described between MHC and mtDNA, there were also similarities regarding, for example, geographic distribution of alleles. Furthermore, the patterns of MHC followed those of nuclear intron sequence data. As described in the introduction, comparing MHC allele distribution patterns with those of a neutral evolving marker is

useful when separating the effects of selection from demographic effects. Concordance between MHC and the neutral markers used in the previous study indicated that drift has been a more important evolutionary force than selection in shaping the patterns of MHC variation in hedgehogs.

However, there was still evidence for how selection has acted on these genes. For both *DQA1* and *DQB1* one allele was shared between the two species, reflecting trans-species polymorphism. There was also a lack of species-specific clades in a neighbor-joining tree. Together, these indicate that alleles have persisted over long evolutionary time and signify evidence of balancing selection over the history of the hedgehog species.

The use of d_N/d_S ratio differences between PBR and non-PBR sites (described in the introduction) only signified the presence of balancing selection for *E. europaeus* and not for *E. concolor*. For both species, there was a strong bias towards PBR variability in the 5' region of the exon, which encodes the β-floor of the peptide binding cleft and there was no or very low variability at PBR sites in the α-helix encoding region. Furthermore, there was higher nucleotide divergence calculated over all sites in the β than the α regions, reaching significance in *E. europaeus*. Both the d_N and the ds value was higher in the β-floor encoding region than in the α-helix encoding region suggesting different evolutionary histories. We identified a previously described recombination motif (Wu et al. 1986), constituting a breakpoint between the β-floor encoding region and the α-helix encoding region. This is probably an example how recombination events can cause relatively recent changes in evolutionary patterns of MHC genes.

Furthermore, there is a possibility that hibernation could affect the selection pressure on MHC genes in hedgehogs. An alternative way of pathogen control could result in reduced selection pressure in hedgehogs compared to non-hibernating mammals and hence result in lower d_N/d_S ratio. Similar studies in other hibernating species would be needed to further test this hypothesis.

Paper II - MHC promoter polymorphism in grey wolves and domestic dogs

Previous studies from humans and mice have shown a significant level of intra-species polymorphism in the proximal promoter regions of MHC genes. Furthermore, they have also shown the importance of this polymorphism for MHC gene expression patterns. We were interested to study MHC promoter polymorphism in an additional species and in natural populations. As dogs have become an important model organism for evolution of human immune-related diseases and the newly released dog genome enabled the region to be amplified in Canids, we chose to detail the level and location of

polymorphism in the proximal promoter sequence of dogs and grey wolves. Furthermore, in dogs and wolves there are single functional genes for *DRB*, *DQA* and *DQB* (Wagner 2003), which simplifies analyses of MHC polymorphism as problems with multi-copy loci are avoided.

Material and Methods

A total of 90 wolf samples from the Finnish/Russian wolf population and 90 samples from ten different breeds of Scandinavian dogs were used in this study. The promoter regions of *DRB1*, *DQA1* and *DQB1* were amplified using dog specific primers derived from previously published MHC sequences (Wagner et al. 1996a; Wagner et al. 1996b; Wagner et al. 1998) and dog genome sequences (Kirkness et al. 2003). All samples were sequenced and alleles from heterozygous individuals were resolved using SSCP (Orita et al. 1989).

Pattern of promoter polymorphisms in dogs and wolves

Among all dogs and wolves, two promoter alleles for *DRB1* were shared between the species. The two alleles differed at only one site, not localized in any region of known importance. For *DQA1*, wolves were fixed for one allele, which was also the most common among dogs. A second rare allele was identified in a few dog breeds. The two *DQA1* alleles were distinguished by a substitution located within 3' end of the T box. In contrast, *DQB1* showed high levels of promoter polymorphism with nine alleles identified in total. Six alleles were shared between wolves and dogs, one allele was identified only in dogs and two alleles were found only in wolves. The alleles differed at six nucleotide sites and a four-bp indel. Two of these sites were localized in the regions known to constitute binding sites for transcription factors, the X1 box and the X2 box respectively (Glimcher and Kara 1992).

Low levels of polymorphism in *DRB1* promoter regions has been described previously in humans (Kruger et al. 2001) and so has high levels of *DQB1* promoter polymorphism (Andersen et al. 1991; Reichstetter et al. 1994). High levels of *DQB1* promoter polymorphism has also been found in mice (Janitz et al. 1997; Mitchison and Roes 2002). The conserved pattern between species may suggest that maintenance of constant *DRB1* expression is of importance for the immune system while varying expression patterns and levels of *DQB1* is beneficial.

It is reasonable that polymorphism identified within the T box of the *DQA1* promoter and within the X1 and X2 box of the *DQB1* promoter regulates expression levels in a similar manner as has been shown in humans and mice (Janitz et al. 1997; Woolfrey and Nepom 1995).

Dogs showed some breed specific patterns, such as fixation of a certain allele. When considering breeds with at least two alleles, wolves showed

higher heterozygosity than dogs. An overall high degree of allele sharing between wolves and dogs is not surprising considering the relatively recent divergence between the species.

Haplotypic associations between particular *DQB1* promoter alleles and exon 2 alleles (available from Seddon and Ellegren (2002)) were suggested from wolves that were homozygous at one or both loci. This was further investigated in the forthcoming study.

Paper III - Allelic combinations of promoter and exon 2 in *DQB*1 in dogs and wolves

We had demonstrated high levels of genetic variation in the proximal promoter region of the *DQB1* gene in dogs and wolves, including polymorphism in regions constituting binding sites for transcription factors. Hence, it is likely that this polymorphism affects expression patterns of MHC genes. High levels of intra-species variation suggested that altering expression patterns of *DQB1* genes may have been evolutionary important. Hence, we hypothesized that balancing selection operates on *DQB1* promoter sequences in a similar manner as on peptide-binding regions of exon 2.

Paper II indicated that there are haplotypic associations between promoters and coding sequences of MHC *DQB1* genes, Haplotypic associations between promoters and coding sequences of MHC genes could result in allele-specific expression patterns. Under such a scenario, established combinations of MHC coding sequence and promoter allele would be selectively favorable. If the promoter sequence affects the level of transcription and hence the resulting concentration of expressed *DQB1* molecules, it could affect the TH1/TH2 balance and the direction of the immune reaction in a way that would be beneficial to combat a pathogen recognized by the PBR region of the linked coding region. Hence, associations between the promoter and exon 2 can result from selective advantageous combinations of alleles. However, associations could also be attributed to any of the other explanations that have been suggested to cause particularly high LD in the MHC region (Raymond et al. 2005) or because of the close chromosomal location (about 2 kb) between the promoter and exon 2. Furthermore, we know that genome-wide LD within dog breeds is expected to be high (Gray et al. 2009; Lindblad-Toh et al. 2005; Sutter et al. 2004).

In this study we compared the patterns of association of the promoter and exon 2 alleles of the *DQB1* gene in dogs and wolves to determine the relative strength of their haplotypic structure. We also surveyed the evolutionary forces that have acted on *DQB1* coding genes, *DQB1* promoters and *DQB1* promoter/exon 2 haplotypes.

Material and Methods

A total of 85 wolves and 89 dogs, for which the *DQB1* promoter information was available from the previous study, were used. Exon 2 of the *DQB1* gene was amplified in all dogs and in those wolves for which this information was not already available (Seddon and Ellegren 2002). Alleles from heterozygote individuals were resolved by comparing the heterozygous sequences with reported DLA *DQB1* alleles. Promoter/exon 2 haplotypes were assigned by hand and also confirmed using Phase v 2 (Stephens and Donnelly 2003; Stephens et al. 2001).

Unbiased data sets of haplotypes as well as random association data sets of alleles were created by resampling haplotypes with replacement. The number of haplotypes, promoter variants and exon 2 alleles, as well as the average number of exon 2 alleles associated with each promoter variant (and vice versa) for observed data and for the unbiased data set, were calculated. The unbiased data set was also compared with the random association data set to test how observed associations differ from random patterns. We included comparisons with a dataset of a single breed, German shepherd dogs, to ensure that values for the dogs were not influenced by admixture of breeds.

Ewens-Watterson statistics was used to determine the strength of selection. Relationships among promoter alleles and the PBR sites of exon 2 were determined using Median Joining networks and phylogenetic relationships among promoter/exon 2 haplotypes were determined in a neighbor-joining tree.

Signature of selection on promoter and exon 2 sequences

The results of the Ewans-Watterson's test suggested that balancing selection is acting on the *DQB1* exon 2 sequences as well on the promoter region. The signal of selection was in fact stronger for the promoter region than for the exon 2 region in both dogs and wolves. These findings suggested that promoter polymorphism might be of at least equal importance for the functionality of the immune system.

Evolutionary effects on promoter/exon 2 associations

We observed fewer promoter/exon 2 haplotypes among dogs and wolves than expected from the random association data set, which was expected considering the close chromosomal location and high LD within the MHC in general. Surprisingly, we found proportionally fewer haplotypes in wolves than in dogs, suggesting weaker associations and linkage within dogs than within wolves. The pattern for a single breed, the German shepherd, showed the same pattern as for the mixed-breed data set, ruling out an admixture of

breed effect. This is in contrast to what is known about the evolutionary history of the species. Dogs show high genome-wide LD levels, particularly within breeds as a consequence of strong bottlenecks associated with domestication and breed formation (Gray et al. 2009; Lindblad-Toh et al. 2005; Sutter et al. 2004) while the historical population size of the Finnish/Russian wolfs is large without reflections of inbreeding and fragmentation (Aspi et al. 2006).

We suggested a few possible explanations for the observed differences among dogs and wolves. Firstly, increased recombination, for example because of a recombination hotspot present in dogs but not wolves, or with varying intensity, could explain less strict association patterns in dogs. Secondly, selection intensity to maintain allelic combinations of promoter and exon 2 could be weaker in dogs. Reduced selection intensity could result from either relaxation of selective constraint as a consequence of artificial selection on other traits (Bjornerfeldt et al. 2006; Cruz et al. 2008) or because bottlenecks associated with domestication and breed formation resulted in a predominance of genetic drift. A third possible, but probably less likely, explanation is that substantial changes to the environment of dogs due to domestication led to a wider range of selectively adapted haplotypes, giving the appearance of reduced linkage in dogs.

Phylogenetic relations among alleles and haplotypes

Promoter/exon 2 haplotypes were analyzed in a phylogenetic tree. The resulting branching pattern showed two main clades. The upper clade was predominately haplotypes from dogs while the lower clade had a greater proportion of wolf haplotypes. Networks of promoter alleles and PBR sites of exon 2 showed wolf alleles in more central positions and dog alleles in more peripheral positions. These observations suggested that the polymorphism observed in wolf alleles has a more ancient origin than the polymorphism observed in dogs. This could support the idea that promoter/exon 2 association patterns in dogs have been affected by altered selection related to domestication.

Paper IV- Evolutionary history of DLA class II haplotypes in canine diabetes mellitus through SNP genotyping

In dogs, as in humans, MHC class II haplotypes are most often defined by the combination of alleles at exon 2 of *DRB1*, *DQA1* and *DQB1* (Kennedy et al. 2002; Kennedy et al. 2007a). Some of these haplotypes have been found in divergent breeds from widespread geographical locations suggesting an ancestral background and/or selectively favourable allele combinations. In this study we analysed whether these exon 2 defined haplotypes share a common genomic background and hence are true conserved haplotypes or whether some exon 2 defined haplotypes are convergent haplotypes, in which combinations of exon 2 alleles have arisen independently. To do this, we constructed extended haplotypes using SNPs across the MHC class II region. We also defined and analysed LD blocks across the region.

Certain MHC class II haplotypes have been associated with diseases such as diabetes type 1 in humans (Caillat-Zucman 2009; Knip and Siljander 2008) and its equivalence in dogs, diabetes mellitus (Kennedy et al. 2006). In dogs, three *DRB1*/*DQA1*/*DQB1* susceptibility haplotypes and one *DQA1*/*DQB1* protective haplotype have been identified (Kennedy et al. 2006). High LD makes it problematic to determine whether the cause of association to disease depends on mutation(s) within the exon 2 of these genes or in nearby regions. If haplotypes were convergent it would be difficult to detect associations with mutations outside exon 2 regions if the haplotypes don't share common genomic background. We used the extended haplotypes to investigate this further.

As previous studies had revealed polymorphism within functional important sites of the *DQB1* promoter gene, this region was also amplified and sequenced in a subset of the dog samples.

Material and methods

A total of 109 Australian pure-bred and cross-bred dogs were used in this study. Of these, 60 had been diagnosed with diabetes mellitus while the remaining 49 dogs were of known healthy state. SNPs were detected by sequencing regions approximately every 6 kb across a region covering the classical MHC class II loci. A total of 20 SNPs were selected for genotyping. SNP genotyping was carried out using a variety of methods. One variable site from exon 2 of *DRB1*, *DQA1* and *DQB1* respectively was also included giving a total of 23 SNPs, which were used to construct extended haplotypes. LD blocks were identified in the Haploview (Barrett et al. 2005). The *DQB1* promoter region was amplified and sequenced in 43 samples of which 34 had been diagnosed with diabetes mellitus.

Conserved and convergent extended MHC class II haplotypes and patterns of LD

Extended haplotypes were constructed from the 23 variable sites extending from the *DRA* gene to beyond *DQB1* covering all classical MHC class II loci. A total of 14 samples were homozygous for exon 2 of all three (*DRB1*, *DQA1* and *DQB1*) genes. A further 24 dogs were heterozygous but for which at least one haplotype was to be found amongst the homozygous individuals and a second haplotype could hence be subtracted. A total of 25 exon 2 defined haplotypes were so identified. Eight of these were found more than once and for each of these the average number of fixed SNP differences was 1.8 (0-6) across the extended SNP haplotypes. Further SNP sites were also heterozygous in some haplotypes. The finding of identical exon 2 haplotypes on different SNP backgrounds suggested convergence of exon 2 haplotypes i.e. that the combination of exon 2 alleles had arisen independently, likely through recombination or gene conversion. Selection may have favored the maintenance of allelic combinations. Another observation was an extended SNP haplotype that was identical across most of the region in two samples, suggesting a common genetic background, but in this case with different *DQB1* alleles, excluding them from being defined as a haplotype based on exon 2 data. The LD blocks were surprisingly short and limited to four smaller regions. This further implied that convergence has been important in shaping the patterns of MHC class II haplotypes.

Extended MHC class II haplotypes and diabetes mellitus

The pattern of LD differed between diabetic dogs and non-diabetic dogs. In both groups two LD blocks were interrupted in the region of the *DRB1* gene and one short block of closely located SNPs was identified upstream from the *DQA1* gene. However in the *DQ* region the pattern differed between the sick and the healthy dogs. In the non-diabetic dogs a short LD block, which only included the exon 2 *DQA1* marker and two closely located SNPs were identified. In the diabetic dogs a larger LD block included both the exon 2 markers of *DQA1* and of *DQB1* and also the intervening SNP markers. Interestingly, the block did not extend all the way to the start of the *DQB1* gene.

Two of the *DRB1*/*DQA1*/*DQB1* susceptibility haplotypes and the *DQA1*/*DQB1* protective haplotype (Kennedy et al. 2006) were found among the samples and their extended SNP haplotypes were compared (*Figure 5*). There were fixed SNP differences between the protective haplotype and the two susceptibility haplotypes across the *DQA1-DQB1* region. A comparison between the two susceptibility haplotypes showed an identical SNP pattern across the *DQ* region in spite of different *DQA1* and *DQB1* exon 2 alleles suggesting that the disease associated mutation(s) may be found outside

	DRB1	*DQA1*	*DQB1*	2	6	10	*DRB1*	12	17	23	28	34	41	42	43	*DQA1*	47	49	65	*DQB1*	67	69	85	91	96	100
S	*00201	*00901	*00101	1	1	2	2	-	2	2	1	1	2	1	1	2	-	2	1	1	2	1	1	h	1	2
S	*00201	*00901	*00101	1	1	2	2	2	2	2	1	1	2	1	1	2	-	2	1	1	2	1	1	h	1	2
S	*01501	*00601	*00301	1	1	2	2	1	2	2	2*	2	2	2	2*	2	1*	2*	1	1	1	1	1	1	1	1
S	*01501	*00601	*02301	1	1	2	2	1	2*	2*	-	h	h	2	2*	2	1*	-	1	1	1	1*	1*	-	1	1
S	*01501	*00601	*02301	1	1	2	2	1	2	2	2	h	2	2	2	2	1	2	1	1	1	1	1	1	1	1
P	*01201	*00401	*01303	2	1	1	1	1	2*	2*	1*	h	2	1	1*	1	2*	1*	2	2	2	1	2	h	1	2
P	*02001	*00401	*01303	2	1	1	2	1	-	-	-	h	2	1	1	1	2	1	2	2	2	1	1	-	1	2
P	*00601	*00401	*01303	1	1	2	2	1	2	1	1	1	2	1	1	1	2	1	2	2	2	1	2	1	1	2
P	*00601	*00401	*01303	2	2	1	2	1	1	2	1	1	2	1	1	1	2	1	2	2	2	1	1	-	-	2

Figure 5. Comparison of extended haplotypes for two exon 2 susceptibility haplotypes (S) and one *DQ* protective haplotype (P). Exon 2 alleles are shown for *DRB1*, *DQB1* and *DQA1*. SNP data are shown as binary alleles, missing data as - and heterozygous sites as h. Alleles 1* and 2* replace missing data where the haplotype was present in a heterozygous dog and the SNP position was homozygous. Boxed region shows fixed differences between susceptibility and protective haplotypes.

exon 2 of these genes. The similarity between the susceptibility haplotypes did not include the 5' region of the *DQB1* gene (marker 67 and 69), which once again indicated weak LD in this region and the potential of disease associated mutations.

DQB1 promoter polymorphism in dogs with diabetes mellitus

A total of 11 promoter alleles were identified among the 48 dogs sequenced. Four of these had not been observed in our earlier studies among Scandinavian dogs and they were only observed among diabetic dogs in the present study. Among the new promoter variants there were three previously unreported polymorphic sites. Two of these were located in the "dyad region", a dyad symmetry element located close to the transcriptional binding site the W box. Polymorphism disrupting the symmetry of the dyad region in humans has been shown to alter the expression patterns of the human *DQB1* gene (Shewey et al. 1992). Hence, it is likely that such variations in the dyad region could affect expression of the associated *DQB1* gene in diabetic dogs.

Susceptibility and protective haplotypes were found among both diabetic and non-diabetic dogs, giving evidence for the complexity of disease associations. Dogs carrying the susceptibility or the protective haplotype in homozygous or heterozygous form were compared in more detail (Table 2). At one polymorphic site, located in the X1 box (site -170), which is a transcriptional factor binding site, the susceptibility haplotypes showed a consistent difference from the protective haplotype in diabetic dogs. All protective haplotypes had an adenine (A) nucleotide at that position while all susceptibility haplotypes contained a guanine (G) nucleotide. However non-diabetic dogs carrying the protective allele had a G nucleotide at the same position. It is therefore possible that the protective haplotype could have different expression patterns in diabetic and non-diabetic dogs, although further analysis in larger sample sizes is required.

Paper V – Linkage disequilibrium and haplotype patterns of the MHC class II region - a comparison between wolves and dogs

From study number IV we learned that regions of high LD were unexpectedly short along extended haplotypes of the MHC class II region in dogs. This was surprising considering the high levels of LD found in the human MHC region (Raymond et al. 2005) and the genome wide high LD observed within dog breeds (Gray et al. 2009; Lindblad-Toh et al. 2005; Sutter et al. 2004). However, the level of LD can be affected by factors such as recombination, population history, and selection on allele combinations (Ardlie et al. 2002).

Table 2. *DQB1* promoter variants associated with protective and susceptibility MHC class II haplotypes for canine diabetes mellitus.

	DRB1	*DQA1*	*DQB1*	DQBp	-272	-256	-213	-170	-154	-105	-87	-71
DQ protective												
Diabetic	*1201/*1201	***0401/*0401**	***1303/*1303**	p*6/p*6	del/del	G/G	C/C	A/A	C/C	A/A	T/T	T/T
Diabetic	*0601/*0601	***0401/*0401**	***1303/*1303**	p*6/p*6	del/del	G/G	C/C	A/A	C/C	A/A	T/T	T/T
Diabetic	*1201/*1501	***0401**/*0601	***1303**/*0301	p*6/p*2	del/ins	G/G	C/C	A/G	C/C	A/A	T/C	T/C
Diabetic	*1501/*2001	***0401**/*0601	***1303**/*2002	p*4/p*5	del/del	T/T	C/C	A/G	C/C	A/A	T/C	T/C
Diabetic	*0201/*2001	***0401**/*0201	***1303**/*1303	p*5/p*7	del/del	G/T	C/C	A/A	C/T	A/A	T/T	T/C
Non-diabetic	*1201/*1201	***0401**/*0401	***1303**/*1701	p*1/p*1	del/del	G/G	C/C	G/G	C/C	A/A	T/T	T/T
Non-diabetic	*1201/*1201	***0401**/*0401	***1303**/*1701	p*1/p*1	del/del	G/G	C/C	G/G	C/C	A/A	T/T	T/T
DR-DQ susceptibility												
Diabetic	***1501/*1501**	***0601/*0601**	***2301/*2301**	p*2/p*10	ins/ins	G/G	C/C	G/G	C/C	A/A	C/T	C/T
Diabetic	***1501**/*2901	***0601**/*0301	***2301**/*0401	p*1/p*10	ins/del	G/G	C/C	G/G	C/C	A/A	T/T	T/T
Diabetic	***1502**/*0101	***0601**/*0101	***2301**/*0201	p*2/p*6	ins/del	G/G	C/C	G/A	C/C	A/A	C/T	C/T
Diabetic		***0601**/*0601	***2301**/*2002	p*4/p*11	ins/del	T/A	C/T	G/G	C/C	A/A	C/C	C/C
Diabetic	***0201**/*0901		***0101**/NEW	p*1/p*3	del/del	G/T	C/C	G/G	C/C	A/T	T/T	T/T
Diabetic	***1501/*0202**		***2301/*0101**	p*2/p*3	ins/del	G/T	C/C	G/G	C/C	A/T	C/T	C/T
Diabetic	***1501**/*0601		***2301**/*2001	p2*/p*4	ins/del	G/T	C/C	G/G	C/C	A/A	C/C	C/C
Diabetic	***0201**/*0801	***0901**/*05011	***0101**/*2801	p*3/p*5	del/del	T/T	C/C	G/A	C/C	A/A	T/T	T/T
Non-diabetic	***1501**/*5201	***0601**/*12011	***2301**/*3501	p*2/p*2	ins/ins	G/G	C/C	G/G	C/C	A/A	C/T	C/T
Non-diabetic	*1201/***1501**	*0401/***0601**	*1701/***2301**	p*2/p*6	ins/del	G/G	C/C	G/A	C/C	A/A	C/T	C/T
Non-diabetic	***0201**/*0601	***0901**/*05011	***0101**/*0701	p*3/p*5	del/del	T/T	C/C	G/A	C/C	A/A	T/T	T/T

Dogs are homozygous or heterozygous for the protective haplotype *DQA1**04/ *DQB1**13 or susceptibility haplotypes *DRB1**15/ *DQA1**06/ *DQB1**23 or *DRB1**02/ *DQA1**09/ *DQB1**01 (shown in bold). DQB1p indicates promoter sequences numbered from the start of exon 1. Sites -170 and -154 are in the X1 and X2 boxes, respectively, and site -213 is in the dyad region. Del, deletion; ins, insertion.

.

To learn more about the evolutionary history of extended SNP haplotypes of the MHC class II region and the influence of population effects and/or altered selection associated with the domestication of dogs and the formation of breeds, we adapted the SNP data set used in paper IV to also capture variation within the MHC class II region of wolves.

Material and Methods

In total, 84 wolves were compared with 84 Scandinavian dogs from different dog breeds. The dog samples were separated into two data sets, one with a mix of all breeds except German shepherds and one with 36 samples of German shepherd samples exclusively. The samples represented the majority of those used to analyze the promoter region and exon 2 of the *DQB1* gene in project II and III. A subset of the wolf samples had also been typed for exon 2 of *DRB1* and *DQA1* (Seddon and Ellegren 2002). The complete exon 2 region of *DRB1* and *DQA1* was amplified in all dog samples and in the wolf samples for which sequence data was not already available. The SNP data set, optimized in project IV to capture variation in the MHC class II region of dogs, was adjusted to capture variation in dogs as well as wolves. A total of 15 SNPs were selected for genotyping, which was carried out in two different ways, by pyrosequencing and by multiplex SNP genotyping with a SNPstream system (analyzed by the SNP Technology Platform, Uppsala, Sweden (www.genotyping.se)). Added to the data set with 15 genotyped SNPs was one position of each of the *DQB1* promoter region and exon 2 of *DRB1*, *DQA1* and *DQB1*. Haploview version 4.1 (Barrett et al. 2005) was used to analyze patterns of LD (*Figure 6*).

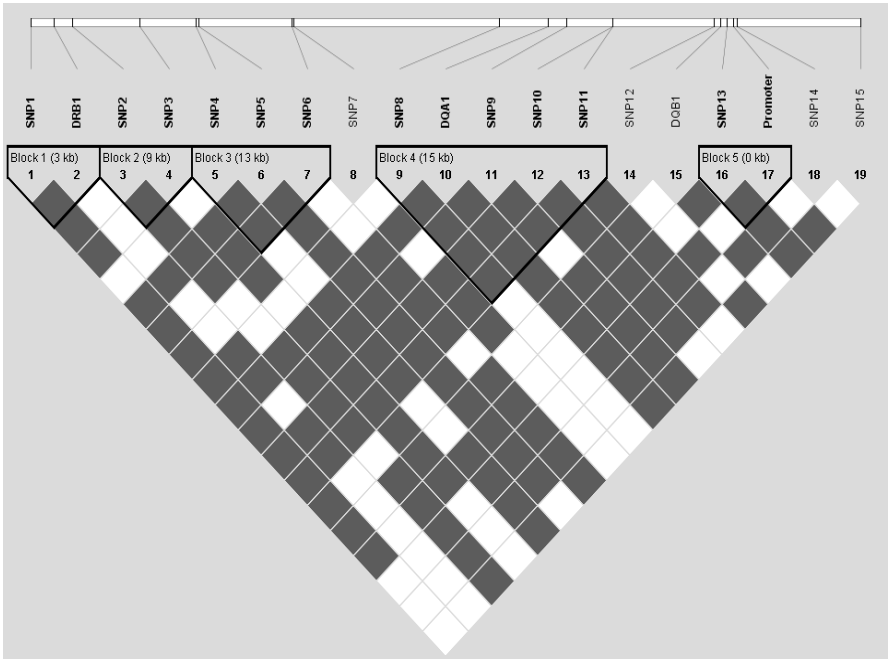
In order to determine the surrounding conservation among extended haplotypes of the MHC class II genes, extended SNP haplotypes were sorted according to exon 2 alleles of *DRB1*, *DQA1* and *DQB1*, based on the combination of alleles of the three genes and by each gene separately. We assessed the conservation of extended haplotypes expanding out from the exon 2 region of each gene (*DRB1*, *DQA1* and *DQB1*) using extended haplotype homozygosity (EHH).

Exon 2 based haplotypes and their genetic backgrounds

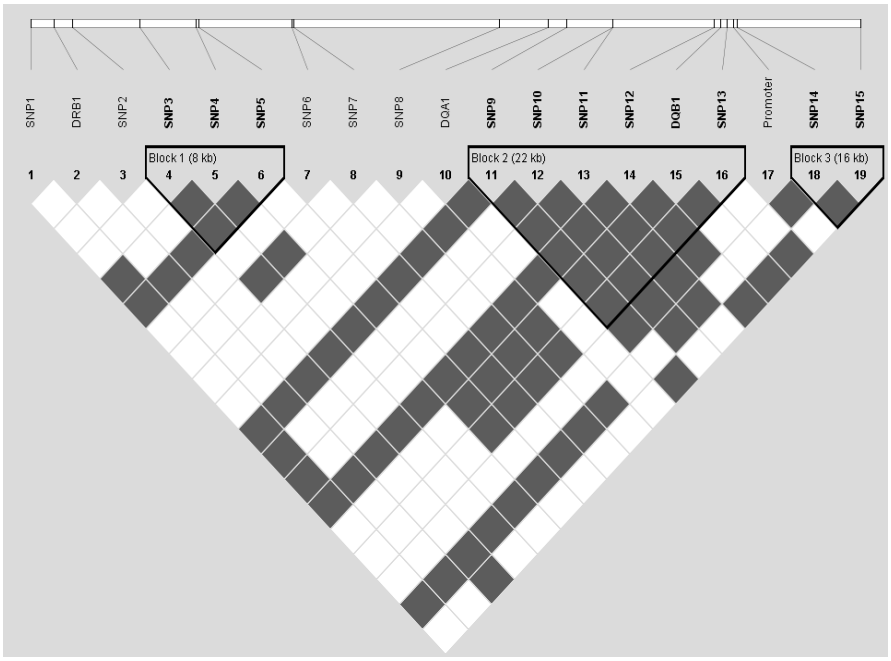
Exon 2 alleles of the *DRB1*, *DQA1* and *DQB1* loci combined to form distinct MHC class II haplotypes. At each locus, every exon 2 allele occurred on one or more *DRB1/DQA1/DQB1* haplotype. Alleles in dogs, in inter-breed as well as intra-breed comparisons, occurred consistently on a higher average number of haplotypes than in wolves. Hence, the association between exon 2 alleles seemed weaker in dogs than wolves.

We found examples of both ancestral and convergent exon 2 based haplotypes among all data sets. The presence of convergent haplotypes suggested

A


SNP1
DRB1
SNP2
SNP3
SNP4
SNP5
SNP6
SNP7
SNP8
DOA1
SNP9
SNP10
SNP11
SNP12
DQB1
SNP13
Promoter
SNP14
SNP15
Block 1 (3 kb)
Block 2 (9 kb)
Block 3 (13 kb)
Block 4 (15 kb)
Block 5 (0 kb)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

B


SNP1
DRB1
SNP2
SNP3
SNP4
SNP5
SNP6
SNP7
SNP8
DOA1
SNP9
SNP10
SNP11
SNP12
DQB1
SNP13
Promoter
SNP14
SNP15
Block 1 (8 kb)
Block 2 (22 kb)
Block 3 (16 kb)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

C

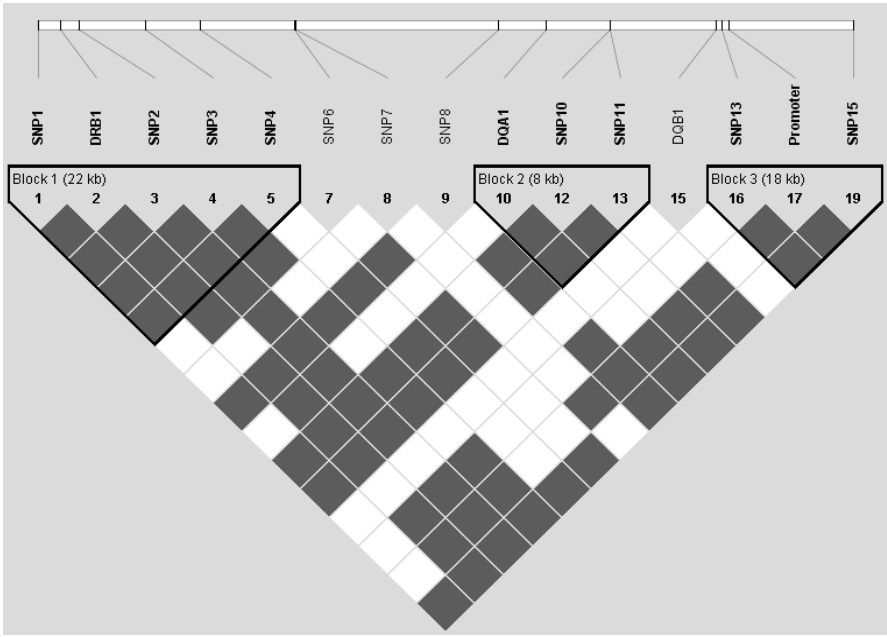


Figure 6. Linkage disequilibrium (LD) across the extended DLA class II haplotypes in A) wolves, B) dogs from mixed breeds and C) German shepherds. Haplotype blocks were defined in Haploview according to the four gamete rule with black colour indicating high LD values (shown by < 4 distinct 2-marker haplotypes) and with white colour indicating low LD values (shown by 4 distinct 2-marker haplotypes).

that recombination events have occurred in the DLA class II region. In the genetic background, LD blocks were relatively short for all data sets, confirming observations in project IV (*Figure 6*). Wolves seemed to have a more restricted recombination pattern, with few but distinct hotspots for recombination, while dogs showed less stringent recombination patterns. For all data sets, regions for recombination were found between all three exon 2 regions.

In the genetic background, LD decreased more rapidly with distance among dogs than among wolves. However, within German shepherds, LD was maintained at a high level over long distances. Therefore, lower LD and recombination patterns in the genetic background in dogs may be explained by breed specific patterns. However, breed specific patterns do not explain weaker associations between exon 2 alleles among *DRB1/DQA1/DQB1* haplotypes in dogs compared to wolves, as weaker patterns were also observed within a breed. In wolves, a higher proportion of convergent haplotypes was found compared to ancestral haplotypes, which suggested that the differenc-

es were not due to higher recombination rates in dogs. We find it likely that a predominance of drift over selection associated with domestication and breed formation of dogs or a relaxation of selective constraint (Bjornerfeldt et al. 2006; Cruz et al. 2008) and continuing inbreeding among dogs may have caused insufficient selection on combinations of alleles. We suggested that this could influence dogs' resistance to pathogens and also predispose them to autoimmune diseases.

MHC haplotype generation and evolution

A larger number of haplotypes maintained EHH values of 1.0, or close to 1.0, across the whole extended haplotype when the *DQB1* gene constituted the selected core region compared to when the *DRB1* gene constituted the selected core region in both wolves and the mixed breed data set. This meant that *DQB1* sorted haplotypes were in general more uniform and could indicate stronger selection on *DQB1* alleles compared to alleles of other genes.
Wolves and dogs shared two *DQA1*/*DQB1* haplotypes. Firstly, the *DQA1*05011/DQB1*00701* haplotype showed a conserved genetic background along the extended haplotype up until the *DRB1* gene, which indicated a likely ancestral haplotype. In contrast, the *DQA1*00301/DQB1*00401* haplotype showed a combination of *DQ* alleles this is likely to have arisen through convergence in both wolves and dogs. The large number of differences between *DRB1* alleles associated with these haplotypes suggested that *DRB1* alleles have been added to the *DQ* haplotype by recombination rather than evolved by mutation.

Wolves and dogs did not share any *DRB1* alleles. The high number of *DRB1* alleles is expected to result in lower allele frequencies. Therefore, *DRB1* alleles may be particularly affected by stochastic events such as drift associated with population bottlenecks. A larger sample size, with more breeds and/or with dogs as well as wolves from geographically diverse regions, is likely to further elucidate these patterns.

Concluding remarks and future perspectives

In this chapter I summarize the most important findings of my research projects and I conclude how I think my research contributes to the understanding of MHC gene and MHC gene expression evolution. I also give suggestions on how additional research can verify the ideas I put forward or provide further knowledge in how to use MHC class II genes in various research fields.

Information from MHC data successfully allowed us to draw more detailed conclusions about the population history of two species of European hedgehogs. Allele distribution estimates for MHC genes in comparison with neutral markers made it possible to separate the effects of selection from demographic effects. The results showed that over a shorter time scale, demographic factors associated with repeated population reductions and expansions have been more important than selection in shaping patterns of MHC genetic variability. However, transspecies polymorphism provided evidence for a long persistent time of alleles and hence signified the presence of balancing selection over the history of species. Weak signal of selection at the nucleotide level (analyzed by d_N/d_S estimates) suggested that the strength of selection could be affected by alternative factors, one of which may be the immunological consequences of hibernation. Further research on hibernating animals could provide answers to this theory. In the hedgehog study we also showed an example of how a recombination event may have shaped genetic patterns of a more recent origin.

Using *DQB1* in dogs and wolves as an example, we learned that the genetic patterns created by balancing selection reaches outside exon 2. In fact, a Ewan-Watterson test suggested stronger selection on the *DQB1* promoter than on exon 2, signifying the importance of a flexible expression pattern of this gene, possibly to maintain an appropriate TH1/TH2 balance. Interestingly, other MHC class II genes did not show polymorphic promoter regions. Hence, it seems that there are different evolutionary strategies for expression of MHC class II genes in spite of an ancient common gene ancestry. The question of why there are different expression strategies for different genes and whether this could be of medical importance are left open. Polymorphism was identified within functional important regions of the *DQB1* promoter and in one case in the *DQA1* promoter. Present investigations are carried out to test how this polymorphism affects gene expression patterns.

Linkage between promoter and coding polymorphism could enable allele-specific expression patterns. We showed weaker associations between the promoter region and exon 2 of *DQB1* in dogs, in inter-breed as well as in intra-breed comparisons, than wolves. We suggest that selection or demographic related factors associated with domestication have broken up promoter-exon 2 allele associations in dogs. Hence, it is possible that adapted allele-specific expression patterns may have been lost. I believe that this could be of great significance.

MHC class II haplotypes have often been defined according to the exon 2 sequences of the classical genes. Some of these haplotypes have been associated with diseases. By using SNP based extended haplotypes, we showed that the exon 2 defined haplotypes do not always reflect an ancestral state but in fact can be convergent haplotypes, in which exon 2 allele combinations have arisen independently. Disease associated mutation(s) located outside exon 2 could be more difficult to detect or more difficult to understand when exon 2 based haplotypes do not reflect common ancestry across the region in which the mutation(s) is located. We showed the example of dogs with diabetes mellitus. Two haplotypes associated with susceptibility to diabetes in dogs showed identical SNP pattern across the *DQ* region in spite of different *DQA1* and *DQB1* exon 2 alleles, suggesting that the disease associated mutation(s) is to be found outside exon 2. Other observations made us suspect that the disease associated mutation(s) is located in the 5' region of the *DQB1* gene. A survey of promoter polymorphism on the *DQB1* gene revealed a polymorphic site, located in a transcriptional factor binding site, which showed consistent differences between diabetes susceptibility and protective haplotypes in diabetic dogs. However in non-diabetic dogs, the difference was not observed suggesting alternative expression patterns of the protective haplotype in sick and healthy dogs. Our results need to be verified with larger data sets and with intra-breed comparisons but the results could point to an important site for MHC class II associations with diabetes mellitus in dogs. Our results also support the importance of studying MHC haplotype ancestry when searching for disease associations.

We constructed extended SNP haplotypes covering the *DRB1-DQA1-DQB1* area also in wolves and Scandinavian dogs and found ancestral as well as convergent exon 2 based haplotypes. This suggested that recombination events have occurred in the DLA class II region. Interestingly, associations between different exon 2 alleles showed a more restricted pattern in wolves compared to dogs. As for the association between exon 2 and the promoter region of *DQB1*, the pattern was consistent both between and within breeds. However, in the genetic background of the MHC class II region (*DRB1-DQA1-DQB1*) as a whole, LD decreased over longer distances in an inter-breed comparison while it remained high within wolves and within a single breed, the German shepherds.

Furthermore, dogs from a mixed breed data set showed a less restricted recombination pattern. This would suggest that in the genetic background, lower LD and recombination patterns in dogs might be explained by breed specific patterns. However, breed specific patterns did not explain weaker associations between exon 2 alleles among *DRB1/DQA1/DQB1* haplotypes in dogs compared to wolves as weaker patterns also were observed within a breed. We suggested that a predominance of drift over selection associated with domestication and breed formation of dogs or a relaxation of selective constraint (Bjornerfeldt et al. 2006; Cruz et al. 2008) and continuous inbreeding among dogs may have caused reduced selection on combinations of DLA class II alleles. Hence, because of stochastic events and/or insufficient selection, beneficial combinations of alleles may not be kept and disadvantageous combinations of alleles may not be removed, as efficiently in dogs as in a large population without inbreeding and genetic differentiation. Consequently, inbred populations, such as dog breeds, could suffer from a deficient immune defense against pathogens and also be more predisposed to autoimmune diseases.

Conservation among extended haplotypes, measured as the maintenance of extended haplotype homozygosity (EHH) among SNP sites over distance from the exon 2 region, showed that *DQB1* sorted haplotypes maintained high EHH values over longer distances than *DRB1* sorted haplotypes. Furthermore, wolves and dogs shared haplotypes only of *DQA1/DQB1* and did not share any *DRB1* alleles. This may indicate that these *DQB1* alleles are more ancestral and/or selectively favorable. Our findings also suggested that *DRB1* alleles have been added by recombination to different *DQA1/DQB1* haplotypes rather than by evolving by mutation after the domestication of dogs.

In my research I, together with colleagues, have analysed the evolutionary patterns affecting genetic variation at single MHC class II coding genes and at the MHC class II region at a larger scale. Specifically, we have analysed variation in regulatory features controlling the expression of MHC class II genes and how this polymorphism associates with polymorphism in coding regions of MHC class II. We also successfully used MHC as a marker in phylogeographical studies. My research challenges the approach of focusing all attention on the coding regions of MHC class II genes, both when it comes to learning about the evolutionary forces that shape patterns in the MHC class II region and when using MHC class II in disease association studies.

Sammanfattning på svenska

Bakgrund

I min avhandling har jag studerat MHC-gener (Major Histocompatibility Complex). Det som gör dessa gener unika är deras exceptionellt höga grad av genetisk variation. Denna genetiska variation yttrar sig genom att varje MHC-gen har många olika varianter, s.k. alleler. Dessa alleler skiljer sig åt sinsemellan med många fixerade mutationer. För att förstå hur genetisk variation uppkommit och varför den bibehålls måste MHC-genernas funktion i genomet förstås.

Första gången MHC-gener definierades var när det upptäcktes att transplantation av vävnad var möjlig mellan möss med identisk genuppsättning i MHC-regionen men att transplanterad vävnad stöttes bort om mössen skiljde sig åt i denna region. Denna upptäckt ledde fram till att funktionen av MHC klargjordes, vilken är att känna igen främmande ämnen och att presentera dem för immunförsvarets T-celler, en typ av vita blodkroppar som då aktiveras. Det finns två typer av antigenpresenterande MHC-molekyler, klass I och klass II. I denna avhandling har jag fokuserat helt och hållet på MHC klass II och de MHC-gener (*DRB1*, *DQA1*och *DQB1*) som är viktiga i denna region.

När en mutation sker är den oftast antingen neutral, då påverkar slumpen om den stannar kvar i populationen, alternativt negativ för individen och då rensas den oftast snabbt bort av negativ selektion. I de regioner som kodar för aminosyror som binder en antigen inom MHC-gener verkar dock positiv selektion och förändringar rensas inte bort. För att variationen ska bibehållas på sikt krävs att selektionen inte är riktad i förmån för en och samma MHC-variant under lång tid. Därför används begreppet balanserande selektion för att beskriva selektion som verkar för att polymorfi bibehålls i populationen.

Det stora antalet skillnader mellan olika MHC-alleler har uppstått genom att MHC-alleler bibehålls inom en population eller art under lång tid så att mutationer har chans att ackumulera. På så vis kan alleler mellan arter vara närmare släkt än alleler inom en art, s.k.”artöverskridande polymorfi”. Detta har t.ex. påvisats i studier hos människa och schimpans. Det finns även många andra sätt att påvisa förekomsten av balanserande selektion och flera av dem används i denna avhandling.

MHC-gener är intressanta ur många perspektiv. De har använts för att studera naturlig selektion och hur selektion integrerar med andra evolutionära processer. Ofta har graden av MHC-variation påverkats även av slump-

mässiga processer som överskuggat effekterna av selektion i små populationer. Genernas höga grad av variation gör också att de med fördel kan användas som genetisk markör i populationsstudier. Inom bevarandegentiken har det talats om att anpassa avelsprogram för utrotningshotade arter baserat på MHC-variation för att stärka deras immuna förutsättningar på lång sikt. I medicinska sammanhang finns en direkt koppling mellan MHC och infektionssjukdomar. Till exempel har kopplingar mellan MHC och sjudomar som HIV och malaria gjorts. MHC-gener är också centrala i forskning runt många immunorelaterade sjukdomar. Nästan samtliga autoimmuna sjukdomar, t.ex. diabetes typ I, har korrelerats med en genetisk faktor kopplat till MHC. Även uttrycksmönster av MHC tros kunna påverka uppkomsten av sjukdomar.

I min forskning har jag studerat vilka evolutionära krafter som påverkar genetisk variation i MHC gener. Genom att arbeta med MHC har jag kunnat kombinera mina forskningsintressen som finns inom evolutionsbiologi, bevarandebiologi och immunogenetik. Ett specifikt intresse har varit att ta reda på om selektion även verkar på de regioner som reglerar uttryck av MHC-gener.

Forskningsprojekten i korthet

Artikel I – MHC-variation hos igelkottar för att avslöja historiska populationsförändraingar

I min första studie använde jag mig av MHC för att studera hur genetisk variation hos två arter av europeiska igelkottar (*Erinaceus europaeus* och *E. concolor*) har påverkats av upprepade istider i centrala och norra Europa. Jag kunde jämföra mina resultat med tidigare studier som gjorts på dessa igelkottar och där en annan genetisk markör (mitokondrie-DNA) användes. Dessa studier visade genetisk differentiering både mellan och inom de båda arterna. Jag var också intresserad av att ta reda på om någon annan orsak, än de som är associerade med populationsförändringar, har påverkat graden av MHC-variation hos igelkottar. Igelkottar går i ide under en stor del av året och under denna period är immunförsvarets kapacitet kraftigt reducerad. Detta skulle kunna påverka styrkan av balanserande selektion.

Mina resultat avvek ifrån de tidigare resultaten av mitokondrie-DNA genom att inte visa någon genetisk differentiering inom en av arterna, *E. europaeus*, medan den andra arten, *E. concolor*, följde tidigare mönster. Detta kan förklaras med att de två arterna påverkats olika av olika istider eller att deras spridningsmönster, framför allt hur snabbt de spred sig, mellan istider varit olika. Mitokondrie-DNA påverkas lättare av en populationsreducering än vad MHC gör, med genetisk differentiering och förlust av genetisk variation som följd.

Först och främst har graden av MHC-variation hos igelkottar påverkats av demografiska faktorer associerade till populationsförändringar under och mellan istider. Förekomsten av artöverskridande polymorfi mellan *E. europaeus* och *E. concolor* visar dock på förekomsten av balanserande selektion över tid. Analysresultaten visade också att selektionstrycket på MHC-gener skulle kunna vara lägre hos igelkottar jämfört med andra däggdjur. Detta skulle kunna korreleras med den immunologiska påverkan av att gå i ide även om konfirmerande studier behövs hos fler arter för att klargöra detta.

Artikel II – Variation i MHC-promotorer hos varg och hund

Framför en MHC-gen finns en promotor. Dit binder transkriptionsfaktorer när genen transkripteras, d.v.s. när den uttrycks till ett protein. Studier hos människa och möss har visat variation i MHC-promotorer och att denna variation i många fall orsakar variationer i uttrycksnivåer hos MHC-genen.

I denna studie sekvenserades promotorregioner hos vargar från Finland och Ryssland och hundar från tio skandinaviska hundraser. Endast *DQB1*-promotorn visade hög grad av genetisk variation, vilket kan tyda på att för *DQB1* är ett flexibelt uttrycksmönster viktigt medan ett mer konserverat uttrycksmönster är viktigt för *DRB1*- och *DQA1*-generna. Inom *DQB1*-promotorn identifierade vi flera variabla positioner varav två fanns i regioner dit transkriptionsfaktorer direkt binder. Totalt hittades åtta *DQB1* promotor-alleler i varg och sju i hund. Sex av dessa delades mellan arterna.

Artikel III – Kombinationer av *DQB1* exon 2- och promotoralleler hos hund och varg

I denna studie var syftet att korrelera polymorfi i *DQB1*-promotorregionen med den i kodande regionen (exon 2). Mängden MHC som uttrycks kan direkt påverka vilken immunorespons som aktiveras. Allelspecifika uttrycksmönster borde vara en evolutionär fördel om en kodande allel som känner igen ett visst antigen är kopplad till en promotor som orsakar en immunreaktion som har förmågan att besegra just den aktuella antigenen.

Resultaten visade att promotorregionen och den kodande regionen oftast är kopplade. Detta var i sak inte överraskande eftersom regionerna ligger förhållandevis nära varandra. Dock visade hundar svagare associationer mellan regionerna jämfört med vargar, både mellan och inom raser. Detta skulle kunna orsakas av att slumpmässiga händelser och/eller selektion på andra saker, såsom beteende hos hundar, har överskuggat effekterna av de selektiva fördelar allelspecifika uttrycksmönster ger. En alternativ förklaring skulle kunna vara förekomsten av en ”hotspot” för rekombination hos hundar men att denna saknas hos vargar.

I denna studie jämfördes också graden av selektion som verkat på de respektive regionerna och det kunde konstateras att selektionstrycket på promotorregionen till och med var högre än på den kodande regionen.

Artikel IV – Evolutionära mönster i MHC-haplotyper hos hundar med diabetes

Hos hundar, liksom som hos människor, definieras MHC klass II haplotyper utifrån exon-alleler av *DRB1*, *DQA1* och *DQB1*. Flera sådana haplotyper har associerats till autoimmuna sjukdomar. Kombinationer av alleler ärvs ofta tillsammans vilket ger upphov till starka associationer mellan alleler. Detta kallas linkage disequlibrium (LD). I denna studie studerades huruvida dessa kombinationer reflekterar konserverade kombinationer av alleler som har ärvts tillsammans utan förekomst av rekombination eller om kombinationer har uppstått oberoende av varandra. I det senare fallet kan det vara svårt att hitta mutationer som direkt påverkar sjukdomsmottagligheten om dessa inte finns inom de kodande regionerna.

I denna studie användes hundar från Australien, varav drygt hälften diagnostiserats med diabetes mellitus (motsvarande diabetes typ I). För att kunna analysera den genetiska bakgrunden i MHC klass II regionen genotypades bestämda enbaspolymorfier utmed hela regionen.

Resultaten visade att vissa exon 2-baserade haplotyper har olika genetisk bakgrund och därmed utgör haplotyper där kombinationer av alleler uppstått oberoende av varandra och troligtvis bibehålls av selektion. I ett annat fall var den genetiska bakgrunden identisk men med olika exon 2-alleler. Tidigare beskrivna haplotyper som kopplats till diabetes hittades bland hundarna. Resultaten visade att den region som medför ökad risk för diabetes troligtvis finns utanför de kodande regionerna.

Även *DQB1*-promotorn analyserades i dessa prover. Fyra alleler, som inte hittats i artikel II, hittades endast bland sjuka hundar i denna studie. Dessutom fanns det bland sjuka hundar en fixerad skillnad mellan haplotyper som medför ökad risk för diabetes och en haplotyp som normalt medför minskad risk för diabetes. Denna skillnad fanns dock inte mellan dessa haplotyper bland friska hundar. Detta skulle kunna betyda att den skyddande haplotypen endast ger skydd om den är kopplad till en viss promotorvariant.

Artikel V – Linkage disequilibrium och haplotypmönster i MHC klass II regionen hos vargar och hundar

Syftet med studien var att vidareutveckla resultaten från artikel IV. Då resultaten där visat på överraskande korta regioner med högt LD i MHC klass II regionen hos hundar ville vi jämföra dem med vargar för att bättre förstå om evolutionära krafter påverkats av domesticering och rasbildning.

I denna studie användes åter skandinaviska hundar och vargar från Finland och Ryssland. Bland både hundar och vargar hittades exon 2-haplotyper där kombinationer av alleler uppstått oberoende av varandra. Troligtvis verkar selektion både för att bibehålla fördelaktiga kombinationer och för att rensa bort dåliga kombinationer. Rekombinationsmönster såväl som associationer mellan olika exon 2-alleler var striktare hos vargar än hundar. Liksom med associationen mellan promotor och kodande regioner har troligtvis slumpmässiga händelser och/eller selektion på andra saker överskuggat effekterna av selektion på kombinationer av exon 2-alleler. Detta skulle kunna medföra att hundar, framför allt inom vissa raser, blir mer infektionskänsliga eller är mer mottagliga för sjukdomar.

Resultaten visade också att *DQ*-alleler i större utsträckning delades mellan vargar och hundar jämfört med *DRB1*-alleler och vidare att *DRB1*-alleler har adderats till konserverade *DQ*-haplotyper genom rekombination.

Mina viktigaste slutsatser

Jag använde framgångsrikt MHC i en populationstudie på igelkottar och visade att MHC, som komplement till andra markörer, kan möjliggöra mer långtdragna slutsatser om hur historiska händelser har påverkat arters genetiska variation. Min forskning har visat den stora vikten av att inte fokusera all uppmärksamhet på de kodande regionerna av MHC-gener. Variation i regioner som styr uttryck av dessa gener kan vara av minst lika stor betydelse. För att förstå hur MHC-haplotyper evolverar krävs att den genetiska bakgrunden inom regionen analyseras. På så vis kan t.ex. sjukdomsassocierade mutationer i MHC-regionen utanför de kodande regionerna hittas. Preliminära resultat påvisade t.ex. en specifik mutation i *DQB1*-promotorn som skulle kunna var viktig för uppkomsten av diabetes mellitus hos hundar. Hundar uppvisar svagare koppling mellan alleler av olika MHC klass II gener så väl som svagare koppling mellan kodande regionen och promotorn av *DQB1*-genen. Detta skulle kunna förklaras med att populationsförändringar associerade till domesticering och rasbildning medfört otillräcklig effekt av selektion på dessa allelkombinationer. Detta skulle kunna göra hundar mer känsliga för infektioner och mer mottagliga för vissa sjukdomar.

Acknowledgements

Thank you so much for in various ways helping me to accomplish this thesis,

Jenny Seddon, my much appreciated supervisor. You have taught me so much about doing and presenting research. You have shown that 15000 kilometers and eight hours time difference isn't reason for not being an excellent supervisor. You are also a great person and friend.
Hans Ellegren, for accepting me as a PhD-student and most of all for trusting me to plan my research projects and follow my research interest all the way through my PhD. Also for letting me have such a big part in undergraduate teaching, something I have really enjoyed.
Carles Vilà and **Jennifer Leonard**, for support and great advice on my research projects and useful comments on manuscripts. For teaching me a lot about the population biology of wolves and the domestication of dogs.
Göran Andersson, for sharing my huge interest in MHC promoters in dogs and for becoming a great collaborator during recent research work.
Coauthors of my papers, Godfrey Hewitt and Linda Fleeman, for contributing with your areas of expertise, and a whole bunch of anonymous **reviewers** who have helped to increase the quality of the papers.
Appreciated **financial support** was obtained from the Swedish and Norwegian Environmental Protection Agencies, The Lawski foundation, The Royal Swedish Academy of Science, The Wallenberg Foundation and The C F Liljewalchs Foundation.
My office mates during the years, most of all **Nilla**, for nice chats and for introducing me to "Herrhagen"; **Susanne**, for all the laughter we had and for making friends with the "ghost"; **Ülo** for being such a friendly person and for patiently putting up with all the girls; **Rebecka** for genuinely being such a sweet person and for lots of encouragement during the end my PhD.
Members of the conservation/ quaternary genetics group, **Carles**, **Jennifer**, **Linda**, **Frank**, **Eva**, **Anna-Karin**, **Susanne**, **Violetta**, **Ülo**, **Santi**, **Robert** and temporary visitors, for great discussions about population genetics in general and for listening to all my MHC talk.
Nilla, **Malin** and **Cecilia** for keeping the lab in such a good shape and for always being kind and helpful. A special thanks to Nilla for help with last minute lab work. To **Carolina** and other administrators for taking care of so many practical issues at the department. **Sofia** and **Niclas** for always being

so friendly and for advice on tricky aspects in molecular evolution. **Frank**, **Mikael**, **Håkan** and **Axel** for much appreciated help with computer issues. **Matthew**, **Hanna**, **Björn** and **Mattias** for organizing educational and enjoyable seminar series.
Tanja for sharing my interest in MHC and for appreciated comments on this thesis as well as on research and conference presentations and constant encouragement while writing this thesis. (I get back to you below :)).
Emma, for sharing "thesis-panic" this spring. For email correspondence late at night while writing our theses. For great company with our families in Barcelona and for being a good friend.
Other past and present members at the department of Evolutionary Biology for making it fun to go to work every day. I could not have wished for nicer people to work with.

And to all of you who made life outside work most important,

All family friends and relatives that have been important throughout my life, most of all my godmother **Kerstin W** and family and next door neighbor **Lennart** and family, as well as all my great aunts and uncles.
The "Herrhagen" families for great dinner parties, and a lot of fun with our kids. Especially all the mums for nice and totally free speaking Monday lunches. To other friends, with whom we often spend nice evenings.
Karin, **Emma** and **Anna Lö**, for maintaining friendship over many years.
Maria and **Thomas** (**Isak** and **Evelina**) for many fun evenings and weekends, for always being helpful in all matters and for being wonderful godparents to Kerstin.
Anna Lu and family. You are such a great friend and connected to so many of my most precious and unforgettable memories in Uppsala.
Tanja, my fantastic friend in better and worse. Your friendship means so much to me. And **Robin**, **Elliot** and **Elvira** for all the great times with our families. Elvira, I will try to be the best "fadder" ever.
My brother **Anders**, **Louise**, **Astrid** and **Harald**, for being important people in our life, for fun times and traditional New Year celebrations.
Mamma and **Pappa**, for always making me feel like number one, for making me trust myself enough to face most challenges in life. From the start making me appreciate outdoor life and all the beauty in it.
STINA and **KERSTIN**, my beautiful, wonderful daughters. There is nothing more fantastic or precious in my life. You make coming home the best moment of the day. I love you so much!
Patrik, my darling husband. Thank you for taking care of everything while this thesis was being produced. For constant encouragement, love and support and for keeping my feet on the ground, during the last month and always.

References

Abbas AK, Lichtman AH, Pober JS (2000) Cellular and Molecular Immunology. W.B. Saunders company, Philadelphia

Andersen LC, Beaty JS, Nettles JW, Seyfried CE, Nepom GT, Nepom BS (1991) Allelic polymorphism in transcriptional regulatory regions of HLA-*DQB* genes. Journal of Experimental Medicine 173:181-192

Ardlie KG, Kruglyak L, Seielstad M (2002) Patterns of linkage disequilibrium in the human genome. Nature Reviews Genetics 3:299-309

Arkush KD, Giese AR, Mendonca HL, McBride AM, Marty GD, Hedrick PW (2002) Resistance to three pathogens in the endangered winter-run chinook salmon (*Oncorhynchus tshawytscha*): effects of inbreeding and major histocompatibility complex genotypes. Canadian Journal of Fisheries and Aquatic Sciences 59:966-975

Aspi J, Roininen E, Ruokonen M, Kojola I, Vila C (2006) Genetic diversity, population structure, effective population size and demographic history of the Finnish wolf population. Molecular Ecology 15:1561-1576

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263-265

Baumgart M, Moos V, Schuhbauer D, Muller B (1998) Differential expression of major histocompatibility complex class II genes on murine macrophages associated with T cell cytokine profile and protective/suppressive effects. Proceedings of the National Academy of Sciences of the United States of America 95:6936-6940

Beck S, Geraghty D, Inoko H, Rowen L, Aguado B, Bahram S, Campbell RD, Forbes SA, Guillaudeux T, Hood L, Horton R, Janer M, Jasoni C, Madan A, Milne S, Neville M, Oka A, Qin S, Ribas-Despuig G, Rogers J, Shiina T, Spies T, Tamiya G, Tashiro H, Trowsdale J, Vu Q, Williams L, Yamazaki M (1999) Complete sequence and gene map of a human major histocompatibility complex. Nature 401:921-923

Benoist C, Mathis D (1990) Regulation of major histocompatibility complex class-II genes: X, Y and other letters of the alphabet. Annual Review of Immunology 8:681-715

Bergstrom TF, Josefsson A, Erlich HA, Gyllensten U (1998) Recent origin of HLA-*DRB1* alleles and implications for human evolution. Nature Genetics 18:237-242

Bernatchez L, Landry C (2003) MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? Journal of Evolutionary Biology 16:363-377

Bjornerfeldt S, Webster MT, Vila C (2006) Relaxation of selective constraint on dog mitochondrial DNA following domestication. Genome Research 16:990-994

Borghans JAM, Beltman JB, De Boer RJ (2004) MHC polymorphism under host-pathogen coevolution. Immunogenetics 55:732-739

Boyer BB, Barnes BM (1999) Molecular and metabolic aspects of mammalian hibernation - Expression of the hibernation phenotype results from the coordinated regulation of multiple physiological and molecular events during preparation for and entry into torpor. Bioscience 49:713-724

Brown JH, Jardetzky T, Saper MA, Samraoui B, Bjorkman PJ, Wiley DC (1988) A hypothetical model of the foreign antigen-binding site of class-II histocompatibility molecules. Nature 332:845-850

Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL, Wiley DC (1993) 3-dimensional structure of the human class-II histocompatibility antigen Hla-*DR1*. Nature 364:33-39

Bruford MW, Bradley DG, Luikart G (2003) DNA markers reveal the complexity of livestock domestication. Nature Reviews Genetics 4:900-910

Burton RS, Reichman OJ (1999) Does immune challenge affect torpor duration? Functional Ecology 13:232-237

Caillat-Zucman S (2009) Molecular mechanisms of HLA association with autoimmune diseases. Tissue Antigens 73:1-8

Carrington M, Nelson GW, Martin MP, Kissner T, Vlahov D, Goedert JJ, Kaslow R, Buchbinder S, Hoots K, O'Brien SJ (1999) HLA and HIV-1: Heterozygote advantage and B*35-Cw*04 disadvantage. Science 283:1748-1752

Clarke B, Kirby DRS (1966) Maintenance of histocompatibility polymorphisms. Nature 211:999-&

Cowell LG, Kepler TB, Janitz M, Lauster R, Mitchison NA (1998) The distribution of variation in regulatory gene segments, as present in MHC class II promoters. Genome Research 8:124-134

Cruz F, Vila C, Webster MT (2008) The legacy of domestication: accumulation of deleterious mutations in the dog genome. Molecular Biology and Evolution 25:2331-2336

de Bakker PIW, McVean G, Sabeti PC, Miretti MM, Green T, Marchini J, Ke XY, Monsuur AJ, Whittaker P, Delgado M, Morrison J, Richardson A, Walsh EC, Gao XJ, Galver L, Hart J, Hafler DA, Pericak-Vance M, Todd JA, Daly MJ, Trowsdale J, Wijmenga C, Vyse TJ, Beck S, Murray SS, Carrington M, Gregory S, Deloukas P, Rioux JD (2006) A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. Nature Genetics 38:1166-1172

De Boer RJ, Borghans JAM, van Boven M, Kesmir C, Weissing FJ (2004) Heterozygote advantage fails to explain the high degree of polymorphism of the MHC. Immunogenetics 55:725-731

Doherty PC, Zinkernagel RM (1975) Enhanced immunological surveillance in mice heterozygous at H-2 gene complex. Nature 256:50-52

Edwards SV, Potts WK (1996) Polymorphism of genes in the major histocompatibility complex (MHC): Implications for conservation genetics of vertebrates. In: Smith TB, Wayne RK (eds) Molecular genetic approaches in conservation. Oxford University Press, Oxford, p 214-237

Egid K, Brown JL (1989) The major histocompatibility complex and female mating preferences in mice. Animal Behaviour 38:548-550

Ellegren H, Hartman G, Johansson M, Andersson L (1993) Major histocompatibility complex monomorphism and low-levels of DNA-fingerprinting variability in a reintroduced and rapidly expanding population of beavers. Proceedings of the National Academy of Sciences of the United States of America 90:8150-8153

Ellegren H, Savolainen P, Rosen B (1996) The genetical history of an isolated population of the endangered grey wolf *Canis lupus*: A study of nuclear and mitochondrial polymorphisms. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 351:1661-1669

Erlich H, Valdes AM, Noble J, Carlson JA, Varney M, Concannon P, Mychaleckyj JC, Todd JA, Bonella P, Fear AL, Lavant E, Louey A, Moonsamy P (2008) HLA *DR-DQ* haplotypes and genotypes and type 1 diabetes risk: analysis of the type 1 diabetes genetics consortium families. Diabetes 57:1084-1092

Figueroa F, Gunther E, Klein J (1988) MHC polymorphism predating speciation. Nature 335:265-267

Frank SA (2002) Immunology and evolution of infectious disease. Princeton University press, Princeton and Oxford

Garrigan D, Hedrick PN (2001) Class I MHC polymorphism and evolution in endangered California Chinook and other Pacific salmon. Immunogenetics 53:483-489

Garrigan D, Hedrick PW (2003) Perspective: detecting adaptive molecular polymorphism: lessons from the MHC. Evolution 57:1707-1722

Glimcher LH, Kara CJ (1992) Sequences and factors: A guide to MHC class-II transcription. Annual Review of Immunology 10:13-49

Graser R, OhUigin C, Vincek V, Meyer A, Klein J (1996) Trans-species polymorphism of class II MHC loci in danio fishes. Immunogenetics 44:36-48

Gray MM, Granka JM, Bustamante CD, Sutter NB, Boyko AR, Zhu L, Ostrander EA, Wayne R (2009) Linkage disequilibrium and demographic history of wild and domestic canids. Genetics 181:1493-1505

Guardiola J, Maffei A, Lauster R, Mitchison NA, Accolla RS, Sartoris S (1996) Functional signficance of polymorphism among MHC class II gene promoters. Tissue Antigens 48:615-625

Gyllensten UB, Erlich HA (1989) Ancient roots for polymorphism at the HLA-*Dq*-Alpha-Locus in primates. Proceedings of the National Academy of Sciences of the United States of America 86:9986-9990

Hedrick PW (1992) Female choice and variation in the major histocompatibility complex. Genetics 132:575-581

Hedrick PW (1994) Evolutionary genetics of the major histocompatibility complex. American Naturalist 143:945-964

Hedrick PW, Kim TJ, Parker KM (2001a) Parasite resistance and genetic variation in the endangered *Gila topminnow*. Animal Conservation 4:103-109

Hedrick PW, Parker KM, Lee RN (2001b) Using microsatellite and MHC variation to identify species, ESUs, and MUs in the endangered *Sonoran topminnow*. Molecular Ecology 10:1399-1412

Hedrick PW, Thomson G (1983) Evidence for balancing selection at HLA. Genetics 104:449-456

Hedrick PW, Thomson G (1988) Maternal-fetal interactions and the maintenance of HLA polymorphism. Genetics 119:205-212

Hedrick PW, Whittam TS, Parham P (1991) Heterozygosity at individual amino-acid sites - extremely high-levels for HLA-A and HLA-B genes. Proceedings of the National Academy of Sciences of the United States of America 88:5897-5901

Heldt C, Listing J, Sozeri O, Blasing F, Frischbutter S, Muller B (2003) Differential expression of HLA class II genes associated with disease susceptibility and progression in rheumatoid arthritis. Arthritis and Rheumatism 48:2779-2787

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. Biological Journal of the Linnean Society 58:247-276

Hewitt GM (1999) Post-glacial re-colonization of European biota. Biological Journal of the Linnean Society 68:87-112

Hill AVS, Allsopp CEM, Kwiatkowski D, Anstey NM, Twumasi P, Rowe PA, Bennett S, Brewster D, McMichael AJ, Greenwood BM (1991) Common west African HLA antigens are associated with protection from severe malaria. Nature 352:595-600

Hill RE, Hastie ND (1987) Accelerated evolution in the reactive center regions of serine protease inhibitors. Nature 326:96-99

Hoelzel AR, Stephens JC, O'Brien SJ (1999) Molecular genetic diversity and evolution at the MHC *DQB* locus in four species of pinnipeds. Molecular Biology and Evolution 16:611-618

Hughes AL (1991) MHC polymorphism and the design of captive breeding programs. Conservation Biology 5:249-251

Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class-I loci reveals overdominant selection. Nature 335:167-170

Hughes AL, Nei M (1989) Nucleotide substitution at major histocompatibility complex class-II loci - evidence for overdominant selection. Proceedings of the National Academy of Sciences of the United States of America 86:958-962

Hughes AL, Yeager M (1998) Natural selection at major histocompatibility complex loci of vertebrates. Annual Review of Genetics 32:415-435

Ibrahim KM, Nichols RA, Hewitt GM (1996) Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. Heredity 77:282-291

Innan H, Kim Y (2004) Pattern of polymorphism after strong artificial selection in a domestication event. Proceedings of the National Academy of Sciences of the United States of America 101:10667-10672

Janitz M, Mitchison A, Reiners-Schramm L, Lauster R (1997) Polymorphic MHC class II promoters exhibit distinct expression pattern in various antigen-presenting cell lines. Tissue Antigens 49:99-106

Jordan WC, Bruford MW (1998) New perspectives on mate choice and the MHC. Heredity 81:239-245

Kalbe M, Eizaguirre C, Dankert I, Reusch TBH, Sommerfeld RD, Wegner KM, Milinski M (2009) Lifetime reproductive success is maximized with optimal major histocompatibility complex diversity. Proceedings of the Royal Society B-Biological Sciences 276:925-934

Kelley J, Walter L, Trowsdale J (2005) Comparative genomics of major histocompatibility complexes. Immunogenetics 56:683-695

Kennedy LJ, Barnes A, Happ GM, Quinnell RJ, Bennett D, Angles JM, Day MJ, Carmichael N, Innes JF, Isherwood D, Carter SD, Thomson W, Ollier WER (2002) Extensive interbreed, but minimal intrabreed, variation of DLA class II alleles and haplotypes in dogs. Tissue Antigens 59:194-204

Kennedy LJ, Barnes A, Short A, Brown JJ, Lester S, Seddon J, Fleeman L, Francino O, Brkljacic M, Knyazev S, Happ GM, Ollier WER (2007a) Canine DLA diversity: 1. New alleles and haplotypes. Tissue Antigens 69:272-288

Kennedy LJ, Barnes A, Short A, Brown JJ, Seddon J, Fleeman L, Brkljacic M, Happ GM, Catchpole B, Ollier WER (2007b) Canine DLA diversity: 3. Disease studies. Tissue Antigens 69:292-296

Kennedy LJ, Davison LJ, Barnes A, Short AD, Fretwell N, Jones CA, Lee AC, Ollier WER, Catchpole B (2006) Identification of susceptibility and protective major histocompatibility complex haplotypes in canine diabetes mellitus. Tissue Antigens 68:467-476

Kim Y, Nielsen R (2004) Linkage disequilibrium as a signature of selective sweeps. Genetics 167:1513-1524

Kirkness EF, Bafna V, Halpern AL, Levy S, Remington K, Rusch DB, Delcher AL, Pop M, Wang W, Fraser CM, Venter JC (2003) The dog genome: survey sequencing and comparative analysis. Science 301:1898-1903

Klein J (1987) Origin of major histocompatibility complex polymorphism - the transspecies hypothesis. Human Immunology 19:155-162
Klein J, Gutknecht J, Fischer N (1990) The major histocompatibility complex and human-evolution. Trends in Genetics 6:7-11
Klein J, Sato A, Nagl S, O'HUigin C (1998) Molecular trans-species polymorphism. Annual Review of Ecology and Systematics 29:1-+
Knip M, Siljander H (2008) Autoimmune mechanisms in type 1 diabetes. Autoimmunity Reviews 7:550-557
Kojola I, Aspi J, Hakala A, Heikkinen S, Ilmoni C, Ronkainen S (2006) Dispersal in an expanding wolf population in Finland. Journal of Mammalogy 87:281-286
Kruger A, Quack P, Schneider PM, Rittner C, Hohler T (2001) Sequence analysis of the *DRB1* promoter reveals limited polymorphism with no influence on gene expression. Genes and Immunity 2:211-215
Langefors A, Lohm J, Grahn M, Andersen O, von Schantz T (2001) Association between major histocompatibility complex class IIB alleles and resistance to *Aeromonas salmonicida* in Atlantic salmon. Proceedings of the Royal Society of London Series B-Biological Sciences 268:479-485
Leonard JA, Vila C, Wayne RK (2005) Legacy lost: genetic variability and population size of extirpated US grey wolves (*Canis lupus*). Molecular Ecology 14:9-17
Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, Clamp M, Chang JL, Kulbokas EJ, Zody MC, Mauceli E, Xie XH, Breen M, Wayne RK, Ostrander EA, Ponting CP, Galibert F, Smith DR, deJong PJ, Kirkness E, Alvarez P, Biagi T, Brockman W, Butler J, Chin CW, Cook A, Cuff J, Daly MJ, DeCaprio D, Gnerre S, Grabherr M, Kellis M, Kleber M, Bardeleben C, Goodstadt L, Heger A, Hitte C, Kim L, Koepfli KP, Parker HG, Pollinger JP, Searle SMJ, Sutter NB, Thomas R, Webber C, Lander ES (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature 438:803-819
Lohm J, Grahn M, Langefors A, Andersen O, Storset A, von Schantz T (2002) Experimental evidence for major histocompatibility complex-allele-specific resistance to a bacterial infection. Proceedings of the Royal Society of London Series B-Biological Sciences 269:2029-2033
Louis P, Vincent R, Cavadore P, Clot J, Eliaou JF (1994) Differential transcriptional activities of HLA-*DR* genes in the various haplotypes. Journal of Immunology 153:5059-5067
Mach B, Steimle V, Martinez-Soria E, Reith W (1996) Regulation of MHC class II genes: lessons from a disease. Annual Review of Immunology 14:301-331
Mach B, Steimle V, Reith W (1994) MHC class II-deficient combined immunodeficiency – a disease of gene regulation. Immunological Reviews 138:207-221

Madden DR (1995) The 3-dimensional structure of peptide-MHC complexes. Annual Review of Immunology 13:587-622
Martinsohn JT, Sousa AB, Guethlein LA, Howard JC (1999) The gene conversion hypothesis of MHC evolution: a review. Immunogenetics 50:168-200
Mayer WE, Ohuigin C, Zaleskarutczynska Z, Klein J (1992) Transspecies Origin of MHC-*DRB* polymorphism in the chimpanzee. Immunogenetics 37:12-23
McConnell TJ, Talbot WS, McIndoe RA, Wakeland EK (1988) The Origin of MHC class-II gene polymorphism within the genus *Mus*. Nature 332:651-654
Meyer-Lucht Y, Sommer S (2005) MHC diversity and the association to nematode parasitism in the yellow-necked mouse (*Apodemus flavicollis*). Molecular Ecology 14:2233-2243
Meyer D, Thomson G (2001) How selection shapes variation of the human major histocompatibility complex: a review. Annals of Human Genetics 65:1-26
Mikko S, Andersson L (1995) Low major histocompatibility complex class-II diversity in European and North-American moose. Proceedings of the National Academy of Sciences of the United States of America 92:4259-4263
Mitchison NA, Roes J (2002) Patterned variation in murine MHC promoters. Proceedings of the National Academy of Sciences of the United States of America 99:10561-10566
Mitchison NA, Schuhbauer D, Muller B (1999) Natural and induced regulation of Th1/Th2 balance. Springer Seminars in Immunopathology 21:199-210
Mueller-Hilke B, Mitchison NA (2006) The role of HLA promoters in autoimmunity. Current Pharmaceutical Design 12:3743-3752
Murray BW, Malik S, White BN (1995) Sequence variation at the major mistocompatibility complex locus *DQ*-Beta in Beluga Whales (*Delphinapterus-Leucas*). Molecular Biology and Evolution 12:582-593
Murray BW, White BN (1998) Sequence variation at the major histocompatibility complex *DRB* loci in beluga (*Delphinapterus leucas*) and narwhal (*Monodon monoceros*). Immunogenetics 48:242-252
Nowak MA, Tarczyhornoch K, Austyn JM (1992) The optimal number of major histocompatibility complex-molecules in an individual. Proceedings of the National Academy of Sciences of the United States of America 89:10896-10899
O´Brien SJ, Roelke ME, Marker L, Newman A, Winkler CA, Meltzer D, Colly L, Evermann JF, Bush M, Wildt DE (1985) Genetic-basis for species vulnerability in the cheetah. Science 227:1428-1434
Oliver MK, Telfer S, Piertney SB (2009) Major histocompatibility complex (MHC) heterozygote superiority to natural multi-parasite infections

in the water vole (*Arvicola terrestris*). Proceedings of the Royal Society B-Biological Sciences 276:1119-1128
Orita M, Suzuki Y, Sekiya T, Hayashi K (1989) Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. Genomics 5:874-879
Ostrander EA, Giniger E (1997) Semper Fidelis: What man's best friend can teach us about human biology and disease. American Journal of Human Genetics 61:475-480
Ostrander EA, Wayne RK (2005) The canine genome. Genome Research 15:1706-1716
Parham P, Lomen CE, Lawlor DA, Ways JP, Holmes N, Coppin HL, Salter RD, Wan AM, Ennis PD (1988) Nature of polymorphism in HLA-A, HLA-B, and HLA-C Molecules. Proceedings of the National Academy of Sciences of the United States of America 85:4005-4009
Parker HG, Kim LV, Sutter NB, Carlson S, Lorentzen TD, Malek TB, Johnson GS, DeFrance HB, Ostrander EA, Kruglyak L (2004) Genetic structure of the purebred domestic dog. Science 304:1160-1164
Paterson S, Wilson K, Pemberton JM (1998) Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population (*Ovis aries L.*). Proceedings of the National Academy of Sciences of the United States of America 95:3714-3719
Perfetto C, Zacheis M, McDaid D, Meador III JW, Schwartz BD (1993) Polymorphism in the promoter region of HLA-*DRB* genes. Human Immunology 36:27-33
Piertney SB, Oliver MK (2006) The evolutionary ecology of the major histocompatibility complex. Heredity 96:7-21
Potts WK, Manning CJ, Wakeland EK (1994) The role of infectious-disease, inbreeding and mating preferences in maintaining MHC genetic diversity - an experimental test. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 346:369-378
Pulliainen E (1980) The status, structure and behavior of populations of the wolf (*Canis-L-Lupus L*) Along the Fenno-Soviet border. Annales Zoologici Fennici 17:107-112
Radosevich M, Ono SJ (2003) Novel mechanisms of class II major histocompatibility complex gene regulation. Immunologic research 27:85-105
Raymond CK, Kas A, Paddock M, Qiu RL, Zhou Y, Subramanian S, Chang J, Palmieri A, Haugen E, Kaul R, Olson MV (2005) Ancient haplotypes of the HLA class II region. Genome Research 15:1250-1257
Reeve N (1994) Hedgehogs. T & AD Poyser (Natural History), London
Reichstetter S, Krellner PH, Meenzen CM, Kalden JR, Wassmuth R (1994) Comparative analysis of sequence variability in the upstream

regulatory region of the HLA-*DQB1* gene. Immunogenetics 39:207-212

Richardson DS, Westerdahl H (2003) MHC diversity in two *Acrocephalus* species: the outbred Great reed warbler and the inbred Seychelles warbler. Molecular Ecology 12:3523-3529

Roy MS, Geffen E, Smith D, Ostrander EA, Wayne RK (1994) Patterns of differentiation and hybridization in North-American wolflike canids, revealed by analysis of microsatellite loci. Molecular Biology and Evolution 11:553-570

Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, Ackerman HC, Campbell SJ, Altshuler D, Cooper R, Kwiatkowski D, Ward R, Lander ES (2002) Detecting recent positive selection in the human genome from haplotype structure. Nature 419:832-837

Saetre P, Lindberg J, Leonard JA, Olsson K, Pettersson U, Ellegren H, Bergstrom TF, Vila C, Jazin E (2004) From wild wolf to domestic dog: gene expression changes in the brain. Molecular Brain Research 126:198-206

Santucci F, Emerson BC, Hewitt GM (1998) Mitochondrial DNA phylogeography of European hedgehogs. Molecular Ecology 7:1163-1172

Savolainen P, Zhang YP, Luo J, Lundeberg J, Leitner T (2002) Genetic evidence for an East Asian origin of domestic dogs. Science 298:1610-1613

Schwaiger FW, Gostomski D, Stear MJ, Duncan JL, McKellar QA, Epplen JT, Buitkamp J (1995) An Ovine major histocompatibility complex *DRB1* allele is associated with low fecal egg counts following natural, predominantly *Ostertagia-Circumcincta* Infection. International Journal for Parasitology 25:815-822

Seddon JM, Baverstock PR (1999) Variation on islands: major histocompatibility complex (MHC) polymorphism in populations of the Australian bush rat. Molecular Ecology 8:2071-2079

Seddon JM, Ellegren H (2002) MHC class II genes in European wolves: a comparison with dogs. Immunogenetics 54:490-500

Seddon JM, Santucci F, Reeve NJ, Hewitt GM (2001) DNA footprints of European hedgehogs, *Erinaceus europaeus* and *E.concolor*. Pleistocene refugia, postglacial expansion and colonization routes. Molecular Ecology 10:2187-2198

Shewey LM, Beaty JS, Andersen LC, Nepom GT (1992) Differential expression of related HLA class II *DQ* beta genes caused by nucleotide variation in transcriptional regulatory elements. Journal of Immunology 148:1265-1273

Shiina T, Inoko H, Kulski JK (2004) An update of the HLA genomic region, locus information and disease associations: 2004. Tissue Antigens 64:631-649

Singal DP, Qiu X (1995) Polymorphism in both X and Y box motifs controls level of expression of HLA-*DRB1* genes. Immunogenetics 43:50-56
Slade RW (1992) Limited MHC polymorphism in the Southern elephant seal - implications for MHC evolution and marine mammal population biology. Proceedings of the Royal Society of London Series B-Biological Sciences 249:163-171
Snell GD (1948) Methods for the study of histocompatibility genes. Journal of Genetics 49:87-108
Spurgin LG, Richardson DS (2010) How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. Proceedings of the Royal Society B-Biological Sciences 277:979-988
Stephens M, Donnelly P (2003) A comparison of Bayesian methods for haplotype reconstruction from population genotype data. American Journal of Human Genetics 73:1162-1169
Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. American Journal of Human Genetics 68:978-989
Streelman JT, Kocher TD (2000) From phenotype to genotype. Evolution & Development 2:166-173
Sutter NB, Eberle MA, Parker HG, Pullar BJ, Kirkness EF, Kruglyak L, Ostrander EA (2004) Extensive and breed-specific linkage disequilibrium in *Canis familiaris*. Genome Research 14:2388-2396
Svartberg K (2006) Breed-typical behaviour in dogs - Historical remnants or recent constructs? Applied Animal Behaviour Science 96:293-313
Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF (1998) Comparative phylogeography and postglacial colonization routes in Europe. Molecular Ecology 7:453-464
Takahata N, Nei M (1990) Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. Genetics 124:967-978
Thomas ML, Harger JH, Wagener DK, Rabin BS, Gill TJ (1985) HLA sharing and spontaneous-abortion in humans. American Journal of Obstetrics and Gynecology 151:1053-1058
Ting JP-Y, Trowsdale J (2002) Genetic control of MHC class II expression. Cell 109 (Suppl):S21-S33
Trowsdale J (1995) Both man and bird and beast - comparative organization of MHC genes. Immunogenetics 41:1-17
Trowsdale J, Groves V, Arnason A (1989) Limited MHC polymorphism in whales. Immunogenetics 29:19-24
Wagner JL (2003) Molecular organization of the canine major histocompatibility complex. Journal of Heredity 94:23-26
Wagner JL, Burnett RC, DeRose SA, Storb R (1996a) Molecular analysis and polymorphism of the DLA-*DQA* gene. Tissue Antigens 48:199-204
Wagner JL, Burnett RC, Storb R (1996b) Molecular analysis of the DLA *DR* region. Tissue Antigens 48:549-553

Wagner JL, Hayes-Lattin B, Works JD, Storb R (1998) Molecular analysis and polymorphism of the DLA-*DQB* genes. Tissue Antigens 52:242-250

van den Elsen PJ, Holling TM, Kuipers HF, van der Stoep N (2004) Transcriptional regulation of antigen presentation. Current Opinion in Immunology 16:67-75

van Oosterhout C (2009) A new theory of MHC evolution: beyond selection on the immune genes. Proceedings of the Royal Society B-Biological Sciences 276:657-665

Watterson GA (1978) Homozygosity test of neutrality. Genetics 88:405-417

Wayne RK, Lehman N, Allard MW, Honeycutt RL (1992) Mitochondrial-DNA variability of the gray wolf - genetic consequences of population decline and habitat fragmentation. Conservation Biology 6:559-569

Wegner KM, Reusch TBH, Kalbe M (2003) Multiple parasites are driving major histocompatibility complex polymorphism in the wild. Journal of Evolutionary Biology 16:224-232

Verardi A, Lucchini V, Randi E (2006) Detecting introgressive hybridization between free-ranging domestic dogs and wild wolves (*Canis lupus*) by admixture linkage disequilibrium analysis. Molecular Ecology 15:2845-2855

Westerdahl H, Waldenstrom J, Hansson B, Hasselquist D, von Schantz T, Bensch S (2005) Associations between malaria and MHC genes in a migratory songbird. Proceedings of the Royal Society B-Biological Sciences 272:1511-1518

Vila C, Amorim IR, Leonard JA, Posada D, Castroviejo J, Petrucci-Fonseca F, Crandall KA, Ellegren H, Wayne RK (1999) Mitochondrial DNA phylogeography and population history of the grey wolf *Canis lupus*. Molecular Ecology 8:2089-2103

Vila C, Savolainen P, Maldonado JE, Amorim IR, Rice JE, Honeycutt RL, Crandall KA, Lundeberg J, Wayne RK (1997) Multiple and ancient origins of the domestic dog. Science 276:1687-1689

Vila C, Seddon J, Ellegren H (2005) Genes of domestic mammals augmented by backcrossing with wild ancestors. Trends in Genetics 21:214-218

Vila C, Walker C, Sundqvist AK, Flagstad O, Andersone Z, Casulli A, Kojola I, Valdmann H, Halverson J, Ellegren H (2003) Combined use of maternal, paternal and bi-parental genetic markers for the identification of wolf-dog hybrids. Heredity 90:17-24

Woolfrey AE, Nepom GT (1995) Differential transcription elements direct expression of HLA-*DQ* genes. Clinical Immunology and Immunopathology 74:119-126

Wu SK, Saunders TL, Bach FH (1986) Polymorphism of human Ia antigens generated by reciprocal intergenic exchange between 2 *DR*-Beta loci. Nature 324:676-679

Yamazaki K, Beauchamp GK, Egorov IK, Bard J, Thomas L, Boyse EA (1983) Sensory distinction between H-2b and H-2bm1 mutant mice. Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences 80:5685-5688

Yuhki N, Beck T, Stephens RM, Nishigaki Y, Newmann K, O'Brien SJ (2003) Comparative genome organization of human, murine, and feline MHC class II region. Genome Research 13:1169-1179

Zinkernagel RM, Doherty PC (1974) Restriction of in-vitro T cell-mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semiallogeneic System. Nature 248:701-702

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 736

Editor: The Dean of the Faculty of Science and Technology

A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-122011

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2010