Alina Niskanen

SELECTION AND GENETIC DIVERSITY IN THE MAJOR HISTOCOMPATIBILITY COMPLEX GENES OF WOLVES AND DOGS
ALINA NISKANEN

SELECTION AND GENETIC DIVERSITY IN THE MAJOR HISTOCOMPATIBILITY COMPLEX GENES OF WOLVES AND DOGS

Academic dissertation to be presented with the assent of the Doctoral Training Committee of Technology and Natural Sciences of the University of Oulu for public defence in the lecture hall of the Botanical Museum (KM 129), Linnanmaa, on 1 November 2014, at 12 noon

UNIVERSITY OF OULU, OULU 2014
Niskanen, Alina, Selection and genetic diversity in the major histocompatibility complex genes of wolves and dogs.
University of Oulu Graduate School; University of Oulu, Faculty of Science, Department of Biology;
University of Oulu, P.O. Box 8000, FI-90014 University of Oulu, Finland

Abstract

Hosts and pathogens are involved in a continuous evolutionary arms race, where pathogens attack and hosts defend themselves. The main tools for winning the race are natural selection and the genetic diversity that selection acts on. However, in small populations natural selection may be ineffective. Therefore, the genes taking part in immune defense may lack adaptability to new or changing pathogens. Major histocompatibility complex (MHC) is an important genomic region that includes highly polymorphic immune defense genes. In this doctoral thesis, I studied the natural selection and genetic diversity of MHC class II genes in dogs and Finnish wolves. I also used dog MHC diversity to estimate the number of founding wolves in dog domestication.

The Finnish wolf population declined rapidly in size due to excessive hunting from the late 19th century until the early 20th century. After decades of a very small population size, the population started recovering in the mid-1990s. This study shows that, despite the fluctuations in population size, the diversity of the MHC loci in the Finnish wolf has remained high and comparable to the larger neighboring Russian Karelian wolf population. Unlike the neutral genetic markers, the MHC loci of the Finnish and Russian Karelian populations have not differentiated. These results indicate similar balancing selection acting on the MHC loci of the two wolf populations.

In dogs, the strength of natural selection is likely weakened by artificial selection and veterinary care. The potential phases of natural selection would be during embryogenesis and fetal development. However, no strong signs of prenatal selection were found in this study.

MHC diversity was estimated to be higher in Asian dogs than in dogs from Europe. A simulation study indicated a minimum of 500 founding wolves for the modern dog population. Dog MHC diversity implies an Asian origin for domestication from a large and diverse wolf population.

Both natural selection and demography have an influence on the genetic diversity of a species. In small populations, random genetic drift is enforced. However, in loci with important fitness impacts, selection may be particularly strong and outweigh drift, as demonstrated in the MHC loci of a small wolf population in this study.

Keywords: Canis familiaris, Canis lupus, domestication, genetic diversity, MHC, natural selection, parasite, prenatal selection
Niskanen, Alina, Luonnonvalinta ja geneettinen monimuotoisuus suuden ja koiran MHC-geeneissä.
Oulun yliopiston tutkijakoulu; Oulun yliopisto, Luonnontieteellinen tiedekunta, Biologian laitos
Oulun yliopisto, PL 8000, 90014 Oulun yliopisto

Tiivistelmä


Keinotekoinen valinta ja eläinlääketieteellinen hoito todennäköisesti heikentävät koirien MHC-geeneihin kohdistuvaa luonnonvalintaa. Luonnonvalinta voisi yhä vaikuttaa alkion- ja sikionkehityksen aikana, mutta tästä ei tutkimuksessa löytynyt todisteita. MHC-muuntelun määrän arvioitiin olevan suurempaa aasialaisissa kuin eurooppalaisissa koirissa. Simulaatiotutkimuksen mukaan nykyisen koirapopulaation perustamiseen olisi tarvittu vähintään 500 sutta. Tulokset viittavat koiran kesyttämisen tapahtumisen Aasiassa suuresta ja monimuotoisesta susipopulaatiosta.

Sekä luonnonvalinta että demografia vaikuttavat lajien geneettiseen monimuotoisuuteen. Pienissä populaatioissa satunnaisajautuminen voimistuu. Valinta voi kuitenkin olla erityisen voimakasta ja voitettaa satunnaisajautumisen geneesissä, joilla on erityisen tärkeä vaikutus yksilön kelpoisuuteen, kuten tutkimuksessa osoitettiin pienen susipopulaation MHC-geenien kohdalla.

Asiakirjat: Canis familiaris, Canis lupus, geneettinen monimuotoisuus, kesyttäminen, loinen, luonnonvalinta, MHC, syntymää edeltävä valinta
Thank you for herding me to my field
Acknowledgements

I started this project in autumn 2007 as a Master’s student and continued it a year later as a PhD student. During these seven years, many people have crossed my path and contributed to the studies. I want to thank you all - even if I forget to mention your name here due to shortage of sleep.

First of all I wish to thank my supervisors: Professor Jouni Aspi, Professor Hannes Lohi, Docent Tanja Pyhäjärvi and Docent Minna Ruokonen. Jouni, thank you for bringing together wolf and dog genetics in the first place and including me in the project. As an experienced scientist you have taught me a great deal about how to do science. Hannes, thank you for giving me the opportunity to work with dogs - a subject that has long been close to my heart. Your scientific enthusiasm and wealth of ideas is exemplary for younger scientists. Tanja, you joined the supervisor team less than two years ago and now I wonder how I ever got along without you! Thank you for being so exact and proficient and never letting me take the easy way out. Unfortunately, Minna passed away all too soon and she is not here today to witness the end of my PhD project. She always found time to discuss both work- and dog-related matters and supported me when I most needed it. I am grateful that I got the opportunity to get to know Minna and I miss her deeply.

I owe my gratitude to all the co-authors for their scientific input to these studies. I especially wish to thank Dr Peter Savolainen for giving me the opportunity to participate in the dog domestication project and Dr Lorna Kennedy for introducing the MHC allele world to me. I am grateful to Professor Philip Hedrick and Docent Anti Vasemägi for reviewing this thesis. I also want to thank the members of my follow-up group, Professor Outi Savolainen, Docent Laura Kvist and Dr Helmi Kuittinen.

I wish to thank all the current and former members of the wildlife genetics group: Jenni, Eeva, Marja, Matti, Liisa, Minna T, Alex, Julia, Sujeet, Veli-Matti and Rodrigo. We have had many interesting conversations about science (and everything else) - thank you for your peer support and friendship! I wish to thank Suvi, Tiina L, Tiina M, Heidi, Kaisa, Tuomo and other great people who I have got to know during my studies. A very special thanks goes to Lumi and Ulla, who shared an office with me for several years. I also want to thank Hannele Parkkinnen for her help with the laboratory work and other staff of the Biology Department in Oulu.
I did a large part of my laboratory work in the canine genetics group at Biomedicum Helsinki in the Research Programs Unit, Molecular Neurology, University of Helsinki, and the Folkhälsan Institute of Genetics. From Helsinki, I especially want to thank Kaisa, Saija, Eija, Päivi, Marjo, Ranja and Minna V. You were always very helpful and very good company during the lunch breaks.

I owe my gratitude to Dr Ilpo Kojola and other people at the Finnish Game and Fisheries Research Institute for providing Finnish wolf samples to our research group. I am grateful to Dr Marja Isomursu and other parasite specialists at Evira for providing us the information on wolf parasites. I also used samples from the dog DNA bank set up by the canine genetics group in Helsinki. I want to thank all dog owners and breed clubs for their contribution to the sample collection - without you, this great resource for scientific studies would not exist! I wish to express my gratitude to all funding parties that have supported this study directly or through my supervisors’ projects: Finnish Population Genetics Graduate Programme, Kone foundation, Kuopion luonnonystävät, University of Oulu Graduate School, the Academy of Finland, Jane and Aatos Erkko Foundation, the Sigrid Juselius Foundation and Biocentrum Helsinki.

I owe my deepest gratitude to my friends, family and relatives. Sakari, my beloved husband, thank you for everything. My precious Akseli and Alda, thank you for limiting my work (mostly) to the office hours and reminding me what’s important in life. I want to thank my in-laws Leila, Esa and Elina. You have been indispensably helpful with everything, especially with childcare! Without you, I could not have completed my thesis within this time frame. I wish to thank my older brother Einari for support and help in science and life. I am grateful to my little sister Ilona for linguistic support, among other things. My warmest thanks go to my parents, Lea and Veikko, for the financial, mental and all the other support you have provided me. Mother, thank you for the boost you gave me during the last months, now I can answer your question: “Kyllä, olen saanut jotain aikaiseksi!”

Oulu, September 2014

Alina Niskanen
Abbreviations

$A$ number of alleles
APC antigen presenting cell
$A_R$ allelic richness
$A_U$ number of unique alleles
BP before present
DLA dog leucocyte antigen
DNA deoxyribonucleic acid
d$_N$ nonsynonymous substitutions
d$_S$ synonymous substitutions
$F_{IS}$ inbreeding coefficient
$G_{ST}$ population differentiation
$G''_{ST}$ standardized population differentiation
$H_e$ expected heterozygosity
$H_o$ observed heterozygosity
HLA human leucocyte antigen
Ii invariant chain
LD linkage disequilibrium
MHC major histocompatibility complex
mtDNA mitochondrial DNA
$N_e$ effective population size
PBR peptide-binding region
$PD$ phylogenetic diversity
$p_N$ nonsynonymous polymorphism
$p_S$ synonymous polymorphism
SNP single nucleotide polymorphism
List of original articles

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


*Equal contribution

Author contributions

<table>
<thead>
<tr>
<th>Original idea</th>
<th>I</th>
<th>II</th>
<th>III</th>
</tr>
</thead>
<tbody>
<tr>
<td>JA, HL</td>
<td>PS</td>
<td>AN</td>
<td></td>
</tr>
<tr>
<td>Data collection</td>
<td>IK, MI</td>
<td>PS, EH</td>
<td>AN, LK, HL</td>
</tr>
<tr>
<td>Laboratory work</td>
<td>AN, EJ, MI</td>
<td>EH</td>
<td>AN</td>
</tr>
<tr>
<td>Data analyses</td>
<td>AN, LK, EJ, MR, TP</td>
<td>AN, EH, JA</td>
<td>AN, LK, TP</td>
</tr>
<tr>
<td>Manuscript preparation</td>
<td>AN, TP, JA</td>
<td>AN, EH, MR, JA, RE, HL</td>
<td>AN, TP, JA</td>
</tr>
</tbody>
</table>

(with help of others)

Alina Niskanen (AN), Lorna Kennedy (LK), Minna Ruokonen (MR), Ilpo Kojola (IK), Hannes Lohi (HL), Marja Isomursu (MI), Eeva Jansson (EJ), Tanja Pyhäjärvi (TP), Jouni Aspi (JA), Erik Hagström (EH), Rodrigo Esparza-Salas (RE), Peter Savolainen (PS),
# Table of contents

Abstract  
Tiivistelmä  
Acknowledgements 9  
Abbreviations 11  
List of original articles 13  
Table of contents 15  

## 1 Introduction 17  
1.1 Natural selection ................................................................. 17  
1.1.1 Selection and demography ................................................ 19  
1.2 Major histocompatibility complex loci ................................... 20  
1.2.1 Form and function of MHC class II proteins .................. 20  
1.2.2 Natural selection in MHC class II loci .............................. 21  
1.3 Grey wolf in Finland ............................................................. 23  
1.4 From wolves to dogs - domestication ................................. 24  
1.5 Aims of the study ................................................................. 26  

## 2 Material and methods 27  
2.1 Samples ............................................................................. 27  
2.2 MHC class II loci ................................................................. 27  
2.3 Microsatellite loci ................................................................. 28  
2.4 Genetic diversity and differentiation ................................. 28  
2.5 Inferring natural selection .................................................... 29  
2.5.1 Prenatal selection ............................................................. 29  
2.6 Parasite associations ............................................................ 30  
2.7 Domestication analyses ........................................................ 31  

## 3 Results and discussion 33  
3.1 MHC class II diversity in wolves ......................................... 33  
3.2 Historical selection in the wolf MHC class II loci ............. 33  
3.3 Recent selection in the canine MHC class II loci ................ 34  
3.3.1 Parasite associations ....................................................... 35  
3.3.2 Prenatal selection ............................................................ 36  
3.4 Domestication and MHC diversity in Asian dogs .............. 37  

## 4 Conclusions 41  
References 43  
Original articles 53
1 Introduction

Genetic diversity is essential for a population’s ability to adapt to a new or changing environment (Barrett & Schluter 2008). Without the capability to adapt, the population or whole species may become extinct. Adaptation takes place through natural selection that may change species gradually and on some occasions cause speciation (Darwin 1859, Via 2009, Langerhans & Riesch 2013).

Genetic diversity is especially important in genomic regions that are involved in immune defense, for example the major histocompatibility complex (MHC). Animals face an endless arms race against pathogens (Van Valen 1973) and need a wide reservoir of weapons in the form of genetic diversity to win the race. In addition to immune defense, MHC diversity is vital also for other purposes: autoimmune diseases (Fernando et al. 2008, Jokinen 2011) and even problems in reproduction (Coulam et al. 1987) have been postulated to be partly a consequence of low diversity in the MHC region. To increase knowledge of the central immunological system in wolves and dogs, this thesis utilizes MHC loci to examine natural selection, genetic diversity and dog domestication.

1.1 Natural selection

There are two or three major types of selection depending on the way of classification: negative selection, positive selection and balancing selection, which can also be thought of as a type of positive selection (e.g. Nielsen 2005). Negative selection removes harmful alleles from the population or decreases their frequencies, after which drift often makes them extinct in the long run. Positive selection favors an allele that brings fitness benefit to its carrier. In an optimal case for a population, positive selection eventually leads to fixation of the favorable allele in the whole population. Balancing selection causes several alleles to be maintained in the population at intermediate frequencies for long periods of time. There are three major types of balancing selection proposed for MHC: overdominant selection (Doherty & Zinkernagel 1975), negative frequency-dependent selection (Clarke & Kirby 1966) and fluctuating selection (Hedrick 2002). The main types of balancing selection are discussed in more detail in chapter 1.2.2.

Selection has an effect on allele frequencies and heterozygosity level. The allele frequency spectrum summarizes the allele frequencies of different polymorphisms. In a neutral situation it is L-shaped (Nielsen 2005). Negative
selection increases the proportion of low frequency alleles while positive selection usually increases the proportion of high frequency derived alleles. Heterozygosity is decreased by negative selection and can be either increased or decreased by positive selection, depending on the pattern of positive selection. Balancing selection makes allele frequencies more equal than expected in a neutral situation, i.e., loci under balancing selection have higher heterozygosity than neutral loci. Many commonly used neutrality tests are based on the allele frequency spectrum. One of the most widely used summary statistics, Tajima’s $D$ (Tajima 1989), is based on the differences of theta (population mutation parameter) estimates that are based on segregating sites and the average number of pairwise nucleotide differences between sequences in a sample.

Selection leaves genetic imprints that can often be detected from sequence polymorphism. A strong sweep of positive selection favors an allele with initially low frequency and raises it quickly to fixation (Maynard Smith & Haigh 1974). At the same time, also linked DNA regions next to selected locus become fixed, which causes lowered sequence polymorphism around the selected locus. A so-called soft sweep is similar to a selective sweep but happens slower or starts with multiple copies of the allele and leaves a less clear pattern of low sequence polymorphism (Hermisson & Pennings 2005). Balancing selection has an opposite impact and it increases sequence polymorphisms in nearby regions (Charlesworth 2006).

If certain alleles of different loci are detected together more often than expected by chance, they are in linkage disequilibrium (LD). Excess LD can be caused, for example, by a recent selective sweep or balancing selection in a transient phase (Nielsen 2005). Old balancing polymorphism reduces LD because recombination breaks the original linkage between the favored allele and its surrounding loci (Charlesworth 2006). As an outcome, the number of haplotypes is higher than expected by chance.

A locus-specific selection pattern that takes a long time to form is the ratio of nonsynonymous substitutions per nonsynonymous site ($d_s$) and synonymous substitutions per synonymous site ($d_S$). Nonsynonymous mutations change the amino acid a codon is coding for, while synonymous substitutions do not change the amino acid. Mainly nonsynonymous variation is thought to be under selection. In a neutral situation, both mutation types are equally common ($d_s/d_S = 1$). If the sequence is under negative selection, synonymous substitutions are more common ($d_s/d_S < 1$, Miyata & Yasunaga 1980), while under positive selection new functional mutations are favored ($d_s/d_S > 1$, Maruyama & Nei 1981).
selection can be thought of as a kind of positive selection where many alleles are favored simultaneously or there is fluctuation in the favored allele through time ($d_{tr}/d_s > 1$, Garrigan & Hedrick 2003). The excess of nonsynonymous versus synonymous substitutions takes a long time to develop and provides a strong evidence of selection. However, as Garrigan and Hedrick (2003) pointed out, the pattern persists a long time after the selective force has disappeared and it is impossible to say when and why the selection happened.

### 1.1.1 Selection and demography

Demographic factors may have similar effects on genetic diversity as natural selection. For example, less severe population bottleneck wipes out low-frequency alleles and thus intermediate frequency alleles become more common than in a stable population. A similar effect can be caused by population subdivision and mixing of different populations. The outcome of these demographic factors resembles balancing selection (e.g. Hein et al. 2005). Positive selection favors a single allele and thus all other alleles are found only at very low frequency resulting in an excess of singletons. However, a similar situation can be caused by exponential growth and a severe population bottleneck (Tajima 1989).

To be able to tell apart the genetic patterns caused by selection from those caused by demographic factors, one solution is to analyze control loci in comparison with the loci of interest. Demography affects the whole genome in a similar way, but selection affects mainly just the loci that are selected. If a large number of control loci or whole genome data is used, null hypotheses could be formulated based on the collected data (e.g. Luikart et al. 2003).

In small populations a stochastic process called genetic drift causes loss of genetic diversity. The efficacy of selection is directly proportional to effective population size ($N_e$). $N_e$ describes the size of an idealized population experiencing the same amount of random drift as the real population in question (Wright 1931). The idealized population has the properties of the Wright-Fisher population model: random mating, constant population size, equal sex ratio and non-overlapping generations. Small $N_e$ can lead to a situation where random drift outweighs selection (Robertson 1962). Many wild populations, for example in grey wolves, are small and fragmented (Boitani 2003) and only very strong selection can be effective. This complicates the detection of selection in small populations.
1.2 Major histocompatibility complex loci

Major histocompatibility complex is a genomic region that contains a cluster of genes related to immunological functions. MHC is widely studied in many organisms because of its extraordinary polymorphism (e.g. Meyer & Thomson 2001). MHC genes code for proteins that take part in immune reactions. MHC genes are divided into classes I, II and III based on their protein function (Trowsdale 1995). Classical class I and II MHC proteins are membrane glycoproteins that present foreign antigens to T cells. Class III proteins have other roles in immune response; for example, some of them belong to the complement system while others are inflammatory cytokines.

1.2.1 Form and function of MHC class II proteins

MHC class II proteins present antigens of extracellular pathogens. They are expressed only on the membranes of antigen-presenting cells (APC), e.g. macrophages, dendritic cells and B cells. All classical class II proteins are heterodimeric glycoproteins consisting of α and β chains (Fig. 1) that are coded by the respective α and β genes. Canine MHC genes are called dog leucocyte antigen (DLA) genes and they are located on the pericentromeric region of chromosome 12 (Yuhki et al. 2007). Four classical class II genes are expressed in canines: DLA-DRA1, DLA-DRB1, DLA-DQA1 and DLA-DQB1. These loci do not usually show copy number variation in domestic dogs and wolves. However, duplicated DLA-DQB locus has rarely been found in dogs, but it is unclear whether both loci are expressed (Kennedy et al. 2007a). For example felines, such as domestic cats, commonly show up to three DRB loci (Yuhki et al. 2007).
In humans, nonclassical class II proteins HLA-DM and HLA-DO assist in loading antigenic peptides to classical class II proteins. After synthesis, the MHC protein is attached to the endoplasmic reticulum and an invariant chain (Ii) is attached to its peptide-binding region (PBR, Ting & Trowsdale 2002). Ii is a chaperone protein helping the MHC protein to achieve correct folding and protecting PBR from being loaded with unspecific proteins. Exogenous proteins are internalized to APC cells, for example by endocytosis, and degraded into antigenic peptides. Antigenic peptide replaces Ii in PBR of the MHC protein and the complex is transferred to cell membrane, where the antigenic peptide is presented to CD4+ T-lymphocytes. The exact mechanism of MHC protein function has not been studied in canines, but due to high sequence similarity (Debenham et al. 2005) it is likely to be very similar to human MHC protein function.

1.2.2 Natural selection in MHC class II loci

According to the traditional view, the mechanisms maintaining high polymorphism in MHC loci are either based on diseases and parasites or reproduction. The form of selection can alter between populations and within a population over time. The pathogen-based selection mechanisms include all major models of balancing selection: overdominant selection (Doherty & Zinkernagel
1975), negative frequency-dependent selection (Clarke & Kirby 1966) and fluctuating selection (Hedrick 2002). Overdominant selection favors heterozygotes and is based on the idea that a heterozygous individual can recognize a wider variety of parasites. Negative frequency-dependent selection in MHC is caused by the arms race between the host and the parasite. The same idea is included in the “Red Queen” hypothesis (Van Valen 1973) that proposes that an organism must evolve all the time to survive in the competition with other organisms. Fluctuating selection can be caused by changing pathogen frequencies in time or space.

Pathogens may cause soft selection on MHC loci but have less impact on population dynamics, that is, they influence the fitness but do not cause the death of the host (Babik et al. 2005). This might explain why many populations seem to flourish despite low MHC diversity (Radwan et al. 2010). It may, however, be that the populations that did not cope low MHC diversity became extinct before they were studied. One extreme example of the consequence of low MHC diversity is the Australian marsupial, Tasmanian devil (Sarcophilus harrisii), which is on the brink of extinction caused by an infectious facial tumor (Siddle et al. 2007). Populations with limited MHC diversity are often small or bottlenecked. Intragenic recombination or gene conversion is thought to increase MHC diversity even tenfold compared to mutation. A threat to the recovery of MHC diversity in small populations may thus be lower effective recombination rate (Richman et al. 2003, Radwan et al. 2010).

It has been shown that assortative mating (Wedekind et al. 1995, Penn 2002) and prenatal selection after mating (e.g. Ober 1999) can also contribute to MHC diversity. In populations where assortative mating occurs, breeding animals choose mates that are similar to or different from (negative assortative mating) themselves (Penn & Potts 1998, Landry et al. 2001). Prenatal selection can take place after mating through sperm selection (Rülicke et al. 1998) or fetal selection (Ober et al. 1998). Both assortative mating and prenatal selection may increase the heterozygosity of the offspring and reduce inbreeding.

Recently, van Oosterhout (2009) suggested a new model of MHC evolution called associate balancing complex evolution. According to the model, recessive deleterious mutations accumulate at the MHC region and are not purged from the population because of the high diversity of MHC genes. Since these deleterious mutations are linked to specific alleles, the outcome is similar to overdominant selection and selection favors individuals heterozygous for the harmful alleles. This process may add to the actual balancing selection targeting MHC alleles.
Despite high diversity within species, MHC class II region is highly conserved among mammals (e.g. Debenham et al. 2005). Alleles are long-living, which leads to so-called trans-species polymorphism - similar alleles are found in closely related species and alleles from different species may cluster together closer than alleles from the same species (Klein 1987, Furlong & Yang 2008). Even though high MHC diversity is advantageous in immune defense, highest possible diversity is not always selected for. For example, in three-spined sticklebacks that have copy number variation in MHC loci, an intermediate rather than maximum number of MHC alleles per individual has been suggested optimal in terms of parasite resistance (Wegner et al. 2003).

Associations between specific MHC alleles and parasite infections have been found in various species (e.g. Paterson et al. 1998, Froeschke & Sommer 2005). In dogs, for example, the risk of visceral leishmaniasis is higher in the carriers of a specific DLA-DRB1 allele (Quinnell et al. 2003). Many autoimmune disorders also associate with MHC diversity (Fernando et al. 2008, Hughes et al. 2010). Diabetes mellitus has for a long time been known to associate with MHC class II alleles in humans (Todd et al. 1987) and similar associations have also been demonstrated in dogs (Kennedy et al. 2006). In addition, homozygosity of the MHC region has been shown to predispose to parasite infections and autoimmune diseases (e.g. Froeschke & Sommer 2005, Jokinen et al. 2011).

1.3 Grey wolf in Finland

The grey wolf (Canis lupus) is one of the top predators in the Northern hemisphere. It used to have the widest distribution area among land mammals (Boitani 2003), but habitat loss and human persecution have led to a strong decrease in wolf numbers, especially during the last 200 years (Wayne et al. 2004, Boitani 2003). However, during the last three decades, the wolf has been protected in many countries and it has returned to parts of its old territories in Europe and North America (Mech 1995, Lucchini et al. 2002). The main reasons for wolf hunting have been fear towards wolves and direct competition with humans for game animals and livestock (Woodroffe 2000, Bisi et al. 2010). Wolf has a very flexible diet including large ungulates, small herbivores, plants and even garbage (Peterson & Ciucci 2003). Diverse diet composition together with good dispersal abilities (e.g. Kojola et al. 2006) enables wolves to adapt and expand into new habitats rather easily, leaving human action as the main reason for their distress (Boitani 2003).
In Finland, similarly to other parts of Europe, wolves were hunted heavily from the late 19th century until the early 20th century (Ermala 2003). The population went through the most severe population bottleneck in the 1920s, when the population size was estimated to be only a few individuals (Pulliainen 1965). Since the mid-1990s the wolf has been strictly protected in Finland and hunting has been allowed only in the reindeer husbandry area in Lapland. Due to protection, the wolf population grew from approximately 100 individuals in the 1990s to 250 individuals in 2006, after which the population size declined again (Jansson et al. 2012). The current census size estimate in 2014 is 140–155 wolves according to the Finnish Game and Fisheries Research Institute. The exact reasons for the recent reduction in population size are not known but one possible factor may be illegal killing (Kojola et al. 2011). Poaching may have major influence on population size. For example, Liberg et al. (2012) estimated that about half of total mortality in the neighboring Scandinavian wolf population was caused by poaching.

The Finnish wolf population has been considered to be a part of the large continuous Russian wolf population (Boitani 2003). This is supported by the finding that the genetic diversity of the Finnish wolf population has been at a similar level as in larger wolf populations in Europe and North America, despite the earlier population bottlenecks (Jansson et al. 2012). However, after the recent reduction in census size the genetic diversity of Finnish wolves has decreased. Simultaneously, gene flow from Russia has decreased and, consequently, the Finnish wolf population has differentiated from the Russian wolf population during the early 21st century (Aspi et al. 2009, Jansson et al. 2012).

1.4 From wolves to dogs - domestication

Domestication can be seen as an ultimate form of selection caused by another species - human. Many domesticated species are far more successful than their wild counterparts. Dog (Canis familiaris) was the first domesticated animal (Clutton-Brock 1995, Raisor 2005) and markedly outnumbers its ancestor, grey wolf (e.g. Vila et al. 1997, Lindblad-Toh et al. 2005). In contrast to wolves, dogs are accepted and liked in most human societies and, consequently, dogs have spread around the globe.

Genetic studies have reached very different estimates for the timing of domestication varying from 11,000 to 135,000 years before present (BP; Vila et al. 1997, Savolainen et al. 2002, Freedman et al. 2014). The most recent studies
have narrowed the domestication timeframe to 11,000–32,000 BP (Thalmann et al. 2013, Freedman et al. 2014). It has been proposed that domestication was initiated in Southeast Asia, Middle East or Europe (Savolainen et al. 2002, Pang et al. 2009, vonHoldt et al. 2010, Thalmann et al. 2013), but the location and number of domestication events is still controversial (vonHoldt et al. 2010, Larson et al. 2012).

Fossil data have also been used to estimate the details of the domestication process. Earliest dog-like fossils have been found in Europe, in Belgium, dating from 36,000 BP (Germonpré et al. 2009). Other important canid fossils have been excavated in Germany (~12–15,000 BP; Nobis 1979, Baales 1996) and in Siberia (33,500 BP; Ovodov et al. 2011). However, it is unclear whether these fossils represent extinct canid species or actual early domestic dogs (Crockford & Kuzmin 2012). Modern dogs show genetic similarity with modern European wolves and dog-like fossils younger than 15,000 BP (Thalmann et al. 2013). This finding and the long time span to the next dog-like fossils suggest that, instead of being ancestors of modern dogs, the oldest specimens may have been parts of separate domestication branches that are now extinct (Ovodov et al. 2011, Thalmann et al. 2013).

The reasons for the early domestication of wolf probably stem from its basic biology and behavioral similarities with humans. The wolf has very flexible dietary and environmental requirements (Peterson & Ciucci 2003, Wayne et al. 2004), which eased the adaptation from a wild predator to a domestic animal. The wolf is a social animal living in packs (Packard 2003, Mech & Boitani 2003), which probably facilitated interspecies communication during hunting or guarding. Wolves might have been tamed actively by taking wolf cubs into human households. Another possibility is that wolves started to approach human dwellings while looking for food in human garbage, and maybe brought some shelter to humans (Coppinger & Coppinger 2002).

Around 340 dog breeds are currently recognized by the World Canine Organisation (FCI). Adding to the official breeds, there are numerous local dog types, mongrels and village dogs around the world. The effective population sizes of different breeds are often very small, less than 100, even though census sizes may be very large, up to tens of thousands of individuals (Calboli et al. 2008). Deliberate crossings of close relatives cause high inbreeding coefficients in breed dogs. Taken together, many dog breeds have very limited genetic diversity and they are highly homozygous (e.g. Parker et al. 2004, Björnerfeldt et al. 2008, vonHoldt et al. 2010).
1.5 Aims of the study

The high diversity of MHC loci in many species is usually thought to be the result of balancing selection. However, genetic drift may outweigh selection in small populations (Robertson 1962, Strand et al. 2012). Thus, I was interested in finding out the possible role of the MHC-related selection in a particular small but natural population such as the Finnish wolf. Parasites are a major selective factor in the MHC diversity but no association studies have been conducted in wolves. I studied associations between two helminth parasites, *Echinococcus canadensis* and *Trichinella* spp., and the Finnish wolf MHC diversity.

The domestic dog is a descendant of the grey wolf, and based on earlier genetic studies, may have originated in South-East Asia or Middle East. Genetic diversity is expected to be highest close to the domestication origin (e.g. Prugnolle et al. 2005). The diversity of the MHC loci has not been studied in relation to the origin question and I addressed it here by examining the MHC diversity of Asian dogs. I also used MHC diversity of modern dogs to estimate the minimum number of the founding wolves.

Natural selection in the MHC loci of wild and domestic animals may differ due to their different living conditions. Modern western pet dogs receive advanced veterinary care and have limited access to mate choice, leaving the fetal development the only potential phase of natural selection without human interference. I investigated the heterozygosity patterns in the MHC class II genes of domestic dog families to test the hypothesis of prenatal selection.

My main research questions were:

1. Has selection in the wolf MHC loci been strong enough to have influenced genetic diversity despite the demographic changes? (I)
2. Are there associations between *Echinococcus canadensis* or *Trichinella* spp. prevalence and the MHC diversity in wolves? (I)
3. What is the level of MHC diversity in Asian dogs? (II)
4. How large a number of founding wolves was needed to explain the current dog MHC diversity? (II)
5. Do dogs show signs of prenatal selection in the MHC class II loci without the opportunity for mate choice? (III)
2 Material and methods

Materials and methods are described only briefly in the following chapters. Detailed information is given in the original papers (I-III).

2.1 Samples

The samples comprised wolves in study I and dogs in studies II and III. Wolf samples for study I were collected from Finland between 1996 and 2011 (n = 242) and Russian Karelia between 1995 and 2010 (n = 37, Fig. 1 in I). Dog samples for study II were collected from dogs of North or East Asian origin (N/E Asia, n = 69) and South or West Asian origin (S/W Asia, n = 59). MHC data from 1484 dogs in Kennedy et al. (2007a) were also included in the study II and the dogs were assigned to their geographic origin: Europe (n = 1051), South America (n = 112), Africa (n = 110), North America (n = 90), S/W Asia (n = 61) and N/E Asia (n = 60).

Family data for the study III comprised dogs from 10 breeds and four crossbred groups (Table 1 in III). Altogether 430 puppies in 110 families were included. The data comprised of totally unexplored dogs, and also dogs from the study of Kennedy et al. (2007b) and Rowan (2007). Dogs from Finland, the UK and USA were sampled. The dog DNA bank resources (> 50,000 samples from 327 breeds) established in the University of Helsinki were utilized in this study.

2.2 MHC class II loci

The second exons of MHC class II loci DLA-DRB1 (270 bp), DLA-DQA1 (246 bp) and DLA-DQB1 (267 bp) were amplified and sequenced in studies I and III. Only the most polymorphic of these loci, DLA-DRB1, was analyzed in study II. The second exon codes for the peptide-binding region of the MHC class II molecule and it is more polymorphic than other exons of the MHC loci. Three-locus haplotypes, consisting of alleles from DLA-DRB1, DLA-DQA1 and DLA-DQB1, were determined for each individual in studies I and III following Kennedy et al. (2007c).

The order of the loci from the centromere is: DLA-DRB1 separated from DLA-DQA1 by 56.6 kb, and DLA-DQA1 separated from DLA-DQB1 by 77.7 kb. The loci usually have only one functional copy in dogs and wolves (Kennedy et al. 2000, Kennedy et al. 2007c) and each primer pair amplifies only one locus.
All dogs and wolves genotyped in this study had single copies of each locus. Previously, a duplicated DLA-DQB1 locus had been found in some dog families that were included in this study (Kennedy et al. 2007b, Rowan 2007). In study III, one crossbred group of six families including 37 offspring was genotyped using microsatellite loci that are linked with the studied MHC loci (Rowan 2007).

2.3 Microsatellite loci

Microsatellites are short tandem DNA repeats and since they do not code for any gene products, they are putatively selectively neutral. To control for demographic effects, the microsatellites were used as reference loci in selection, population divergence and parasite association analyses of study I. Seventeen selectively neutral microsatellite loci were amplified from the same wolves that were analyzed for MHC in study I. Microsatellite data were obtained for 187 Finnish and 30 Russian wolf individuals from a previous study (Jansson et al. 2012). An additional 54 individuals were genotyped for the same microsatellite loci for study I. Detailed information about the microsatellite loci and amplification methods can be found in Jansson et al. (2012).

2.4 Genetic diversity and differentiation

Genetic diversity of MHC II loci was studied in the Finnish and Karelian wolf populations (study I) and in dogs from different geographical regions (study II). For the latter, diversity estimations were used to identify the region where domestication started; these analyses are explained with more detail in chapter 2.6. In study I, basic diversity parameters were estimated: number of alleles ($A$), observed ($H_o$) and expected ($H_e$) heterozygosity, inbreeding coefficient ($F_{IS}$) and allelic richness ($A_R$). The same parameters were also estimated for the microsatellites with the aim to compare neutral genetic variation and MHC diversity in wolves. To monitor changes in genetic diversity through time, the Finnish wolf samples were divided into 3-year periods: 1995-1997, 1998-2000, 2001-2003, 2004-2006 and 2007-2009. Three years correspond approximately to the generation interval of wolves in Finland (Aspi et al. 2006).

Differentiation between Finnish and Russian Karelian wolves was studied using MHC II loci and microsatellite markers. Since neutral variation and differentiation has been studied in more detail by Aspi et al. (2006, 2009) and Jansson et al. (2012), the main reason for using both markers was to find out
whether it was possible to recognize putative selection patterns in MHC loci despite the genomewide effects of a population bottleneck. Differentiation was studied with $G_{ST}$ and standardized $G''_{ST}$ which takes into account the small number of populations and different diversity levels of the markers studied (Meirmans & Hedrick 2011).

2.5 Inferring natural selection

Natural selection in MHC loci was the main focus of studies I and III. Historical and recent selection produce different kinds of genetic patterns (Garrigan & Hedrick 2003), which is why both nucleotide and allele level analyses were conducted. Prenatal selection (study III) was also examined, and will be discussed in more detail in the next chapter. In study I, long-term selection was studied from the ratio of nonsynonymous ($p_N$) to synonymous ($p_S$) polymorphisms (written $d_N/d_S$ in I). Selection was estimated using coalescence-based population genetics approximation as implemented in omegaMap (Wilson & McVean 2006). Recombination and different mutation rates (synonymous transversion, transition/transversion and insertion/deletion) were included in the analysis as priors. In study I, all priors were chosen to represent neutrality.

Sequence level tests included also Tajima’s $D$ (1989) and Fu & Li $D^*$ (1993), which depict historical selection but can be influenced by demographic effects (Simonsen et al. 1995). Both tests are based on the allele frequency spectrum of nucleotide polymorphisms. Contemporary selection was studied with Ewens-Watterson test (Ewens 1972, Watterson 1978) which compares the observed allele frequencies (estimated as the expected homozygosity) to expected allele frequencies. Ewens-Watterson test was conducted for each locus at MHC allele level. A compound test, DHEW (Zeng et al. 2007), was also used, since it should be more robust to demographic changes and the presence of recombination than the individual tests it is based on: Tajima’s $D$, Ewens-Watterson and Fay & Wu $H$ (2000).

2.5.1 Prenatal selection

Study III concentrated on putative prenatal selection in MHC loci by examining genotype proportions in dog families. The main questions were: (i) do dog puppies show excess of heterozygosity and (ii) do puppies inherit a dissimilar allele from their father than the one carried by their mother, when parents share
one allele. For the first time, the canine MHC alleles were clustered into supertypes according to their amino acid properties at the PBR sites. Supertype classification was done for DLA-DRB1 and DLA-DQB1 loci. The analyses of prenatal selection were done at allele, three-locus haplotype, amino-acid (PBR) and supertype levels. Families were selected from the original data according to their parental genotypes so that there were potentially both homo- and heterozygotes in the offspring. Four cross types were included in the heterozygosity analyses: A1A1 × A1AX, A1AZ × A1A1, A1A2 × A1A2 and A1AZ × A1AX (where: mother × father and A1 ≠ A2 ≠ A1 ≠ A2). To answer the question of inheriting maternal or paternal type allele from the father, two cross types were included in the analyses: A1A2 × A1A1 and A1A1 × A1AX.

2.6 Parasite associations

To analyze the association between MHC diversity and parasite prevalence, Finnish wolves were screened for two parasitic worms, Trichinella spp. and Echinococcus canadensis (strain G10, study I). The number of Trichinella larvae was counted per gram of striated muscle using mechanically assisted pooled sample digestion method (sedimentation technique, EU Reg. No 2075/2005). The presence and absence of E. canadensis was examined first by testing Echinococcus coproantigens with ELISA test and confirming positive samples with sedimentation and counting technique (SCT, Eckert et al. 2001).

The association between MHC diversity and the prevalence of Trichinella spp. or E. canadensis was studied at single nucleotide polymorphism (SNP), allele, three-locus haplotype and heterozygosity levels. General linear model (GLM) and mixed linear model (MLM) were used for SNP level association analyses in each MHC locus. Since latitude or age of the wolf might influence the infection status (Oppelt et al. 2010), both were included in the analyses as cofactors. Kinship coefficients (Loiselle et al. 1995) were also estimated for each individual pair using microsatellite data and included in the MLM analyses.

Association analyses between each MHC allele or three-locus haplotype and the prevalence of each parasite were conducted with Fisher’s exact test. Similarly, associations between neutral microsatellites and the prevalence of each parasite were studied to evaluate the potential confounding effects of population demography on the MHC association results. The association between MHC or microsatellite heterozygosity and parasite prevalence was studied with logistic regression.
2.7 Domestication analyses

The history of dog domestication has been studied with diverse genetic methods. In study II, the DLA-DRB1 variation of dogs from different geographical regions was analyzed. Since higher diversity is expected at the domestication center or source population (Prugnolle et al. 2005), the number of alleles, number of unique alleles ($A_U$) and Faith phylogenetic diversity ($PD$, Faith 1992, Crozier 1997) were measured or estimated for each region. The number of wolf individuals founding the dog population was also of interest. A simulation program following Vincek et al. (1997) was coded to conduct simulations on the minimum number of wolf founders needed to achieve the current MHC diversity.
3 Results and discussion

3.1 MHC class II diversity in wolves

The Finnish wolf population showed similar MHC diversity as the Russian Karelian wolf population (e.g. $H_e = 0.805$ and 0.803, respectively; Table 2 in I). The populations shared most of the MHC alleles, including two new DLA-DRB1 alleles, and three-locus haplotypes (Table 1 in I). One new DLA-DQB1 allele was found, and it was limited to the Finnish population. This may be a sampling effect, since more samples were collected from Finland than from Karelia. Other European wolf populations show comparable level of MHC diversity, for example expected heterozygosity was 0.84 in Croatian population, 0.67 in Estonian population and 0.78 in Latvian population (Seddon & Ellegren 2004, Arbanasic et al. 2013). Accordingly, the MHC diversity of the Finnish population is still remarkably high, despite the bottleneck and currently small population size.

3.2 Historical selection in the wolf MHC class II loci

Long-term balancing selection has caused an excess of nonsynonymous polymorphisms compared to synonymous polymorphisms in MHC class II loci of many species (Hughes & Nei 1989, Luetkemeier et al. 2009, Radwan et al. 2010). This implication of historical selection was also found in Finnish wolves at the codon level (Fig. 3 in I). Synonymous polymorphisms were completely lacking in some codons, which is why subtraction $p_N - p_S$ was used instead of the more familiar ratio $p_N / p_S$ in depicting the polymorphism patterns (Fig. 3 in I).

The excess of nonsynonymous polymorphisms is usually more common in codons that code for peptide-binding regions compared to other regions (Hughes & Nei 1989, Hughes & Yeager 1998). However, some codons in PBR seem to be conserved and show a deficiency of nonsynonymous polymorphisms (Furlong & Yang 2008). Codons with both excess and deficiency of nonsynonymous polymorphisms were found in the putative PBR of wolf DLA-DRB1 and DLA-DQB1 loci (Fig. 3 in I). Still, the codons with an excess of nonsynonymous polymorphisms were located more frequently in PBR than in other regions ($t = 3.54–3.97$, $df = 29.8–30.7$, $p = 0.0004–0.001$; in I). Varying polymorphism patterns may indicate diverse roles for the codons in the PBR. For example, some codons might function in stabilizing the peptide-binding pocket and thus remain
unchanged (Furlong & Yang 2008). It is also possible that the codons assigned to PBR for human MHC molecules are not entirely correct for wolves, despite high sequence similarity between the species (Debenham et al. 2005).

Balancing selection is not always detected at codon level. For example, Marsden et al. (2012) found only evidence of negative selection in DLA-DRB1 of the African wild dog. It would be interesting to know whether the result of no balancing selection is caused by different analysis methods or a real difference in selection patterns. The ratio of nonsynonymous substitutions to synonymous substitutions was originally used to estimate selection in diverged sequences between species. The usage of within-population \( p_n/p_s \) as a selection measure has been criticized, since the ratio does not behave monotonically according to expectations and is less sensitive to selection than between-species \( d_n/d_s \) (Kryazhimskiy & Plotkin 2008). The anomaly may cause strong positive selection to go undetected and lead to false interpretation of neutrality or even negative selection. However, \( p_n/p_s \) is distorted in the direction of too small values and therefore conservative when testing for positive or balancing selection.

### 3.3 Recent selection in the canine MHC class II loci

Selection can change population differentiation patterns among adaptive and neutral loci. If selection pressures differ between two populations, MHC loci may differentiate more than neutral loci that are mainly affected by drift (e.g. Landry & Bernatchez 2001, Heath et al. 2006, Ekblom et al. 2007, Marsden et al. 2012). On the other hand, similar selection pressures in different geographical regions may drive the MHC allele frequencies of populations to the same direction (Takahata 1993, Aguilar et al. 2004, van Oosterhout et al. 2006). The latter was also detected between the Finnish and Russian Karelian wolf populations: the MHC loci of the Finnish and Russian wolf populations were not differentiated (over loci \( G''_{ST} = 0.047, p = 0.377 \) in I), but the microsatellite loci were differentiated (over loci \( G''_{ST} = 0.148, p = 0.008 \) in I and Jansson et al. 2012). Common pathogens in these adjacent regions have possibly caused similar selection pressures among the populations. Such pathogen-mediated selection has been found e.g. in guppy populations, where Fraser & Neff (2010) detected less divergence at a MHC gene than at neutral microsatellite loci. A likely cause of the homogenizing selection was that one MHC allele gave better resistance against a common parasitic worm.
Another piece of evidence for a recent selection in the Finnish wolf population was the fact that allele frequencies in MHC loci were more even than expected by chance. The homozygosity value estimated in Ewens-Watterson test was lower than expected for DLA-DQA1 and DLA-DQB1 loci (Table 3 and Fig. 4 in I). A population bottleneck can also cause loss of low frequency alleles and, consequently, result in low expected homozygosity. This would give a false result of balancing selection in the Ewens-Watterson test (Watterson 1986). However, the MHC homozygosity level of the Finnish wolves was even lower than the microsatellite homozygosity level that is affected only by demography (Fig. 4 in I). Thus, the even MHC allele frequencies are most likely caused by balancing selection acting on these loci.

Balancing selection must have been very strong to be effective in the Finnish wolf population, since it has recently gone through a population bottleneck and the effective population size is still small ($N_e = 20-65$; Jansson et al. 2012). Another example of strong MHC selection in a bottlenecked canine population comes from red wolves in North America, where Hedrick (2002) found more amino acid differences and more even distribution of MHC alleles than expected by chance. It thus seems that strong balancing selection may be rather common in the MHC loci of canid populations.

### 3.3.1 Parasite associations

Since parasites are a major selective factor modifying MHC diversity, it is not surprising that many studies have linked MHC loci to parasite infections (e.g. Paterson et al. 1998, Quinnell et al. 2003, Froeschke & Sommer 2005). In this study, the wolves carrying the DLA-DRB1*10101 allele and the three-locus haplotype including the allele had less often *Trichinella* spp. infection than individuals not carrying the allele (Fig. 5 and Table 4 in I). Also wolves heterozygous for MHC loci had less often *Trichinella* spp. infection than homozygous wolves (Table 4 in I). No association between *E. canadensis* and MHC alleles or heterozygosity was found. The parasite has a rather low prevalence that decreases the power to detect associations; in the examined individuals the prevalence was 14 % for *E. canadensis* and 35 % for *Trichinella* spp.

Studies of selection agents are challenging to conduct in wild animals due to multiple potentially confounding environmental factors in natural circumstances. In study I, latitude and age of the wolf were included as cofactors in the GLM and
MLM analyses as non-genetic factors (Table 1 in I). Since the wolf sample consisted of related animals from a small population, kinship coefficients were also included in the MLM analyses. Under this model, genetic association was found between less *Trichinella* spp. infections and SNP positions 211, 219 and 220 in DLA-DRB1 ($R^2 = 0.0363$, $p = 0.0058$, $q = 0.039$ and Fig. 6 in I). The SNP loci were in complete LD and formed a haplotype that was shared by the DLA-DRB1*10101 and DLA-DRB1*05301 alleles. These DLA-DRB1 alleles were inherited in three-locus haplotypes associated with low *Trichinella* infection prevalence (Table 4).

Despite the efforts to include non-genetic factors in the analyses, the association does not necessarily mean a causal relationship between the MHC alleles and parasite prevalence. This matter would require further analyses, possibly including physiological studies examining parasite antigens and the peptide-binding specificities of different MHC alleles. Another subject for further study could be the association of MHC diversity to parasite infection intensity, since immune response may control parasite loads rather than parasite prevalence (Westerdahl et al. 2012).

### 3.3.2 Prenatal selection

MHC-related prenatal selection events including mate selection, sperm selection and selective abortion may add to parasite-mediated selection in MHC loci. Pet dogs do not usually have the possibility to choose their mates, which leaves only post-copulatory possibilities for selection. In study III, the observed genotype proportions in dog families followed Mendelian expectations at allele (Fig. 3 in III) and three-locus haplotype levels, suggesting lack of prenatal selection. The DLA-DRB1 and DLA-DQB1 alleles were divided into supertypes based on the chemical properties of the amino acids in PBR. Five supertype classes were determined for DLA-DRB1 (Fig. 1 in III) and three for DLA-DQB1 (Fig. 2 in III). Only the cross type A1A2 × A1A2 had a slight excess of homozygous offspring at the supertype level ($\chi^2 = 4.55$, df = 1, $p = 0.05$; Fig. 4 in III). The deviation towards homozygosity was surprising, since selection is expected to favor heterozygous individuals. However, the offspring number in the analysis was small, only 22 individuals, and a replication in a larger cohort is warranted for the confirmation of the preliminary result.

Segregation of dog MHC genotypes was also studied at putative PBR-sites. The genotypes of PBR-sites 10-12 in DLA-DRB1, corresponding to codons 57,
60 and 67, deviated from Mendelian expectations when studied with combined chi-squared statistics over all cross types (Fig. 5 in III). Site 12 deviated most and one cross type, $A_1A_2 \times A_1A_1$, showed an excess of heterozygous offspring ($\chi^2 = 10.62$, df = 1, $p = 0.001$). The significance of this finding is unclear, since the expression of MHC genes during dog pregnancy remains poorly characterized. It is known that MHC class I gene DLA-88 and class II gene DLA-DRA are not expressed in early canine embryos (Schäfer-Somi et al. 2008). Even if the fetal MHC class II genes were not expressed in the maternal-fetal interface, fetal cell debris may be available for recognition by the maternal MHC molecules (Trowsdale & Betz 2006). Madeja et al. (2011) showed that histoincompatibility in a MHC class I locus increases dilation of uterine vessels in mice. Thus, heterozygous offspring carrying the combination of maternal and paternal type alleles could have an advantage early in their development (Lenz 2011). Such studies have not been conducted in dogs but a similar heterozygote advantage may exist in canines.

Many studies of prenatal selection in MHC, including this study, lack neutral control loci. Excessive genomewide homozygosity might, for example, lead to lowered fertility (e.g. Keller & Waller 2002). It would be important to include neutral loci in the segregation analyses to verify that the earlier results of prenatal selection are actually caused by MHC selection and not by misinterpreted inbreeding depression.

In study I, the genetic diversity of the Finnish wolf population was examined at three-year intervals corresponding approximately to the wolves’ generation interval (Aspi et al. 2006). During the last study period (2007-2009), an excess of observed heterozygosity compared to expected heterozygosity in the form of negative $F_{IS}$ was detected in MHC loci while microsatellites showed an opposing pattern and positive $F_{IS}$ (Fig. 2a in I). Jansson et al. (2012) observed that the connectivity of Finnish wolf packs was low during that same time period. If wolves mate with close relatives, they might prefer MHC dissimilar mates. This result could indicate disassortative mate choice based on MHC - or pure chance - but is definitely worth further study.

### 3.4 Domestication and MHC diversity in Asian dogs

Recent studies have reached the conclusion that dog domestication originated in some part(s) of Asia or Europe. In line with the earlier genetic studies that were based on mitochondrial DNA (mtDNA; Savolainen et al. 2002, Pang et al. 2009),
SNP (vonHoldt et al. 2010) and whole-genome sequencing data (Wang et al. 2013), also MHC diversity was higher in Asian than in European dogs (Fig. 1 in II). The dogs from N/E Asia and S/W Asia had a higher number of alleles, a higher number of unique alleles and greater phylogenetic diversity in the DLA-DRB1 locus than dogs from Europe ($A = 36/35/30$, $A_u = 16/14/9$, $PD = 0.584/0.585/0.489$, respectively; Table 2 in II). No substantial differences between the MHC diversities of N/E and S/W Asian regions were found (Table 1 & Fig. 1 in II). Altogether 17 novel DLA-DRB1 alleles were identified in Asian dogs. These results support an Asian origin of domestic dog, since higher genetic diversity is expected at or close to domestication center.

Natural selection has most likely also influenced the MHC diversity of domestic dogs, because MHC molecules take part in recognizing foreign peptides. There are many different selective factors that may contribute to the differences in MHC diversity between regions. One of the major factors is parasite abundance that varies depending on latitude and population density, for example (Lindenfors et al. 2007). However, Prugnolle et al. (2005) showed that in humans, MHC diversity is more affected by founder effect than parasite burden. The intensity of veterinary care might also differ between Europe and Asia. A high level of veterinary care may ease the selection pressure on MHC loci. The sampling of Asian breed dogs was mostly done in Europe, so the recent environmental conditions are similar between Asian and European breed dogs. Mate choice may take place in non-breed dogs, whereas breed dogs seldom have the freedom to choose their mates. Asian breed and non-breed dogs showed similar MHC diversities (Table 1 in II), so there was no indication of differences caused by recent veterinary care or mate choice.

Using simulations, it was estimated that to achieve the current DLA-DRB1 diversity in dogs, a minimum of 500 wolves must have contributed to dog domestication (Fig. 3 in II). This result is based on the number of 100 dog DLA-DRB1 alleles. New alleles are constantly being discovered and 200 alleles have already been reported in canines, mostly for dogs in study III. The population growth rate was kept constant in the simulations. In reality, the strength of drift has altered through time; the dog population size was small at the beginning of the domestication process, but Thalmann et al. (2013) also showed that the dog population size has followed the fluctuations of the human population. Recently, Freedman et al. (2014) came to the conclusion that the ancestral dog population ($N_0$) was smaller than 2,000 individuals. This size estimate is very plausible also
in the light of our results based on MHC diversity, when the large allele number and varying strength of drift is considered.

Since modern dog mtDNA is more closely related to wolf than to fossil dog mtDNA, it seems that not all of the domestication processes have led to successful dog lines that still have descendants (Thalmann et al. 2013). This result may also be at least partly influenced by the small effective population size and maternal inheritance of mtDNA. Beneficial MHC alleles may have been transferred between different domestication lines and, in contrast to mtDNA, they have survived until the present day. Also backcrossing with wolves has probably added to the allele pool (Vilà et al. 2005, Verardi et al. 2006). The possibility to sequence MHC loci from fossil dogs and wolves would shed light on the questions of MHC allele origin and dog domestication. The scattered fossil data, controversial genetic results and large founding population imply that domestication may have been a rather common practice in a wide region and over a long period of time.
4 Conclusions

I conclude this thesis by answering my main research questions:

1. Has selection in the wolf MHC loci been strong enough to have influenced genetic diversity despite the demographic changes?

Strong balancing selection in the MHC loci of the Finnish wolves was implied in many analyses. Historical selection was evident in amino acid polymorphism patterns and more recent selection in allele frequencies of the wolf population. Another indication of balancing selection was the fact that the MHC loci of the Finnish and Russian Karelian wolf populations were less differentiated than the neutral microsatellite loci. These results support the hypothesis that regardless of demographic changes, balancing selection is effective, and they highlight the importance of maintaining the MHC class II diversity.

2. Are there associations between *Echinococcus canadensis* or *Trichinella* spp. infection status and the MHC diversity in wolves?

Associations between wolf MHC diversity and parasite prevalence were found. Heterozygous wolves had less often *Trichinella* spp. infection than homozygous wolves. Also a specific MHC allele and a SNP haplotype were associated with a decreased prevalence of *Trichinella* spp. However, these findings require further physiological studies to verify a causal relationship between the locus and parasite prevalence. According to the hypothesis, MHC diversity and particular alleles may improve defense against parasites, and our results are in line with the idea.

3. What is the level of MHC diversity in Asian dogs?

Asian dogs showed higher MHC diversity than European dogs, which is in line with the earlier studies of genetic diversity in Asian dogs. Higher diversity is expected at and close to the domestication origin, and our results support the suggested Asian origin of domestic dogs.

4. How large a number of founding wolves was needed to explain the current dog MHC diversity?

At least 500 founding wolves were needed to achieve the current dog MHC diversity. The high number of founding wolves suggests that domestication was a common practice or that backcrossing to wolves has been rather frequent. It also
supports the idea that wolves were actively seeking human contacts rather than being only objects of domestication, which would make such a large number of founding wolves more realistic. These results support the hypothesis that a large number of founder wolves were required for the generation of the current dog populations and the spectrum of MHC alleles.

5. Do dogs show signs of prenatal selection in the MHC class II loci without the opportunity for mate choice?

In the dog families studied, no strong prenatal selection was detected. There was an excess of heterozygotes in one peptide-binding site in DLA-DRB1 but the meaning of this finding is unclear. Significant results were found at SNP and amino acid level in the wolf parasite association and the dog family analyses. In both studies, the polymorphisms were shared by more than one MHC allele. This emphasizes the importance of fine-scale sequence analyses, in addition to MHC allele level analyses, also in future studies. It was expected that the lack of mate choice may add pressure in post-copulatory prenatal selection, but our data do not support the hypothesis. However, further studies are necessary in larger study cohorts.

This study showed that the current Finnish wolf MHC diversity is at a level comparable to other European wolf populations. However, if the population were to decrease in size, balancing selection might not remain effective and the risk for pathogen outbreaks would increase. The global wolf population should be investigated to lend support to the results presented here and to find out whether the strong balancing selection is a common phenomenon in small bottlenecked natural populations. In contrast to wolves, the selection patterns in dogs remain unknown. Studies of mate choice would be important to enhance the knowledge of natural selection in the MHC loci of dog and wolf populations.

In this study, a number of new DLA-DRB1 alleles were revealed, which indicates that there could be many more to be discovered in a global dog sample. Dogs possess high overall MHC diversity, but a large proportion of the diversity is harbored within isolated breeds and populations. It is likely that MHC loci are among the most strongly selected loci but still represent only a fraction of the adaptive loci. To get a wider perspective on the strength of selection, genome-wide approaches are warranted to investigate the role and relationship of neutral and functional markers.
References


Original articles


Reprinted with permission from John Wiley and Sons (I) and Nature Publishing Group (II).

Original publications are not included in the electronic version of the dissertation.
615. Holm, Jana (2013) Catalytic pretreatment and hydrolysis of fibre sludge into reducing sugars
616. Kemi, Ulla (2013) Adaptation to growing season length in the perennial Arabidopsis lyrata
618. Rodríguez, Pilar (2013) Combining lean thinking and agile software development: how do software-intensive companies use them in practice?
619. Vatka, Emma (2014) Boreal populations facing climatic and habitat changes
620. Isomursu, Marja (2014) Host-parasite interactions of boreal forest grouse and their intestinal helminth parasites
622. Matusek, Florian (2014) Selective privacy protection for video surveillance
623. Virtanen, Elina (2014) Effects of haulm killing and gibberellic acid on seed potato (Solanum tuberosum L.) and techniques for micro- and minituber production in northern latitudes
624. Kopatz, Alexander (2014) Genetic structure of the brown bears (Ursus arctos) in Northern Europe
625. Loukola, Olli (2014) Information networks among species: adaptations and counter-adaptations in acquiring and hiding information
626. Langrial, Sitwat (2014) Exploring the influence of persuasive reminders and virtual rehearsal on the efficacy of health behavior change support system
630. Stibe, Agnis (2014) Socially influencing systems: persuading people to engage with publicly displayed Twitter-based systems

Book orders:
Granum: Virtual book store
http://granum.uta.fi/granum/
Alina Niskanen

SELECTION AND GENETIC DIVERSITY IN THE MAJOR HISTOCOMPATIBILITY COMPLEX GENES OF WOLVES AND DOGS