Assessment of MHC diversity in Central and Southeastern European grey wolf (Canis lupus) population

Pavlinec, Željko

Doctoral thesis / Disertacija

2021

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Science / Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:217:583354

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-07-10**

Repository / Repozitorij:

Repository of the Faculty of Science - University of Zagreb

University of Zagreb

FACULTY OF SCIENCE
DEPARTMENT OF BIOLOGY

Željko Pavlinec

ASSESSMENT OF MHC DIVERSITY IN CENTRAL AND SOUTHEASTERN EUROPEAN GREY WOLF (*Canis lupus*) POPULATIONS

DOCTORAL DISSERTATION

Supervisors:
Assoc. prof. Ana Galov, PhD
Assoc. prof. Pavel Hulva, PhD

Zagreb, 2021

PRIRODOSLOVNO MATEMATIČKI FAKULTET
BIOLOŠKI ODSJEK

Željko Pavlinec

PROCJENA MHC RAZNOLIKOSTI U POPULACIJAMA VUKA (*Canis lupus*) SREDNJE I JUGOISTOČNE EUROPE

DOKTORSKI RAD

Mentori:
Izv. prof. dr. sc. Ana Galov
Izv. prof. dr. sc. Pavel Hulva

Zagreb, 2021

This doctoral thesis was made in the Department of Biology, Faculty of Science, University of Zagreb, under the supervision of Assoc. Prof. Ana Galov and Assoc. Prof. Pavel Hulva, Department of Zoology, Faculty of Science, Charles University, Prague, Czech Republic. The doctoral thesis was made as part of the University postgraduate doctoral study of Biology, at the Department of Biology, Faculty of Science, University of Zagreb.

Supervisors

Ana Galov, PhD is an associate professor at Department of Biology, Faculty of Science, University of Zagreb where she is a course teacher in Fundamentals of molecular ecology, Methods in immunology and Fundamentals of animal physiology. She obtained her PhD in 2007 at the Faculty of Science, University of Zagreb working on population genetics of bottlenose dolphin in the Adriatic Sea. Her main areas of interest are wildlife and domestic animal populations in Croatia, particularly in the field of molecular ecology where she is using neutral and adaptive genetic markers in her research. She was a mentor of 20 diploma and 7 PhD theses, and published 50 peer-reviewed papers in the international scientific journals.

Pavel Hulva, PhD is an associate professor at Department of Zoology, Faculty of Science, Charles University, where he is a course teacher in Vertebrate zoology, Evolutionary and systematic zoology, Evolution of vertebrates, Biogeography, Molecular applications in Zoology and Field course of zoology. He obtained his PhD in 2006 at the Faculty of Science, Charles University working on molecular evolution of bats. Fields of his expertise include phylogenetics, phylogeography and population, landscape and conservation genetics. He studies mainly speciation, hybridization and molecular ecology of mammals (especially from the group Laurasiatheria). He was a mentor of 20 diploma and 3 PhD theses, and published over 50 peer-reviewed papers in the international scientific journals.

Acknowledgements

To my mentors Ana Galov and Pavel Hulva, thank you for giving me the opportunity to work on this research. I am aware how rare and precious these types of samples are, and I am grateful for your trust in me to handle them. Thank you for your guidance these last few years and for all the help you provided when it was needed.

For the Carpathian samples, I'm also grateful to Vladimír Antal and State nature conservancy of the Slovak Republic.

To Gordana Žakman, thank you for all the help, but especially for the pleasant company during the many hours we spent in the laboratory.

To Damjan Franjević, thank you for all of your academic and moral support, and for some of the most interesting conversations we had during my studies.

To Ivana, thank you for keeping everything together, and making it possible for me to split my time between my work and my studies. Thank you for all your love and endless support.

University of Zagreb
Faculty of Science
Department of Biology

Doctoral thesis

ASSESSMENT OF MHC DIVERSITY IN CENTRAL AND SOUTHEASTERN EUROPEAN GREY WOLF (*Canis lupus*) POPULATIONS

ŽELJKO PAVLINEC

University of Zagreb, Faculty of Science

Major histocompatibility complex diversity affects individual fitness, population viability and prospects for its long-term survival and can thereby serve as a valuable genetic indicator of population fitness and its ability to adapt to environmental changes. Ten populations of grey wolf (*Canis lupus*) have been identified in Europe today, the two largest being Carpathian and Dinaric-Balkan, with their sizes estimated to 4,000 and 5,000 individuals, respectively. In this thesis I assessed MHC class II allelic diversity in Central and Southeastern European grey wolves, namely from Carpathian and Dinaric Mountains. I used sequencing-cloning procedure to analyse exon 2 of DLA-DQA1, DQB1 and DRB1 genes in 99 samples from Carpathian population and 79 samples from Dinaric-Balkan population. I found four DQA, seven DQB and seven DRB alleles, with ten haplotypes in the Carpathian population, and eight DQA, 11 DQB and 12 DRB alleles, with 18 haplotypes in the Dinaric-Balkan population. Two alleles in Dinaric-Balkan population are unknown from previous research. There are strong indications of positive selection acting upon all examined loci in both populations, and a strong linkage disequilibrium between all pairs of loci. Dinaric-Balkan population seems to be the most diverse European population with regards to MHC genes.

(132 pages, 17 figures, 68 tables, 121 references, original in English)

Keywords: dog leukocyte antigen, Carpathian population, Dinaric-Balkan population, balancing selection

Supervisors: Ana Galov, PhD, Associate Professor
Pavel Hulva, PhD, Associate Professor

Reviewers:

Sveučilište u Zagrebu Doktorski rad

Prirodoslovno-matematički fakultet

Biološki odsjek

PROCJENA MHC RAZNOLIKOSTI U POPULACIJAMA VUKA (*Canis lupus*) SREDNJE I JUGOISTOČNE EUROPE

ŽELJKO PAVLINEC

Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet

Raznolikost glavnog sustava tkivne podudarnosti smatra se dobrim genetičkim pokazateljem fitnesa populacije i sposobnosti prilagodbe na promjene u okolišu, s obzirom da utječe na fitnes jedinke, vijabilnost populacije i njene izglede za dugoročno preživljavanje. Danas u Europi postoji deset populacija vuka (*Canis lupus*), od kojih su dvije najveće populacije karpatska i dinarsko-balkanska koje broje oko 4000, odnosno 5000 jedinki. U disertaciji je istražena raznolikost lokusa MHC klase II u populacijama vuka srednje i jugoistočne Europe, to jest populacija Karpatskog i Dinarskog gorja. Eksoni 2 lokusa DLA-DQA1, DQB1 i DRB1 analizirani su metodama sekvenciranja i molekularnog kloniranja u 99 uzoraka karpatske populacije i 79 uzoraka dinarsko-balkanske populacije. Identificirano je četiri DQA, sedam DQB i sedam DRB alela te deset haplotipova u karpatskoj populaciji, dok je u dinarsko-balkanskoj populaciji identificirano osam DQA, 11 DQB i 12 DRB alela, te čak 18 haplotipova. Dva alela nađena u potonjoj populaciji nisu od ranije poznata. U obje populacije rezultati ukazuju na postojanje pozitivne selekcije na svim istraženim lokusima te na snažnu neravnotežu vezanosti alela između svih parova lokusa. Dinarsko-balkanska populacija pokazuje najvišu razinu genetičke raznolikosti procijenjenu na MHC lokusima od svih europskih populacija vukova.

(132 stranica, 17 slika, 68 tablica, 121 literaturna navoda, jezik izvornika engleski)

Ključne riječi: pseći leukocitni antigen, Karpatska populacija, Dinarsko-Balkanska populacija, ravnotežna selekcija

Mentori: dr. sc. Ana Galov, izvanredni profesor
dr. sc. Pavel Hulva, izvanredni profesor

Ocjenjivači:

Table of Contents

1. Introduction 1
   1.1. Objectives of research 2
   1.2. Methodology 3
2. Literature review 4
   2.1. The Major Histocompatibility Complex 4
   2.2. Grey wolf 13
      2.2.1. European grey wolf 13
      2.2.2. Carpathian population 15
      2.2.3. Dinaric-Balkan population 15
   2.3. Previous research on DLA in wolves 16
   2.4. Tests of selection 18
3. Materials and methods 21
   3.1. Samples 21
   3.2. Polymerase chain reaction 31
   3.3. Electrophoresis 31
   3.4. Sequencing 32
   3.5. Molecular cloning 32
   3.6. BioEdit 36
   3.7. SeqScape 36
   3.8. MEGA 7 37
   3.9. iTOL 40
   3.10. Arlequin 3 40
   3.11. EasyCodeML 41
   3.12. Inference of haplotypes 42

3.13. Mapping 42
4. Results 43
4.1. Carpathian population alleles and haplotypes 43
4.2. Dinaric-Balkan population alleles and haplotypes 48
4.3. Evolutionary distances 54
4.4. Hardy-Weinberg equilibrium 56
4.5. Selection 57
4.6. Linkage disequilibrium 70
4.6.1. Carpathian population 70
4.6.2. Dinaric-Balkan population 73
4.7. Comparison of DLA alleles from all explored European populations 77
5. Discussion 91
6. Conclusions 102
7. References 103
8. Appendix 117
8.1. Carpathian population tables 117
8.2. Dinaric-Balkan population tables 122
8.3. List of abbreviations 129
9. Biography 131

1. Introduction

Most research on genetic diversity for the purpose of wildlife conservation, exploring biodiversity on a population level, as well as population structuring has so far been conducted using microsatellite DNA, since this marker offers high level of polymorphism (Arif et al., 2011; Bruford et al., 2017). The downside of using this marker is that microsatellite DNA is part of non-coding genome, and thus cannot be used to explore adaptive genetic diversity. The genes of the major histocompatibility complex provide us with a unique opportunity to circumvent the downside of using microsatellite loci.

The major histocompatibility complex (MHC) has a central role in the immune regulation and immune response in vertebrates (Andreis et al., 2010; Kindt et al., 2007). This set of genes can be found in all higher vertebrates (Britannica, 2018) and at least some fish (Yamaguchi & Dijkstra, 2019). Because of the nature in which MHC provides resistance to pathogens, which will be discussed in detail in the following chapter, some of the genes belonging to this complex are the most polymorphic coding parts of DNA (Lokki & Paakkanen, 2019). Looking at the currently available data regarding the polymorphism of human MHC, we can see that there are over 23 thousand known alleles (Robinson et al., 2015). This polymorphism allows us to use MHC loci as a marker in intraspecies studies of genetic diversity. The exact origin of this allelic richness is still under debate, but the overwhelming support is on the side of mechanisms of balancing selection (Llaurens et al., 2017; Takahata & Nei, 1990), which will also be discussed in more detail in the following chapter.

Grey wolf (*Canis lupus*) belongs to the family Canidae of the order Carnivora. Historically it inhabited the entire northern hemisphere, but today it has been extirpated in Mexico, most of the USA and most of Western Europe. Currently, it is not considered under threat of extinction since the global population estimate is around 250 thousand individuals. There is a high level of variation on MHC loci in wild populations of grey wolves (Arbanasić et al., 2013; Galaverni et al., 2013; Kennedy et al., 2007), with observed signatures of balancing selection. Genetic diversity research in European grey wolf, using MHC as a molecular marker, has so far covered northern European populations (Niskanen et al., 2014; Seddon & Ellegren, 2004), and southern European populations (Arbanasić et al., 2013; Galaverni et al., 2013), but to my knowledge, no research has been done on MHC diversity in grey wolf in central or eastern Europe. Hulva et al. (2018) gave novel data regarding distribution and genetic structure

of the grey wolf population in Central Europe. Their research included samples from the Carpathian population, but was conducted using neutral markers, so no conclusions about adaptive genetic diversity could be inferred from it. This thesis provides an opportunity to fill this knowledge gap, and compare MHC diversity in European grey wolf populations over much broader geographic area.

1.1. Objectives of research

The main objective of this dissertation is to assess adaptive genetic diversity of grey wolves (*Canis lupus*) from Carpathian and Dinaric-Balkan populations by analysing second exon of the major histocompatibility class II DQA, DQB and DRB loci.

Specific aims of this research are to:

- Determine and analyse alleles and three-locus haplotypes on MHC class II DQA, DQB and DRB loci in the population of European grey wolf from Carpathian Mountains.
- Determine and analyse alleles and three-locus haplotypes on MHC class II DQA, DQB and DRB loci in the population of European grey wolf from Serbian part of Dinaric Mountains.
- Test for the presence of selection on examined loci.
- Compare obtained data on genetic diversity with published data on European grey wolf populations.
- Make a phylogeographic analysis of newly obtained and previously available data on examined loci.

1.2. Methodology

To fulfil the objectives of this thesis I have used 99 samples of genomic DNA from grey wolves from the Carpathian population, and 79 samples of genomic DNA from grey wolves from the Dinaric-Balkan population. Polymerase chain reaction (PCR) was used to selectively amplify second exon of DLA-DQA1, DQB1 and DRB1 genes. SeqScape software (Applied Biosystems) was used to compare each new sequence to a library of known allele sequences to identify the closest allele or combination of alleles in each sample. For heterozygous wolves whose exact alleles couldn't be determined using in-silico methods, and for those with presumed new alleles I performed molecular cloning using pGEM®-T Vector System (Promega) with JM109 Competent Cells (Promega). Once the allele composition of the entire sample of both populations was determined, I performed several in-silico analyses to test for the presence of selection, to calculate several parameters of molecular evolution and to calculate and visualise the phylogenetic trees. The software used for those analyses were: Arlequin 3 (Excoffier et al., 2007), EasyCodeML (Gao et al., 2019), MEGA 7 (Kumar et al., 2016) and iTOL (Letunic & Bork, 2019).

2. Literature review

2.1 The Major Histocompatibility Complex

The major histocompatibility complex is a set of genes that controls the immune response in vertebrates. This set of genes codes for membrane molecules whose main function is to present processed peptide antigens to T lymphocyte cells (Klein, 1986). Since T lymphocytes recognize foreign antigens exclusively within the MHC molecule – foreign antigen complex, the MHC determines the clonal selection of T lymphocytes and therefore has a central immunoregulatory role (Kindt et al., 2007). By their primary function MHC genes, and molecules are divided into two main groups, MHC class I and MHC class II. The MHC class I molecules regulate the function of CD8+ cytotoxic T lymphocytes mediating the immune response against endogenous antigens, while the MHC class II molecules are involved in the presentation of exogenous antigens to helper T lymphocytes (Andreis et al., 2010). To explain how MHC molecules present antigens to lymphocytes, we first need to look at their structure. Both class I and class II molecules have intracellular part, transmembrane part and extracellular part (figure 1).

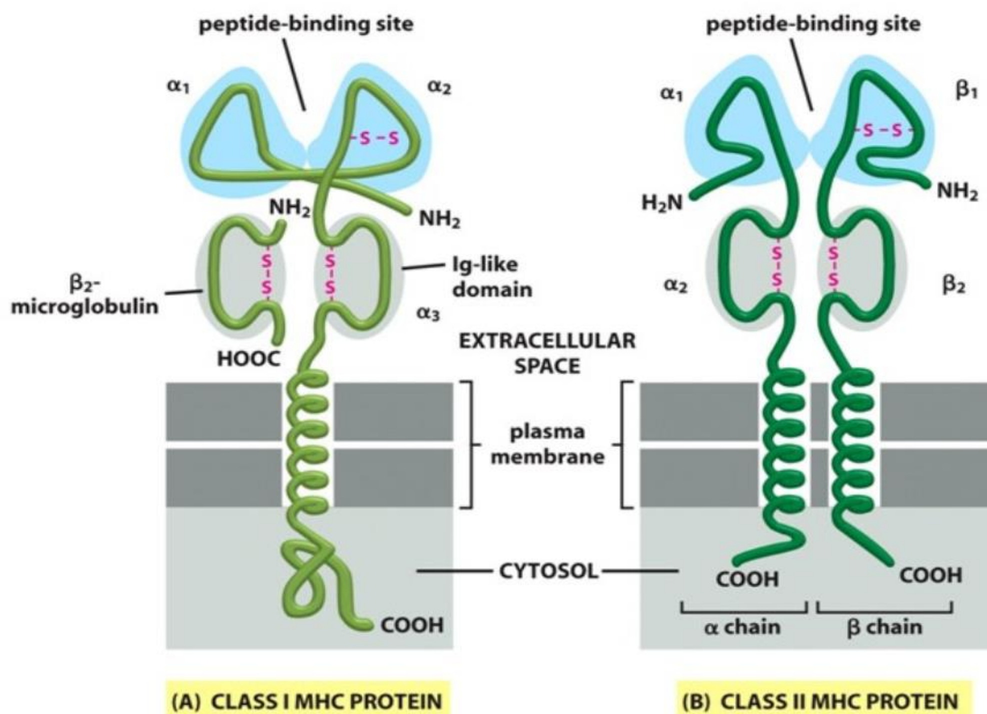

Figure 1. Schematic of MHC I and MHC II molecules. Taken from Alberts et al. (2008).

MHC I molecules are built from one smaller globular polypeptide called β2 microglobulin, and one larger polypeptide chain called α chain, which is divided into three separate extracellular domains, α1, α2 and α3. The MHC II molecules are built from two polypeptide chains, called α and β chain, and both chains are divided into 2 extracellular domains, α1 and α2 for the α chain and β1 and β2 for the β chain (Andreis et al., 2010). The two extracellular domains positioned farthest from the cell membrane form on their contact site a peptide binding groove. This groove is the position in which processed peptides are held for the presentation to T lymphocytes, and which peptides will be able to be presented depends on the chemical composition of the peptide binding groove. The degraded peptides coming from inside the cell, either from cellular metabolism or from intracellular pathogens are, as already stated, presented by MHC I molecules, so these molecules can be found on almost all cells in every tissue in vertebrate animals.

Regular pathway is presented in figure 2, which shows how after a peptide is degraded by a proteasome, it's fragments are transferred inside the endoplasmic reticulum (ER) where they are bound by MHC I molecules found in the ER membrane. Once the MHC I – peptide complex is formed, it will be transferred through the cell's secretory system, and end up positioned on the surface of the cell. Once such complex is recognized as foreign by CD8+ T lymphocytes, the cell usually undergoes apoptosis to reduce the risk of infection spreading to neighbouring cells (Andersen et al., 2006).

Peptides coming from outside of the cells, from the cellular environment and extracellular pathogens, are presented by MHC II molecules, so they are exclusively found on specialized antigen presenting cells. The pathway to presenting the MHC II – peptide complex on the surface of the antigen presenting cells is similar to the one described above for MHC I molecules, with some major differences. First difference is that the foreign protein is not intracellular, so it has to enter the antigen presenting cell, which is usually done by phagocytosis. The second major difference is that the complex MHC II – peptide is not formed in ER, but in the endosome once the invariant chain, a small peptide which blocks the peptide binding groove of MHC II molecules, is degraded. The MHC II molecules are formed in ER already complexed with the invariant chain so that they wouldn't bind peptides from the endogenous (MHC I) pathway (Roche & Furuta, 2015).

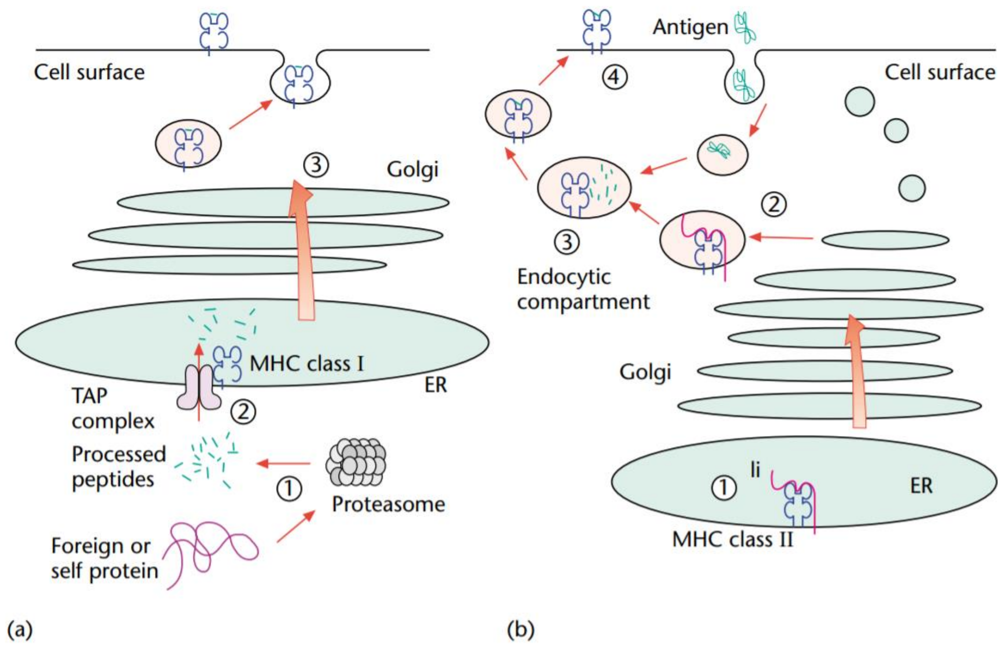

Figure 2. Schematic representation of MHC class I (a) and class II (b) pathways for antigen processing and presentation. MHC – the major compatibility complex, ER – endoplasmic reticulum, TAP – peptide transporter, Ii – invariant chain. Taken from Sandberg and Glas (2001).

Since peptide binding groove is the primary place of recognition of foreign antigens, which starts the immune response, the amino acid composition of the groove has large variations even between individuals of the same species. A single MHC molecule can bind several peptides, different in their primary structure, which contain similar key amino acid residues at the same positions, but no MHC molecule can bind all kinds of foreign peptides (Klein et al., 2007). That is why it is important for a population to maintain high diversity of its MHC molecules, so that it can recognize and defend itself from many different pathogens. For this reason MHC class I and II are the most polymorphic coding parts of DNA, since this high degree of polymorphism allows for the recognition of a wider spectrum of pathogens, so MHC diversity is important for species survival (Heimeier et al., 2009). The extent to which MHC diversity contributes to ensuring the survival of a population is still one of the fundamental questions of conservation genetics, but it is considered that a lack of diversity can increase population susceptibility to infectious diseases with potentially deleterious consequences (Yang et al., 2007). The level of diversity on MHC loci, and the number of different alleles even in the same species is astonishing when compared to the rest of the coding parts of the vertebrate genome. Early population studies in mice (*Mus musculus*), using MHC as a marker, showed

that even the small subpopulations have over two dozen different alleles at each of the two main MHC I loci, and that different subpopulations contain different alleles (Klein et al., 2007). If we look at the data from humans, which are by far the most studied species when it comes to MHC region, we can see extreme allelic richness with over 17,100 alleles at class I loci, and over 6,700 alleles at class II loci (Robinson et al., 2015). Also, comparing different populations of humans, we can see extreme differences in allelic composition and frequencies, especially between populations from different continents or from different ethnic groups (Gonzalez-Galarza et al., 2015).

It is considered that high diversity on MHC loci is maintained by balancing selection, which is a broad term covering different types of mechanisms of natural selection which act in such a way that no one allele has the greatest fitness, so the result is a high polymorphism and/or a high proportion of heterozygotes in a population (Llaurens et al., 2017). While the exact mechanisms through which balancing selection acts on MHC loci is still a matter of debate in evolutionary biology, the best explanation given so far is that it is a pathogen mediated selection, and acts through mechanisms of heterozygote advantage, rare allele advantage and by temporally and spatially fluctuating selection. The mechanism of heterozygote advantage will maintain high number of alleles in a given population because heterozygotes on MHC loci can recognize more peptides than any given homozygote, so they can initiate immunological response against a larger number of pathogens (Hughes & Nei, 1988). The rare allele advantage is easiest to explain when observed in context of host – pathogen co-evolution. In a constant evolutionary race there is a strong selective pressure on pathogens to evolve resistance to host's immune response. This is why often new alleles grant higher fitness to the individuals, since they can grant resistance to newly evolved pathogens. The similar can be applied to old, rare alleles whose frequency has decreased due to lower presence of pathogens for which they granted resistance. If the same or similar pathogens start to increase in numbers there will again be a higher positive selection on these old, rare alleles (Spurgin & Richardson, 2010). The mechanism of temporally and spatially fluctuating selection is similar to the rare allele advantage mechanism in that the selective pressure on a given allele in a population can increase or decrease based on the composition of pathogens in an environment the population is currently occupying. When the population moves in time, whether seasonally or over longer periods, or when the population moves in space, if the pathogen composition changes (either with time or in the new environment) the advantage conferred by specific alleles will also change (Hill, 1991). The major difference between fluctuating selection mechanism and the rare allele

advantage mechanism is that in the former the selective pressure of pathogens on their hosts is not determined by their co-evolution, but by biotic and abiotic environment, chance dispersal and extinction events (Spurgin & Richardson, 2010).

In addition to balancing selection, MHC based sexual selection or MHC based disassortative mating is another force acting to increase heterogeneity in MHC region. The exact way in which mate choice is based on differences in MHC alleles is greatly dependent of species, but some of the mechanisms are olfactory recognition of mates with different alleles, increasing the chance of producing heterozygous offspring, and visual and/or olfactory recognition of diseased individuals, or parasite carriers, thus eliminating these individuals from mating and consequently reducing the frequency of their alleles in offspring (Milinski, 2006). MHC based disassortative mating can explain how functional MHC variants which are not under current selective pressure are protected against genetic drift, and also how allele frequencies and genetic variation can be stabilized even in smaller populations (Ejsmond et al., 2014).

Considering all the reasons previously stated for increased diversity on MHC loci, one can wonder why a proportion of homozygotes in populations is still relatively high. One reason for this, besides random pairing, is that if there is a strong selective pressure on the population caused by certain pathogen, the number of individuals carrying the allele which provides protection against it will increase in future generations, and subsequently the number of individuals homozygous for this allele will also increase. But another major reason is that many MHC alleles are linked to increased susceptibility to infectious diseases caused both by viral and by bacterial pathogens, as well as increased susceptibility to most autoimmune diseases (Matzaraki et al., 2017), so depending on the selective pressure on the population in any given time and space some alleles can cause negative fitness. To make matters even more complicated sometimes the same MHC allele can grant increased resistance to one disease while granting increased susceptibility to another disease. This can even occur with similar diseases caused by different pathogens, so for example Loiseau et al. (2008) have shown how the same MHC I allele can grant resistance to *Haemoproteus* while increasing susceptibility to *Plasmodium*, similar parasites causing malaria in house sparrows (*Passer domesticus*).

In addition to high numbers of different alleles on MHC loci, what contributes to the total MHC diversity is the level of allele divergence, i.e. how different the alleles on the same loci are, when compared to each other. It is common that two MHC alleles on the same loci

differ by 20 substitutions (Klein et al., 2007), and some can differ by up to 20 amino acids (Janeway et al., 2001). Most of this diversity is located in the domains forming peptide binding region, and in MHC II molecules mostly in the β chain (figure 3).

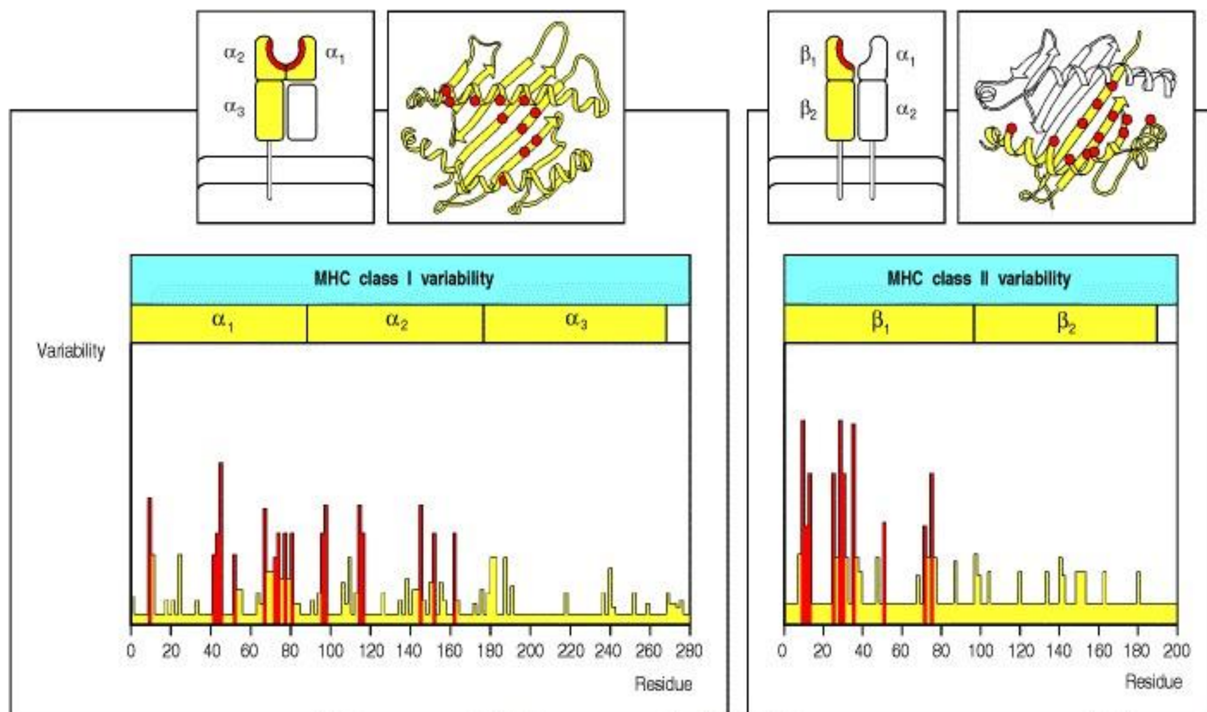

Figure 3. Schematic of allelic variation at specific sites within MHC molecules. In lower boxes the variability plots of amino acid sequences are shown, and in the upper boxes are schematic representations of MHC molecules and secondary structures of peptide binding groove. Different positions of amino acids in peptide binding groove which are key for antigen recognition are highlighted in red. Taken from Janeway et al. (2001).

The evolution of the MHC is characterized by repetitive gene duplications and losses, and the phylogenetic patterns can show whether duplications occurred in the modern species or in an ancestor (Strand, 2011). The MHC gene organization, their number and level of expression are highly variable between different vertebrate groups, but also between taxonomically close species (Kelley et al., 2005). The best studied MHC region is the one in humans, so it is also the best example on which to show MHC organization (figure 4). In humans this region is also called human leukocyte antigen (HLA). HLA region is located on the short arm of chromosome 6, position 6p21, and contains over 130 protein coding genes that have important roles in immune system regulation (Shiina et al., 2009).

HLA I and HLA II regions contain what is referred to as „classical“ HLA genes which code for previously described MHC molecules. HLA III region contains numerous protein coding genes not directly involved with cell immunity, like complement proteins (e.g. C2 or C4), tumour necrosis factors (TNF) or heat shock proteins (e.g. Hsp70). They serve as mediators in many regulatory pathways of immunological response. The reason why HLA region has been explored so extensively is not only because it controls our immune response to pathogens, but because of many associations between certain diseases, especially autoimmune diseases, and specific HLA alleles (Gough & Simmonds, 2007). Another major reason is because some HLA genes determine tissue compatibility in transplantation, where MHC molecules produced in donor tissue act as foreign antigens to recipient MHC molecules, triggering the immune response, which results in rejection of donor tissue (Montgomery et al., 2018).

One of the major structural features of MHC is a strong linkage disequilibrium, or a non-random association between alleles on different MHC loci. This is probably caused by differences in recombination rates between some loci, and due to this, MHC is organized in haplotype blocks (Wall & Pritchard, 2003). How strong this link between certain alleles can be is easily depicted in almost any study on MHC allele and haplotype diversity that was done on a larger sample size. For example, in a recent study of HLA allele haplotype frequencies on 120,926 volunteers from the Italian Bone Marrow Donor Registry, Sacchi et al. (2019) found the allele HLA-A*01:01 with frequency 11.5281%, HLA-B*08:01 with 5.7597 %, HLA-C*07:01 with 17.1076%, and HLA-DRB1*03:01 with frequency 9.4723%. If we would assume random pairing between these alleles, with no linkage, we would expect to see haplotype HLA-A*01:01/B*08:01/C*07:01/DRB1*03:01 with total frequency of 0.01%, while in reality it was found in frequency of over 2.5%.

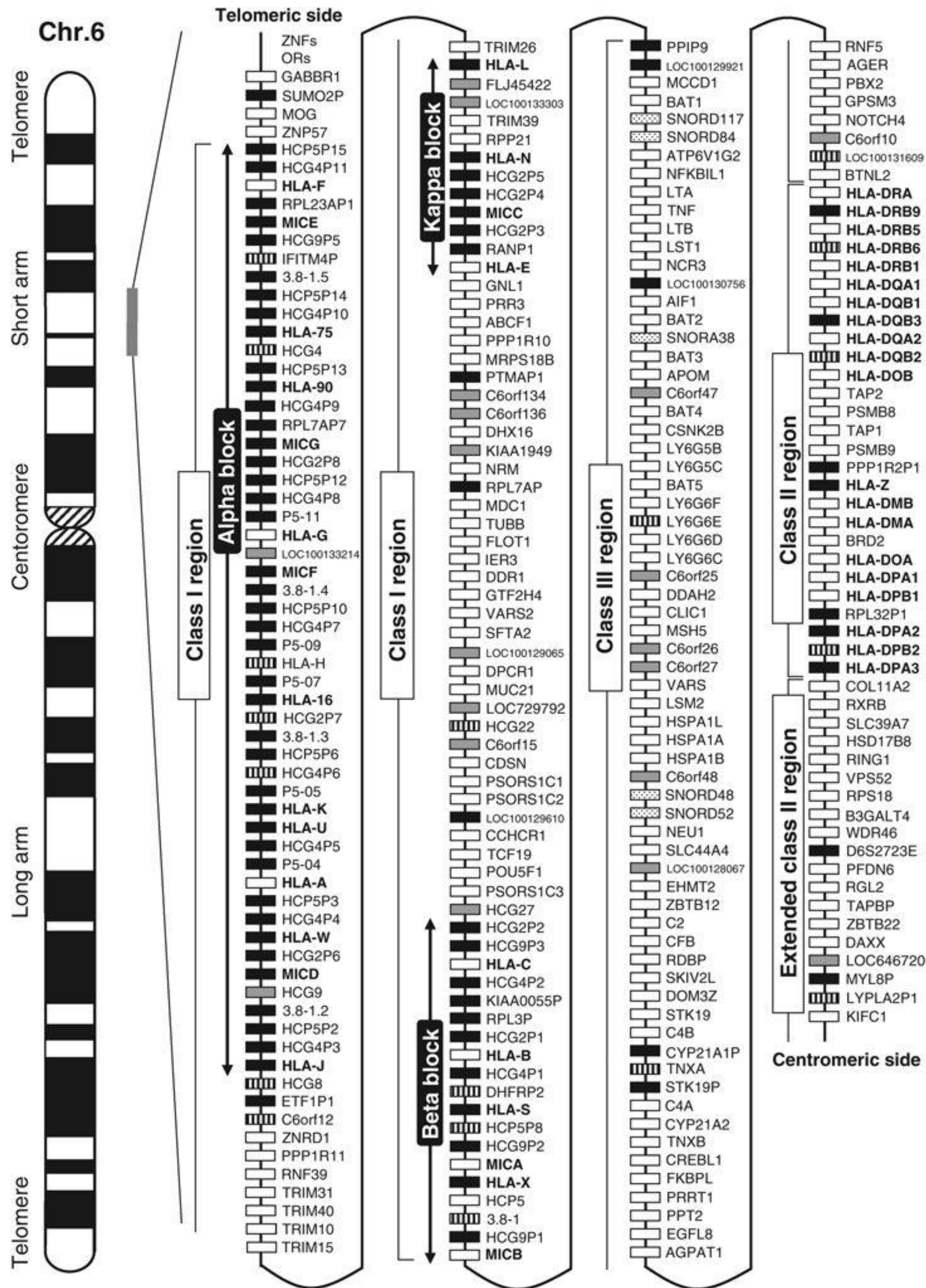

Figure 4 Gene map of the HLA region. White, grey, striped and black boxes show expressed genes, gene candidates, non-coding genes and pseudogenes, respectively. The location of the alpha, beta and kappa blocks containing the cluster of duplicated HLA class I genes in the class I region are indicated. Taken from Shiina et al. (2009).

Some MHC alleles can be preserved through a long evolutionary period and their origin may predate the speciation of the species in which they are found, resulting in a similarity or even identity between alleles of closely related but different species (Klein et al., 1998). This is usually referred to as trans-species polymorphism, and it occurs only in genetic systems under balancing selection which is, as already described, a typical mode of evolution of the MHC region. It was previously documented in Primates (Huchard et al., 2006), Perissodactyla (Radwan et al., 2007), Rodentia (Cutrera & Lacey, 2007), Carnivora (Kennedy et al., 2007), Didelphimorphia (Meyer-Lucht et al., 2008), and Artiodactyla (Xu et al., 2009).

2.2. Grey wolf

Rare are the examples of animals that have as complex relationship with humans as the wolves. They have been feared and revered; humans have domesticated them for labour, security and friendship, while at the same time hunted them for food, materials and safety. Thankfully, todays perceptions of wolves are shifting form dangerous predator to an endangered species (Masius & Sprenger, 2015). And even in popular culture we can observe the shift from the villainous Big Bad Wolf of Aesop's Fables and Grimms' Fairy Tales to the fatherly protector Bigby Wolf of Willingham's Fables.

Wolves have adapted to extreme natural habitat conditions, so their distribution range is very broad, and mostly linked to prey densities and influenced by anthropogenic factors. As carnivores they typically require three to five kilograms of meat per day, and their diet is as diverse as their habitats, including mammals of all sizes, birds, invertebrates, vegetation and carcasses. Although they are not limited to forests, in Europe they prefer large forest areas (Boitani, 2018).

2.2.1. European grey wolf

European grey wolf (*Canis lupus*) can currently be found over much of the Europe, and is considered as a resident species in 36 out of 51 European countries (Boitani, 2018). In a comprehensive review on the status of large carnivores in Europe, Chapron et al. (2014) show that even though global biodiversity crisis caused mainly by human overpopulation and overconsumption is increasing, European continent is succeeding in maintaining, and even restoring, viable large carnivore populations on a continental scale. They further show that the current populations of European grey wolf can be clustered into ten populations (figure 5): Scandinavian (1), Karelian (2), Baltic (3), Central European Lowlands (4), Carpathian (5), Dinaric-Balkan (6), Italian peninsula (7), Alpine (8), Sierra Morena (9) and North-West Iberian (10); numbering in total over 12,000 individuals, while according to Boitani (2018) they exceed 17,000 individuals. Even though in the first half of 20th century most of the populations of grey wolf in Europe were in decline, and some even threatened with extinction, in recent decades, excluding the Sierra Morena population, the grey wolf populations in Europe have been either stable or in constant increase in number of individuals. Observations made by Chapron et al.

(2014) were that the main reason for this is coordinated legislation to specific management practices and institutional arrangements shared by many European countries since the end of World War II.

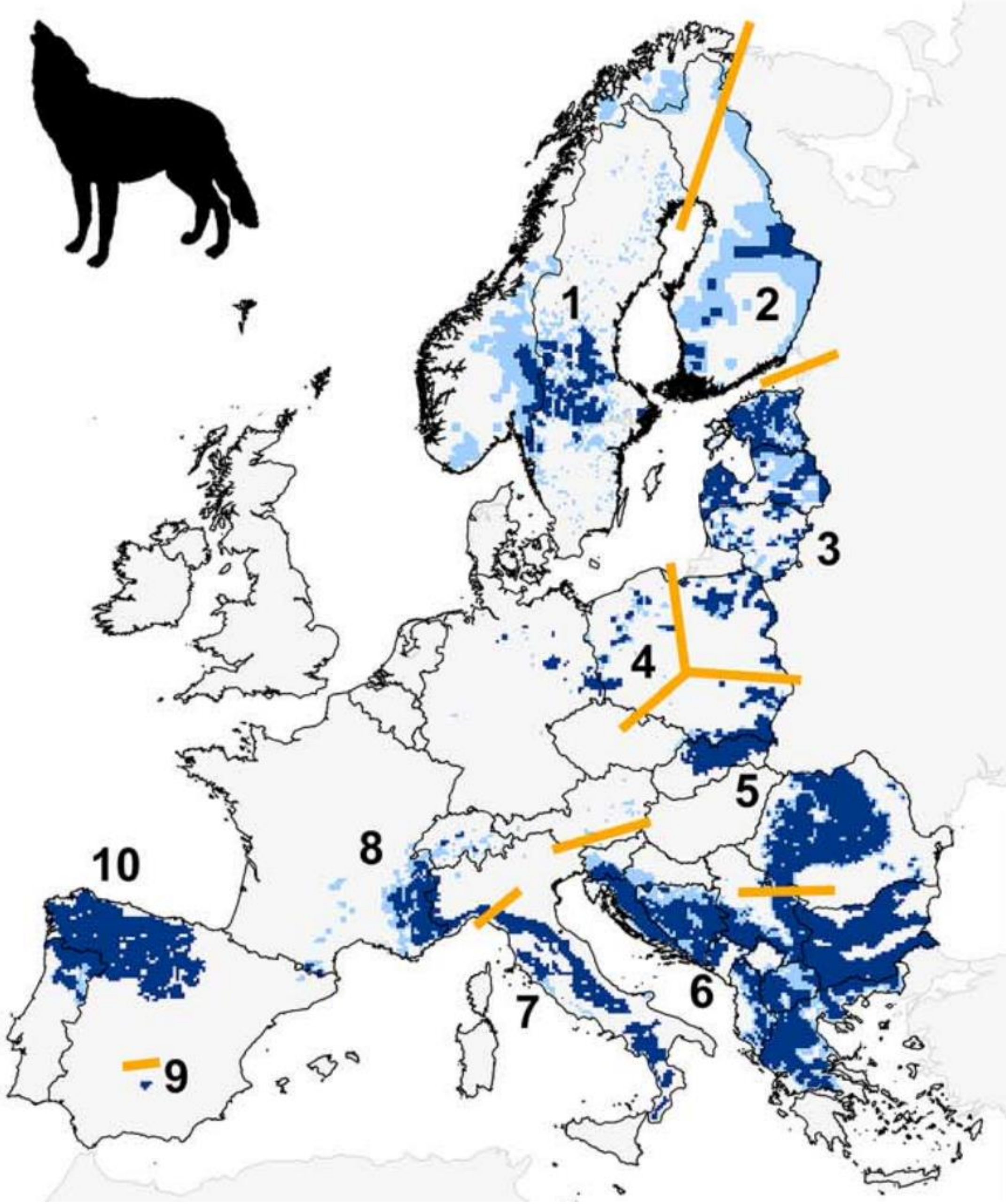

Figure 5 Wolf populations in Europe in 2011 (taken from Chapron et al. (2014)). Dark blue indicates areas of permanent occurrence; light blue indicates areas of sporadic occurrence; orange lines are boundaries between populations; numbers refer to population identifications in text.

2.2.2. Carpathian population

The Carpathian population of grey wolf, currently numbers between 3,500 and 4,000 individuals. Today this population inhabits the area from northern Bulgaria, northeast Serbia, Romania and southwest Ukraine in the east of Europe, through Slovakia, north Hungary, and south Poland, and ending in Czech Republic in the central Europe (Boitani, 2018; Chapron et al., 2014; Kutal et al., 2017). Historically it inhabited a continuous range including wolves from Poland and Belarus, but is now constrained by large areas where wolves have been exterminated by humans (Boitani, 2018). It is likely that some genetic exchange occurs with the Dinaric-Balkan population in Bulgaria and Serbia (Rigg et al., 2014), and with the Baltic population in eastern-central Poland (Boitani, 2018). The main problem when dealing with the conservation of this population is the major differences in the legislative frame regarding the protection status of wolves in different countries of this region. So, while for example in Poland the wolves are strictly protected species, in Ukraine they are not protected, and sometimes hunting clubs even offer bounties on them (Salvatori et al., 2002). Molecular study using neutral markers (Hulva et al., 2018) showed that the population is sub-structured into four distinct subpopulations, likely due to anthropogenic influence.

2.2.3. Dinaric-Balkan population

The Dinaric-Balkan population also numbers close to 4,000 individuals, but is thought to be severely sub-structured because of the elongated geographical range. It spans from Slovenia in the northwest, through Croatia, Bosnia and Herzegovina, Serbia, Kosovo, Montenegro, North Macedonia and Albania to southwest Bulgaria in the east and Greece in the south (Boitani, 2018; Chapron et al., 2014). On the eastern border there may be exchange of genetic material with the large wolf population of the Carpathians which extends as far as into northern Bulgaria and eastern Serbia (Boitani, 2018). Since the area is politically fragmented, there are vast differences in conservation status, demographic trends and even available data throughout the range. Slovenia and Croatia have implemented an Action Plan for monitoring and management of wolves and since the 1990s the subpopulation in these two countries has significantly recovered. In the rest of the range the legislative frame for protection of wolves is either poor or non-existent, and there is limited research and development in wildlife management. In Bosnia and Herzegovina and in Serbia, because of overharvesting, the

subpopulation has recently decreased. Throughout the range there are major threats in the form of nonregulated or illegal hunting and habitat fragmentation due to anthropogenic activities such as land use, deforestation and transportation (Boitani, 2018).

The grey wolf is a game species in most of its range in Serbia, being legally protected only in smaller, northern part of its range. Its population size was estimated to 700-800 individuals, with population that is stable or slowly increasing (Milenković et al., 2007). Based on mtDNA analysis Djan et al. (2014) detected a clear genetic subdivision within the Dinaric-Balkan wolf population with samples from Serbia and Macedonia clustering into „eastern“ subpopulation, while samples from Croatia and Bosnia and Herzegovina clustered into „western“ subpopulation. They also confirmed relatively high genetic diversity within Dinaric-Balkan wolves although they detected signal of recent population bottleneck (which was not supported by other available data, on the contrary, constant population size increase was reported for grey wolves of Serbia (Milenković et al., 2007). Lastly, they did not obtain support for presumed subdivision within the Serbian wolf population. The notion that grey wolves from Serbia could be considered as one continuous population without detectable population genetic structure was further confirmed by microsatellite analyses (Đan et al., 2016).

2.3. Previous research on DLA in wolves

When trying to start molecular studies in most non-model model organisms, scientists will often encounter the same problem, high initial cost compared to perceived gain, which is often hard to justify. For this reason MHC research in many vertebrate species is non-existent. Fortunately, wolves have almost identical MHC gene structure and organization as dogs (*Canis lupus familiaris*), economically more relevant canid subspecies. So far, a lot of the research on canid MHC region has been performed in dogs (Wagner, 2003), and in the last decade most of it with a focus on application in veterinary medicine (Pedersen et al., 2011; Safra et al., 2011; Venkataraman et al., 2017). This research was mainly focused on the second exon of class II genes, which in early studies showed high polymorphism in the positions coding for the peptide binding site (Wagner et al. 1996). Similar to HLA, the DLA or dog leukocyte antigen (Kennedy et al., 1999) is grouped into three polygenic classes, with the difference being that the DLA complex is split and situated on two chromosomes (Yuhki et al., 2007). Trans-species polymorphism on DLA loci in dogs, coyotes (*Canis latrans*), and several species of wolves has

been documented more than two decades ago (Hedrick et al., 2000; Kennedy et al., 2001). Some DLA alleles, and increased homozygosity on DLA loci have already been linked with certain diseases and resistance to pathogens in different dog breeds (Barber et al., 2011; It et al., 2010; Jokinen et al., 2011; Kennedy et al., 2006; Quinnell et al., 2003), but also in the endangered Mexican wolf (*Canis lupus baileyi*) (Hedrick et al., 2003) and Ethiopian wolf (*Canis simensis*) (Kennedy et al., 2011).

Seddon and Ellegren (2004) explored the MHC diversity of wolves in northern Europe. Using samples of 90 wolves from Scandinavia, 25 wolves from Estonia, 22 wolves from Finland, 15 wolves from Latvia and 51 wolves from northwest Russia they found in total seven DQA, 10 DQB and 15 DRB alleles with extremely varied distribution among different regions. Niskanen et al. (2014) again explored the populations from Finland and northwest Russia using samples of 242 and 37 wolves, respectively. They found seven DQA, nine DQB and 11 DRB alleles in total, with only one rare DQB allele unique to samples from Finland.

In southern Europe, wolf MHC diversity has been explored in animals from Croatia (Arbanasić et al., 2013) and Italy (Galaverni et al., 2013). In 77 wolves from Croatia they found 7, 11 and 13 alleles on DQA, DQB and DRB loci, respectively. In 71 wolves and three wolf x dog first generation hybrids from Italy they found six DQA, eight DQB and nine DRB alleles. Some alleles are unique to each population, and there are large differences when comparing the northern and southern populations. This will be further explored once I present my results on the populations from central and south-eastern Europe.

2.4. Tests of selection

Using the data obtained by sequencing we can answer some interesting questions regarding the evolution of our subject of interest. In the previous chapter it was stated that high number of MHC alleles in the same species is maintained by balancing selection. It is a special form of positive selection which aims to increase intraspecies genetic diversity on a given loci (Charlesworth, 2015; Koenig et al., 2019; Llaurens et al., 2017; Mérot et al., 2020). There are several tests we can perform to test whether there is some type of selective pressure acting on the loci under investigation.

One of the first thing we can look at is the Hardy-Weinberg Theorem (Hardy, 1908; Weinberg, 1908), or whether the genotype frequencies in our population of interest deviate from Hardy-Weinberg expectations. Hardy-Weinberg theorem states that the allele and genotype frequencies remain constant from generation to generation if all of the assumptions of the model are true. In this case we can say that the population is in Hardy-Weinberg equilibrium. If the population is not in equilibrium, but the assumptions of the model are true, it will take only one generation of random mating to achieve it. The assumptions of the model are: the population is panmictic, the population is large enough that it can be treated as if it were infinite, there is no gene flow, there is no mutation in the genes and there is no natural selection affecting the locus (Futuyma, 2009). If there is any violation in these assumptions we can expect deviation from Hardy-Weinberg equilibrium.

When conducting research on multiple loci in the same organism it is important to check whether they are in linkage equilibrium, which assumes random association of alleles at different loci, similar to Hardy-Weinberg equilibrium. Contrary to Hardy-Weinberg equilibrium, one generation of random mating will not suffice to reach linkage equilibrium, but it will happen slowly for closely linked loci (Lewontin & Kojima, 1960). Genetic drift, gene flow, mutation and selection will likewise affect linkage equilibrium, so there are many conditions under which linkage disequilibrium will persist (Slatkin, 2008).

When looking at allelic composition of a population, over time we can expect a different outcome of mutation in the presence of selection, from the one in which no selection is present. Probably the simplest way of testing whether a certain locus is under positive selection is to compare the synonymous and nonsynonymous substitution rates on a given locus. A telling sign of positive selection would be if we observe a ratio of nonsynonymous and synonymous

substitution rates significantly greater than one (Yang, 1998). This test can be done over an entire sequence of the genomic region of interest, or it can be done on a codon by codon basis. For a more powerful test we can employ a Z-test of selection to test the null hypothesis $H_0 : d_N = d_S$, where d_N is the number of nonsynonymous substitutions per nonsynonymous site and d_S is the number of synonymous substitutions per synonymous site (Nei & Kumar, 2000). The test is calculated as $Z = \frac{(d_N - d_S)}{\sqrt{(Var(d_S)+Var(d_N))}}$, and its significance depends on the alternative hypothesis (HA). The possible alternative hypotheses are: 1. HA : dN ≠ dS which tests for non-neutrality, 2. HA : dN > dS which tests for positive selection and 3. HA : dN < dS which tests for negative or purifying selection.

Another way to check whether our loci of interest evolve neutrally or not is to perform Tajima's Test of Neutrality (Tajima, 1989). If applied on an infinite-site model, if a population is under no selective pressure, and if the population is in a mutation – genetic drift balance, there are two ways to approximate the product of the effective population size (N) and the mutation rate per site (μ). One way is to calculate the observed number of segregating sites per nucleotide site (p_s) from our data set and, if the assumptions of the model are true, equalize it to the expected number of segregating sites per nucleotide site ($E(p_s)$). Since $E(p_s) = a_1 * \theta$; where $a_1 = 1 + 2^{-1} + 3^{-1} + \cdots + (m - 1)^{-1}$, m equals the total number of sequences in a population and $\theta = 4 * N * \mu$; it follows that $N * \mu = \frac{E(p_s)}{4a_1}$. Another way is to calculate the observed nucleotide diversity (π) from the data set as $\pi = \sum_{ij}^{q} x_i x_j d_{ij}$; where q is the total number of alleles (different sequences) in the population, x_i is the frequency of the i-th allele in the population, x_j is the frequency of the j-th allele in the population and d_{ij} is the number of substitutions per site between the i-th and j-th allele. Again, if the assumptions of the model are true we can equate the calculated nucleotide diversity with the expected nucleotide diversity ($E(\pi)$). Since $E(\pi) = 4 * N * \mu$, it follows that $N * \mu = \frac{E(\pi)}{4}$. When we compare the observed value of the segregating sites per nucleotide site with the observed value of nucleotide diversity, the obtained difference gives us an indication of neutral or non-neutral evolution. If both values are the same it is a good indicator of neutral evolution, if p_s/a_1 is higher compared to π it is an indicator of many deleterious alleles in the population and if π is higher than p_s/a_1 it is an indicator of a force which increases allele frequencies, such as balancing selection (Nei & Kumar, 2000).

When looking for signs of positive selection in coding sequences we can rightfully ask ourselves whether selection works equally on the entire gene, or are there only some parts that are under selective pressure. This can be especially true with genes coding for proteins that bind other molecules, and so form particular binding sites, such as the one formed by peptide binding domains of MHC chains (described above). For this reason we can try to detect positive selection affecting only certain sites in our sequences using CodeML analysis (Gao et al., 2019). In this analysis, we again look at the ratio of nonsynonymous and synonymous substitution rates (ω) as a measure of natural selection, but instead of averaging ω ratio over all sites in the sequence, we use site-specific codon-based substitution models that allow ω ratio to vary among sites (Yang & Nielsen, 2002). CodeML employs seven different site-specific codon-based substitution models that vary in their assumption of the distribution of ω ratio: M0 (one-ratio), M1a (nearly neutral), M2a (positive selection), M3 (discrete), M7 (beta), M8 (beta and $\omega > 1$) and M8a (beta and $\omega = 1$) (Gao et al., 2019; Nielsen & Yang, 1998; Yang et al., 2000; Yang & Nielsen, 2002). Using likelihood ratio tests we can compare the fit of these models to the sequence data. If model M2a shows a better fit than M1a, or if model M8 shows a better fit than M7 or M8a it is a strong indicator of positive selection (Gao et al., 2019).

3. Materials and methods

3.1. Animal samples

Within this study, a total number of individual animals was 178. The animals belonged to two distinct geographical populations, namely the Carpathian and the Dinaric population. The sample collection did not require additional institutional permits or authorisations according to the laws and regulations on protected species acts in the countries of wolves sample origins.

3.1.1. Wolves from the Carpathian population

To analyse the MHC diversity of European grey wolf population from Carpathian population I used 99 samples of genomic DNA from grey wolves found in different areas of Czech Republic and Slovakia. Most of them belonged to western Carpathian population, although certain degree of admixture with central European population was referred to (Hulva et al. 2018). The samples come from tissues from roadkill (Czech Republic) or legally culled (Slovakia) grey wolves. Table 1 provides detailed data with the sample ID, the exact date when the sample was obtained, and the exact location. The local geographic distribution of Carpathian samples can be seen in figure 6. Two samples (VYS and X1) were sampled outside of the Carpathian mountain range. The location of the population samples in Europe can be seen in figure 8.

Table 1. Samples from the Carpathian population

Sample ID	Year	Country	Region	Lat	Long
CL 308	2011	Slovakia	Ondavská vrchovina	50.64690	14.65664
CL 309	2011	Slovakia	Ondavská vrchovina	49.50530	17.99620
CL 310	2011	Slovakia	Ondavská vrchovina	49.50281	15.47380
CL 333	2011	Slovakia	Runina	48.87750	18.57450
CL 353	2011	Slovakia	Poĺana	48.91240	18.72790
CL 364	2011	Slovakia	Volovské vrchy	49.31810	18.88020
CL 369	2011	Slovakia	Poĺana	49.30720	18.90090
CL 372	2011	Slovakia	Volovské vrchy	49.33790	18.90270
CL 374	2011	Slovakia	Nízké Tatry	48.75970	19.14250

Table 1. continued

Sample ID	Year	Country	Region	Lat	Long
CL 375	2011	Slovakia	Nízké Tatry	48.65560	19.21500
CL 638	2013	Slovakia	Veporské vrchy	49.26580	19.22470
CL 639	2013	Slovakia	Veporské vrchy	49.15810	19.23750
CL 640	2013	Slovakia	Malá Fatra	49.25800	19.24660
CL 642	2013	Slovakia	Levočské vrchy	49.24030	19.24860
CL 646	2013	Slovakia	Slovenský kras	49.19640	19.29470
CL 647	2013	Slovakia	Krupinská planina	49.39850	19.30700
CL 648	2013	Slovakia	Veporské vrchy	48.38230	19.31510
CL 649	2013	Slovakia	Veporské vrchy	49.12860	19.36980
CL 653	2013	Slovakia	Levočské vrchy	48.62220	19.37560
CL 656	2013	Slovakia	Oravské Beskydy	49.01300	19.40070
CL 657	2013	Slovakia	Oravské Beskydy	49.11780	19.40680
CL 658	2013	Slovakia	Nízké Tatry	48.65000	19.41310
CL 659	2014	Slovakia	Nízké Tatry	48.63970	19.41310
CL 660	2013	Slovakia	Poĺana	48.74580	19.42690
CL 661	2013	Slovakia	Oravská Magura	48.65000	19.43330
CL 662	2013	Slovakia	Oravská Magura	48.65390	19.44060
CL 663	2013	Slovakia	Chočské vrchy	49.42790	19.45030
CL 664	2014	Slovakia	Volovské vrchy	48.67700	19.45220
CL 665	2013	Slovakia	Levočské vrchy	48.67610	19.45540
CL 666	2013	Slovakia	Levočské vrchy	49.02430	19.54840
CL 667	2013	Slovakia	Levočské vrchy	48.97330	19.54950
CL 679	2014	Slovakia	Levočské vrchy	48.85450	19.57030
CL 680	2014	Slovakia	Levočské vrchy	48.99570	19.58640
CL 685	2014	Slovakia	Poĺana	48.68000	19.59390
CL 686	2014	Slovakia	Levočské vrchy	48.97400	19.68080
CL 688	2014	Slovakia	Poĺana	49.03730	19.71520
CL 692	2014	Slovakia	Slovenské rudohorie	48.71830	19.72290
CL 693	2014	Slovakia	Oravská Magura	48.73860	19.74330
CL 700	2014	Slovakia	Poĺana	49.10380	19.78010

Table 1. continued

Sample ID	Year	Country	Region	Lat	Long
CL 701	2015	Slovakia	Poĺana	48.94100	19.80600
CL 702	2014	Slovakia	Veporské vrchy	48.68330	19.85000
CL 703	2014	Slovakia	Poĺana	48.68970	19.85800
CL 704	2014	Slovakia	Poĺana	48.99480	19.86020
CL 706	2015	Slovakia	Slovenské rudohorie	49.10300	19.96820
CL 708	2014	Slovakia	Nízké Tatry	48.85950	20.04720
CL 709	2014	Slovakia	Nízké Tatry	48.85950	20.04720
CL 710	2014	Slovakia	Tatry	48.74920	20.10020
CL 711	2014	Slovakia	Nízké Tatry	48.59220	20.11860
CL 712	2014	Slovakia	Nízké Tatry	48.87060	20.17450
CL 713	2015	Slovakia	Veporské vrchy	48.92090	20.18260
CL 714	2015	Slovakia	Chočské vrchy	49.09500	20.22350
CL 715	2014	Slovakia	Tatry	49.00980	20.29590
CL 716	2014	Slovakia	Tatry	48.83320	20.39010
CL 717	2014	Slovakia	Chočské vrchy	48.86670	20.48330
CL 718	2014	Slovakia	Javorníky	49.27570	20.82550
CL 719	2014	Slovakia	Javorníky	48.75810	20.50630
CL 720	2014	Slovakia	Javorníky	49.11920	20.54180
CL 721	2014	Slovakia	Strážovské vrchy	48.74790	20.54390
CL 723	2014	Slovakia	Nízké Tatry	48.84530	20.54500
CL 724	2014	Slovakia	Nízké Tatry	48.94590	20.56240
CL 725	2014	Slovakia	Nízké Tatry	49.38670	20.62390
CL 726	2014	Slovakia	Volovské vrchy	49.35350	20.62530
CL 727	2014	Slovakia	Slovenské rudohorie	48.77930	20.62560
CL 728	2014	Slovakia	Nízké Tatry	48.77930	20.62560
CL 729	2014	Slovakia	Nízké Tatry	48.87720	20.62610
CL 730	2014	Slovakia	Levočské vrchy	49.22290	20.63460
CL 731	2014	Slovakia	Slovenské rudohorie	49.02050	20.64430
CL 732	2015	Slovakia	Volovské vrchy	49.01880	20.64610
CL 733	2015	Slovakia	Volovské vrchy	49.01280	20.64680

Table 1. continued

Sample ID	Year	Country	Region	Lat	Long
CL 734	2015	Slovakia	Levočské vrchy	49.20750	20.70470
CL 735	2014	Slovakia	Ľubovnianská vrchovina	49.20760	20.70520
CL 736	2015	Slovakia	Levočské vrchy	49.21250	20.73560
CL 737	2014	Slovakia	Spišská Magura	48.72430	20.74070
CL 738	2015	Slovakia	Levočské vrchy	49.08080	20.74560
CL 739	2015	Slovakia	Levočské vrchy	49.25600	20.75900
CL 740	2015	Slovakia	Levočské vrchy	48.90930	20.76480
CL 742	2014	Slovakia	Levočské vrchy	49.04654	20.77798
CL 743	2014	Slovakia	Oravská Magura	48.72880	20.79760
CL 744	2014	Slovakia	Čergov	48.84940	20.80250
CL 745	2015	Slovakia	Čergov	49.23370	20.81580
CL 746	2015	Slovakia	Nízké Beskydy	49.07790	20.85120
CL 747	2015	Slovakia	Čergov	49.06111	20.85778
CL 748	2015	Slovakia	Čergov	49.07890	20.86860
CL 749	2015	Slovakia	Slánské vrchy	48.84420	20.93970
CL 750	2014	Slovakia	Ondavská vrchovina	48.66160	21.06590
CL 753	2014	Slovakia	Volovské vrchy	49.21900	21.07500
CL 754	2015	Slovakia	Volovské vrchy	49.13580	21.07740
CL 755	2015	Slovakia	Volovské vrchy	49.13630	21.18320
CL 756	2015	Slovakia	Nízké Beskydy	49.11260	21.20810
CL 757	2014	Slovakia	Ondavská vrchovina	49.30380	21.23680
CL 759	2015	Slovakia	Spišská Magura	48.84080	21.47840
CL 762	2015	Slovakia	Ondavská vrchovina	49.24170	21.69500
CL 763	2015	Slovakia	Nízké Tatry	49.24170	21.69500
CL 764	2016	Slovakia	Volovské vrchy	49.24170	21.69510
CL 765	2015	Slovakia	Volovské vrchy	49.08970	21.74800
CL 766	2015	Slovakia	Volovské vrchy	49.31160	21.83250
M.B.	2012	Czech Republic	Moravskoslezské Beskydy	49.14110	21.89380
VYS	2017	Czech Republic	Vysočina	49.17090	22.02480
X1		Czech Republic	Kokořínsko	49.02200	22.31140

Lat, Long – latitude and longitude coordinates in WGS 84 / Pseudo-Mercator projected coordinate system

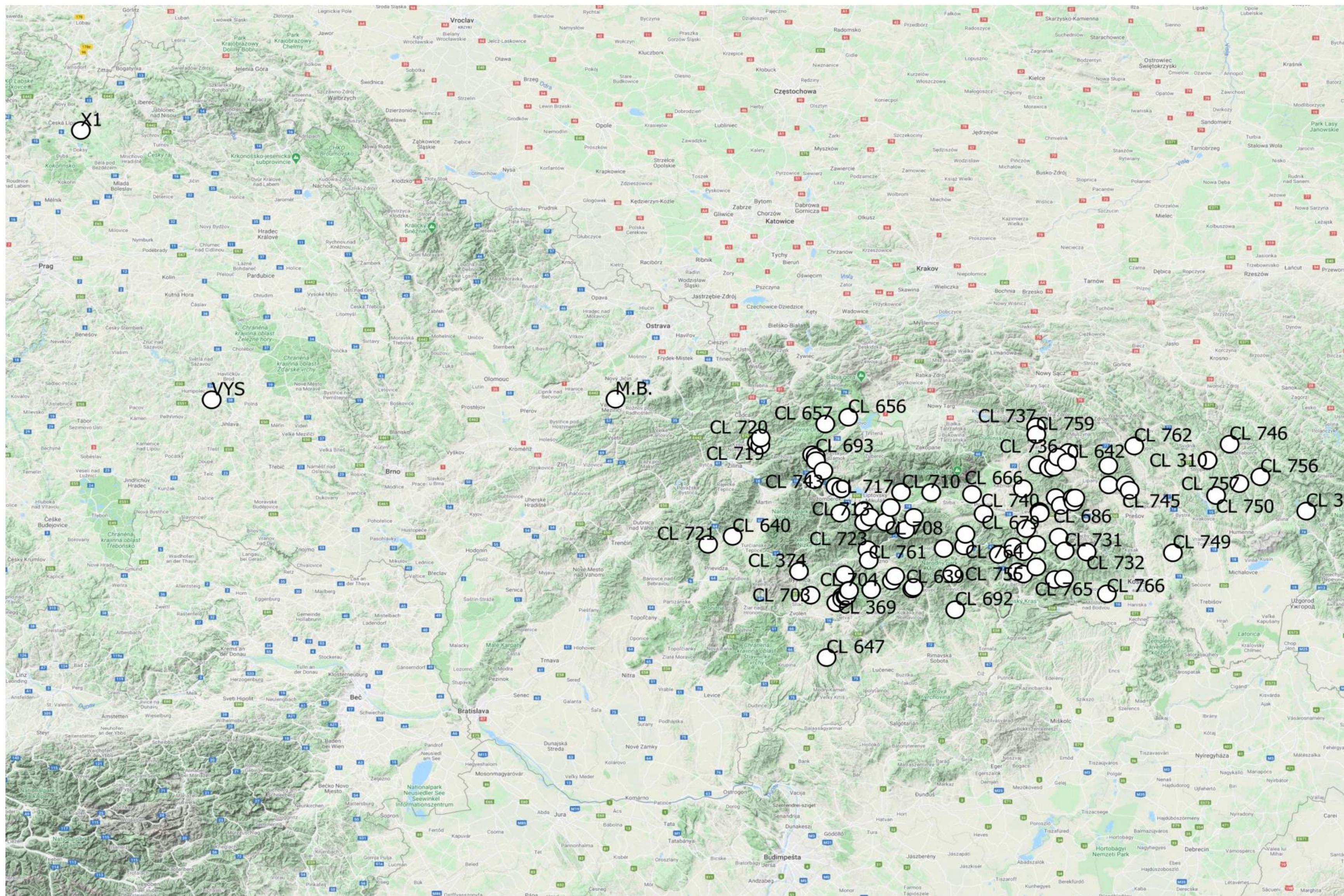

Figure 6. Map of sampling sites from the Carpathian population. Numbers represent sample ID (table 1).

To analyse the MHC diversity of European grey wolf from the eastern Dinaric-Balkan population I used 79 samples of genomic DNA from grey wolves found in different parts of Serbia (Table 2). The samples come from tissues from roadkill or were collected during legal hunts in Serbia. The local geographic distribution of Serbian samples can be seen in figure 7. The location of the population samples in Europe can be seen in figure 8.

Table 2. Samples from Serbia

Sample ID	Year	Country	Region	Lat	Long
1	2006	Serbia	Sedlare (Svijlajnac)	44.188083	21.339499
4	2003	Serbia	Đurinac-Sedlare-Svijlajnac	44.195278	21.301111
8	2004	Serbia	Pokleštica-Stara planina-Pirot	43.206328	22.740724
9	2005	Serbia	Južni Kučaj	44.075546	21.825824
13	2009	Serbia	Bela Palanka	43.21666	22.316667
18	2011	Serbia	Sjenica	43.274665	20.003886
19	2010	Serbia	Ćovdin	44.258157	21.4511
22	2011	Serbia	Županjevac	43.846588	20.937772
23	2011	Serbia	Sjenica	43.274665	20.003886
28	2010	Serbia	Južni Kučaj	44.075546	21.825824
31	2011	Serbia	Kopaonik-Raška	43.276991	20.784974
35	2010	Serbia	Stara planina-Knjaževac	43.568716	22.243023
41	2010	Serbia	Južni Kučaj	43.276991	20.784974
43	2009	Serbia	Bela Palanka	43.21666	22.316667
49	2011	Serbia	Jelica planinina-Čačak	43.820695	20.318745
51	2011	Serbia	Javor planina-Ivanjica	43.471599	20.09939
57	2009	Serbia	Bela Palanka	43.21666	22.316667
62	2007	Serbia	Raška	43.276991	20.784974
65	2010	Serbia	Bela Palanka	43.21666	22.316667
75	2010	Serbia	Negotin	44.227535	22.564748
76	2009	Serbia	Sjenica	43.274665	20.003886
78	2010	Serbia	Kuršumlija	43.131401	21.269711
84	2006	Serbia	Jagoštica-Tara-Bajna Bašta	43.975834	19.279547
85	2006	Serbia	Jagoštica-Tara-Bajna Bašta	43.975834	19.279547

Table 2. continued

Sample ID	Year	Country	Region	Lat	Long
86	2006	Serbia	Jagoštica-Tara-Bajna Bašta	43.975834	19.279547
95	2010	Serbia	Zlatibor-Čajetina	43.75611	19.725134
96	2009	Serbia	Dobroselica-Zlatibor-Čajetina	43.624827	19.700905
98	2010	Serbia	Zlatibor-Čajetina	43.75611	19.725134
100	2009	Serbia	Kopaonik-Leposavić	43.25	20.833333
101	2009	Serbia	planina Rogozna-Leposavić	43.193213	20.66222
104	2009	Serbia	Knjaževac (Stara planina)	43.568716	22.243023
150	2011	Serbia	Ćovdin	44.258157	21.4511
151	2011	Serbia	Kremna	43.848878	19.581427
158	2012	Serbia	Gornja Vrdnica-Aleksandrovac Župski	43.458567	21.046972
159	2012	Serbia	Bela Palanka	43.21666	22.316667
163	2012	Serbia	Planina Jelica-Čačak	43.820695	20.318745
166	2012	Serbia	Tutin	42.994838	20.334918
170	2012	Serbia	Južni Kučaj	44.075546	21.825824
171	2012	Serbia	Kremna	43.848878	19.581427
175	2012	Serbia	Sjenica	43.274665	20.003886
177	2012	Serbia	Leskovac-Zaječar	43.81154	22.21058
185	2013	Serbia	Sjenica	43.274665	20.003886
193	2013	Serbia	Bela Palanka	43.21666	22.316667
196	2013	Serbia	Željin-Aleksandrovac Župski	43.453587	21.049798
199	2013	Serbia	Stanišinci-planina Golija	43.335702	20.278597
204	2014	Serbia	Javorak-Južni Kučaj planina	43.915405	21.692176
205	2014	Serbia	Javorak-Južni Kučaj planina	43.915405	21.692176
206	2014	Serbia	Rudno-Kraljevo	43.405372	20.468629
208	2014	Serbia	Šljivovica-Zlatibor-Čajetina	43.813772	19.633641
210	2013	Serbia	Kremna	43.848878	19.581427
264	2016	Serbia	Rasnica-Pirot	43.177428	22.132383
265	2015	Serbia	Sjenica	43.274665	20.003886
268	2014	Serbia	Selište-Resavica	44.045445	21.594862
269	2014	Serbia	Busur-Ćovdin	44.224807	21.385648

Table 2. continued

Sample ID	Year	Country	Region	Lat	Long
271	2016	Serbia	Novi Pazar	43.137863	20.515331
275	2016	Serbia	Novi Pazar	43.137863	20.515331
279	2016	Serbia	Mali Suvodol-Pirot	43.201458	22.474841
280	2016	Serbia	Mali Suvodol-Pirot	43.201458	22.474841
281	2015	Serbia	Sjenica	43.274665	20.003886
295	2017	Serbia	Manastirište-Vlasotince	42.96	22.1526
299	2017	Serbia	Negotinac-Novi Pazar	43.1028	20.6361
315	2016	Serbia	Golija-Stup-Sjenica	43.287047	20.151988
316	2016	Serbia	Bujanovačko jezero-Bujanovac	42.430852	21.779676
329	2016	Serbia	Okruglica-Svrljig	43.3835	22.221666
333	2017	Serbia	Braničevo	44.714142	21.559062
342	2016	Serbia	Boljevac	43.824833	21.951833
343	2016	Serbia	Vodice-Zlatibor-Čajetina	43.632074	19.720484
350	2017	Serbia	Zlatibor	43.75611	19.725134
352	2017	Serbia	Despotovac	44.088889	21.441389
356	2016	Serbia	Južni Kučaj	44.075546	21.825824
378	2017	Serbia	Ponikve-Čovdin	44.225735	21.498537
387	2017	Serbia	Lovačke bare-Ćićevac	43.711108	21.495275
390	2017	Serbia	Bregovi-Zlatibor-Čajetina	43.705297	19.665458
395	2017	Serbia	Petka-Radaljica-Novi Pazar	43.255833	20.318666
404	2017	Serbia	Gradski dol-Preslap-Crna trava	42.820278	22.404722
411	2018	Serbia	Jabuka-Prijepolje	43.333302	19.506955
417	2018	Serbia	Alabana-Blace	43.293105	21.348493
420	2018	Serbia	Dragodol-Suvodanje-Osecina	44.248496	19.624883
428	2018	Serbia	Akmačići-Nova Varoš	43.424167	19.879318

Lat, Long – latitude and longitude coordinates in WGS 84 / Pseudo-Mercator projected coordinate system

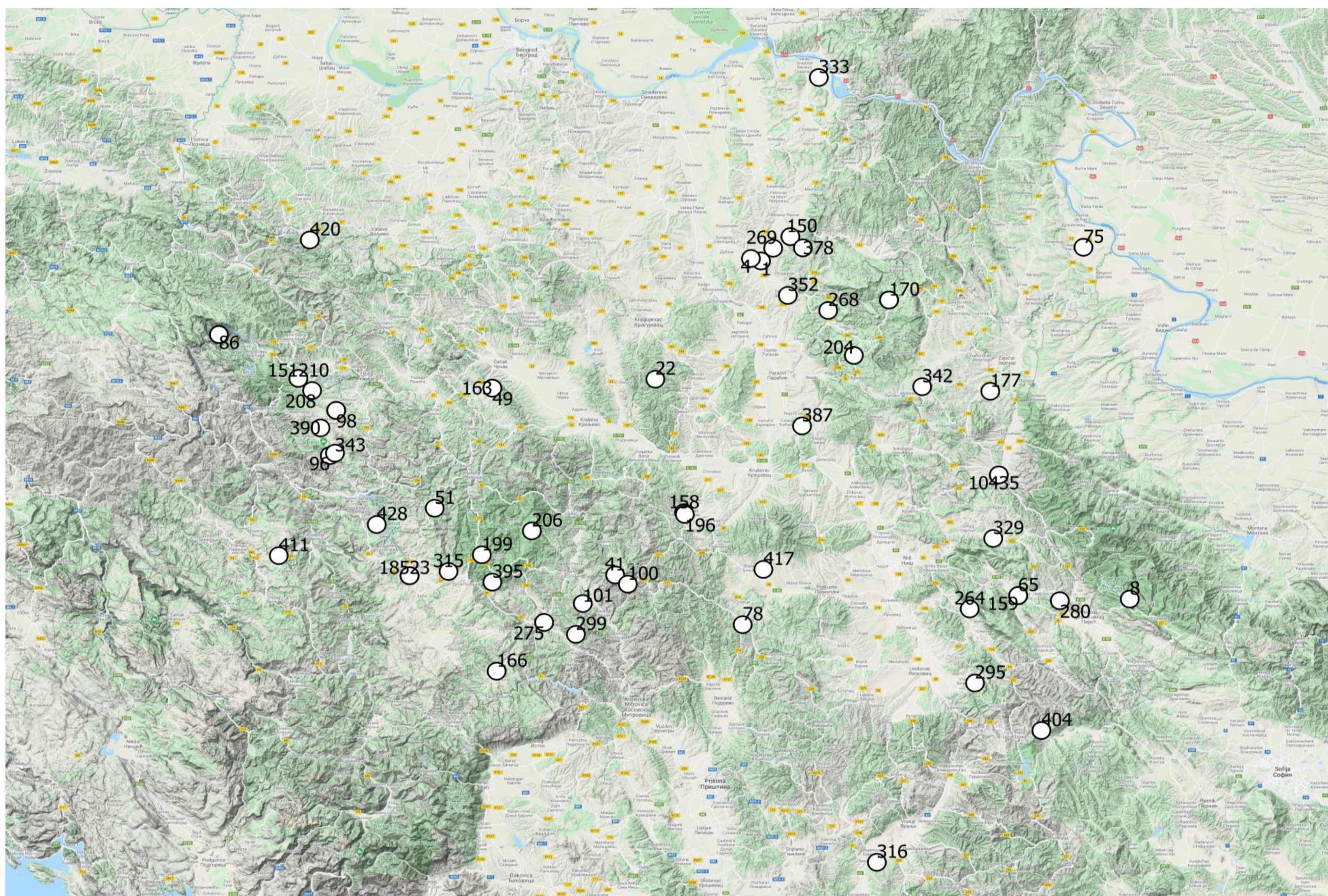

Figure 7. Map of the sampling sites from Serbia. Numbers represent sample ID (table 2).

Figure 8. Map of the sampling sites from both explored populations.
Red dots represent locations of samples from Carpathian population.
Orange dots represent locations of samples from Dinaric-Balkan population.
Diagonal lines represent the area of geographic distribution of grey wolf. The shapefile of the global distribution of grey wolf was downloaded from IUCN Red List of Threatened Species website (Boitani et al., 2018).

3.2. Polymerase chain reaction

Polymerase chain reaction (PCR) was used to selectively amplify second exon of DLA-DQA1, DQB1 and DRB1 genes. For PCR I used specific primers with HotStarTaq Master Mix Kit (Qiagen). Final volume of each reaction was 25 µl, and consisted of 12.5 µl of HotStarTaq Master Mix, 10.5 µl of nuclease free water, 0.5 µl of each primer (final concentration of 0.2 µM) and 1 µl of DNA. For the amplification of second exon of DLA-DQA1 gene I used primers: DQAin1 5'-TAA GGT TCT TTT CTC CCT CT-3' (forward) and DQAin2 5'-GGA CAG ATT CAG TGA AGA GA-3' (reverse) (Wagner, Burnett, DeRose, et al., 1996), for DLA-DQB1 gene primers: DQB1BT7 5'-CTC ACT GGC CCG GCT GTC TC-3' (forward) (Kennedy et al., 2007) and DQBR3 5'-ACC TGG GTG GGG AGC CCG-3' (reverse) (Galov et al., 2015), and for DLA-DRB1 gene primers: DRBF 5'-GAT CCC CCC GTC CCC ACA G-3' (forward) (Kennedy et al., 2005) and DRB1R 5'-TGT GTC ACA CAC CTC AGC ACC A-3' (reverse) (Wagner, Burnett, Works, et al., 1996). All of the primers were designed to align in the intronic regions of their specific loci. I used the temperature protocol described by Kennedy *et al.* (2007): polymerase activation at 95°C for 15 min, then 14 touchdown cycles with denaturation at 95°C for 30 s, annealing for 1 min starting at 54°C for DLA-DQA1, 73°C for DLA-DQB1 or 62°C for DLA-DRB1 decreasing by 0.5°C each cycle and extension at 72°C for 1 min. Touchdown cycles were followed by 20 cycles of denaturation at 95°C for 30 s, annealing at 47°C (DQA1), 66°C (DQB1) or 55°C (DRB1) for 1 min and extension 72°C for 1 min, with the final extension at 72°C for 10 min.

3.3. Electrophoresis

To check for the presence of amplification products after the PCR reaction, after the purification of PCR products, and after molecular cloning (chapter 2.5) I used gel electrophoresis on 1% agarose gels. In one of the lines I used DNA size marker to check the approximate size of the PCR products. To visualize the DNA fragments I used Midori Green Advance colour (Nippon Genetics) in ratio 0.5 µl of colour on 10 ml of gel. Electrophoresis was done at 100V for 30 minutes, and after, the gels were observed under UV light.

3.4. Sequencing

All of the products obtained by specific DLA-DQA, DQB and DRB PCR (chapter 3.2) that were of the expected length, and all of the plasmids obtained by molecular cloning (chapter 3.5) were sequenced using Macrogen sequencing service (Macrogen Europe). All of the products were sequenced using the Standard-seq service and the primers used for sequencing were the same as the ones used for PCR. Samples were sequenced first only from one direction, forward for DLA-DQA1 and reverse for DLA-DQB1 and DLA-DRB1, and only in the cases in which the sequences weren't completely clear using one direction, I ordered the sequencing from the other direction to obtain unambiguous sequences. Some examples of those are sequences with heterozygous sites near the primer binding region, or sequences with multiple heterozygous sites in a row. Before sending the PCR products for sequencing I performed enzymatic purification using ExoSAP-IT PCR Product Cleanup Reagent (Applied Biosystems) to remove residual nucleotides, primers and unfinished PCR products. Purification of plasmids is described in chapter 3.5.

3.5. Molecular cloning

For heterozygous wolves whose exact alleles couldn't be determined using in-silico methods, and for those with presumed new alleles I performed molecular cloning using pGEM®-T Vector System (Promega) with JM109 Competent Cells (Promega). pGEM®-T is a plasmid which carries a gene conferring ampicillin resistance to the host cells, has a multiple cloning region and a lactose operon, and it comes already linearized in the *lacZ* gene with a thymine added at both 3' ends to ensure easier insertion of DNA fragments into the plasmid. These characteristics are vital for the selection of those bacteria that have been transformed with the plasmid which carries the insert of interest, in this case the amplified product of the exon 2 of a DLA allele.

The molecular cloning procedure starts first with the amplification of the allele using PCR described in chapter 3.2. After the PCR, I performed the purification of PCR products with the Wizard® SV Gel and PCR Clean-Up System (Promega) using the following procedure. In the finished PCR reaction I added an equal volume (25 µl) of Membrane Binding Solution. Then I put the SV Minicolums in clean Collection Tubes and added the solutions with the

amplified DNA to the SV Minicolumns. After 1 min incubation, I centrifuged the tubes at 16,000 g for 1 minute, discarded the flow-through and reinserted the SV Minicolumns into the same Collection Tubes. I added 700 µl of Wash Solution to the SV Minicolumns, then centrifuged at 16,000 g for 1 min. I discarded the flow-through, added 500 µl of Wash Solution and centrifuged at 16,000 g for 5 min. Again, I discarded the flow-through, and centrifuged just the column assembly at 16,000 g for 2 min to evaporate any residual fluid. I transferred the SV Minicolumns to new, sterile 1.5 ml tubes and added 20 µl of Nuclease Free Water. After 1 min incubation I centrifuged the tubes at 16,000 g for 1 minute. The flow-through contained the purified products of my PCR reactions, which was checked using electrophoresis described in chapter 3.3.

After obtaining the purified PCR amplicons of my alleles of interest (inserts), I performed the ligation reaction. The pGEM®-T Vector System is declared to work successfully in a range from 8 : 1 to 1 : 8 insert : vector molar ratios. So the first step in setting up the ligation reaction was to measure the concentrations of my purified inserts, which I did using DS-11 Series Fluorometer (DeNovix) with DeNovix dsDNA Broad Range Fluorescent Assay Kit. To determine the volume of insert to be added to the ligation reaction one can use the general equation given in the manual:

$$mass\ of\ insert\ (ng) = \frac{mass\ of\ vector\ (ng) * length\ of\ insert\ (kb)}{length\ of\ vector\ (kb)} * insert{:}\ vector\ molar\ ratio.$$

In my, and my colleagues previous work we found that the reaction works best for us when using 3 : 1 insert : vector molar ration, and since the size of insert is approximately 0.3 kb, size of vector 3 kb and concentration of vector is 50 ng/µl I used the following equation:

$$mass\ of\ insert\ (ng) = \frac{50\ ng * 0.3kb}{3kb} * 3{:}\,1.$$

This way I calculated that the recommended mass of my inserts is around 15 ng, so using the concentrations measured with the fluorometer it was easy to calculate the required volume using the equation: $volume\ (\mu l) = \frac{mass\ (ng)}{concentration\ (ng/\mu l)}$

After determining the required volume of insert I set up the ligation reaction. In a 0.2 mL tube I mixed 5 µl of 2X Rapid Ligation Buffer, 1 µl of pGEM®-T plasmid (50 ng), 1 µl of T4 DNA Ligase (3 Weiss units/µl), required volume of purified insert and nuclease free water

to a final volume of 10 µl. I also prepared two control reactions, one negative control which had the same reaction but without the insert, and control of transformation which had 2 µl of control insert instead of my insert of interest. The control insert is a 542 bp DNA which contains multiple stop codons in all reading frames so it should ensure a low background of blue colonies. All ligation reactions were left overnight at 4°C.

On the next day I performed the transformation of JM109 Competent Cells with the ligation reactions. For each ligation reaction I used two agar plates with 30 mL of LB media. The growth media was prepared according to the following composition: 1 L H_20, 10 g Bacto Tryptone, 5 g Bacto Yeast Extract, 5 g NaCl and 15 g agar, titrated to 7.0 pH using NaOH. After mixing, the media was autoclaved, Ampicillin was added to a final concentration of 100 µl/ml, and then the media was poured into Petri dishes. After the media cooled on room temperature I added 100 µl of 100 mM IPTG and 20 µl of X-Gal (50 mg/ml) on top, and left the plates for 30 minutes on 37°C to absorb the solutions. Ligation reactions from yesterday were briefly centrifuged and transferred to new, sterile 1.5 mL tubes. Tubes containing JM109 Competent Cells, were taken out of the freezer and put in an ice bath for 5 min to melt. After light mixing of the cells I pipetted 50 µl of cells on top of each ligation reaction. The tubes were lightly mixed, then put on ice for 20 min. To ease the transformation the cells were heat shocked by putting the tubes in a warm bath at 42°C for 45 s, and then back on ice for 2 min. Then I added 800 µl of liquid LB media to the tubes, and incubated them in a thermomixer at 37°C for 90 min with a mixing speed of 300 rpm. The liquid LB media was prepared in a same way as the one for the plates, only without adding agar to the mix. After the incubation, I pipetted 100 µl from each reaction on two separate LB/Ampicillin/IPTG/X-Gal plates. All of the plates were then incubated at 37°C for 24 h.

On the following day for each cloned sample I prepared and marked ten 15 mL tubes and added 3 mL of liquid LB media to each tube, prepared with the same recipe as above. After the plate incubation finished, to select and multiply the transformed cells I transferred bacterial colonies from the plates to the tubes, one colony per tube, selecting only white colonies. All of the tubes were then incubated overnight on 37°C with light shaking.

The logic behind selecting only white colonies is as follows. If the bacteria grows on the plates with ampicillin it means that it was transformed with our plasmid, which confers resistance to ampicillin. But not all of the plasmid have our insert of interest in them. So, if the insert didn't combine with the plasmid the plasmid will close with the working lactose operon.

Since IPTG, which I added to the plates serves as a lactose operon inducer, *lacZ* gene will be transcribed, and the bacteria will produce β-galactosidase. X-Gal, which I also added to the plates is an analogue of lactose, and serves as one of the substrates of β-galactosidase. When cleaved by β-galactosidase, X-Gal, which is colourless, produces galactose and 5-bromo-4-chloro-3-hydroxyindole. 5-bromo-4-chloro-3-hydroxyindole than spontaneously dimerizes and is oxidised to produce 5,5'-dibromo-4,4'-dichloro-indigo, which is blue, giving the bacterial colonies blue colour (Sandhu, 2010).

After the overnight incubation of bacterial cultures in liquid LB medium, I proceeded with extraction and purification of plasmid DNA from the bacterial cells using PureYield Plasmid Miniprep System (Promega). To a 2 mL micro centrifuge tube I transferred 1.5 mL of overnight culture and pelleted the bacteria using the micro centrifuge for 30 s at 16,000 g. I discarded the supernatant, added the remaining 1.5 mL of cultured cells to the tube with the pellet, and again centrifuged the tube for 30 s at 16,000 g. After discarding the supernatant I resuspended the bacterial pellet in 600 μl of TE buffer. To the resuspension I added 100 μl of Cell Lysis Buffer and mixed by inverting the tube several times. Once the cell solution started clearing, indicating complete lysis, I added 350 μl of chilled Neutralization Solution and mixed by inverting the tube several times. Once the solution turned yellow, indicating that the neutralization is nearly done, I inverted the tube a few times to ensure complete neutralisation, and then centrifuged the tube at 16000 g for 3 min. I pipetted the supernatant to a PureYield Minicolumn, threw away the remaining pellet, put the minicolumn in the PureYield Collection Tube and centrifuged the minicolumn assembly at 16,000 g for 15 s. After discarding the flow-through I added 200 μl of Endotoxin Removal Wash solution to the minicolumn and centrifuged the assembly at 16,000 g for 15 s. Then I added 400 μl of Column Wash Solution to the minicolumn and centrifuged the assembly at 16,000 g for 30 s, after which I transferred the minicolumn to a clean 1.5 mL tube, added 30 μl of Elution Buffer and incubated everything for 1 min at room temperature. After 15 s of centrifugation at 16,000 g I obtained purified plasmid DNA in the eluate. Before sending the samples for sequencing, the presence of plasmid DNA was checked using electrophoresis as described in chapter 3.3.

While performing laboratory research for this thesis I cloned in total 16 samples. From the Carpathian population samples CL364 and CL656 were cloned to obtain DLA-DQA1 alleles, samples CL738 and CL764 for the DLA-DQB1 alleles and sample X1 was cloned for the DLA-DRB1 alleles. From the eastern Dinarid population samples: 4, 166, 175, 199 and 208

were cloned to obtain DLA-DQA1 alleles, samples 19 and 76 for the DLA-DQB1 alleles and samples: 49, 75, 265 and 277 were cloned for the DLA-DRB1 alleles.

3.6. BioEdit

BioEdit (Hall, 1999) is a programme package which contains many options for analysing and editing of nucleotide and amino acid sequences. Each of the chromatograms obtained by sequencing of samples processed in this research was first analysed using BioEdit. It was used to check the overall quality of obtained reads, to check if the samples were homo- or heterozygotes on explored loci, and for editing the sequences which included deleting the low quality data on each end of the sequences, deleting the parts of the sequences not in alignment with the referent allele (for example, plasmid sequence) and obtaining reverse complement of sequences for aligning reads obtained from sequencing from opposite directions. Furthermore, BioEdit contains ClustalW Multiple alignment plugin which was used to make all of the alignments required for further processing using SeqScape (chapter 3.7.) and MEGA (chapter 3.8.) software packages. These alignments include: alignments of sequenced alleles with the reference allele, alignments of sequenced plasmids with the reference allele, alignments of sequences from the same sample obtained by cloning to determine allele composition, alignments of allele libraries and alignments of alleles used for phylogenetic reconstruction. The parameters used for the alignment were: default gap open penalty, default gap extend penalty, and do full multiple alignment.

3.7. SeqScape

The SeqScape software (Applied Biosystems) is a resequencing package designed for mutation detection and analysis, SNP discovery and validation, pathogen sub-typing, allele identification, and sequence confirmation. In this research SeqScape software was used for library searching, meaning it was used to compare each new sequence to a library of known allele sequences to identify the closest allele or combination of alleles in each sample. This way I avoided using molecular cloning for each heterozygous sample. All usable sequences from heterozygous wolves were processed using SeqScape software. At first I used allele libraries made of all available confirmed DLA-DQA1, DLA-DQB1 and DLA-DRB1 alleles, one library

per locus. Libraries were created based on the data found in the IPD-MHC database (Maccari et al., 2017). For DLA-DQA1 the library contained 32 sequences of known alleles, for DLA-DQB1 154 sequences, and for DLA-DRB1, 161 sequences of known alleles. Once I encountered samples which couldn't be resolved using SeqScape software, either because they contained new alleles, or the result had multiple possible allele combinations, I performed molecular cloning on these samples (chapter 3.5.). This process (sequencing – in-silico analysis – cloning) was iterative, meaning I cloned just a few samples, usually one or two, out of a group of similar heterozygotes. If the molecular cloning provided any new alleles, they were added to the respective DLA allele library I was currently using, and all of the samples were reanalysed using the new library. If there were still any unresolved samples I cloned more of the unresolved ones. Also, once I achieved complete resolution of known alleles in a few samples with multiple possible combinations, I presumed the same alleles will be found in other samples with same combinations.

The parameters for the analysis using SeqScape software were: use standard codon table, use KB.bcp basecaller, use true profile, do not assign N's to basecalls, use mixed base identification, call mixed base if second highest peak is equal or greater than 25% of the first highest, remove bases from the ends until fewer than four bases out of 20 have quality value lower than 20, use reference trimming, maximum number of mixed bases is 20%, minimum clear length is 50 bp and minimum sample score is 20. For the reference alleles I used: DLA-DQA1*00101, DLA-DQB1*00101 and DLA-DRB1*00101, each one on their respective locus.

3.8. MEGA 7

Programme package MEGA 7 (Eng. Molecular Evolutionary Genetics Analysis) (Kumar et al., 2016) contains necessary tools for analysing and comparing amino acid and nucleotide sequences for the inference of patterns of molecular evolution of certain species or genetic loci. The software contains programmes for creating alignments of DNA and protein sequences, for phylogenetic analysis of molecular data and for the calculation of parameters of molecular evolution. MEGA 7.0.21 was used in this research to calculate:

1. best models of nucleotide and amino acid substitution,
2. nucleotide and amino acid evolutionary distances,

3. total number of variable positions in all sequenced DLA alleles,
4. rates of synonymous and nonsynonymous nucleotide substitutions,
5. Z-test of selection,
6. Tajima's test of neutrality,
7. maximum likelihood phylogenetic tree,
8. neighbour joining phylogenetic tree, and
9. maximum parsimony phylogenetic tree

Calculations were performed on each locus independently. Best models were calculated using the Model Selection analysis with the following parameters: use automatic (Neighbor-joining) tree, Maximum Likelihood statistical method, use all sites, use all codon positions and very strong branch swap filter.

Evolutionary distances were calculated using Overall Mean Distance Estimation analysis selecting the best calculated model and same pattern among lineages. Total number of variable positions in sequences was observed using Mark variable sites option in MEGA Sequence Data Explorer.

Rates of synonymous and nonsynonymous nucleotide substitutions on individual codons were calculated using Estimate Selection at Codons via HyPhy analysis (Pond et al., 2005) with Felsenstein 1981 model (Felsenstein, 1981). It is a maximum likelihood analysis of natural selection that estimates the numbers of inferred synonymous and nonsynonymous substitutions for each codon in a DNA sequence. Average rates of synonymous nucleotide substitutions (rates over the entire sequence) were calculated as the average of individual synonymous nucleotide substitution rates (per codon rates) by counting method. In the same way I calculated the average rate of nonsynonymous nucleotide substitution rates.

For codon based Z-test of selection, I used Overall Average scope and modified Nei-Gojobori method to allow for difference in transversional and transitional substitution frequencies (Zhang et al., 1998) with Jukes-Cantor correction to account for multiple substitutions at the same site (Nei & Kumar, 2000). Test hypothesis (or null hypothesis) was neutrality (same rates of synonymous and nonsynonymous substitutions). It was tested separately against three alternative hypotheses:

1. non-neutrality (different rates of synonymous and nonsynonymous substitutions),

2. positive selection (rate of nonsynonymous substitutions larger than rate of synonymous substitutions) and
3. purifying selection (rate of nonsynonymous substitutions smaller than rate of synonymous substitutions).

For each test I applied a bootstrap variance estimation method using 10,000 bootstrap replications.

Tajima's test of neutrality (Tajima, 1989) compares the number of segregating sites per nucleotide site in a sequence alignment with the nucleotide diversity of the same sequence alignment. As stated in chapter one Tajima's D test statistic can give us an indication of non-neutral evolution. For the calculation I used nucleotide substitutions option through all sites. Confidence limit of Tajima's D test statistic was taken from (Tajima, 1989).

For phylogenetic reconstruction I used all of the alleles obtained in this research, and alleles found in previous publications exploring the MHC diversity of wolves in Europe (Arbanasić et al., 2013; Galaverni et al., 2013; Niskanen et al., 2014; Seddon & Ellegren, 2004). Maximum likelihood trees were calculated using the following parameters: nucleotide substitutions type, with the best model as calculated by Model Selection analysis (described above), subtree-pruning-regrafting extensive heuristic method and a very strong branch swap filter. When required by the best model, number of discrete gamma categories was set to five. For calculations of Neighbor-joining trees the parameters were set in the equivalent way. When required by the best model, gamma parameter was calculated using the Estimate Rate Variation among Sites analysis. For the maximum parsimony statistical method on the DQA1 and DQB1 loci, the tree inference search method used was max-mini branch-&-bound, while on the DRB1 locus, because of the computational constraints, I used subtree-pruning-regrafting with 1000 initial trees and search level 5. All of the trees were tested using the bootstrap method with 1000 bootstrap replications.

3.9. iTOL

Final editing of obtained phylogenetic trees was done in Interactive Tree Of Life software (iTOL) version 5.6.3 (Letunic & Bork, 2019). iTOL is an online tool for the display, annotation and editing of phylogenetic trees. All of the changes made in iTOL were purely of cosmetic nature, to obtain images of better quality, and no structural changes were made.

3.10. Arlequin 3

Arlequin (Excoffier et al., 2007) is an integrated software package for population genetics data analysis. This software provides users with statistical tests which can be used to explore genetic features of a collection of population samples. Test in Arlequin have been implemented in the form of permutation tests or exact tests in such a way as to minimize hidden assumptions and to be as powerful as possible. In this work I used the following tests in Arlequin version 3.5.2.2:

1. Exact test of Hardy-Weinberg equilibrium (Guo & Thompson, 1992)
2. Pairwise linkage disequilibrium test (Slatkin & Excoffier, 1996)

For exact test of Hardy-Weinberg equilibrium the null hypothesis is that the observed diploid genotypes are the product of a random union of gametes. The test is an analogous test to Fisher's exact test but extended to a contingency table of arbitrary size. The number of steps in Markov chain (the maximum number of alternative tables to explore) was set to 5,000,000, and the number of dememorization steps (the number of steps to perform before beginning to compare the alternative table probabilities to that of the observed table) to 500,000.

Pairwise linkage disequilibrium was tested using a likelihood-ratio test. The likelihood of the data is computed by using the fact that assuming linkage equilibrium, the haplotype frequencies are obtained as the product of the allele frequencies. The likelihood of the data not assuming linkage equilibrium is obtained using the Expectation-Maximization algorithm to estimate haplotype frequencies. This procedure is an iterative process aiming at obtaining maximum-likelihood estimates of haplotype frequencies from multi-locus genotype data (Dempster et al., 1977). Under the null hypothesis (linkage equilibrium) the following permutation procedure is used to approximate the distribution of the likelihood-ratio statistic:

1. Permute the alleles between individuals at one locus
2. Re-estimate the likelihood of the data by the Expectation-Maximization algorithm. The likelihood of the data assuming linkage equilibrium is unaffected by this permutation.
3. Repeat steps one and two a large number of times to get the null distribution of likelihood of the data not assuming linkage equilibrium.

For the calculation I have set the number of random permuted samples to generate to 16,000 and the number of random initial conditions from which the Expectation-Maximization algorithm is started to repeatedly estimate the sample likelihood to five.

3.11. EasyCodeML

EasyCodeML (Gao et al., 2019) is a software for detecting selection in molecular evolutionary analysis. It implements the major codon-based models of CodeML in a visual and user-friendly interface. For input files it uses a sequence alignment file in PAML format and a tree file in plain Newick format. For my calculations I used version 1.21 (latest version). For the Newick format input file, tree was made in MEGA 7.0.21 (Kumar et al., 2016) using maximum likelihood method in the same way as described above. I used the preset running mode, which has all the key parameters of the nested models built-in, and the site model setup with seven codon substitution models: M0, M1a, M2a, M3, M7, M8 and M8a. Fit of these models to the sequence data was compared using likelihood-ratio tests. Additionally, to infer exact codons that are under positive selection I used Bayes Empirical Bayes analysis (Yang et al., 2005).

3.12. Inference of haplotypes

Once I determined exact allele composition of each loci in all samples, I inferred three-locus DLA-DQA1/DQB1/DRB1 haplotypes using a method described by Kennedy *et al.* (2002). First I established which wolves are homozygous at all three DLA loci, and identified haplotypes from those wolves. Then I checked if any of the wolves homozygous at two DLA loci had any of the haplotypes already identified from the first group. If they did, I inferred the other haplotype. The same procedure I then repeated for wolves homozygous at only one DLA locus, and finally in wolves heterozygous at all three DLA loci. So, sequentially more and more haplotypes were established, until I was able to identify almost all haplotypes in studied populations despite the absence of family data.

3.13. Mapping

All of the maps created for this thesis were done in QGIS 3.4.12 Geographic Information System software (2018). The coordinates of the samples were taken at the time of sampling in WGS 84 / Pseudo-Mercator coordinate system. For the local maps I used the Google Terrain map, and for the larger map I used ESRI Physical map with ESRI Reference Overlay. The shapefile of the global distribution of grey wolf was downloaded from IUCN Red List of Threatened Species website (Boitani et al., 2018).

4. Results

Nucleotide sequence lengths obtained in this study were 246 bp for DQA1 locus, 267 bp for DQB1 locus and 270 bp for DRB1 locus. None of the detected sequences contained any stop codons.

4.1. Carpathian population alleles and haplotypes

In the sample of 99 grey wolves from the western Carpathian population I found a total of four DLA-DQA1 alleles, seven DLA-DQB1 alleles and seven DLA-DRB1 alleles. Their nucleotide sequences are presented in the appendix (tables 61, 62 and 63, respectivelly), and are transcribed into the identical number of amino acid sequences, respectively (tables 3, 4 and 5). All of the detected alleles were known from previous research and could be found in the IPD-MHC database. Still, alleles DQB1*03502, DQB1*04901 and DRB1*04902 were found for the first time in any of the European grey wolf populations. Note that in the presented DLA-DRB1 allele sequences in the appendix (table 63) codon start is on the third position.

Table 3. Amino acid sequences of DLA-DQA1 alleles found in grey wolves from the Carpathian population (dots represent the identical nucleotide as in the top sequence).

Allele name	Allele amino acid sequence	Position
DQA1*005011	D H V A Y Y G I N V Y Q S Y G P S G Q F T H E F D G D E E F	[30]
DQA1*00601	. . . . Y	[30]
DQA1*012011	. . . . Y	[30]
DQA1*00301	. . . . Y	[30]
DQA1*005011	Y V D L E K K E T V W R L P V F S T F T S F D P Q G A L R N	[60]
DQA1*00601	. . . . R	[60]
DQA1*012011	. . . . A	[60]
DQA1*00301	. . . .	[60]
DQA1*005011	L A I T K Q N L N I M T K R S N K T A A T N	[82]
DQA1*00601	. . . I . . . . . . L . . . . . Q	[82]
DQA1*012011	. . . A . . . . Q	[82]
DQA1*00301	. . R A . . . . . . L . . S . . Q	[82]

Table 4. Amino acid sequences of DLA-DQB1 alleles found in grey wolves from the Carpathian population (dots represent the identical nucleotide as in the top sequence).

Allele name	Allele amino acid sequence	Position
DQB1*00401	D F V F Q F K G E C Y F T N G T E R V R L L T K Y I Y N R E	[30]
DQB1*02002	. . . Y . . . A . . . . R S	[30]
DQB1*03501	. . . Y . . . F . . . . A R D	[30]
DQB1*03502	. . . Y . . . F . . . . A R D	[30]
DQB1*03901	. . . . F . . . . A	[30]
DQB1*04101	. . . .	[30]
DQB1*04901	. . . Y . . . A . . . . R S	[30]
DQB1*00401	E Y V R F D S D V G E Y R A V T E L G R P W A E Y W N P Q K	[60]
DQB1*02002	. F . . . . F . . . . V . . . . . G . .	[60]
DQB1*03501	. H . . . . D . . . . . G . .	[60]
DQB1*03502	. H . . . . D . . . . . G . .	[60]
DQB1*03901	. F	[60]
DQB1*04101	. F	[60]
DQB1*04901	. F . . . . F . . . . V . . . . . G . .	[60]
DQB1*00401	D E M D R V R A E L D T V C R H N Y G L E E L Y T L Q R R	[89]
DQB1*02002	E I L E . K . . A V . R . . . . R . . . T	[89]
DQB1*03501	E L L E Q R	[89]
DQB1*03502	E L L E Q R . . . . V	[89]
DQB1*03901	. . . . V	[89]
DQB1*04101	. . . .	[89]
DQB1*04901	E I L E . K . . . V . R . . . . R . . . T	[89]

Table 5. Amino acid sequences of DLA-DRB1 alleles found in grey wolves from the Carpathian population (dots represent the identical nucleotide as in the top sequence).

Allele name	Allele amino acid sequence	Position
DRB1*03601	H F L E M L K S E C H F T N G T E R V R F V E R Y I H N R E	[30]
DRB1*04901	. . . K . V . F . . . . L . . . D . Y . . .	[30]
DRB1*04902	. . . K . V . F . . . . L . . . D . Y . . .	[30]
DRB1*05401	. . . . V A . . . . Y	[30]
DRB1*092011	. . . . F	[30]
DRB1*09901	. . . . V A . . . . Y	[30]
DRB1*12801	. . . K . V . F . . . . L . . . D . Y . . .	[30]
DRB1*03601	E N V R F D S D V G E Y R A V T E L G R R D A E S W N R Q K	[60]
DRB1*04901	. Y . . . . F . . . . P S . . Y . . G . .	[60]
DRB1*04902	. Y . . . . P S . . Y . . G . .	[60]
DRB1*05401	. . . . P . . . . . . G . .	[60]
DRB1*092011	. F . . . . P	[60]
DRB1*09901	. . . .	[60]
DRB1*12801	. Y . . . . F . . . . P S . . Y . . G . .	[60]
DRB1*03601	E L L E R K R A E V D T Y C R H N Y G V I E S F T V Q R R	[89]
DRB1*04901	. . . . Q R . . . . . . V . . . . A	[89]
DRB1*04902	. . . . Q R . . . . . . V . . . . A	[89]
DRB1*05401	. . . . Q R . . A	[89]
DRB1*092011	. F . . Q R . . . . G	[89]
DRB1*09901	. . . .	[89]
DRB1*12801	. . . . Q R . . . . . . V . . . . G	[89]

The frequency of individual alleles and the number of homozygotes in Carpathian population are given in tables 6, 7 and 8 for DQA, DQB and DRB loci, respectively. On DQA1 locus the total number of detected alleles is four. The most frequent allele at DQA1 locus is DQA1*00301 (41.4%), while the least frequent allele is DQA1*00601 (7.6%). The proportion of homozygous wolves for this locus is 37.4%.

Table 6. Frequency of DLA-DQA1 alleles in the sample from Carpathian population (N=99)

Allele name	#	f	H (h)
DQA1*00301	82	0.414	0.182 (18)
DQA1*005011	30	0.152	0.061 (6)
DQA1*00601	15	0.076	0 (0)
DQA1*012011	71	0.359	0.131 (13)
total	198	1	0.374 (37)

- absolute number of a given allele found in the sample; f – frequency of a given allele in the sample; H – homozygosity of a given allele in the sample (h/N); h – number of homozygotes of a given allele in the sample

DQB1 locus is more polymorphic as 7 alleles are found and the most frequent allele is DQB1*03501 (32.3%), followed by DQB1*00401 (29.3%), while the least frequent allele is DQB1*04901 (0.5%) found only once in a heterozygous wolf. Two other alleles on this locus, DQB1*02002 and DQB1*03502, have no homozygotes and their frequencies are below 10%. The proportion of homozygotes is lower on this locus with 30.3%

Table 7. Frequency of DLA-DQB1 alleles in the sample from Carpathian population (N=99)

Allele name	#	f	H (h)
DQB1*00401	58	0.293	0.081 (8)
DQB1*02002	14	0.071	0 (0)
DQB1*03501	64	0.323	0.131 (13)
DQB1*03502	7	0.035	0 (0)
DQB1*03901	30	0.152	0.061 (6)
DQB1*04101	24	0.121	0.030 (3)

Table 7. continued

Allele name	#	f	H (h)
DQB1*04901	1	0.005	0 (0)
total	198	1	0.303 (30)

- absolute number of a given allele found in the sample; f – frequency of a given allele in the sample; H – homozygosity of a given allele in the sample (h/N); h – number of homozygotes of a given allele in the sample

The same number of alleles (7) is found in Carpathian population at DRB1 locus, with two most frequent alleles being DRB1*05401 and DRB1*03601, found with frequency of 39.9% and 35.9%, respectively. Again, the least frequent allele DRB1*12801 is found only once (0.5%), and again, two more alleles are found only in heterozygous animals (DRB1*092011 and DRB1*09901). The proportion of homozygotes is 34.3%, which is higher than on the DQB1 locus, but lower then on the DQA1 locus.

Table 8. Frequency of DLA-DRB1 alleles in the sample from Carpathian population (N=99)

Allele name	#	f	H (h)
DRB1*03601	71	0.359	0.131 (13)
DRB1*04901	6	0.030	0.010 (1)
DRB1*04902	23	0.116	0.040 (4)
DRB1*05401	79	0.399	0.162 (16)
DRB1*092011	15	0.076	0 (0)
DRB1*09901	3	0.015	0 (0)
DRB1*12801	1	0.005	0 (0)
total	198	1	0.343 (34)

- absolute number of a given allele found in the sample; f – frequency of a given allele in the sample; H – homozygosity of a given allele in the sample (h/N); h – number of homozygotes of a given allele in the sample

From the Carpathian population sample, I have inferred 10 distinct haplotypes, with 72 heterozygous wolves. The frequency of each detected haplotype is presented in table 9. The most frequent three-locus haplotype found in the western Carpathian population is haplotype DQA1*012011 / DQB1*03501 / DRB1*03601 found with frequency of 32.3% and, unsurprisingly, in the largest number of homozygous individuals (13). The next frequent haplotype DQA1*00301 / DQB1*00401 / DRB1*05401 is found in only 6 homozygotes, although it is found in quite high frequency (27.8%). The least frequent haplotypes are DQA1*005011 / DQB1*03901 / DRB1*12801 and DQA1*00601 / DQB1*04901 / DRB1*092011 found only once each (0.5%). The total proportion of homozygous wolves is 27.3%. Interestingly, haplotype DQA1*005011 / DQB1*03901 / DRB1*04901 is found in a homozygous animal, even though its population frequency is only 3.0%. Of note is also the haplotype DQA1*005011 / DQB1*03901 / DRB1*12801 found only in wolf X1 outside of the Carpathian mountain range, and could belong to the population of European lowlands. Haplotype distribution in individual wolves is presented in table 64 in the appendix.

Table 9. Individual haplotype frequency in the sample from the Carpathian population (N=99)

Haplotype			#	f	H (h)
DQA1*012011	DQB1*03501	DRB1*03601	64	0.323	0.131 (13)
DQA1*00301	DQB1*00401	DRB1*05401	55	0.278	0.061 (6)
DQA1*00301	DQB1*04101	DRB1*05401	24	0.121	0.030 (3)
DQA1*005011	DQB1*03901	DRB1*04902	23	0.116	0.040 (4)
DQA1*00601	DQB1*02002	DRB1*092011	14	0.071	0 (0)
DQA1*012011	DQB1*03502	DRB1*03601	7	0.035	0 (0)
DQA1*005011	DQB1*03901	DRB1*04901	6	0.030	0.010 (1)
DQA1*00301	DQB1*00401	DRB1*09901	3	0.015	0 (0)
DQA1*005011	DQB1*03901	DRB1*12801	1	0.005	0 (0)
DQA1*00601	DQB1*04901	DRB1*092011	1	0.005	0 (0)
total			198	1	0.273 (27)

- absolute number of a given haplotype found in the sample; f – frequency of a given haplotype in the sample; H – homozygosity of a given haplotype in the sample (h/N); h – number of homozygotes of a given haplotype in the sample

4.2. Dinaric-Balkan population alleles and haplotypes

In the sample of 79 grey wolves from Serbia I found a total of 8 DLA-DQA alleles, 11 DLA-DQB alleles and 12 DLA-DRB alleles. Their nucleotide sequences are presented in the appendix (tables 65, 66 and 67, respectivelly). On DLA-DQA1 and DLA-DQB1 loci, the nucleotide sequences are transcribed into the identical number of amino acid sequences (tables 10, 11), while on DLA-DRB1 locus alleles DRB1*092011 and DRB1*092012 code for the same amino acid chain, so there are 11 different amino acid sequences in total (table 12). Note that in the presented DLA-DRB1 allele sequences in the appendix (table 67) codon start is on the third position.

Table 10. Amino acid sequences of DLA-DQA1 alleles found in grey wolves from Serbia (dots represent the identical nucleotide as in the top sequence).

Allele name	Allele amino acid sequence	Position
DQA1*00201	D H V A Y Y G I N V Y Q S Y G P S G Q Y T H E F D G D E E F	[30]
DQA1*00301	. . . .	[30]
DQA1*00401	. . . .	[30]
DQA1*00402	. . . . L .	[30]
DQA1*005011	. . . . F	[30]
DQA1*00601	. . . .	[30]
DQA1*00701	. . . .	[30]
DQA1*012011	. . . .	[30]
DQA1*00201	Y V D L E K K E T V W R L P V F S T F T S F D P Q G A L R N	[60]
DQA1*00301	. . . .	[60]
DQA1*00401	. . . .	[60]
DQA1*00402	. . . .	[60]
DQA1*005011	. . . .	[60]
DQA1*00601	. . . . R	[60]
DQA1*00701	. . . .	[60]
DQA1*012011	. . . . A	[60]
DQA1*00201	L A I T K Q N L N I M T K R S N K T A A T N	[82]
DQA1*00301	. . R A . . . . . . L . . S . . Q	[82]
DQA1*00401	. . . I . . . . . . L . . . . . Q	[82]
DQA1*00402	. . . I . . . . . . L . . . . . Q	[82]
DQA1*005011	. . . .	[82]
DQA1*00601	. . . I . . . . . . L . . . . . Q	[82]
DQA1*00701	. . . . Q	[82]
DQA1*012011	. . . A . . . . Q	[82]

Alleles DQB1*039** and DRB1*036** are previously undiscovered DLA alleles. To avoid future confusion regarding nomenclature of DLA alleles, for now I decided not to assign any five digit allele designations, and wait for them to be assigned by future DLA nomenclature committee. Allele DQB1*039** differs from allele DQB1*03901 by a single nonsynonymous transversion (adenine to thymine) on nucleotide position 125 in the explored sequence. This

mutation caused the change of codon 42 from TAC coding for tyrosine to TTC coding for phenylalanine. Similarly, allele DRB1*036** differs from allele DRB1*03601 by a single nonsynonymous transversion (cytosine to guanine) on nucleotide position 174 in the explored sequence. This mutation caused the change of codon 58 from CGG coding for arginine to GGG coding for glycine. The rest of the detected alleles were all known from previous research, but alleles: DQA1*00402, DQA1*00701, DQB1*02305, DQB1*02902, DQB1*04901, DRB1*03201 and DRB1*08001 were found for the first time in any of the European grey wolf populations.

Table 11. Amino acid sequences of DLA-DQB1 alleles found in grey wolves from Serbia (dots represent the identical nucleotide as in the top sequence).

Allele name	Allele amino acid sequence	Position
DQB1*00401	D F V F Q F K G E C Y F T N G T E R V R L L T K Y I Y N R E	[30]
DQB1*02002	. . . Y . . . A . . . . R S	[30]
DQB1*02301	. . . Y	[30]
DQB1*02305	. . . Y	[30]
DQB1*02901	. . . Y . C . A . . . . F . A	[30]
DQB1*02902	. . . Y . C . A . . . . F . A	[30]
DQB1*03501	. . . Y . . . F . . . . A R D	[30]
DQB1*03801	. . . . F	[30]
DQB1*03901	. . . . F . . . . A	[30]
DQB1*04901	. . . Y . . . A . . . . R S	[30]
DQB1*039**	. . . . F . . . . A	[30]
DQB1*00401	E Y V R F D S D V G E Y R A V T E L G R P W A E Y W N P Q K	[60]
DQB1*02002	. F . . . . F . . . . V . . . . . G . .	[60]
DQB1*02301	. . . . S	[60]
DQB1*02305	. . . . S	[60]
DQB1*02901	. F . . . . F . . . . D . . . . . G . .	[60]
DQB1*02902	. F . . . . F . . . . D . . S . . G . .	[60]
DQB1*03501	. H . . . . D . . . . . G . .	[60]
DQB1*03801	. . . .	[60]
DQB1*03901	. F	[60]
DQB1*04901	. F . . . . F . . . . V . . . . . G . .	[60]
DQB1*039**	. F . . . . F	[60]
DQB1*00401	D E M D R V R A E L D T V C R H N Y G L E E L Y T L Q R R	[89]
DQB1*02002	E I L E . K . . A V . R . . . . R . . . T	[89]
DQB1*02301	. . . . T	[89]
DQB1*02305	. . . . T S	[89]
DQB1*02901	E L L E Q E . . T V . . . . V	[89]
DQB1*02902	E L L E Q E . . T V . . . . V	[89]
DQB1*03501	E L L E Q R	[89]
DQB1*03801	. . . . T	[89]
DQB1*03901	. . . . V	[89]
DQB1*04901	E I L E . K . . . V . R . . . . R . . . T	[89]
DQB1*039**	. . . . V	[89]

Table 12. Amino acid sequences of DLA-DRB1 alleles found in grey wolves from Serbia (dots represent the identical nucleotide as in the top sequence).

Allele name	Allele amino acid sequence	Position
DRB1*01501	H F L E M V K F E C H F T N G T E R V R L L V R D I Y N R E	[30]
DRB1*03201	. . V Y Q F . P . . . . F V E . Y . H . . .	[30]
DRB1*03202	. . V Y Q F . P . . . . F V E . Y . H . . .	[30]
DRB1*03601	. . . . . L . S . . . . F V E . Y . H . . .	[30]
DRB1*04301	. . . . . L . . . . Y	[30]
DRB1*04901	. . . K . . . . V E	[30]
DRB1*05401	. . . . V A . S . . Y . . . . F V E . Y . H . . .	[30]
DRB1*08001	. . . . V L . . . . V E . Y	[30]
DRB1*090012	. . . . . L . . . . Y . L . H	[30]
DRB1*092011	. . . . . L . . . . F V E . Y . H . . .	[30]
DRB1*092012	. . . . . L . . . . F V E . Y . H . . .	[30]
DRB1*036**	. . . . . L . S . . . . F V E . Y . H . . .	[30]
DRB1*01501	E H V R F D S D V G E Y R A V T E L G R P D A E Y W N G Q K	[60]
DRB1*03201	. F . . . . V . . S	[60]
DRB1*03202	. F . . . . V . . S	[60]
DRB1*03601	. N . . . . R . . . S . . R . .	[60]
DRB1*04301	. . . . R . .	[60]
DRB1*04901	. Y . . . . F . . . . S	[60]
DRB1*05401	. N . . . . S	[60]
DRB1*08001	. Y . . . . R . . . S . . R . .	[60]
DRB1*090012	. I L . . . . F . . . . S . . R . .	[60]
DRB1*092011	. F . . . . S . . R . .	[60]
DRB1*092012	. F . . . . S . . R . .	[60]
DRB1*036**	. N . . . . R . . . S	[60]
DRB1*01501	E L L E Q R R A E V D T V C R H N Y G V I E S F T V Q R R	[89]
DRB1*03201	. I . . . E . . T	[89]
DRB1*03202	. I . . . E . . T . . . Y	[89]
DRB1*03601	. . . . R K . . . . . . Y	[89]
DRB1*04301	. . . . R . . . . R . G	[89]
DRB1*04901	. . . . A	[89]
DRB1*05401	. . . . A . . . Y	[89]
DRB1*08001	. . . . A . . . Y . . . . . R . G	[89]
DRB1*090012	. . . . . K . . A . . . Y . . . . . R . G	[89]
DRB1*092011	. F . . . . Y . . . . G	[89]
DRB1*092012	. F . . . . Y . . . . G	[89]
DRB1*036**	. . . . R K . . . . . . Y	[89]

The frequency of individual alleles and the number of homozygotes are given in tables 13, 14 and 15 for DQA, DQB and DRB loci, respectively. On DQA1 locus the total number of detected alleles is eight. The most frequent allele at DQA1 locus is DQA1*00601 (29.1%), while the least frequent alleles are DQA1*00402 and DQA1*00701 (0.6%), both found only once in heterozygous wolves. Likewise, allele DQA1*00401 is found only in heterozygotes. The proportion of homozygous wolves for this locus is 24.1%.

Table 13. Frequency of DLA-DQA1 alleles in the sample from Serbia (N= 79)

Allele name	#	f	H (h)
DQA1*00201	15	0.095	0.013 (1)
DQA1*00301	38	0.241	0.063 (5)
DQA1*00401	3	0.019	0 (0)
DQA1*00402	1	0.006	0 (0)
DQA1*005011	22	0.139	0.038 (3)
DQA1*00601	46	0.291	0.089 (7)
DQA1*00701	1	0.006	0 (0)
DQA1*012011	32	0.203	0.038 (3)
total	158	1	0.241 (19)

- absolute number of a given allele found in the sample; f – frequency of a given allele in the sample; H – homozygosity of a given allele in the sample (h/N); h – number of homozygotes of a given allele in the sample

DQB1 locus is more polymorphic as 11 alleles are found and the most frequent allele is DQB1*02002 (29.7%), while the least frequent alleles are DQB1*02301, DQB1*02305 and DQB1*04901 (0.6%), all found only once. Two more alleles, DQB1*02902 and DQB1*039** were found only in heterozygous wolves. The proportion of homozygotes is slightly lower on this locus with 22.8%

Table 14. Frequency of DLA-DQB1 alleles in the sample from Serbia (N=79)

Allele name	#	f	H (h)
DQB1*00401	24	0.152	0.038 (3)
DQB1*02002	47	0.297	0.089 (7)
DQB1*02301	1	0.006	0 (0)
DQB1*02305	1	0.006	0 (0)
DQB1*02901	12	0.076	0.013 (1)
DQB1*02902	3	0.019	0 (0)
DQB1*03501	32	0.203	0.038 (3)
DQB1*03801	15	0.095	0.025 (2)
DQB1*03901	16	0.101	0.025 (2)

Table 14. continued

Allele name	#	f	H (h)
DQB1*04901	1	0.006	0 (0)
DQB1*039**	6	0.038	0 (0)
total	158	1	0.228 (18)

- absolute number of a given allele found in the sample; f – frequency of a given allele in the sample; H – homozygosity of a given allele in the sample (h/N); h – number of homozygotes of a given allele in the sample

The largest number of alleles (12) was found at DRB1 locus, with the most frequent allele found with frequency of 28.5% (allele DRB1*092011), and the least frequent alleles are DRB1*01501 and DRB1*08001 (0.006%), again found only once. The proportion of homozygotes is 20.3%, which is the lowest out of three examined loci, and in total there are six alleles found only in heterozygous animals (table 15).

Table 15. Frequency of DLA-DRB1 alleles in the sample from Serbia (N=79)

Allele name	#	f	H (h)
DRB1*01501	1	0.006	0 (0)
DRB1*03201	3	0.019	0 (0)
DRB1*03202	10	0.063	0.013 (1)
DRB1*03601	24	0.152	0 (0)
DRB1*04301	15	0.095	0.025 (2)
DRB1*04901	22	0.139	0.038 (3)
DRB1*05401	24	0.152	0.038 (3)
DRB1*08001	1	0.006	0 (0)
DRB1*090012	8	0.051	0 (0)
DRB1*092011	45	0.285	0.089 (7)
DRB1*092012	3	0.019	0 (0)
DRB1*036**	2	0.013	0 (0)
total	158	1	0.201 (16)

- absolute number of a given allele found in the sample; f – frequency of a given allele in the sample; H – homozygosity of a given allele in the sample (h/N); h – number of homozygotes of a given allele in the sample

There are in total 18 distinct haplotypes in the Dinaric-Balkan population sample, two of which cannot be inferred since they both come from the same animal and are unique to the population. The frequency of each detected haplotype in presented in table 16. The most frequent three-locus haplotype found in the sample is haplotype DQA1*00601 / DQB1*02002 / DRB1*092011 with frequency of 26.6% and, unsurprisingly, found in the largest number of homozygous individuals (7). The next frequent haplotype DQA1*00301 / DQB1*00401 / DRB1*05401 is found with frequency of 15.2% and is found in three homozygous wolves.

Table 16. Individual haplotype frequency in the sample from Serbia (N=79)

Haplotype			#	f	H (h)
DQA1*00601	DQB1*02002	DRB1*092011	42	0.266	0.089 (7)
DQA1*00301	DQB1*00401	DRB1*05401	24	0.152	0.038 (3)
DQA1*012011	DQB1*03501	DRB1*03601	22	0.139	0 (0)
DQA1*005011	DQB1*03901	DRB1*04901	16	0.101	0.025 (2)
DQA1*00301	DQB1*03801	DRB1*04301	14	0.089	0.013 (1)
DQA1*00201	DQB1*02901	DRB1*03202	11	0.070	0.013 (1)
DQA1*012011	DQB1*03501	DRB1*090012	8	0.051	0 (0)
DQA1*005011	DQB1*039**	DRB1*04901	6	0.038	0 (0)
DQA1*00201	DQB1*02901	DRB1*03601	2	0.013	0 (0)
DQA1*00201	DQB1*02902	DRB1*03201	2	0.013	0 (0)
DQA1*00401	DQB1*02002	DRB1*092011	2	0.013	0 (0)
DQA1*00601	DQB1*02002	DRB1*092012	2	0.013	0 (0)
DQA1*012011	DQB1*03501	DRB1*036**	2	0.013	0 (0)
DQA1*00401	DQB1*02002	DRB1*092012	1	0.006	0 (0)
DQA1*00601	DQB1*04901	DRB1*092011	1	0.006	0 (0)
DQA1*00701	DQB1*03801	DRB1*04301	1	0.006	0 (0)
could not resolve			2	0.013	0 (0)
total			158	1	0.177 (14)

- absolute number of given haplotype found in the sample; f – frequency of a given haplotype in the sample; H – homozygosity of a given haplotype in the sample (h/N); h – number of homozygotes of a given haplotype in the sample

There are five least frequent haplotypes, all found with a frequency of 0.6% (once). Three of those five could be resolved using the inference method: DQA1*00401 / DQB1*02002 / DRB1*092012, DQA1*00601 / DQB1*04901 / DRB1*092011 and DQA1*00701 / DQB1*03801 / DRB1*04301. The total proportion of homozygous wolves is 17.7%. Haplotype distribution in individual wolves is presented in table 68 in the appendix. Note that for animal 299, haplotypes could not be resolved. Wolf 299 has the following genotype: DQA1*00402/*00601 // DQB1*02301/*02305 // DRB1*01501/*08001. So on both DQB1 and DRB1 loci this wolf has alleles not found in any of the other animals in the population sample, and thus its genotype cannot be reliably separated into haplotypes using the inference method.

4.3. Evolutionary distances

Table 17 shows the calculated overall mean evolutionary distances obtained by separately comparing nucleotide sequences and amino acid sequences in the samples from the Carpathian population, while table 18 shows the results of the same analysis for the sample from Serbia. Results are obtained and presented for each locus independently. The locus DQB1 has the highest proportion of variable nucleotide sites (12.4%), followed by the locus DRB1 (11.1%), while locus DQA1 has only 4.1% of variable positions in the Carpathian population. Similarly, in population from Serbia, substantially higher proportion of variable nucleotide sites show loci DQB1 and DRB1 (14.6% and 17.4%, respectively), while locus DQA1 has only 4.9% of variable positions. However, it should be noted that all those values are higher for population from Serbia than Carpathian population. Nucleotide distances are comparable in both populations, as well as amino acid distances, with the exception of amino acid distance for the locus DQB1 which is estimated to 0.9 for Carpathian population but only to 0.245 for population from Serbia.

Table 17. Number of variable positions, average nucleotide and amino acid distances and the number of unique amino acid sequences for DLA-DQA, DLA-DQB and DLA-DRB alleles found in the sample of 99 grey wolves from the Carpathian population

Locus	Sequence length	Number of variable positions	Nucleotide distance		Amino acid distance		Number of unique amino acid sequences
			Substitution model	d	Substitution model	d	
DQA1	246	10	JC	0.022	JTT	0.055	4
DQB1	267	33	JC + G	0.383	JTT + G	0.900	7
DRB1	270	30	T92 + G	0.208	JTT + G	0.186	7

d – overall mean distance; JC – Jukes-Cantor; T92 – Tamura 3-parameter; JTT – Jones-Taylor-Thornton; G – Gamma distribution

Table 18. Number of variable positions, average nucleotide and amino acid distances and the number of unique amino acid sequences for DLA-DQA, DLA-DQB and DLA-DRB alleles found in the sample of 79 grey wolves from Serbia

Locus	Sequence length	Number of variable positions	Nucleotide distance		Amino acid distance		Number of unique amino acid sequences
			Substitution model	d	Substitution model	d	
DQA1	246	12	JC	0.016	JTT	0.040	8
DQB1	267	39	JC + G	0.380	JTT + G	0.245	11
DRB1	270	47	T92 + G	0.321	JTT + G	0.230	11

d – overall mean distance; JC – Jukes-Cantor; T92 – Tamura 3-parameter; JTT – Jones-Taylor-Thornton; G – Gamma distribution

4.4. Hardy-Weinberg equilibrium

The values of observed and expected heterozygosity calculated in Arlequin software can be seen in tables 19 and 20. Values are presented separately for Carpathian population and Dinaric-Balkan population and for each locus. Expected heterozygosity is higher than observed heterozygosity at each locus and in both populations. However, the excess was not significant at most loci, except for locus DQB1 in Carpathian population ($P_{HWE}<0.05$).

Table 19. Observed and expected heterozygosity in regards to Hardy-Weinberg equilibrium for each studied locus in the sample from Carpathian population

Locus	#	H_o	H_e	P_{HWE}
DQA1	99	0.626	0.675	0.221
DQB1	99	0.697	0.770	0.000*
DRB1	99	0.657	0.695	0.059

- number of sequences analysed; H_o – observed heterozygosity; H_e – expected heterozygosity under Hardy-Weinberg equilibrium; P_{HWE} – p value assuming Hardy-Weinberg equilibrium; * $P<0.05$

Table 20. Observed and expected heterozygosity in regards to Hardy-Weinberg equilibrium for each studied locus in the sample from Serbia

Locus	#	H_o	H_e	P_{HWE}
DQA1	79	0.759	0.793	0.370
DQB1	79	0.772	0.825	0.342
DRB1	79	0.797	0.841	0.098

- number of sequences analysed; H_o – observed heterozygosity; H_e – expected heterozygosity under Hardy-Weinberg equilibrium; P_{HWE} – p value assuming Hardy-Weinberg equilibrium

4.5. Selection

For both population samples I performed several analyses to test the type of selection acting on the MHC loci. In the tables 21 and 22 we can see the results of Estimate Selection at Codons (via HyPhy) analysis for the samples from Carpathian population and Dinaric-Balkan population, respectively. The tables include dN/dS ratio as an important indicator of evolution (Yang, 1998). The dN-dS values indicate positive selection at all three loci and in both populations, as rates of nonsynonymous nucleotide substitutions are higher than those of synonymous substitutions. The difference between dN and dS is especially significant at DQB1 (P=0.000) and DRB1 (P=0.001) loci, and less so at DQA1 locus where the p value is lower than 5% only in the sample from Serbia.

Table 21. Rates of nucleotide substitutions in the sample from Carpathian population

Locus	dS	dN	dN-dS	P	dN/dS
DQA1	0.012	0.055	0.043	0.074	4.52
DQB1	0.039	0.202	0.163	0.000*	5.13
DRB1	0.082	0.179	0.097	0.001*	2.18

dS – average rate of synonymous substitutions; dN – average rate of nonsynonymous substitutions; P – p value of dN-dS; * $P<0.05$

Table 22. Rates of nucleotide substitutions in the sample from Serbia

Locus	dS	dN	dN-dS	P	dN/dS
DQA1	0.012	0.066	0.054	0.045*	5.45
DQB1	0.064	0.285	0.222	0.000*	4.48
DRB1	0.172	0.389	0.216	0.000*	2.25

dS – average rate of synonymous substitutions; dN – average rate of nonsynonymous substitutions; P – p value of dN-dS; * $P<0.05$

To further explore whether DLA loci evolve under neutral selection or not I performed a codon based Z-test of selection under the null hypothesis of neutrality, meaning that the average rates of mutations are similar for nonsynonymous and synonymous substitutions (H0 : $d_N = d_S$). Three alternative hypothesis were tested: evolution is not neutral ($d_N \neq d_S$), evolution is under positive selection ($d_N > d_S$), and evolution is under purifying selection ($d_N < d_S$). The results are presented in table 23. The analysis strongly indicates that the loci DQB1 and DRB1 are evolving non-neutrally and under positive selection in both populations. For the DQA1 loci the p value of the non-neutrality test is high in both populations, so the null hypothesis cannot be rejected based on this analysis.

Table 23. Results of codon based Z-test of selection

Population	Locus	#s	#c	P (H_A : $d_N \neq d_S$)	P (H_A : $d_N > d_S$)	P (H_A : $d_N < d_S$)
Carpathian	DQA1	198	82	0.343	0.177	1.000
	DQB1	198	89	0.003*	0.000*	1.000
	DRB1	198	89	0.003*	0.001*	1.000
Dinaric-Balkan	DQA1	158	82	0.172	0.086	1.000
	DQB1	158	89	0.001*	0.000*	1.000
	DRB1	158	89	0.010*	0.005*	1.000

#s – number of sequences included in analysis; #c – number of codons in the final dataset;
P – p-value under the null hypothesis; H_A – alternative hypothesis;
dN - number of nonsynonymous substitutions per site;
dS – number of synonymous substitutions per site; * $P<0.05$

Furthermore, I performed Tajima's test of neutrality, results of which can be seen in table 24. Positive Tajima’s D found at each locus in both populations signifies balancing selection. Confidence limit is lowest on the DQA1 locus in the Dinaric-Balkan population, but was still significant at 95%. On the rest of the tested loci the confidence limit is at least 99%.

Table 24. Results from Tajima's test of neutrality

Population	Locus	m	n	S	p_s	θ	π	D	confidence limit of D
Carpathian	DQA1	198	246	10	0.041	0.007	0.015	2.566	99%
	DQB1	198	267	33	0.124	0.021	0.051	4.052	99.9%
	DRB1	198	270	30	0.111	0.019	0.034	2.262	99%
Dinaric-Balkan	DQA1	158	246	12	0.049	0.009	0.016	2.081	95%
	DQB1	158	267	39	0.146	0.026	0.065	4.498	99.9%
	DRB1	158	270	47	0.174	0.031	0.054	2.297	99%

m – number of sequences; n – total number of sites; S – number of segregating sites; p_s – S/n; θ - p_s/a_1, a1 = 1 + 2-1 + 3-1 + . . . + (m - 1)-1; π - nucleotide diversity; D – Tajima's test statistic;

As an additional test of selection I used EasyCodeML to calculate and compare likelihoods of seven different site models of evolution: M0, M1a, M2a, M3, M7, M8 and M8a. The results for the Carpathian population sample is presented in tables 25, 26 and 27 for DQA1, DQB1 and DRB1 loci, respectively. The results for the Dinaric-Balkan population sample is presented in tables 28, 29 and 30 for DQA1, DQB1 and DRB1 loci, respectively. As models M3, M2a and M8 show a better fit to the sequence data (higher likelihood) than models M0, M1a, M7 and M8a (with $p < 0.05$ for each likelihood-ration test), it is a strong indicator of positive selection at each locus and in each population. Again, the indication for positive selection is much stronger on the DQB1 and DRB1 loci, than on the DQA1 locus. But this analysis shows significant difference for all tested models in favour of non-neutral and positive selection on DQA1 locus as well, in both populations.

Table 25. Results of likelihood ratio test comparing different site models for DQA1 locus in the sample from Carpathian population

Model	$P_{\#}$	Ln L	M vs. M	$\Delta P_{\#}$	2 Δ Ln L	LRT p-value
M3	11	-397.926	M0 vs. M3	4	11.313	0.023*
M0	7	-403.582				
M2a	10	-397.926	M1a vs. M2a	2	11.313	0.003*
M1a	8	-403.582				
M8	10	-397.926	M7 vs. M8	2	12.557	0.002*
M7	8	-404.204				
M8a	9	-403.582	M8a vs. M8	1	11.313	0.001*

$P_{\#}$ - number of parameters; Ln L – log-likelihood value of tested model; LRT – likelihood ratio test; * $P<0.05$

Table 26. Results of likelihood ratio test comparing different site models for DQB1 locus in the sample from Carpathian population

Model	$P_{\#}$	Ln L	M vs. M	$\Delta P_{\#}$	2 Δ Ln L	LRT p-value
M3	17	-517.417	M0 vs. M3	4	60.181	0.000*
M0	13	-547.508				
M2a	16	-518.229	M1a vs. M2a	2	35.084	0.000*
M1a	14	-535.771				
M8	16	-518.229	M7 vs. M8	2	35.086	0.000*
M7	14	-535.771				
M8a	15	-535.771	M8a vs. M8	1	35.084	0.000*

$P_{\#}$ - number of parameters; Ln L – log-likelihood value of tested model; LRT – likelihood ratio test; * $P<0.05$

Table 27. Results of likelihood ratio test comparing different site models for DRB1 locus in the sample from Carpathian population

Model	$P_{\#}$	Ln L	M vs. M	$\Delta P_{\#}$	2 Δ Ln L	LRT p-value
M3	17	-528.508	M0 vs. M3	4	49.131	0.000*
M0	13	-553.074				
M2a	16	-536.457	M1a vs. M2a	2	15.369	0.000*
M1a	14	-544.141				
M8	16	-537.030	M7 vs. M8	2	16.249	0.000*
M7	14	-545.155				
M8a	15	-544.141	M8a vs. M8	1	14.223	0.000*

$P_{\#}$ - number of parameters; Ln L – log-likelihood value of tested model; LRT – likelihood ratio test; * $P<0.05$

Table 28. Results of likelihood ratio test comparing different site models for DQA1 locus in the sample from Serbia

Model	$P_{\#}$	Ln L	M vs. M	$\Delta P_{\#}$	2 Δ Ln L	LRT p-value
M3	19	-415.637	M0 vs. M3	4	12.536	0.014*
M0	15	-421.905				
M2a	18	-415.637	M1a vs. M2a	2	12.109	0.002*
M1a	16	-421.692				
M8	18	-415.637	M7 vs. M8	2	14.112	0.001*
M7	16	-422.693				
M8a	17	-421.692	M8a vs. M8	1	12.109	0.001*

$P_{\#}$ - number of parameters; Ln L – log-likelihood value of tested model; LRT – likelihood ratio test; * $P<0.05$

Table 29. Results of likelihood ratio test comparing different site models for DQB1 locus in the sample from Serbia

Model	$P_{\#}$	Ln L	M vs. M	$\Delta P_{\#}$	2 Δ Ln L	LRT p-value
M3	25	-634.268	M0 vs. M3	4	79.227	0.000*
M0	21	-673.881				
M2a	24	-634.290	M1a vs. M2a	2	33.488	0.000*
M1a	22	-651.034				
M8	24	-634.290	M7 vs. M8	2	33.493	0.000*
M7	22	-651.037				
M8a	23	-651.034	M8a vs. M8	1	33.488	0.000*

$P_{\#}$ - number of parameters; Ln L – log-likelihood value of tested model; LRT – likelihood ratio test; * P<0.05

Table 30. Results of likelihood ratio test comparing different site models for DRB1 locus in the sample from Serbia

Model	$P_{\#}$	Ln L	M vs. M	$\Delta P_{\#}$	2 Δ Ln L	LRT p-value
M3	27	-756.860	M0 vs. M3	4	109.079	0.000*
M0	23	-811.400				
M2a	26	-760.443	M1a vs. M2a	2	20.513	0.000*
M1a	24	-770.700				
M8	26	-760.457	M7 vs. M8	2	20.487	0.000*
M7	24	-770.700				
M8a	25	-770.700	M8a vs. M8	1	20.486	0.000*

$P_{\#}$ - number of parameters; Ln L – log-likelihood value of tested model; LRT – likelihood ratio test; * P<0.05

The following tables show individual per codon rates of nucleotide substitutions for each codon which has any variation in the entire dataset, obtained by HyPhy analysis. From these tables we can read which codons in each population underwent any change, and if changes on those codons are more often synonymous or nonsynonymous. Tables 31, 32 and 33 show results for the Carpathian population, for DQA1, DQB1 and DRB1 loci, respectively. On DQA1 locus (table 31) there are eight variable codons, seven of which are showing only non-synonymous substitutions and one codon with a synonymous substitution. On DQB1 locus (table 32) there is a much greater variability, with a total of 20 variable codons. Sixteen codons show only non-synonymous substitutions, while on four codons we find both synonymous and non-synonymous substitutions, but with an excess of non-synonymous. DRB1 locus (table 33) again shows high diversity, with 22 variable codons and much greater number of non-synonymous than synonymous substitutions. Eighteen codons show only non-synonymous substitutions, and just one codon has only synonymous substitution. The other three codons have both types of substitutions, but with an excess of non-synonymous.

Table 31. Results of maximum likelihood analysis of natural selection codon-by-codon for DQA1 locus in the sample from the Carpathian population (only codons with any changes in the dataset are presented)

Codon	Syn	Nonsyn	Syn sites	Nonsyn sites	dS	dN	dN-dS	p
20	0	1	0.316	2.070	0	0.483	0.483	0.868
50	0	2	0.958	2.000	0	1.000	1.000	0.457
62	1	0	1.000	2.000	1	0.000	-1.000	1.000
63	0	1	0.676	2.267	0	0.441	0.441	0.770
64	0	2	0.960	2.040	0	0.980	0.980	0.463
71	0	2	0.391	2.609	0	0.767	0.767	0.756
74	0	1	0.650	2.350	0	0.425	0.425	0.783
77	0	1	0.322	2.360	0	0.424	0.424	0.880

Codon – order number of the codon in the sequence alignment; Syn – number of inferred synonymous substitutions; Nonsyn – number of inferred nonsynonymous substitutions; Syn sites – number of sites that are estimated to be synonymous; Nonsyn sites – number of sites that are estimated to be nonsynonymous; dS – number of synonymous substitutions per synonymous site (Syn / Syn sites); dN – number of nonsynonymous substitutions per nonsynonymous site (Nonsyn / Nonysyn sites); p – p value of rejecting the null hypothesis

Table 32. Results of maximum likelihood analysis of natural selection codon-by-codon for DQB1 locus in the sample from the Carpathian population (only codons with any changes in the dataset are presented)

Codon	Syn	Nonsyn	Syn sites	Nonsyn sites	dS	dN	dN-dS	p
4	0	1	0.229	2.245	0	0.446	0.446	0.908
8	0	4	0.471	2.525	0	1.584	1.584	0.505
23	1	2	1	2	1	1	0	0.741
24	0	1	0.682	2.097	0	0.477	0.477	0.754
25	0	3	0.272	2.484	0	1.208	1.208	0.732
32	0	3	0.241	2.738	0	1.096	1.096	0.776
42	0	1	0.229	2.147	0	0.466	0.466	0.904
52	0.25	3.75	0.410	2.385	0.610	1.572	0.962	0.622
58	0	2	1.019	1.981	0	1.010	1.010	0.436
61	0	1	0.332	2.456	0	0.407	0.407	0.881
62	0.5	3.5	0.708	2.218	0.707	1.578	0.871	0.541
63	0	1	0.416	2.392	0	0.418	0.418	0.852
64	0	1	0.332	2.456	0	0.407	0.407	0.881
65	0	1	1.217	1.767	0	0.566	0.566	0.592
66	0.833	3.167	0.701	2.238	1.188	1.415	0.226	0.687
69	0	1	0.388	2.306	0	0.434	0.434	0.856
70	0	1	1.185	1.815	0	0.551	0.551	0.605
72	0	1	0.940	2.060	0	0.485	0.485	0.687
80	0	4	0.871	2.107	0	1.898	1.898	0.251
84	0	2	0.376	2.043	0	0.979	0.979	0.714

Codon – order number of the codon in the sequence alignment; Syn – number of inferred synonymous substitutions; Nonsyn – number of inferred nonsynonymous substitutions; Syn sites – number of sites that are estimated to be synonymous; Nonsyn sites – number of sites that are estimated to be nonsynonymous; dS – number of synonymous substitutions per synonymous site (Syn / Syn sites); dN – number of nonsynonymous substitutions per nonsynonymous site (Nonsyn / Nonysyn sites); p – p value of rejecting the null hypothesis

Table 33. Results of maximum likelihood analysis of natural selection codon-by-codon for DRB1 locus in the sample from the Carpathian population (only codons with any changes in the dataset are presented)

Codon	Syn	Nonsyn	Syn sites	Nonsyn sites	dS	dN	dN-dS	p
4	0	1	0.405	2.327	0	0.430	0.430	0.852
5	0	1	0.177	2.823	0	0.354	0.354	0.941
6	0	3	0.855	1.750	0	1.714	1.714	0.303
8	0	1	0.441	2.559	0	0.391	0.391	0.853
11	0	1	0.212	2.649	0	0.378	0.378	0.926
21	0	1	0.342	2.565	0	0.390	0.390	0.882
25	0	1	0.212	2.254	0	0.444	0.444	0.914
27	0	1	0.212	2.534	0	0.395	0.395	0.923
32	0	2	0.212	2.279	0	0.877	0.877	0.837
42	0	1	0.212	2.090	0	0.479	0.479	0.908
51	0	2	1	2	0	1	1	0.444
52	0.75	2.25	0.509	2.400	1.473	0.938	-0.535	0.829
55	0	1	0.746	2	0	0.5	0.5	0.728
58	0	2	1.177	1.823	0	1.097	1.097	0.369
62	0	1	0.950	2.050	0	0.488	0.488	0.683
65	0	2	0.606	2.231	0	0.897	0.897	0.618
66	0	2	0.640	2.313	0	0.865	0.865	0.613
69	0	1	0.443	2.279	0	0.439	0.439	0.837
72	1	0	1	2	1	0	-1	1
73	0.5	2.5	0.442	2.087	1.132	1.198	0.066	0.741
81	2	4	0.546	2.454	3.664	1.630	-2.034	0.922
85	0	2	1	2	0	1	1	0.444

Codon – order number of the codon in the sequence alignment; Syn – number of inferred synonymous substitutions; Nonsyn – number of inferred nonsynonymous substitutions; Syn sites – number of sites that are estimated to be synonymous; Nonsyn sites – number of sites that are estimated to be nonsynonymous; dS – number of synonymous substitutions per synonymous site (Syn / Syn sites); dN – number of nonsynonymous substitutions per nonsynonymous site (Nonsyn / Nonysyn sites); p – p value of rejecting the null hypothesis

Tables 34, 35 and 36 show results of the HyPhy analysis for the population sample from Serbia, for DQA1, DQB1 and DRB1 loci, respectively. On DQA1 locus (table 34) there are nine variable codons, eight of which are showing only non-synonymous substitutions and one codon with a synonymous substitution. As before, on DQB1 locus (table 35) we see greater variability, with a total of 25 variable codons. Twenty codons show only non-synonymous substitutions, and three codons have both synonymous and non-synonymous substitutions, but with a much higher number of non-synonymous. For the remaining two codons, one has an equal amount of both types of substitutions, and one has only synonymous. DRB1 locus (table 36) shows highest diversity, with 28 variable codons, again with a much greater number of non-synonymous than synonymous substitutions. Twenty one codons show only non-synonymous substitutions, and two codons have only synonymous substitutions. Remaining five codons have both types of substitutions, again with an excess of non-synonymous.

Table 34. Results of maximum likelihood analysis of natural selection codon-by-codon for DQA1 locus in the sample from Serbia (only codons with any changes in the dataset are presented)

Codon	Syn	Nonsyn	Syn sites	Nonsyn sites	dS	dN	dN-dS	p
20	0	1	0.317	2.026	0	0.494	0.494	0.865
29	0	2	0.426	2.275	0	0.879	0.879	0.710
50	0	2	0.987	2	0	1	1	0.448
62	1	0	1	2	1	0	-1	1
63	0	1	0.676	2.271	0	0.440	0.440	0.771
64	0	2	0.901	2.099	0	0.953	0.953	0.489
71	0	2	0.608	2.392	0	0.836	0.836	0.636
74	0	1	0.654	2.346	0	0.426	0.426	0.782
77	0	1	0.321	2.359	0	0.424	0.424	0.880

Codon – order number of the codon in the sequence alignment; Syn – number of inferred synonymous substitutions; Nonsyn – number of inferred nonsynonymous substitutions; Syn sites – number of sites that are estimated to be synonymous; Nonsyn sites – number of sites that are estimated to be nonsynonymous; dS – number of synonymous substitutions per synonymous site (Syn / Syn sites); dN – number of nonsynonymous substitutions per nonsynonymous site (Nonsyn / Nonysyn sites); p – p value of rejecting the null hypothesis

Table 35. Results of maximum likelihood analysis of natural selection codon-by-codon for DQB1 locus in the sample from Serbia (only codons with any changes in the dataset are presented)

Codon	Syn	Nonsyn	Syn sites	Nonsyn sites	dS	dN	dN-dS	p
4	0	2	0.227	2.232	0	0.896	0.896	0.824
6	0	1	0.227	2.751	0	0.363	0.363	0.924
8	0	6	0.596	2.400	0	2.500	2.500	0.264
21	0	1	0.939	2.061	0	0.485	0.485	0.687
23	2	2	1	2	2	1	-1	0.889
24	0	2	0.652	2.129	0	0.939	0.939	0.586
25	0	3	0.227	2.130	0	1.408	1.408	0.738
32	0	3	0.235	2.635	0	1.139	1.139	0.774
42	0	2	0.227	2.315	0	0.864	0.864	0.830
52	0.25	4.75	0.356	2.427	0.702	1.957	1.255	0.597
55	0	1	0.233	2	0	0.5	0.5	0.896
58	0	2	1.013	1.987	0	1.007	1.007	0.439
61	0	1	0.327	2.471	0	0.405	0.405	0.883
62	0.5	3.5	0.727	2.182	0.688	1.604	0.916	0.527
63	0	1	0.397	2.421	0	0.413	0.413	0.859
64	0	1	0.327	2.471	0	0.405	0.405	0.883
65	0	2	0.795	2.083	0	0.960	0.960	0.524
66	0.833	4.167	0.656	2.238	1.271	1.862	0.591	0.615
68	1	0	1	2	1	0	-1	1
69	0	3	0.435	2.284	0	1.313	1.313	0.593
70	0	1	1.136	1.864	0	0.537	0.537	0.621
72	0	1	0.964	2.036	0	0.491	0.491	0.679
80	0	4	0.616	2.150	0	1.860	1.860	0.365
84	0	4	0.388	2.029	0	1.971	1.971	0.497
85	0	1	1	1.998	0	0.501	0.501	0.666

Codon – order number of the codon in the sequence alignment; Syn – number of inferred synonymous substitutions; Nonsyn – number of inferred nonsynonymous substitutions; Syn sites – number of sites that are estimated to be synonymous; Nonsyn sites – number of sites that are estimated to be nonsynonymous; dS – number of synonymous substitutions per synonymous site (Syn / Syn sites); dN – number of nonsynonymous substitutions per nonsynonymous site (Nonsyn / Nonysyn sites); p – p value of rejecting the null hypothesis

Table 36. Results of maximum likelihood analysis of natural selection codon-by-codon for DRB1 locus in the sample from Serbia (only codons with any changes in the dataset are presented)

Codon	Syn	Nonsyn	Syn sites	Nonsyn sites	dS	dN	dN-dS	p
3	0	1	0.659	2.092	0	0.478	0.478	0.760
4	0	3	0.368	2.304	0	1.302	1.302	0.641
5	0	4	0.123	2.850	0	1.403	1.403	0.844
6	0	5	0.735	1.814	0	2.756	2.756	0.182
8	0	2	0.489	2.511	0	0.796	0.796	0.701
11	0	1	0.217	2.763	0	0.362	0.362	0.927
21	0	6	0.277	2.617	0	2.293	2.293	0.547
22	0	2	1.096	1.904	0	1.050	1.050	0.403
23	2	3	0.724	2.107	2.761	1.424	-1.337	0.890
25	0	2	0.217	2.287	0	0.875	0.875	0.834
27	0	1	0.217	2.340	0	0.427	0.427	0.915
32	0	6	0.227	2.339	0	2.565	2.565	0.574
33	0	1	1.012	1.988	0	0.503	0.503	0.663
42	0	2	0.217	2.088	0	0.958	0.958	0.821
51	1	2	1	2	1	1	0	0.741
52	0.75	3.25	0.372	2.612	2.018	1.244	-0.773	0.838
55	0	1	0.755	2	0	0.5	0.5	0.726
58	0	4	1.108	1.892	0	2.114	2.114	0.158
62	0	2	0.918	2.082	0	0.961	0.961	0.482
65	0	2	0.455	2.343	0	0.854	0.854	0.701
66	0	4	0.643	2.304	0	1.736	1.736	0.374
68	1	0	1	2	1	0	-1	1
69	0	4	0.776	2.112	0	1.894	1.894	0.286
72	2	0	1	2	2	0	-2	1
73	1	5	0.469	2.039	2.134	2.452	0.319	0.688
79	0	3	1.082	1.918	0	1.564	1.564	0.261
81	3	6	0.676	2.324	4.439	2.582	-1.857	0.877
85	0	1	1	2	0	0.5	0.5	0.667

Codon – order number of the codon in the sequence alignment; Syn – number of inferred synonymous substitutions; Nonsyn – number of inferred nonsynonymous substitutions; Syn sites – number of sites that are estimated to be synonymous; Nonsyn sites – number of sites that are estimated to be nonsynonymous; dS – number of synonymous substitutions per synonymous site (Syn / Syn sites); dN – number of nonsynonymous substitutions per nonsynonymous site (Nonsyn / Nonysyn sites); p – p value of rejecting the null hypothesis

In addition to the analysis of selection I used EasyCodeML to identify codons that have potentially evolved under selection. In table 37 we can see the codons identified as under positive selection, and the posterior probability, by applying Bayes empirical Bayes method. For the Carpathian population, the analysis shows seven codons under selection on the DQA1 locus, with three having high posterior probability (> 95%). On the DQB1 locus there are 20 codons identified as under positive selection, ten of which with high probability. Finally on the DRB1 locus there are 18 detected codons, six of them having posterior probability higher than 95%. For the Dinaric-Balkan sample, on DQA1 locus the analysis detects eight codons, four of which show high posterior probability. On DQB1 locus there are 19 codons detected as under positive selection, 11 with high probability, and on DRB1 locus there are 23 codons detected, 13 with high probability.

Table 37. Results of Bayes empirical Bayes method in EasyCodeML

Population	Loci	Codons (posterior probability)
Carpathian	DQA1	20 (0.836), **50** (0.983), 63 (0.851), **64** (0.983), **71** (0.982), 74 (0.843), 77 (0.848)
	DQB1	4 (0.838), **8** (1.000), **23** (0.988), 24 (0.904), **25** (0.999), **32** (0.999), 42 (0.838), **52** (1.000), **58** (0.986), 61 (0.750), **62** (1.000), 63 (0.841), 64 (0.750), 65 (0.871), **66** (0.999), 69 (0.753), 70 (0.825), 72 (0.845), **80** (1.000), **84** (0.988)
	DRB1	**6** (0.999), 8 (0.622), 11 (0.776), 21 (0.596), 25 (0.519), 27 (0.789), **32** (0.963), 42 (0.943), **51** (0.964), **52** (0.991), 55(0.666), 58 (0.913), 62 (0.767), 65 (0.919), 66 (0.887), **73** (0.990), **81** (1.000), 85 (0.532)
Dinaric-Balkan	DQA1	20 (0.841), **29** (0.982), **50** (0.986), 63 (0.861), **64** (0.985), **71** (0.984), 74 (0.852), 77 (0.853)
	DQB1	4 (0.935), 6 (0.803), **8** (1.000), 21 (0.862), **23** (0.982), **24** (0.988), **25** (0.997), **32** (0.996), 42 (0.945), **52** (1.000), 58 (0.913), **62** (0.999), 65 (0.912), **66** (1.000), **69** (0.993), 72 (0.523) **80** (1.000), **84** (1.000), 85 (0.509)
	DRB1	4 (0.935), **5** (0.996), **6** (1.000), 8 (0.925), 11 (0.756), **21** (1.000), 22 (0.825), **23** (0.994), 25 (0.826), 27 (0.841), **32** (1.000), 42 (0.892), **51** (0.951), **52** (0.991), 55 (0.575), **58** (0.967), 62 (0.928), **65** (0.956), **66** (0.995), **69** (0.995), **73** (1.000), 79 (0.790), **81** (0.999)

Codons with posterior probability over 95% are in bold. Numbers outside of parentheses are the codon order numbers in the obtained sequences of a given loci starting from the 5' end. Note that for the DRB1 loci the first codon starts on base 3 in the nucleotide sequences given in appendix (tables 63 and 67). Numbers in parentheses are the posterior probablities of the Bayes empirical Bayes analysis.

4.6. Linkage disequilibrium

Exact test of linkage disequilibrium showed the presence of significant association between pairs of loci for all examined loci in both populations. The results of linkage disequilibrium tests are presented as observed contingency tables for each pair of loci independently and expected contingency tables under the hypothesis of random association between alleles at different loci.

4.6.1. Carpathian population

For Carpathian population observed contingency tables are table 38 for DQA1 and DQB1 loci, table 40 for DQA1 and DRB1 loci and table 42 for DQB1 and DRB1 loci. Expected contingency tables for Carpathian population are table 39 for DQA1 and DQB1 loci, table 41 for DQA1 and DRB1 loci and table 43 for DQB1 and DRB1 loci.

When testing the linkage of DQA1 and DQB1 loci in the Carpathian population, we see that the expected number of two-loci haplotypes is at least 24 (table 39), while the observed number is seven (table 38). Under the null hypothesis the p value is equal to 0.000, so we can reject the possibility of random association of alleles between the DQA1 and DQB1 loci.

Table 38. Observed contingency table for DQA1 and DQB1 loci in the sample from Carpathian population (N= 99)

DQA1\\DQB1	*00401	*03901	*03501	*04101	*02002	*03502	*04901
*00301	58	0	0	24	0	0	0
*005011	0	30	0	0	0	0	0
*012011	0	0	64	0	0	7	0
*00601	0	0	0	0	14	0	1

Table 39. Expected contingency table for DQA1 and DQB1 loci in the sample from Carpathian population (N= 99)

DQA1\\DQB1	*00401	*03901	*03501	*04101	*02002	*03502	*04901
*00301	24.02	12.42	26.51	9.94	5.80	2.90	0.41
*005011	8.79	4.55	9.70	3.64	2.12	1.06	0.15
*012011	20.80	10.76	22.95	8.61	5.02	2.51	0.36
*00601	4.39	2.27	4.85	1.82	1.06	0.53	0.08

When testing the linkage of DQA1 and DRB1 loci in the Carpathian population, we see that the expected number of two-loci haplotypes is at least 21 (table 41), while the observed number is seven (table 40). Under the null hypothesis the p value is equal to 0.000, so we can reject the possibility of random association of alleles between the DQA1 and DRB1 loci.

Table 40. Observed contingency table for DQA1 and DRB1 loci in the sample from Carpathian population (N= 99)

DQA1\\DRB1	*05401	*04902	*03601	*092011	*09901	*04901	*12801
*00301	79	0	0	0	3	0	0
*005011	0	23	0	0	0	6	1
*012011	0	0	71	0	0	0	0
*00601	0	0	0	15	0	0	0

Table 41. Expected contingency table for DQA1 and DRB1 loci in the sample from Carpathian population (N= 99)

DQA1\\DRB1	*05401	*04902	*03601	*092011	*09901	*04901	*12801
*00301	32.72	9.53	29.40	6.21	1.24	2.48	0.41
*005011	11.97	3.48	10.76	2.27	0.45	0.91	0.15
*012011	28.33	8.25	25.46	5.38	1.08	2.15	0.36
*00601	5.98	1.74	5.38	1.14	0.23	0.45	0.08

Finally, when testing the linkage of DQB1 and DRB1 loci in the Carpathian population, we see that the expected number of two-loci haplotypes is at least 27 (table 43), while the observed number is nine (table 42). Again, since the p value is equal to 0.000, we can reject the null hypothesis of random association of alleles between the DQB1 and DRB1 loci.

Table 42. Observed contingency table for DQB1 and DRB1 loci in the sample from Carpathian population (N= 99)

DQB1\\DRB1	*05401	*04902	*03601	*092011	*09901	*04901	*12801
*00401	55	0	0	0	3	0	0
*03901	0	23	0	0	0	6	1
*03501	0	0	64	0	0	0	0
*04101	24	0	0	0	0	0	0
*02002	0	0	0	14	0	0	0
*03502	0	0	7	0	0	0	0
*04901	0	0	0	1	0	0	0

Table 43. Expected contingency table for DQB1 and DRB1 loci in the sample from Carpathian population (N= 99)

DQB1\\DRB1	*05401	*04902	*03601	*092011	*09901	*04901	*12801
*00401	23.14	6.74	20.80	4.39	0.88	1.76	0.29
*03901	11.97	3.48	10.76	2.27	0.45	0.91	0.15
*03501	25.54	7.43	22.95	4.85	0.97	1.94	0.32
*04101	9.58	2.79	8.61	1.82	0.36	0.73	0.12
*02002	5.59	1.63	5.02	1.06	0.21	0.42	0.07
*03502	2.79	0.81	2.51	0.53	0.11	0.21	0.04
*04901	0.40	0.12	0.36	0.08	0.02	0.03	0.01

4.6.2. Dinaric-Balkan population

Observed contingency tables for Dinaric-Balkan population are table 44 for DQA1 and DQB1 loci, table 46 for DQA1 and DRB1 loci and table 48 for DQB1 and DRB1 loci. Expected contingency tables for Dinaric-Balkan population are table 45 for DQA1 and DQB1 loci, table 47 for DQA1 and DRB1 loci and table 49 for DQB1 and DRB1 loci.

When testing the linkage of DQA1 and DQB1 loci in the Dinaric-Balkan population, we see that the expected number of two-loci haplotypes is at least 40 (table 45), while the observed number is 13 (table 44). Under the null hypothesis the p value is equal to 0.000, so we can reject the possibility of random association of alleles between the DQA1 and DQB1 loci.

Table 44. Observed contingency table for DQA1 and DQB1 loci in the sample from Serbia (N=79)

DQA1 \ \ DQB1	*035 01	*020 02	*004 01	*039 **	*029 01	*039 01	*038 01	*023 01	*023 05	*029 02	*049 01
*012011	32	0	0	0	0	0	0	0	0	0	0
*00601	0	44	0	0	0	0	0	0	1	0	1
*00301	0	0	24	0	0	0	14	0	0	0	0
*005011	0	0	0	6	0	16	0	0	0	0	0
*00201	0	0	0	0	13	0	0	0	0	2	0
*00401	0	3	0	0	0	0	0	0	0	0	0
*00701	0	0	0	0	0	0	1	0	0	0	0
*00402	0	0	0	0	0	0	0	1	0	0	0

Table 45. Expected contingency table for DQA1 and DQB1 loci in the sample from Serbia (N=79)

DQA1 \ \ DQB1	*035 01	*020 02	*00 401	*039 **	*029 01	*039 01	*038 01	*023 01	*023 05	*029 02	*049 01
*012011	6.48	9.52	4.86	1.22	2.63	3.24	3.04	0.20	0.20	0.41	0.20
*00601	9.32	13.68	6.99	1.75	3.78	4.66	4.37	0.29	0.29	0.58	0.29
*00301	7.70	11.30	5.77	1.44	3.13	3.85	3.61	0.24	0.24	0.48	0.24
*005011	4.46	6.54	3.34	0.84	1.81	2.23	2.09	0.14	0.14	0.28	0.14
*00201	3.04	4.46	2.28	0.57	1.23	1.52	1.42	0.09	0.09	0.19	0.09
*00401	0.61	0.89	0.46	0.11	0.25	0.30	0.28	0.02	0.02	0.04	0.02
*00701	0.20	0.30	0.15	0.04	0.08	0.10	0.09	0.01	0.01	0.01	0.01
*00402	0.20	0.30	0.15	0.04	0.08	0.10	0.09	0.01	0.01	0.01	0.01

When testing the linkage of DQA1 and DRB1 loci in the Dinaric-Balkan population, we see that the expected number of two-loci haplotypes is at least 42 (table 47), while the observed number is 16 (table 46). Under the null hypothesis the p value is equal to 0.000, so we can reject the possibility of random association of alleles between the DQA1 and DRB1 loci.

Table 46. Observed contingency table for DQA1 and DRB1 loci in the sample from Serbia (N=79)

DQA1 \ \ DRB1	*036 01	*090 012	*092 011	*054 01	*049 01	*032 02	*043 01	*092 012	*036 **	*015 01	*080 01	*032 01
*012011	22	8	0	0	0	0	0	0	2	0	0	0
*00601	0	0	43	0	0	0	0	2	0	0	1	0
*00301	0	0	0	24	0	0	14	0	0	0	0	0
*005011	0	0	0	0	22	0	0	0	0	0	0	0
*00201	2	0	0	0	0	11	0	0	0	0	0	2
*00401	0	0	2	0	0	0	0	1	0	0	0	0
*00701	0	0	0	0	0	0	1	0	0	0	0	0
*00402	0	0	0	0	0	0	0	0	0	1	0	0

Table 47. Expected contingency table for DQA1 and DRB1 loci in the sample from Serbia (N=79)

DQA1 \ \ DRB1	*036 01	*090 012	*092 011	*054 01	*049 01	*032 02	*043 01	*092 012	*036 **	*015 01	*080 01	*032 01
*012011	4.86	1.62	9.11	4.86	4.46	2.23	3.04	0.61	0.41	0.20	0.20	0.41
*00601	6.99	2.33	13.10	6.99	6.41	3.20	4.37	0.87	0.58	0.29	0.29	0.58
*00301	5.77	1.92	10.82	5.77	5.29	2.65	3.61	0.72	0.48	0.24	0.24	0.48
*005011	3.34	1.11	6.27	3.34	3.06	1.53	2.09	0.42	0.28	0.14	0.14	0.28
*00201	2.28	0.76	4.27	2.28	2.09	1.04	1.42	0.28	0.19	0.09	0.09	0.19
*00401	0.46	0.15	0.85	0.46	0.42	0.21	0.28	0.06	0.04	0.02	0.02	0.04
*00701	0.15	0.05	0.28	0.15	0.14	0.07	0.09	0.02	0.01	0.01	0.01	0.01
*00402	0.15	0.05	0.28	0.15	0.14	0.07	0.09	0.02	0.01	0.01	0.01	0.01

Finally, when testing the linkage of DQB1 and DRB1 loci in the Dinaric-Balkan population, we see that the expected number of two-loci haplotypes is at least 48 (table 49), while the observed number is 15 (table 48). Again, since the p value is equal to 0.000, we can reject the null hypothesis of random association of alleles between the DQB1 and DRB1 loci.

Table 48. Observed contingency table for DQB1 and DRB1 loci in the sample from Serbia (N=79)

DQB1 \ \ DRB1	*036 01	*090 012	*092 011	*054 01	*049 01	*032 02	*043 01	*092 012	*036 **	*015 01	*080 01	*032 01
*03501	22	8	0	0	0	0	0	0	2	0	0	0
*02002	0	0	44	0	0	0	0	3	0	0	0	0
*00401	0	0	0	24	0	0	0	0	0	0	0	0
*039**	0	0	0	0	6	0	0	0	0	0	0	0
*02901	2	0	0	0	0	11	0	0	0	0	0	0
*03901	0	0	0	0	16	0	0	0	0	0	0	0
*03801	0	0	0	0	0	0	15	0	0	0	0	0
*02301	0	0	0	0	0	0	0	0	0	1	0	0
*02305	0	0	0	0	0	0	0	0	0	0	1	0
*02902	0	0	0	0	0	0	0	0	0	0	0	2
*04901	0	0	1	0	0	0	0	0	0	0	0	0

Table 49. Expected contingency table for DQB1 and DRB1 loci in the sample from Serbia (N=79)

DQB1 \ \ DRB1	*036 01	*090 012	*092 011	*054 01	*049 01	*032 02	*043 01	*092 012	*036 **	*015 01	*080 01	*032 01
*03501	4.86	1.62	9.11	4.86	4.46	2.23	3.04	0.61	0.41	0.20	0.20	0.41
*02002	7.14	2.38	13.39	7.14	6.54	3.27	4.46	0.89	0.59	0.30	0.30	0.59
*00401	3.65	1.22	6.84	3.65	3.34	1.67	2.28	0.46	0.30	0.15	0.15	0.30
*039**	0.91	0.30	1.71	0.91	0.84	0.42	0.57	0.11	0.08	0.04	0.04	0.08
*02901	1.97	0.66	3.70	1.97	1.81	0.91	1.23	0.25	0.16	0.08	0.08	0.16
*03901	2.43	0.81	4.56	2.43	2.23	1.11	1.52	0.30	0.20	0.10	0.10	0.20
*03801	2.28	0.76	4.27	2.28	2.09	1.04	1.42	0.28	0.19	0.09	0.09	0.19
*02301	0.15	0.05	0.28	0.15	0.14	0.07	0.09	0.02	0.01	0.01	0.01	0.01
*02305	0.15	0.05	0.28	0.15	0.14	0.07	0.09	0.02	0.01	0.01	0.01	0.01
*02902	0.30	0.10	0.57	0.30	0.28	0.14	0.19	0.04	0.03	0.01	0.01	0.03
*04901	0.15	0.05	0.28	0.15	0.14	0.07	0.09	0.02	0.01	0.01	0.01	0.01

4.7. Comparison of DLA alleles from all explored European populations

The frequency distribution of all currently known alleles from the populations of European grey wolf is presented in tables 50 for the DQA loci, 51 for the DQB loci and 52 for the DRB loci. From the tables we can see which alleles can be found in which populations, and their frequencies in the examined samples. Data for Estonia, Latvia, Scandinavia and Finland and Russia from year 2004 are taken from Seddon and Ellegren (2004). Data for Finland and Russia from year 2014 are taken from Niskanen et al. (2014). Data for Croatia are taken from Arbanasić et al. (2013). Data for Italy are taken from Galaverni et al. (2013).

Phylogenetic trees were calculated for each loci using the sequences of all alleles represented in European populations of grey wolf (tables 50, 51 and 52). For each loci, the trees were calculated using three different methods. Figures 9, 10 and 11 show trees for DQA locus calculated using maximum parsimony, maximum likelihood and neighbor-joining method, respectfully. In the same way the data is presented for DQB locus in figures 12, 13 and 14, and for DRB locus in figures 15, 16 and 17.

For DQA locus the best model was calculated to be JC + G. Estimated value of the shape parameter for the discrete gamma distribution is 0.05.

For DQB locus the best model was calculated to be JC + G + I. Estimated value of the shape parameter for the discrete gamma distribution is 0.3297. The proportion of invariant sites is 64.591%.

For DRB locus the best model was calculated to be JC + G. Estimated value of the shape parameter for the discrete gamma distribution is 0.0516.

Table 50. Frequency distribution of DLA-DQA1 alleles

	NORTHERN POPULATIONS							SOUTHERN POPULATIONS			
	1 Scandinavian	2 Karelian				3 Baltic		5 Carpathian	6 Dinaric-Balkan		7, 8 Italian (peninsula + Alps)
Allele DQA1*	Scandinavia 2n = 180	Finland 2004 2n = 44	Russia 2004 2n = 102	Finland 2014 2n = 484	Russia 2014 2n = 74	Estonia 2n = 50	Latvia 2n = 30	Carpathian 2n = 198	Serbia 2n = 158	Croatia 2n = 154	Italy 2n = 148
005011	0.41	0.16	0.09	0.06	0.06	-	0.17	0.15	0.14	0.29	0.68
012011	0.01	0.20	0.25	0.29	0.27	0.50	0.23	0.36	0.20	0.10	0.14
00301	0.01	0.23	0.38	0.21	0.34	0.42	0.43	0.41	0.24	0.41	-
00201	-	0.02	-	0.12	0.13	-	-	-	0.09	0.03	0.04
01101	0.58	0.30	0.04	0.25	0.14	0.08	0.10	-	-	-	-
014012	-	0.09	0.23	0.04	0.01	-	0.03	-	-	-	-
01001	-	-	0.01	0.02	0.04	-	0.03	-	-	-	-
00601	-	-	-	-	-	-	-	0.08	0.29	0.14	0.04
00101	-	-	-	-	-	-	-	-	-	0.01	0.02
00401	-	-	-	-	-	-	-	-	0.02	-	0.07
00402	-	-	-	-	-	-	-	-	0.01	-	-
00701	-	-	-	-	-	-	-	-	0.01	-	-
00901	-	-	-	-	-	-	-	-	-	0.02	-

Table 51. Frequency distribution of DLA-DQB1 alleles

	NORTHERN POPULATIONS							SOUTHERN POPULATIONS			
	1 Scandinavian	2 Karelian				3 Baltic		5 Carpathian	6 Dinaric-Balkan		7, 8 Italian (peninsula + Alps)
Allele DQB1*	Scandinavia 2n = 180	Finland 2004 2n = 44	Russia 2004 2n = 102	Finland 2014 2n = 484	Russia 2014 2n = 74	Estonia 2n = 50	Latvia 2n = 30	Carpathian 2n = 198	Serbia 2n = 158	Croatia 2n = 154	Italy 2n = 148
03501	0.01	0.20	0.25	0.29	0.28	0.5	0.23	0.32	0.20	0.10	0.15
03901	0.41	0.14	0.06	0.06	0.04	-	0.07	0.15	0.10	0.01	0.37
00401	0.01	0.23	0.38	0.21	0.32	0.4	0.43	0.29	0.15	0.11	-
01303	-	0.02	0.01	0.03	0.03	-	0.10	-	-	-	0.06
02901	-	0.09	-	0.12	0.14	-	-	-	0.08	0.03	0.04
00701	-	-	0.02	1	-	-	-	-	-	0.28	0.31
03801	-	-	-	-	-	0.02	-	-	0.09	0.22	-
04001	0.58	0.23	0.04	0.23	0.12	0.08	0.10	-	-	-	-
04401	-	0.09	0.23	0.04	0.03	-	0.03	-	-	-	-
05601	-	-	0.01	0.02	0.04	-	0.03	-	-	-	-
04102	-	-	-	0.01	-	-	-	-	-	-	-
02002	-	-	-	-	-	-	-	0.07	0.30	0.14	0.01
04901	-	-	-	-	-	-	-	0.01	0.01	-	-
04101	-	-	-	-	-	-	-	0.12	-	0.08	-
03502	-	-	-	-	-	-	-	0.04	-	-	-
02301	-	-	-	-	-	-	-	-	0.01	0.01	-
02305	-	-	-	-	-	-	-	-	0.01	-	-
02902	-	-	-	-	-	-	-	-	0.02	-	-
039**	-	-	-	-	-	-	-	-	0.04	-	-
00101	-	-	-	-	-	-	-	-	-	0.02	-
00802	-	-	-	-	-	-	-	-	-	0.01	-
00201	-	-	-	-	-	-	-	-	-	-	0.02
00301	-	-	-	-	-	-	-	-	-	-	0.03

Table 52. Frequency distribution of DLA-DRB1 alleles

	NORTHERN POPULATIONS							SOUTHERN POPULATIONS			
	1 Scandinavian	2 Karelian				3 Baltic		5 Carpathian	6 Dinaric-Balkan		7, 8 Italian (peninsula + Alps)
Allele DRB1*	Scandinavia 2n = 180	Finland 2004 2n = 44	Russia 2004 2n = 102	Finland 2014 2n = 484	Russia 2014 2n = 74	Estonia 2n = 50	Latvia 2n = 30	Carpathian 2n = 198	Serbia 2n = 158	Croatia 2n = 154	Italy 2n = 148
03601	0.01	0.20	0.26	0.29	0.27	0.52	0.23	0.36	0.15	0.08	0.14
04901	0.41	0.14	0.03	0.06	0.04	-	0.07	0.03	0.14	0.01	-
05401	0.01	0.18	0.25	0.16	0.23	0.28	0.40	0.40	0.15	0.21	-
09901	-	-	0.01	0.02	0.07	0.02	-	0.02	-	-	-
04301	-	-	-	-	-	0.02	-	-	0.09	0.21	-
03101	0.58	0.23	0.04	0.23	0.13	0.08	0.10	-	-	-	-
02002	-	-	0.01	0.02	0.04	-	0.03	-	-	-	-
03801	-	0.02	0.01	0.01	0.01	-	0.10	-	-	-	-
05301	-	0.05	0.13	0.06	0.04	-	-	-	-	-	-
05601	-	0.09	0.23	0.04	0.03	-	0.03	-	-	-	-
02901	0.01	-	-	-	-	-	-	-	-	-	-
10001	-	0.07	-	0.06	0.11	-	-	-	-	-	-
10101	-	0.02	-	0.06	0.01	-	-	-	-	-	-
calu-drb1*13	-	-	0.03	-	-	-	-	-	-	-	-
03602	-	-	-	-	-	0.08	0.03	-	-	-	-
092011	-	-	-	-	-	-	-	0.08	0.28	0.07	0.01
12801	-	-	-	-	-	-	-	0.01	-	-	0.38

Table 52 - continued

	NORTHERN POPULATIONS							SOUTHERN POPULATIONS			
	1 Scandinavian	2 Karelian				3 Baltic		5 Carpathian	6 Dinaric-Balkan		7, 8 Italian (peninsula + Alps)
Allele DRB1*	Scandinavia 2n = 180	Finland 2004 2n = 44	Russia 2004 2n = 102	Finland 2014 2n = 484	Russia 2014 2n = 74	Estonia 2n = 50	Latvia 2n = 30	Carpathian 2n = 198	Serbia 2n = 158	Croatia 2n = 154	Italy 2n = 148
01501	-	-	-	-	-	-	-	-	0.01	0.02	0.03
03202	-	-	-	-	-	-	-	-	0.06	0.03	0.04
03701	-	-	-	-	-	-	-	-	-	0.28	0.03
04902	-	-	-	-	-	-	-	0.12	-	-	-
03201	-	-	-	-	-	-	-	-	0.02	-	-
036**	-	-	-	-	-	-	-	-	0.01	-	-
08001	-	-	-	-	-	-	-	-	0.01	-	-
090012	-	-	-	-	-	-	-	-	0.05	0.01	-
01502	-	-	-	-	-	-	-	-	-	0.01	-
01801	-	-	-	-	-	-	-	-	-	0.01	-
04302	-	-	-	-	-	-	-	-	-	0.01	-
092012	-	-	-	-	-	-	-	-	0.02	0.05	-
02001	-	-	-	-	-	-	-	-	-	-	0.04
092013	-	-	-	-	-	-	-	-	-	-	0.27
00101	-	-	-	-	-	-	-	-	-	-	0.02

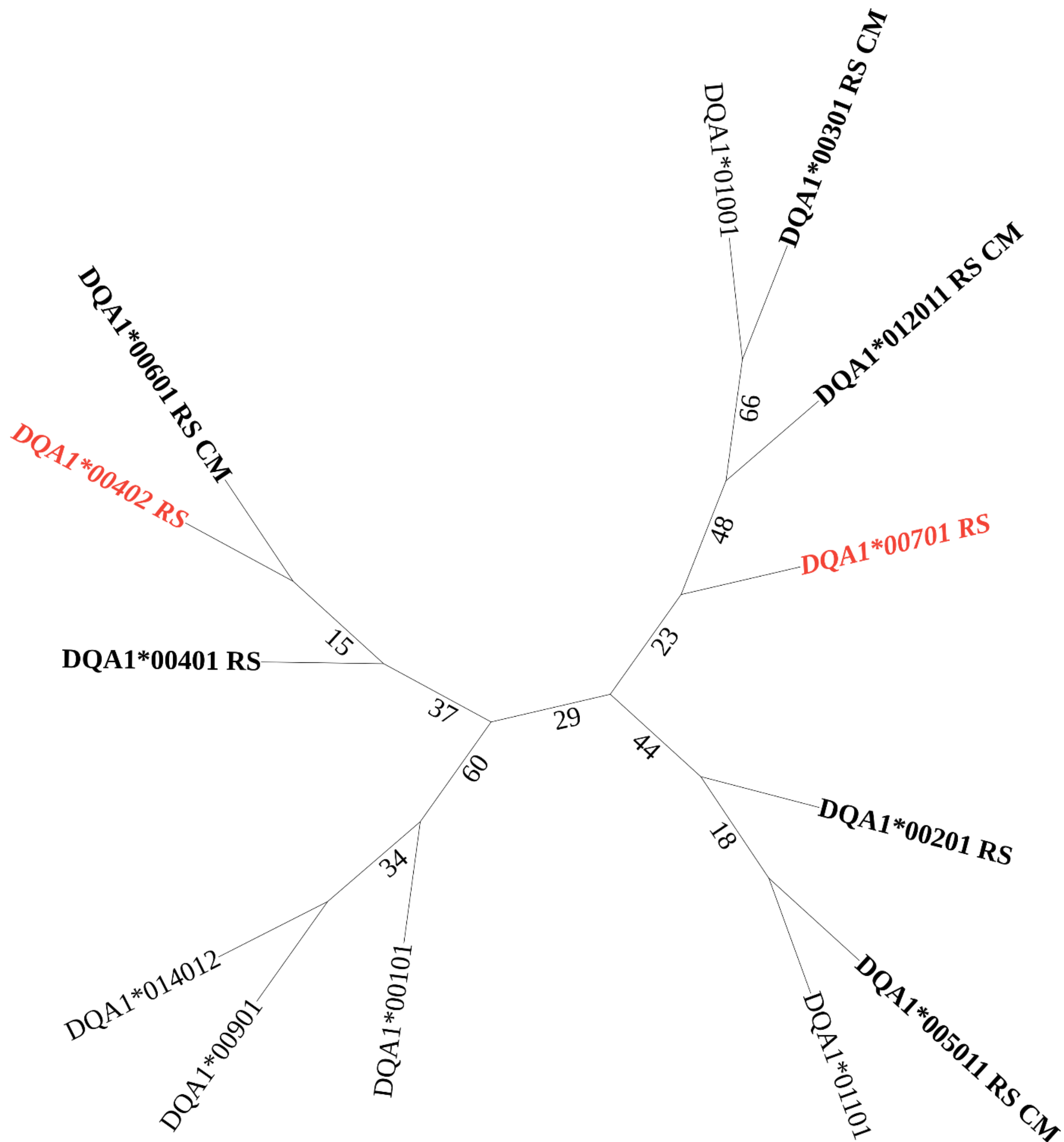

Figure 9. Maximum parsimony tree for DLA-DQA1 locus

In bold are alleles found in samples explored in this research; RS – allele found in the sample from Serbia; CM – allele found in the sample from Carpathian population;

In italic are alleles found for the first time in any European population of grey wolf (red colour).

Black numbers are bootstrap values scaled by a factor of 100.

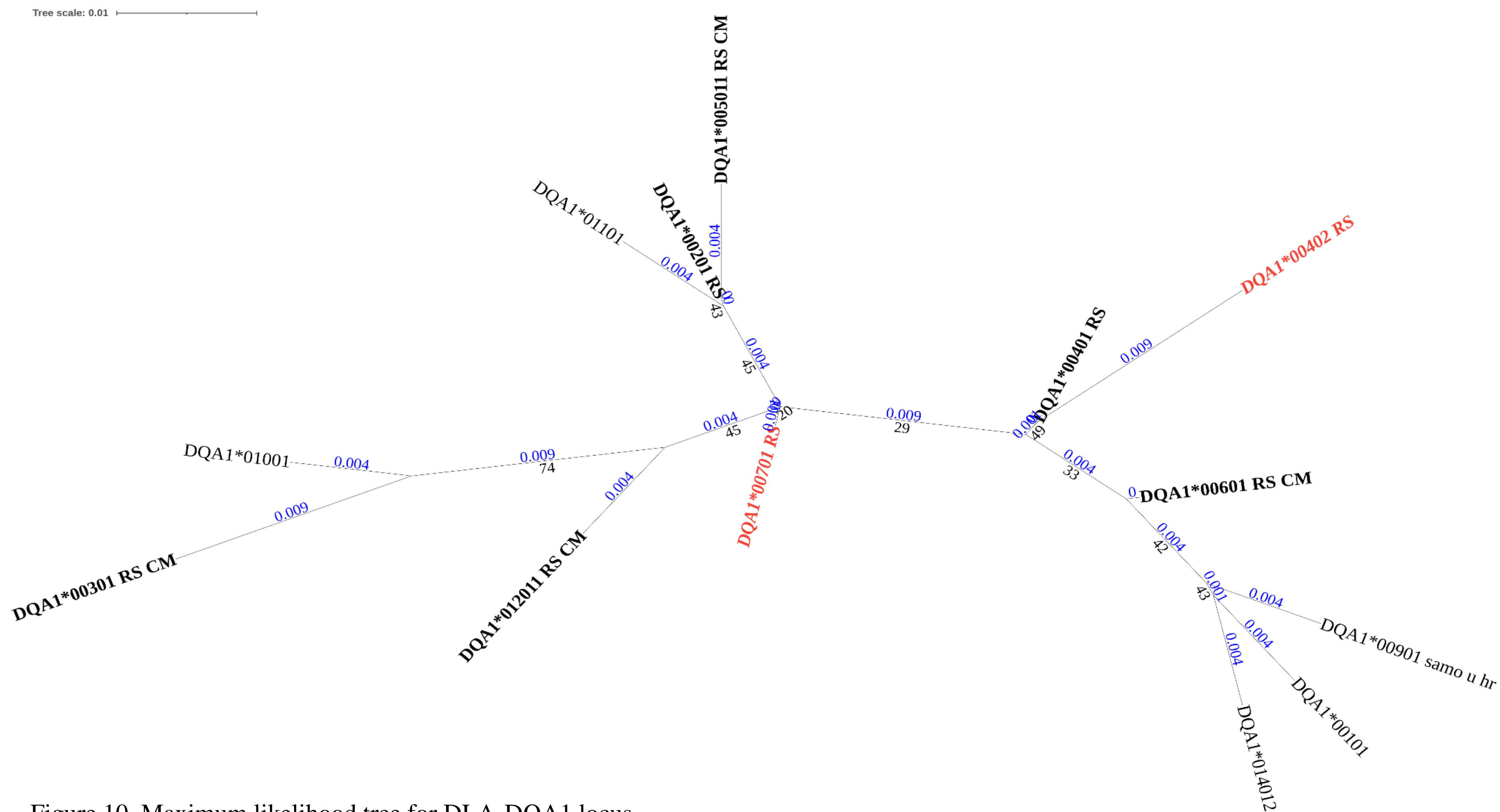

Figure 10. Maximum likelihood tree for DLA-DQA1 locus

In bold are alleles found in samples explored in this research; RS – allele found in the sample from Serbia; CM – allele found in the sample from Carpathian population. In italic are alleles found for the first time in any European population of grey wolf (red colour).

Black numbers are bootstrap values scaled by a factor of 100; blue numbers are branch lengths rounded to 3 decimals.

Figure 11. Neighbor-joining tree for DLA-DQA1 locus

In bold are alleles found in samples explored in this research; RS – allele found in the sample from Serbia; CM – allele found in the sample from Carpathian population. In italic are alleles found for the first time in any European population of grey wolf (red colour).

Black numbers are bootstrap values scaled by a factor of 100; blue numbers are branch lengths rounded to 3 decimals.

Figure 12. Maximum parsimony tree for DLA-DQB1 locus

In bold are alleles found in samples explored in this research; RS – allele found in the sample from Serbia; CM – allele found in the sample from Carpathian population;

In italic are alleles found for the first time in any European population of grey wolf (red colour).

Black numbers are bootstrap values scaled by a factor of 100.

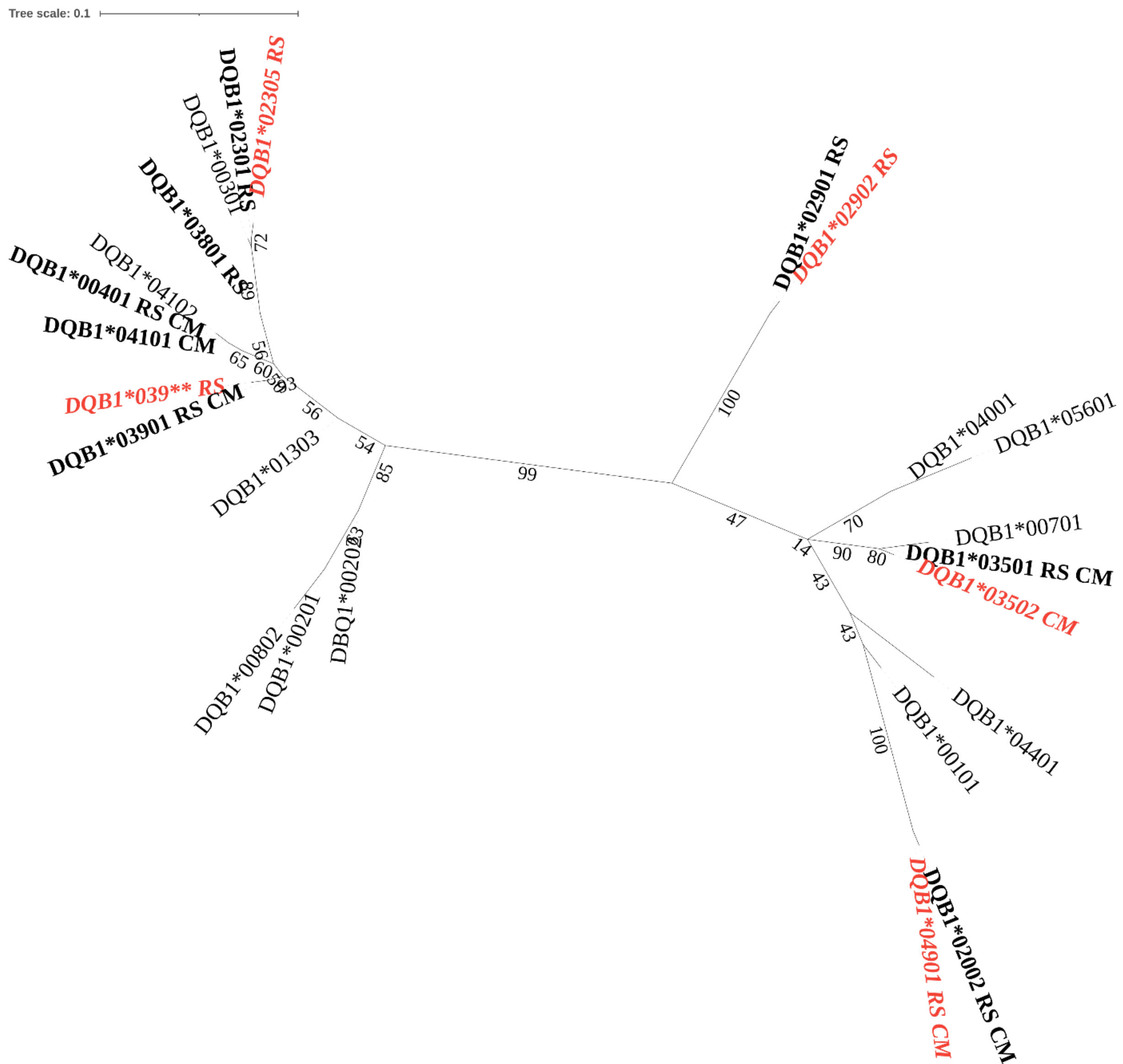

Figure 13. Maximum likelihood tree for DLA-DQB1 locus

In bold are alleles found in samples explored in this research; RS – allele found in the sample from Serbia; CM – allele found in the sample from Carpathian population.

In italic are alleles found for the first time in any European population of grey wolf (red colour).

Black numbers are bootstrap values scaled by a factor of 100.

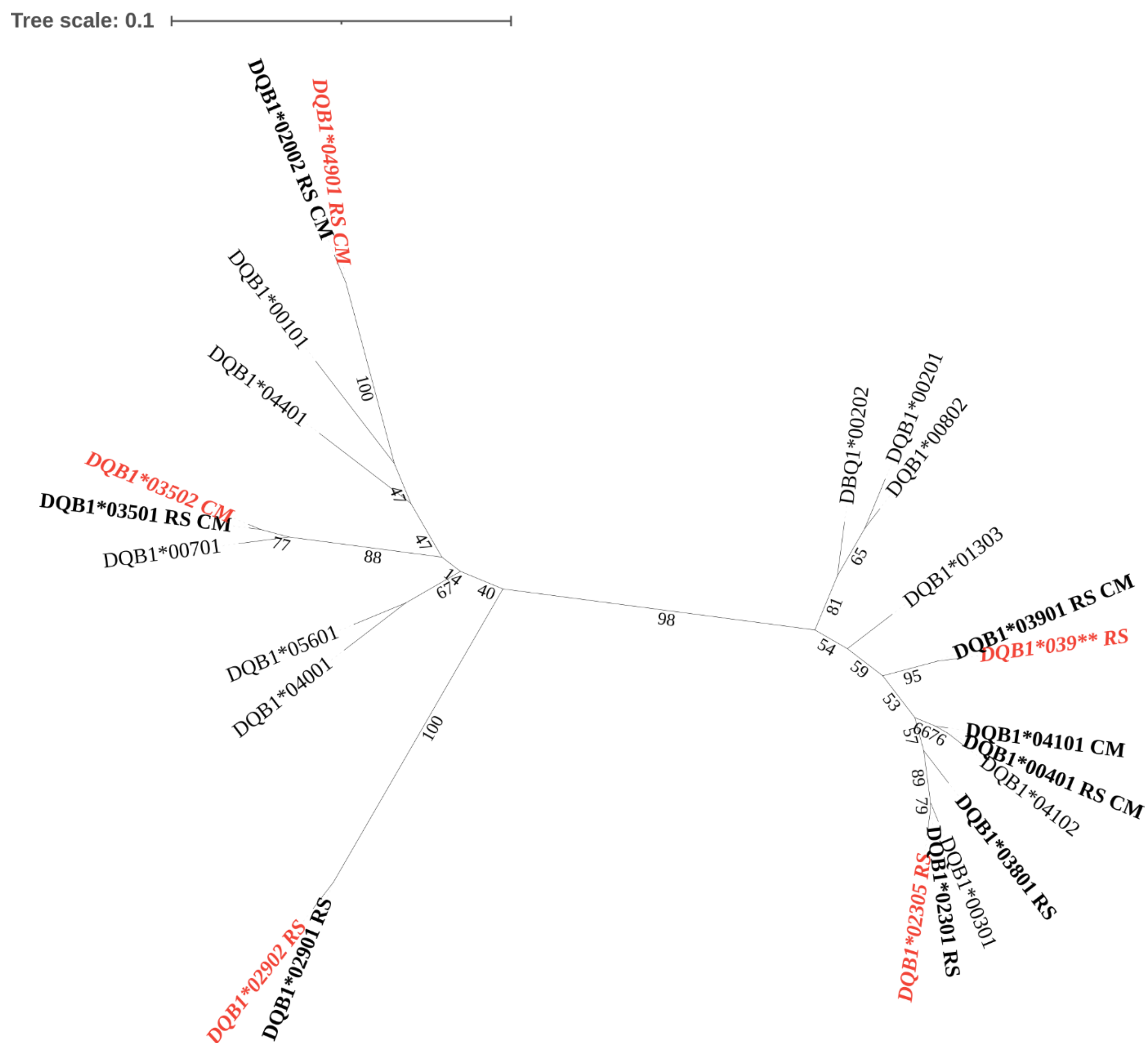

Figure 14. Neighbor-joining tree for DLA-DQB1 locus

In bold are alleles found in samples explored in this research; RS – allele found in the sample from Serbia; CM – allele found in the sample from Carpathian population.

In italic are alleles found for the first time in any European population of grey wolf (red colour).

Black numbers are bootstrap values scaled by a factor of 100.

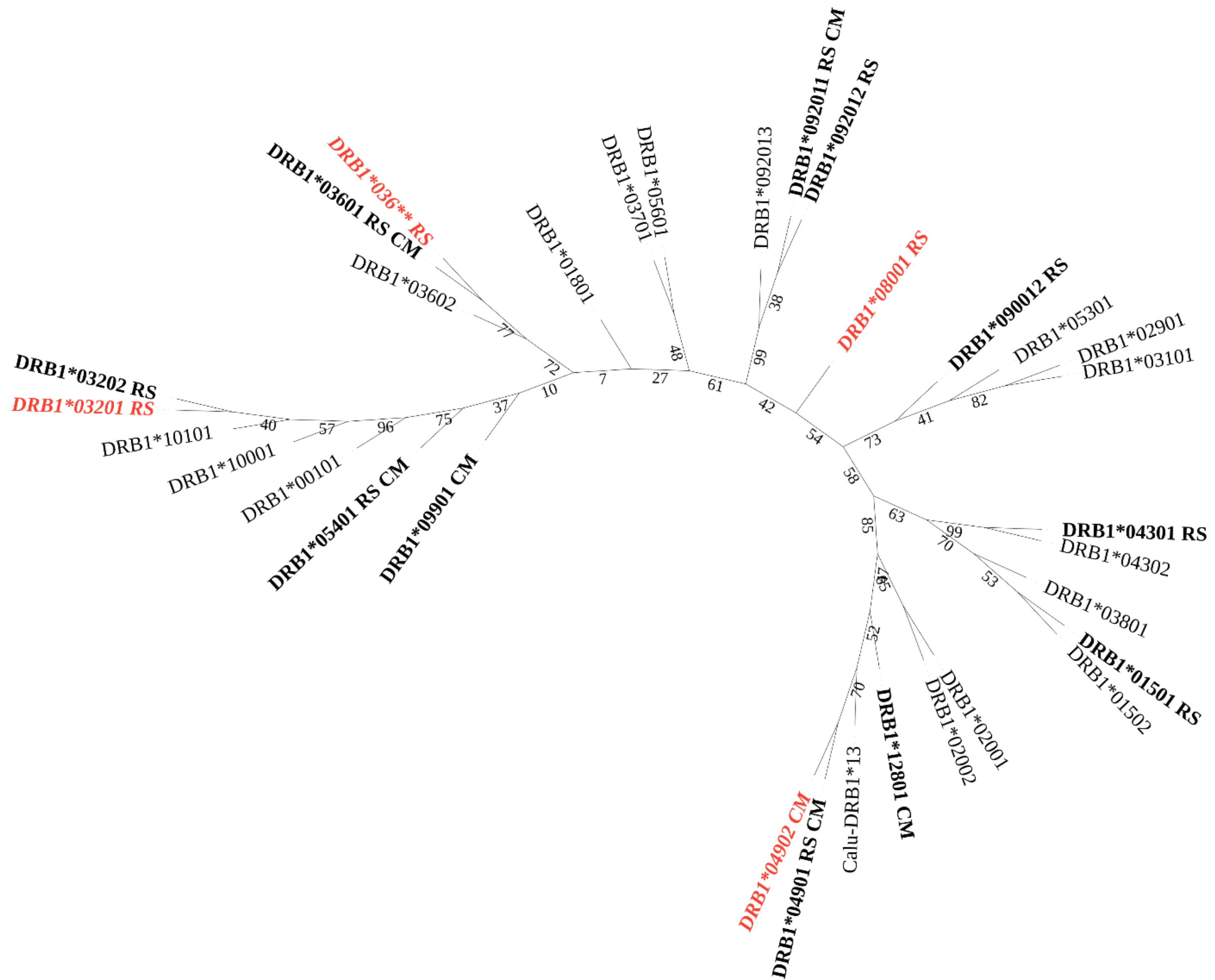

Figure 15. Maximum parsimony tree for DLA-DRB1 locus

In bold are alleles found in samples explored in this research; RS – allele found in the sample from Serbia; CM – allele found in the sample from Carpathian population;

In italic are alleles found for the first time in any European population of grey wolf (red colour).

Black numbers are bootstrap values scaled by a factor of 100.

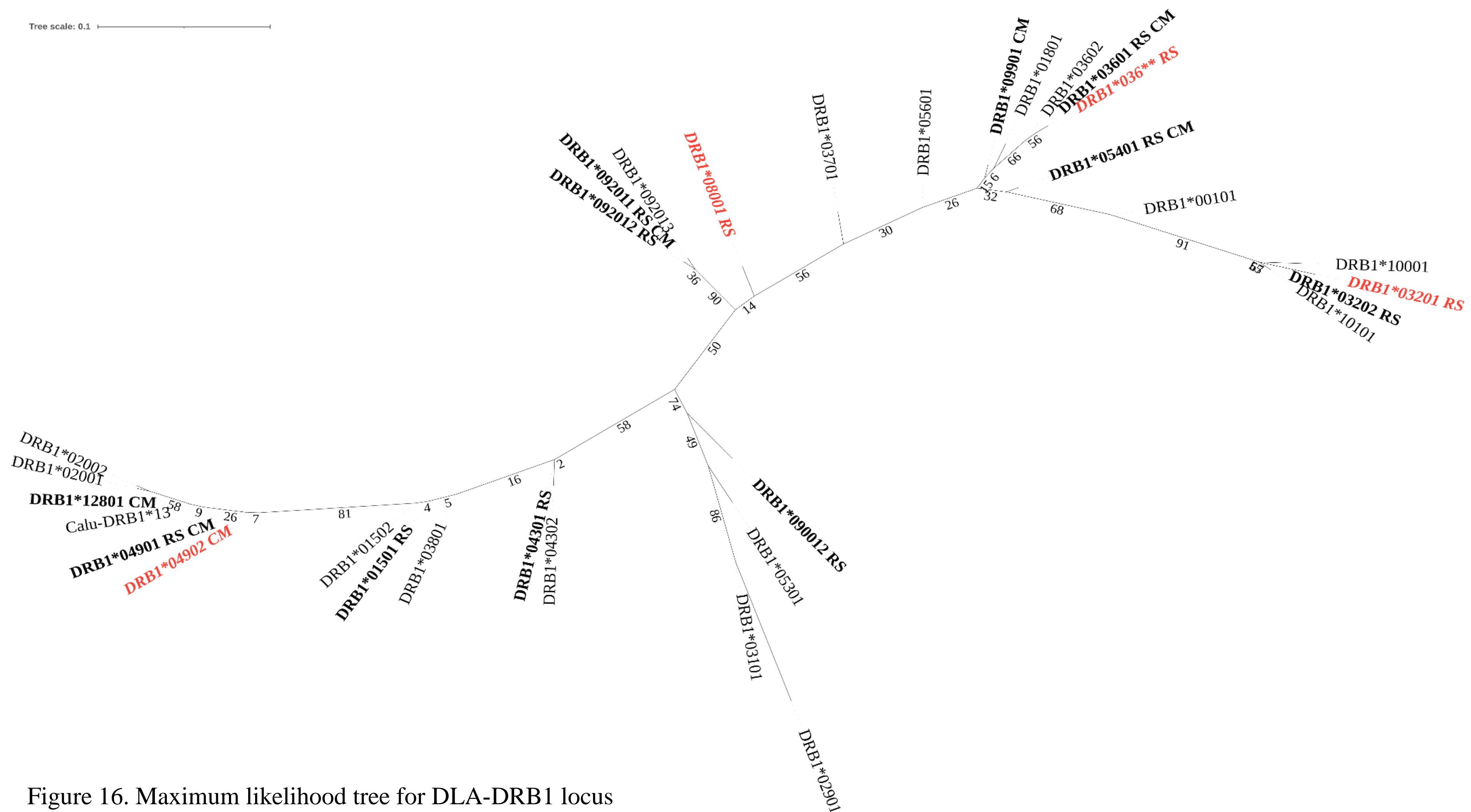

Figure 16. Maximum likelihood tree for DLA-DRB1 locus

In bold are alleles found in samples explored in this research; RS – allele found in the sample from Serbia; CM – allele found in the sample from Carpathian population. In italic are alleles found for the first time in any European population of grey wolf (red colour).

Black numbers are bootstrap values scaled by a factor of 100; blue numbers are branch lengths rounded to 3 decimals.

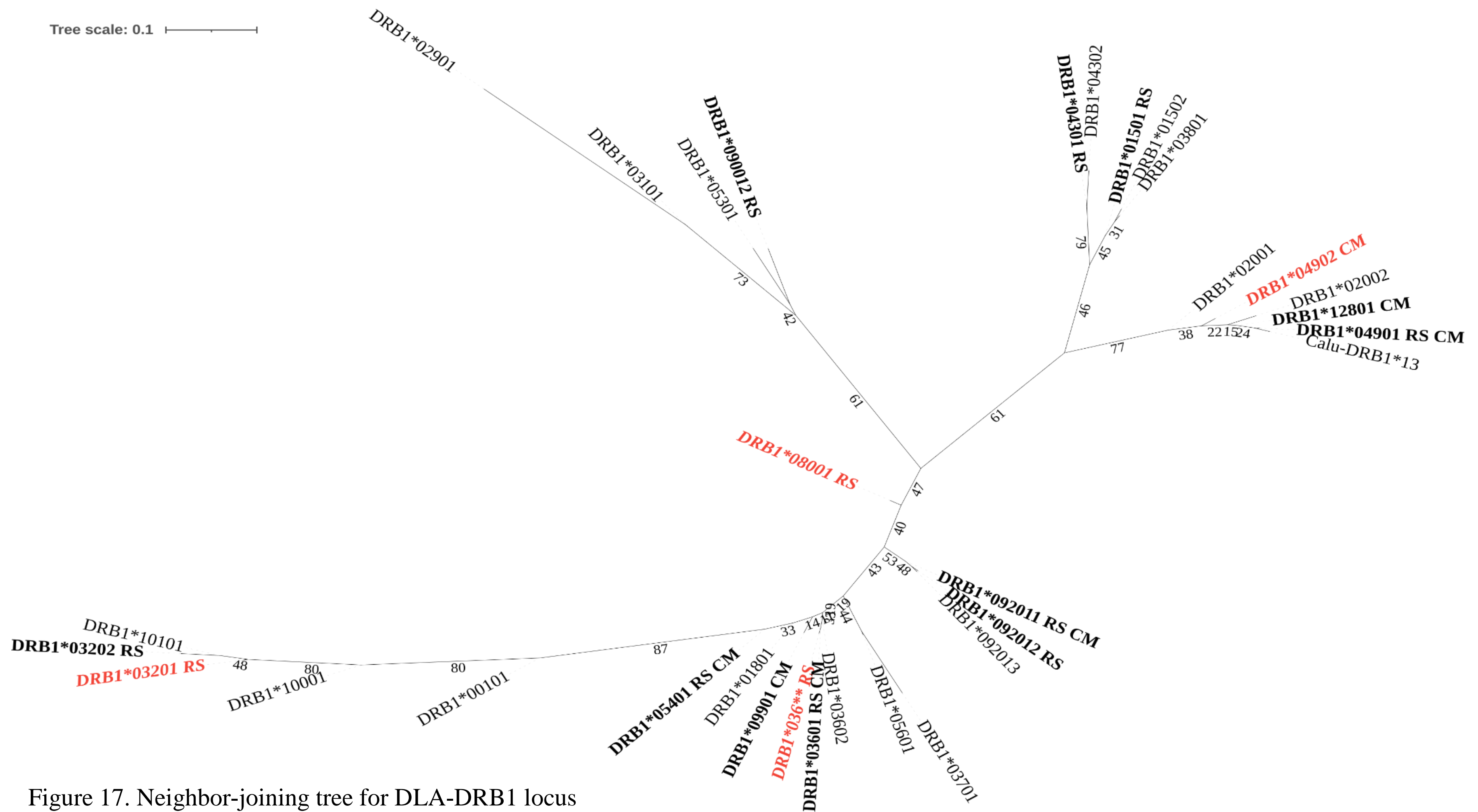

Figure 17. Neighbor-joining tree for DLA-DRB1 locus

In bold are alleles found in samples explored in this research; RS – allele found in the sample from Serbia; CM – allele found in the sample from Carpathian population. In italic are alleles found for the first time in any European population of grey wolf (red colour).

Black numbers are bootstrap values scaled by a factor of 100; blue numbers are branch lengths rounded to 3 decimals.

5. Discussion

In this thesis I have examined the adaptive genetic diversity of two populations of European grey wolf, using DLA as a molecular marker. Looking at the examined loci, there is a clear evidence of greater diversity in the beta chains of MHC molecules, than in the alpha chains. DQB1 locus has a greater number of unique amino acid sequences, seven in the Carpathian sample (table 4) and eleven in the Serbian sample (table 11), compared to DQA1 locus with four and eight sequences in the western Carpathian population (table 3) and Serbian subpopulation (table 10), respectively. Furthermore, the amino acid distances are several times greater on the DQB1 locus than on the DQA1 locus: 0.900 compared to 0.055 in the Carpathian population (table 17), and 0.245 compared to 0.040 in the Serbian subpopulation (table 18). Likewise, the total number of variable nucleotide positions is greater on the beta locus than on the alpha locus, with 33 compared to 10 variable positions in the Carpathian population (table 17) and 39 compared to 12 variable positions in the sample from Serbia (table 18).

When we compare the two explored populations, all of the results point to a greater diversity in the Dinaric-Balkan population samples from Serbia, than that of the Carpathian population. We can see that there are more alleles found in Dinaric-Balkan sample on all three loci: eight compared to four on DQA1 locus (tables 3, 6 and 61 vs tables 10, 13 and 65), eleven compared to seven on DQB1 locus (tables 4, 7, and 62 vs tables 11, 14 and 66) and twelve compared to seven on DRB1 locus (tables 5, 8 and 63 vs tables 12, 15 and 67). Likewise, on each locus there are more heterozygotes in the Dinaric-Balkan population than in the Carpathian population with observed heterozygosity ranging from 75.9% to 79.7% in the Dinaric-Balkan population (table 20), and from 62.6% to 69.7% in the Carpathian population (table 19). Also, there are two new alleles found in the Dinaric-Balkan sample, one on DQB1 locus (tables 11 and 66), and one on DRB1 locus (tables 12 and 67). Furthermore, there are more inferred haplotypes (18 vs 10, tables 9 and 64 vs tables 16 and 68) and the haplotype heterozygosity is higher in the Dinaric-Balkan sample (82.3%), compared to the Carpathian sample (72.7%). Given the fact that the Carpathian population sample consist of 99 wolves, and the Dinaric-Balkan population sample of 79 wolves, we can see that the adaptive genetic diversity is greater in the Dinaric-Balkan population. Contributing to the same conclusion is the fact that the number of variable positions in the sequences of the alleles across all three examined loci is larger in the sample from Serbia, than in the sample from the Carpathian (tables 17 and 18). Even though the diversity is grater in the Dinaric-Balkan population, one can still claim that the

Carpathian population shows a substantial functional diversity on the MHC loci. That is indicated by the nucleotide distances of the alleles across all three loci. On DRB1 locus the nucleotide distance is greater in the Dinaric-Balkan population (0.346), but in the Carpathian population it is still large (0.208). On DQA1 and DQB1 loci the nucleotide distances are larger in the Carpathian population than in the Dinaric-Balkan. This shows that even though there are less alleles in the Carpathian population there is a greater difference between them, than between the alleles in the Dinaric-Balkan population. Furthermore, considering that amino acid distances follow the same pattern, this is clear indication that substantial functional diversity in Carpathian population is still retained. This is especially true for the DQB1 loci on which the amino acid distance in the Carpathian population sample is 0.9 (table 17).

As stated previously, in the sample from Serbia I have found two previously undescribed DLA alleles, one on DQB1 locus, and one on DRB1 locus, with neither of them being the least common allele. Allele DQB1*039** present exclusively in the examined sample from Dinaric-Balkan population, compared to allele DQB1*03901 present in almost all described European populations (table 51), changes 42nd codon (table 66) of the second exon of the DQB1 gene from TAC coding for tyrosine to TTC coding for phenylalanine (table 11). Tyrosine, compared to phenylalanine, contains an extra hydroxyl group that can form hydrogen bonds, which makes it an important functional group of some enzymes (Nelson & Cox, 2005). This structural change from tyrosine to phenylalanine could affect the binding site of the MHC molecule for which allele DQB1*039** codes, indicating a local change in the selection pressure. Likewise, new allele DRB1*036**, compared to allele DRB1*03601 present in all described European populations (table 52), changes 58th codon (table 67), from CGG to GGG, changing the amino acid in the produced protein from arginine to glycine (table 12). This change is chemically even bigger than in the new DQB allele, since arginine compared to glycine contains a large and positively charged R group. The described change from arginine to glycine would make any potential protein less hydrophilic, but possibly sterically less constricted, and so would have a large impact on any binding interactions. Again, this indicates a local change in the selection pressure.

Results of the Hardy-Weinberg equilibrium test, taken at face value, would suggest that the Dinaric-Balkan population is evolving neutrally, while the western Carpathian population is not, at least on some loci. But, here we have to be very careful with the interpretation, for several reasons. The fact that there is no statistical deviation from the Hardy-Weinberg

expectations is not the proof that there is no selection, or even that all of the assumptions of the theorem are true. Balancing selection usually produces allelic frequency distributions that conform to Hardy-Weinberg expectations, since it maintains allelic diversity in a population (Andrews, 2010). Also, the fact that the population does not conform to the Hardy-Weinberg expectations means only that one or some of the assumptions of the theorem are violated, not necessarily that the population is under selection. For example one of the assumptions is that the populations is large enough to be considered infinite, and one could claim that regarding wolves this is not the case. To further complicate this analysis, there is some evidence that mate choice in vertebrates is non-random, but influenced by MHC, usually in such a way that females select males that produce offspring with greater genetic diversity. This was documented in fish (Neff et al., 2008), reptiles (Han et al., 2019), birds (Rymešová et al., 2017) and mammals (Dandine-Roulland et al., 2019), but is still an issue of some debate (Cretu-Stancu et al., 2018; Promerová et al., 2017). If true, this would violate the assumption of non-random mating of the Hardy-Weinberg theorem.

One strong indicator of positive selection is an excess of nonsynonymous mutations compared to synonymous mutations in the protein coding regions (Yang, 1998). More precisely, if the rates of nonsynonymous mutations (dN) exceed the rates of synonymous mutations (dS) across a coding sequence, we can assume that the locus is under positive selection (Nei & Kumar, 2000). This excess exists in both of the population samples I explored, across all three loci (tables 21 and 22). The dN/dS ratio is smallest on the DRB1 locus, but still significantly larger than one, with 2.42 in the sample from Serbia, and 2.18 in the sample from the Carpathian. On the other two loci it is notably larger with between four and a half and five and a half times greater dN than dS, so one cannot dismiss the possibility of non-neutral evolution. To test this in more detail, I performed the codon based Z-test of selection, testing the null hypothesis of neutral evolution, or dN = dS, against three alternative hypotheses: non neutral evolution, or dN ≠ dS, evolution under positive selection, or dN > dS and evolution under purifying selection, or dN < dS. From the results (table 23) we can clearly see that the null hypothesis is easily dismissed in favour of either the non-neutral, or positive selection hypotheses, at least on DQB1 and DRB1 loci, while the purifying selection is unlikely on all three loci. Out of the two likely hypotheses (non-neutral and positive selection), a stronger case can be made for the one including positive selection, since the p values are lower on all tested loci in both populations. Still, the dN/dS relationship alone is not enough to accept the hypothesis that the examined loci evolve under positive selection. One aspect I am missing to

improve this analysis is time (Mugal et al., 2014), since there are no available data regarding the MHC alleles of the previous generations of these populations. Second aspect is the lack of demographic data, specifically if and how the effective population size changed in the past several generations. Without these data it is impossible to dismiss the possibility that the effective population size is decreasing, in which case the increase in dN/dS ratio can be attributed to reduced efficiency of purifying selection, *i.e.* that in small populations the genetic drift has a much greater effect than selection on mildly deleterious mutations (Ohta, 1992). Another issue with the Z-test of selection is that it averages the dN/dS ratio over the entire sequence, which can mask the evidence of selection on specific sites. Selection pressure cannot be the same on all sites in a protein, since different structural and functional roles are performed by different sites (Nielsen & Yang, 1998). For these reasons I have included two more analyses to test the type of evolution on the examined loci.

Tajima's test of neutrality (Tajima, 1989) clearly shows the excess of nucleotide diversity in relation to the number of segregating sites per nucleotide site, so in both populations across all three loci the Tajima test statistic has high positive values (D>2) (table 24). This again indicates either that selection is maintaining variation in these populations, which is a characteristic of balancing selection, or a recent contraction of both populations. One drawback of Tajima's test is the assumption that all nucleotides in a sequence have the same mutation rate, and are equally subjected to the potential selection pressure. Given the variation in polymorphism on the first, second and third codon base, and the codon usage bias we know that the mutation pattern is more complicated. Site model tests included in CodeML analyses account for the changing substitution rates among different codons in the sequence (Yang & Nielsen, 2002). The M0 vs. M3 test, while not being an indicator of positive selection, shows that there is a variability in the dN/dS ratio among sites in all tested cases with the difference in likelihood (2 Δ Ln L) ranging from 11.313 on DQA1 locus in the Carpathian population (table 25) to 109.079 on DRB1 locus in the Dinaric-Balkan population (table 30). The rest of the likelihood ratio tests performed, clearly indicate the presence of positive selection on the explored loci in both populations with the likelihood differences always favouring models which include the positive value of dN-dS. Also, from the likelihood ratios it appears that the selection is strongest on the DQB1 loci, where 2 Δ Ln L of M7 vs. M8 test is highest (35.086, table 26, and 33.493, table 29, for the Carpathian and Dinaric-Balkan population, respectively) and weakest on the DQA1 loci, where the same values are lowest (12.557, table 25, and 14.112, table 28, for the Carpathian and Dinaric-Balkan population, respectively).

Two different analyses were used to identify the codons that are possible targets of selection: the maximum likelihood approach developed by Muse and Gaut (1994) and Bayes empirical Bayes approach developed by Yang et al. (2005). As we can see from the p values in tables 31-36, the maximum likelihood analysis was expectedly unreliable, due to the small numbers of inferred synonymous and nonsynonymous substitutions (Nei & Kumar, 2000). We can compare the two methods locus by locus. First, on DQA1 locus in both populations, both methods found evidence of positive selection on the same codons, with high posterior probability (>95%) of Bayesian analysis (table 37) on those codons for which the dN-dS statistic also shows high values (> 0.7, tables 31 and 34). On DQB1 locus in Carpathian population, Bayesian analysis (table 37) identified one codon more than the maximum likelihood analysis (table 32). We again see the correlation of high posterior probability and dN-dS test statistic, with two exceptions: for codon 23 the dN-dS test statistic equals zero, while on the same codon the Bayesian analysis gives high posterior probability of positive selection (0.988), and for codon 66 the dN-dS statistic is small relative to the rest of the identified codons (0.226), but the posterior probability is 99.9%. On DQB1 locus in the Dinaric-Balkan population we can see a larger difference between the two analyses. All accept one codons identified by Bayesian analysis (table 37) also have positive dN-dS statistic (table 35), though the correlation with the posterior probability is not as clear as on the other loci. For example on codon 58 the dN-dS test statistic is higher than one (1.007), but the posterior probability is 0.913, while codon 62 has higher posterior probability of 0.999 but lower dN-dS of 0.916. Here again we have a discrepancy in two analyses regarding codon 23; while maximum likelihood analysis calculates that the dN-dS statistic equals -1, which would suggest purifying selection, the Bayesian analysis assigns a posterior probability of 0.982 of positive selection. Furthermore, codons: 55, 61, 63 and 64 all have positive values of dN-dS, but are undetected as positively selected codons using Bayes empirical Bayes approach, which would imply they could be false positives. The biggest differences between the two analyses are observed on DRB1 locus in the Carpathian population sample (tables 33 and 37). While only three codons: 4, 5 and 69 with positive dN-dS statistic, are undetected by Bayesian analysis, the correlation between the higher posterior probability and higher dN-dS is even looser than on the previously described locus. For example codon 73 is assigned 0.99 posterior probability of being positively selected, while it's dN-dS equals only 0.066. Even grater difference is on codons 52 and 81, which have negative dN-dS values of -0.535 and -2.034, respectively, but are both assigned high posterior probabilities of being under positive selection, of 99.1% and 100%, respectively. In the other explored population, on the DRB1 locus there are five codons with major differences between

the two analyses (tables 36 and 37). Codons 3, 33 and 85 have positive dN-dS values, but no positive selection on these codons is detected by the Bayesian analysis. Opposite that, codon 51 has been assigned a high posterior probability (95.1%) of being under positive selection by the Bayesian analysis, while the maximum likelihood approach calculates the dN-dS value of zero. A larger difference is observed on codons: 23, 52 and 81, which all have negative dN-dS values of -1.326, -0.763 and -1.885, respectively, but by the Bayesian analysis all are assigned high posterior probabilities of 99.4%, 99.1% and 99.9%, respectfully. Still, one can see that the final results are similar between the two methods. Especially, most of the codons with high posterior probabilities of being under positive selection, as calculated by the Bayes empirical Bayes approach, also have high dN-dS test statistic, as calculated by the maximum likelihood approach. So, those are clear indications that these codons evolve influenced by positive selection.

Looking through the results of all the analyses I performed to test the null hypothesis of neutrality on the examined loci, we can see one common theme. In both populations it appears that the selection is strongest on the DQB1 locus, slightly weaker on DRB1 locus, and a lot weaker on the DQA1 locus. Because of this reason, and since I was exploring loci closely related both in function and in their position, it was interesting to see if there is any evidence of linkage disequilibrium. It turns out that the linkage is so strong that not only is the hypothesis of random association between alleles rejected with p values lower than 0.000 in all three combinations of loci, but from the results seen in observed and expected contingency tables (tables 38 to 49) one can easily and intuitively see that the pairing of alleles is not random. Considering this result, I am faced with a dilemma which I unfortunately cannot answer without more data. Is the evidence that points to weak selection on DQA1 locus result of strong selection on DQB1 and DRB1 loci and their strong linkage, or is the linkage increased because of the selection on all loci? As Slatkin (2008) beautifully states: “If alleles at two loci are in linkage disequilibrium and they both affect reproductive fitness, the response to selection on one locus might be accelerated or impeded by selection affecting the other.”, but we also know that if fitnesses of two loci are more than multiplicative selection can increase linkage disequilibrium (Felsenstein, 1965; Slatkin, 2008).

When analysing the allelic distribution of MHC genes among the different populations of European grey wolf, we have to be careful in making any comparisons between Scandinavian population and the rest of studied populations. As we can see from tables 50, 51 and 52

Scandinavian population is composed almost entirely of wolves carrying the same two alleles (DRB1*03101 and *04910; DQA1*00501 and *01101; DQB1*03901 and *04001). This population was founded in the early 1980s from a single mating pair, and is a great example in molecular ecology of the consequences of a strong founder effect (Seddon & Ellegren, 2004). The authors also explain that the other seven alleles, numbering in one to two copies in the entire population, originate from unsuccessful immigrants. Comparing the rest of the populations we can make some interesting observations. Examining the number of alleles on each locus across different populations (tables 50, 51 and 52), we can conclude that the adaptive genetic diversity is highest in the Dinaric-Balkan population (Croatia and Serbia), with 10 alleles on DQA1 locus, 15 on DQB1 locus and 16 alleles on DRB1 locus. This is followed by the Karelian population (Finland and Russia), with seven, ten and twelve alleles on DQA1, DQB1 and DRB1 loci, respectively. Then the Baltic population (Estonia and Latvia) with six, eight and ten alleles on DQA1, DQB1 and DRB1 loci, respectively, closely followed by the Italian peninsula population with only one less allele on the DRB1 locus. Finally, the lowest diversity seems to be in the Carpathian population, with four, seven and seven alleles on DQA1, DQB1 and DRB1 loci, respectively. Here, we have to be careful and note that the sample from the Carpathian population included animals mainly from Slovakia and some from Czech Republic, but unfortunately none from Ukraine or Romania.

Across all three loci, some alleles are (almost) ubiquitous in the entire explored range (e.g. alleles DRB1*03601, DQB1*03501 and DQA1*005011 are shared across each studied grey wolf population, while some alleles are present in majority of studied populations, e.g. DLA-DRB1*05401), possibly maintained due to shared selection pressure, e.g. a common pathogen present in all populations. On DQA1 locus, alleles: DLA-DQA1*00301, DLA-DQA1*005011 and DLA-DQA1* 012011 are present in high frequencies throughout Europe, with subpopulation from Estonia the only one not containing allele DLA-DQA1*005011. On DQB1 locus, alleles DLA-DQB1*03501 and DLA-DQB1*03901 are found in all populations, with again only the Estonian subpopulation not containing the DLA-DQB1*03901 allele. The range of allele DLA-DQB1*00401 is also extensive, since it is found in all populations except the one from Italian peninsula. On DRB1 locus, allele DLA-DRB1*03601 can be found in all populations with frequencies somewhat higher in the north. Allele DLA-DRB1*05401 can be found in high frequencies in all populations except the Italian peninsula, while allele DLA-DRB1*04901 in all except the Italian peninsula and the Estonian subpopulation, but in highly varying frequencies (from 0.01 to 0.14).

Other alleles are private to certain regions or even populations, pointing to local changes in selection pressure. In particular, there is a clear division between north (Estonia, Latvia, Finland and Russia) and south (Carpathian, Serbia, Croatia and Italy) of Europe (tables 50, 51 and 52). On DQA1 locus alleles: DLA-DQA1*01001, DLA-DQA1*01101 and DLA-DQA1*014012 are found exclusively in the north populations, while alleles: DLA-DQA1*00101, DLA-DQA1*00401, DLA-DQA1*00402, DLA-DQA1*00601, DLA-DQA1*00701 and DLA-DQA1*00901 are found only in south populations. Similarly, on DQB1 locus alleles: DLA-DQB1*00101, DLA-DQB1*00201, DLA-DQB1*00301, DLA-DQB1*00802, DLA-DQB1*02002, DLA-DQB1*02301, DLA-DQB1*02305, DLA-DQB1*02902, DLA-DQB1*04101, DLA-DQB1*04901 and the newly discovered allele DLA-DQB1*039** are present only in south populations. On the other end, alleles: DLA-DQB1*04001, DLA-DQB1*04102, DLA-DQB1*04401 and DLA-DQB1*05601 are found only in north populations. On DRB1 locus, the north to south divergence in allelic distribution is again present, with alleles: DLA-DRB1*00101, DLA-DRB1*01501, DLA-DRB1*01502, DLA-DRB1*01801, DLA-DRB1*02001, DLA-DRB1*03201, DLA-DRB1*03202, DLA-DRB1*03701, DLA-DRB1*04302, DLA-DRB1*08001, DLA-DRB1*090012, DLA-DRB1*092011, DLA-DRB1*092012, DLA-DRB1*092013, DLA-DRB1*12801 and newly discovered DLA-DRB1*036** found only in south populations of Europe. Furthermore, alleles: DLA-DRB1*02002, DLA-DRB1*03101, DLA-DRB1*03602, DLA-DRB1*03801, DLA-DRB1*05301, DLA-DRB1*05601, DLA-DRB1*10001, DLA-DRB1*10101 and calu-drb1*13 were found only in north of Europe. Allele DLA-DRB1*09901, found mainly in north, is present also in the Carpathian population, possibly due to migration from Russia.

Alleles DLA-DQA1*00201, DLA-DQB1*01303 and DLA-DQB1*02901 are present sporadically through Europe. These usually have higher frequencies in one or two neighbouring populations and lower frequencies elsewhere. These could belong to the wolves that have migrated in the past, originating from the populations with higher frequencies of these "sporadic" alleles, and integrating themselves in other populations, thus contributing to the genetic diversity and increasing the overall fitness of grey wolf in Europe. Alleles DLA-DQB1*00701, DLA-DQB1*03801 and DLA-DRB1*04301 are predominantly found in south Europe, but are also present in low frequencies in Russia, Estonia and Estonia, respectively. The causes of this can be many, for example: a rare migration, a rare local shift in selection pressure, artificially caused by humans (either by introduction of a pathogen in a new environment or transport of wolves from different parts of the continent) or even random

mutation. Unfortunately, without more data, especially regarding individual wolves carrying these alleles, I cannot give a clear answer.

Another observation we can make is that there are only three private alleles in the northern populations. Allele DLA-DQB1*04102 and allele calu-drb1*13 are private to Finland and Russia, respectively, while allele DLA-DRB1*03602 is private to Baltic population (Finland and Latvia). Similarly, Carpathian population has only two private alleles, one on DQB1 and one on DRB1 locus. Compared to this, the Italian peninsula population has a wealth of five private alleles, two on DQB1 locus and three on DRB1 locus. This could be explained by the large agricultural areas surrounding the Po river in the northern Italy, which prevent the free migration form the peninsula to the other parts of Europe (Boitani, 2018). More extreme is the abundance of private alleles in the Dinaric-Balkan population, with a total of 17 private alleles. On DQA1 locus, the Dinaric-Balkan population has three private alleles, two in Serbian subpopulation and one in Croatian subpopulation. On DQB1 locus, there are six more private alleles, three of which are only found in Serbia, two only in Croatia, and only one in both subpopulations. DRB1 locus has the greatest number of private alleles, three in Croatian subpopulation, three in Serbian subpopulation and two in both subpopulations. This abundance of private alleles indicates that the wolf migrations from this area of Europe are not common, probably due to large, and mainly undisturbed forest areas still existing through the region. The fact that many of the alleles are private to only one subpopulation is pointing to strong sub structuring of Dinaric-Balkan population, caused mainly by geography and anthropogenic factors.

Using phylogenetic analysis I wanted to compare all of the DQA1, DQB1, and DRB1 alleles that can currently be found in the European population of grey wolf. Phylogenetic trees were calculated for each loci using tree different methods, with each of them having fundamentally different assumptions of what the "best" tree is: minimum number of evolutionary steps to display the tree (maximum parsimony, figures 9, 12 and 15) , highest probability of the sequences with a given model of evolution (maximum likelihood, figures 10, 13, and 16) , or minimizing the branch length of the tree at each stage of clustering of operational taxonomic units (neighbor-joining, figures 11, 14 and 17)(Saitou & Nei, 1987)(Saitou & Nei, 1987)(Saitou & Nei, 1987)(Saitou & Nei, 1987)(Saitou & Nei, 1987)(Saitou & Nei, 1987). No matter the method used to obtain the trees, they all lead us to the same conclusion. It appears that there is no obvious clustering of alleles based on their

sequence similarity and geographic origin. So, despite the fact that many of the alleles can be found exclusively in the north or south populations of Europe, which was described in detail above, the same alleles are not related to each other more closely than the alleles not found in the same populations. This also contributes to the conclusion that there is a high intrapopulation variation on the MHC loci. Another thing that is nicely visualized on all of the trees obtained by the likelihood and distance methods is that there is a large number of alleles inside a short evolutionary distance. As a consequence of this, some of the branches on the (best) calculated trees are statistically unreliable which can be seen in some of the low bootstrap values.

To get a global perspective on the level of adaptive genetic diversity of European grey wolf we can compare the number of alleles on each explored locus with the grey wolves from North America. Kennedy et al. (2007) examined the MHC diversity of six populations of grey wolves from Alaska and Canada, on a total of 175 wolves. Comparing their data with the data presented in tables 50, 51 and 52 we can see that the European grey wolf has a much greater MHC diversity than North American grey wolf. On DQA1 locus the number of alleles is closest, with 11 different alleles in wolves from North America and 13 in wolves from Europe. Eight of those alleles are the shared on both continents (DQA1*00101, DQA1*00201, DQA1*00301, DQA1*005011, DQA1*00601, DQA1*01001, DQA1*01101 and DQA1*012011), three alleles are private for North America (DQA1*01301, DQA1*014011 and DQA1*01701) and five alleles are private for Europe (DQA1*00401, DQA1*00402, DQA1*00701, DQA1*00901, DQA1*014012). On DQB1 locus the difference is more pronounced with wolves from North America having 15 different alleles, while the ones from Europe have 23. The differences between alleles are also greater, with only six shared alleles (DQB1*00701, DQB1*01303, DQB1*02002, DQB1*02901, DQB1*03501 and DQB1*04001), nine alleles private for North America (DQB1*008011, DQB1*01401, DQB1*03201, DQB1*03301, DQB1*03401, DQB1*03802, DQB1*04101, DQB1*04201 and DQB1*05501) and 17 private alleles for Europe (DQB1*00101, DQB1*00201, DQB1*00301, DQB1*00401, DQB1*00802, DQB1*02301, DQB1*02305, DQB1*02902, DQB1*03502, DQB1*03801, DQB1*03901, DQB1*039**, DQB1*04001, DQB1*04102, DQB1*04401, DQB1*04901 and DQB1*05601). The biggest difference in the number of different alleles is on the DRB1 locus where the wolves from North America have 17, while the wolves from Europe have 32. Most of the alleles on DRB1 locus are private, as was the case on DQB1 locus. Only eight alleles are shared between continents (DRB1*02901, DRB1*03101, DRB1*03202, DRB1*03601, DRB1*03701, DRB1*03801, DRB1*04902 and DRB1*092011), nine alleles are private for

North America (DRB1*00601, DRB1*00901, DRB1*03501, DRB1*04101, DRB1*04401, DRB1*04501, DRB1*06501, DRB1*09101 and DRB1*09301) and 24 alleles are private for Europe (DRB1*00101, DRB1*01501, DRB1*01502, DRB1*01801, DRB1*02001, DRB1*02002, DRB1*03201, DRB1*03602, DRB1*036**, DRB1*04301, DRB1*04302, DRB1*04901, DRB1*05301, DRB1*05401, DRB1*05601, DRB1*08001, DRB1*090012, DRB1*092012, DRB1*092013, DRB1*09901, DRB1*10001, DRB1*10101, DRB1*12801 and calu-drb1*13). If we assume that the balancing selection is the mechanism by which most of the MHC alleles originate and are maintained in populations, a hypothesis for which there are strong indications in both my thesis and the cited literature, the large number of private alleles from both continents suggest a large geographic differences in selection pressures. The most probable explanation for this is the difference in biogeography of pathogens, something that is well documented in humans (Murray et al., 2015).

Most of European and North American grey wolf populations today number enough individual animals so as not to be considered endangered, and consequently have plenty of genetic diversity. In stark contrast to them is the Mexican grey wolf subspecies (*C. l. baileyi*). Mexican grey wolves are thought to be extinct in the wild, with the extant wolves living in captivity (Hedrick et al., 2000). We can look at the published data on the MHC diversity of this subspecies to see the consequences of extirpation and near extermination of natural populations of wild mammals. In the samples of three separate lineages of Mexican grey wolf numbering 18, 10 and 8 individuals Hedrick et al. (2000) found two, one and three DRB1 alleles, respectively, with the total of five different alleles. Wolves suffer extreme exposure to pathogens, both in North America (Jara et al., 2016) and in Europe (Hodžić et al., 2020). Knowing that, if we compare the large difference in adaptive genetic diversity and the possible immune response to different pathogens between the Mexican wolf and the grey wolf (with 17 different DRB1 alleles in North American populations of grey wolf, or 32 different DRB1 alleles in the European grey wolf), it is easy to imagine the possible consequences of exposure of Mexican grey wolves to some of the pathogens co-existing with other wolves in the wild. To avoid this happening to the existing populations of European grey wolves, any management systems should consider not only their total numbers, but should also keep in mind to preserve their genetic diversity so as not to allow any single introduction of a novel pathogen to threaten their extinction.

6. Conclusions

Both examined populations show relatively high diversity of the major histocompatibility complex.

Dinaric-Balkan population shows more adaptive genetic diversity than western Carpathian population.

Two previously undescribed DLA alleles were found in the sample from Dinaric-Balkan population, one on DQB locus and one on DRB locus.

Observed diversity was higher on DQB1 locus, coding for beta chains of MHC molecules, than on DQA1 locus, coding for alpha chains.

There is a strong indication that the MHC loci in both studied populations evolve under positive selection.

There is an extreme linkage disequilibrium detected between all three examined loci.

There is an abundance of private alleles in the Dinaric-Balkan population, suggesting low migrations of wolves from this region.

There are private alleles in both the Croatian and Serbian subpopulations, indicating sub structuring of the Dinaric-Balkan population.

Dinaric-Balkan population seems to be the most diverse European population with regards to MHC genes.

There is a clear difference in the geographic distribution of certain alleles in Europe, especially between the north and south of Europe.

There is no obvious clustering of alleles based on their sequence similarity and geographical origin.

The adaptive genetic diversity is much greater in the European grey wolf than in the North American grey wolf (according to literature sources).

7. References

Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. *Molecular Biology of the Cell*; 5th Edition; Garland Science, **2008**; ISBN 0815341059.

Andersen, M.H.; Schrama, D.; thor Straten, P.; Becker, J.C. Cytotoxic T Cells. *Journal of Investigative Dermatology* **2006**, *126*, 32–41, https://doi.org/10.1038/sj.jid.5700001.

Andreis, I.; Batinić, D.; Čulo, F.; Grčević, D.; Lukinović-Škudar, V.; Marušić, M.; Taradi, M.; Višnjić, D. *Imunologija*; 7th edn.; Medicinska naklada, Zagreb, **2010**;

Andrews, C.A. The Hardy-Weinberg Principle. *Nature Education Knowledge* **2010**, *3(10)*, 65. Accessed from: https://www.nature.com/scitable/knowledge/library/the-hardy-weinberg-principle-13235724/

Arbanasić, H.; Huber, Đ.; Kusak, J.; Gomerčić, T.; Hrenović, J.; Galov, A. Extensive Polymorphism and Evidence of Selection Pressure on Major Histocompatibility Complex DLA-DRB1, DQA1 and DQB1 Class II Genes in Croatian Grey Wolves. *Tissue Antigens* **2013**, *81*, 19–27, https://doi.org/10.1111/tan.12029.

Arif, I.A.; Khan, H.A.; Bahkali, A.H.; Al Homaidan, A.A.; Al Farhan, A.H.; Al Sadoon, M.; Shobrak, M. DNA Marker Technology for Wildlife Conservation. *Saudi Journal of Biological Sciences* **2011**, *18*, 219–225, https://doi.org/10.1016/j.sjbs.2011.03.002.

Barber, R.M.; Schatzberg, S.J.; Corneveaux, J.J.; Allen, A.N.; Porter, B.F.; Pruzin, J.J.; Platt, S.R.; Kent, M.; Huentelman, M.J. Identification of Risk Loci for Necrotizing Meningoencephalitis in Pug Dogs. *Journal of Heredity* **2011**, *102*, S40-S46, https://doi.org/10.1093/jhered/esr048.

Boitani, L. Canis Lupus (Errata Version Published in **2019**). *The IUCN Red List of Threatened Species* 2018, e.T3746A144226239 (accessed on 7 October 2020). Available online: https://www.iucnredlist.org/species/3746/144226239.

Boitani, L.; Phillips, M.; Jhala, Y. Canis Lupus (Errata Version Published in **2020**). *The IUCN Red List of Threatened Species* 2018, e.T3746A163508960 (accessed on 3 March 2021). dx.doi.org/10.2305/IUCN.UK.2018-2.RLTS.T3746A163508960.en

Britannica, The Editors of Encyclopaedia. Major Histocompatibility Complex. *Encyclopedia Britannica* **2018**, https://www.britannica.com/science/major-histocompatibility-complex.

Bruford, M.W.; Davies, N.; Dulloo, M.E.; Faith, D.P.; Walters, M. Monitoring Changes in Genetic Diversity. In: Walters, M., Scholes, R.J., (eds) *The GEO Handbook on Biodiversity Observation Networks*. Springer, Cham, **2017,** 107–128, ISBN: 978-3-319-27288-7.

Chapron, G.; Kaczensky, P.; Linnell, J.D.C.; Von Arx, M.; Huber, D.; Andrén, H.; López-Bao, J.V.; Adamec, M.; Álvares, F.; Anders, O.; et al. Recovery of Large Carnivores in Europe's Modern Human-Dominated Landscapes. *Science* **2014**, *346*, 1517–1519, https://doi.org/10.1126/science.1257553.

Charlesworth, B. Causes of Natural Variation in Fitness: Evidence from Studies of Drosophila Populations. *Proceedings of the National Academy of Sciences USA* **2015**, *112*, 1662–1669, https://doi.org/10.1073/pnas.1423275112.

Cretu-Stancu, M.; Kloosterman, W.P.; Pulit, S.L. No Evidence That Mate Choice in Humans Is Dependent on the MHC. *bioRxiv* **2018**, 339028, https://doi.org/10.1101/339028.

Cutrera, A.P.; Lacey, E.A. Trans-Species Polymorphism and Evidence of Selection on Class II MHC Loci in Tuco-Tucos (Rodentia: Ctenomyidae). *Immunogenetics* **2007**, *59*, 937–948, https://doi.org/10.1007/s00251-007-0261-3.

Dandine-Roulland, C.; Laurent, R.; Dall'Ara, I.; Toupance, B.; Chaix, R. Genomic Evidence for MHC Disassortative Mating in Humans. *Proceedings of the Royal Society B: Biological Sciences* **2019**, *286*, 20182664, https://doi.org/10.1098/rspb.2018.2664.

Dempster, A.P.; Laird, N.M.; Rubin, D.B. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society: Series B* **1977**, *39*, 1–38.

Djan, M.; Maletic, V.; Trbojević, I.; Popović, D.; Velićković, N.; Burazerović, J.; Ćirović, D. Genetic Diversity and Structuring of the Grey Wolf Population from the Central Balkans Based on Mitochondrial DNA Variation. *Mammalian Biology* **2014**, *79*, 277–282, https://doi.org/10.1016/j.mambio.2014.03.001.

Đan, M.; Šnjegota, D.; Veličković, N.; Stefanović, M.; Obreht Vidaković, D.; Ćirović, D. Genetic Variability and Population Structure of Grey Wolf (*Canis lupus*) in Serbia. *Russian Journal of Genetics* **2016**, *52*, 821–827, https://doi.org/10.1134/S1022795416080044.

Ejsmond, M.J.; Radwan, J.; Wilson, A.B. Sexual Selection and the Evolutionary Dynamics of the Major Histocompatibility Complex. *Proceedings of the Royal Society B: Biological Sciences* **2014**, *281*, 20141662, https://doi.org/10.1098/rspb.2014.1662.

Excoffier, L.; Laval, G.; Schneider, S. Arlequin (Version 3.0): An Integrated Software Package for Population Genetics Data Analysis. *Evolutionary Bioinformatics Online* **2007**, *1*, 47–50.

Felsenstein, J. Evolutionary Trees from DNA Sequences: A Maximum Likelihood Approach. *Journal of Molecular Evolution* **1981**, *17*, 368–376, https://doi.org/10.1007/BF01734359.

Felsenstein, J. The Effect of Linkage on Directional Selection. *Genetics* **1965**, *52*, 349–363.

Futuyma, D.J. Variation. In *Evolution*; 2nd edn.; Sinauer Associates, Inc.: Sunderland, MA, **2009**, 215–253 ISBN 978-0-87893-223-8.

Galaverni, M.; Caniglia, R.; Fabbri, E.; Lapalombella, S.; Randi, E. MHC Variability in an Isolated Wolf Population in Italy. *Journal of Heredity* **2013**, *104*, 601–612, https://doi.org/10.1093/jhered/est045.

Galov, A.; Fabbri, E.; Caniglia, R.; Arbanasić, H.; Lapalombella, S.; Florijančić, T.; Bošković, I.; Galaverni, M.; Randi, E. First Evidence of Hybridization between Golden Jackal (*Canis aureus*) and Domestic Dog (*Canis familiaris*) as Revealed by Genetic Markers. *Royal Society Open Science* **2015**, *2*, https://doi.org/10.1098/rsos.150450.

Gao, F.; Chen, C.; Arab, D.A.; Du, Z.; He, Y.; Ho, S.Y.W. EasyCodeML: A Visual Tool for Analysis of Selection Using CodeML. *Ecology and Evolution* **2019**, *9*, 3891–3898, https://doi.org/10.1002/ece3.5015.

Gonzalez-Galarza, F.F.; Takeshita, L.Y.C.; Santos, E.J.M.; Kempson, F.; Maia, M.H.T.; da Silva, A.L.S.; Teles e Silva, A.L.; Ghattaoraya, G.S.; Alfirevic, A.; Jones, A.R.; et al. Allele Frequency Net 2015 Update: New Features for HLA Epitopes, KIR and Disease and HLA Adverse Drug Reaction Associations. *Nucleic Acids Research* **2015**, *43*, D784-D788, https://doi.org/10.1093/nar/gku1166.

Gough, S.C.L.; Simmonds, M.J. The HLA Region and Autoimmune Disease: Associations and Mechanisms of Action. *Current Genomics* **2007**, *8*, 453–465, https://doi.org/10.2174/138920207783591690.

Guo, S.W.; Thompson, E.A. Performing the Exact Test of Hardy-Weinberg Proportion for Multiple Alleles. *Biometrics* **1992**, *48*, 361–372. https://doi.org/10.2307/2532296

Hall, T.A. BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **1999**, *41*, 95–98.

Han, Q.-H.; Sun, R.-N.; Yang, H.-Q.; Wang, Z.-W.; Wan, Q.-H.; Fang, S.-G. MHC Class I Diversity Predicts Non-Random Mating in Chinese Alligators (*Alligator sinensis*). *Heredity* **2019**, *122*, 809–818, https://doi.org/10.1038/s41437-018-0177-8.

Hardy, G.H. Mendelian Proportions in a Mixed Population. *Science* **1908**, *28*, 49-50, https://doi.org/10.1126/science.28.706.49.

Hedrick, P.W.; Lee, R.N.; Buchanan, C. Canine Parvovirus Enteritis, Canine Distemper, and Major Histocompatibility Complex Genetic Variation in Mexican Wolves. *Journal of Wildlife Diseases* **2003**, *39*, 909–913, https://doi.org/10.7589/0090-3558-39.4.909.

Hedrick, P.W.; Lee, R.N.; Parker, K.M. Major Histocompatibility Complex (MHC) Variation in the Endangered Mexican Wolf and Related Canids. *Heredity* **2000**, *85*, 617–624, https://doi.org/10.1046/j.1365-2540.2000.00805.x.

Heimeier, D.; Baker, C.S.; Russell, K.; Duignan, P.J.; Hutt, A.; Stone, G.S. Confirmed Expression of MHC Class I and Class II Genes in the New Zealand Endemic Hector's Dolphin (*Cephalorhynchus hectori*). *Marine Mammal Science* **2009**, *25*, 68–90, https://doi.org/10.1111/j.1748-7692.2008.00244.x.

Hill, A.V.S. HLA Associations with Malaria in Africa: Some Implications for MHC Evolution. In: Klein J., Klein D. (eds) *Molecular Evolution of the Major Histocompatibility Complex*. NATO ASI Series (Series H: Cell Biology), 59, Springer, Berlin, Heidelberg, **1991**, 403–420, https://doi.org/10.1007/978-3-642-84622-9_33.

Hodžić, A.; Georges, I.; Postl, M.; Duscher, G.G.; Jeschke, D.; Szentiks, C.A.; Ansorge, H.; Heddergott, M. Molecular Survey of Tick-Borne Pathogens Reveals a High Prevalence and Low Genetic Variability of Hepatozoon Canis in Free-Ranging Grey Wolves (*Canis lupus*) in Germany. *Ticks an Tick Borne Diseases* **2020**, *11*, 101389, https://doi.org/10.1016/j.ttbdis.2020.101389.

Huchard, E.; Cowlishaw, G.; Raymond, M.; Weill, M.; Knapp, L.A. Molecular Study of Mhc-DRB in Wild Chacma Baboons Reveals High Variability and Evidence for Trans-Species Inheritance. *Immunogenetics* **2006**, *58*, 805–816, https://doi.org/10.1007/s00251-006-0156-8.

Hughes, A.L.; Nei, M. Pattern of Nucleotide Substitution at Major Histocompatibility Complex Class I Loci Reveals Overdominant Selection. *Nature* **1988**, *335*, 167–170, https://doi.org/10.1038/335167a0.

Hulva, P.; Bolfíková, B.Č.; Woznicová, V.; Jindřichová, M.; Benešová, M.; Mysłajek, R.W.; Nowak, S.; Szewczyk, M.; Niedźwiecka, N.; Figura, M.; et al. Wolves at the Crossroad: Fission-Fusion Range Biogeography in the Western Carpathians and Central Europe. *Diversity and Distributions* **2018**, 24, 179–192, https://doi.org/10.1111/ddi.12676.

It, V.; Barrientos, L.; López Gappa, J.; Posik, D.; Díaz, S.; Golijow, C.; Giovambattista, G. Association of Canine Juvenile Generalized Demodicosis with the Dog Leukocyte Antigen System. *Tissue Antigens* **2010**, *76*, 67–70, https://doi.org/10.1111/j.1399-0039.2010.01463.x.

Janeway, C.A.J.; Travers, P.; Walport, M.; Shlomchik, M.J. The major histocompatibility complex and its functions. In *Immunobiology: The Immune System in Health and Disease*; 5th edn.; Garland Science: New York, **2001** ISBN 0-8153-3642-X.

Jara, R.F.; Wydeven, A.P.; Samuel, M.D. Gray Wolf Exposure to Emerging Vector-Borne Diseases in Wisconsin with Comparison to Domestic Dogs and Humans. *PLoS One* **2016**, *11*, e0165836–e0165836, https://doi.org/10.1371/journal.pone.0165836.

Jokinen, P.; Rusanen, E.M.; Kennedy, L.J.; Lohi, H. MHC Class II Risk Haplotype Associated with Canine Chronic Superficial Keratitis in German Shepherd Dogs. *Veterinary Immunology and Immunopathology* **2011**, *140*, 37–41, https://doi.org/10.1016/j.vetimm.2010.11.007.

Kelley, J.; Walter, L.; Trowsdale, J. Comparative Genomics of Major Histocompatibility Complexes. *Immunogenetics* **2005**, *56*, 683–695, https://doi.org/ 10.1007/s00251-004-0717-7

Kennedy, L.J.; Altet, L.; Angles, J.M.; Barnes, A.; Carter, S.D.; Francino, O.; Gerlach, J.A.; Happ, G.M.; Ollier, W.E.; Polvi, A.; et al. Nomenclature for Factors of the Dog Major Histocompatibility System (DLA), 1998. First Report of the ISAG DLA Nomenclature Committee. *Tissue Antigens* **1999**, *54*, 312–321, https://doi.org/10.1034/j.1399-0039.1999.540319.x.

Kennedy, L.J.; Angles, J.M.; Barnes, A.; Carmichael, L.E.; Radford, A.D.; Ollier, W.E.R.; Happ, G.M. DLA-DRB1, DQA1, and DQB1 Alleles and Haplotypes in North American Gray Wolves. *Journal of Heredity* **2007**, *98*, 491–499, https://doi.org/10.1093/jhered/esm051.

Kennedy, L.J.; Angles, J.M.; Barnes, A.; Carter, S.D.; Francino, O.; Gerlach, J.A.; Happ, G.M.; Ollier, W.E.R.; Thomson, W.; Wagner, J.L. Nomenclature for Factors of the Dog Major Histocompatibility System (DLA), 2000: Second Report of the ISAG DLA Nomenclature Committee. *Tissue Antigens* **2001**, *58*, 55–70.

Kennedy, L.J.; Barnes, A.; Happ, G.M.; Quinnell, R.J.; Bennett, D.; Angles, J.M.; Day, M.J.; Carmichael, N.; Innes, J.F.; Isherwood, D.; et al. Extensive Interbreed, but Minimal Intrabreed, Variation of DLA Class II Alleles and Haplotypes in Dogs. *Tissue Antigens* **2002**, *59*, 194–204, https://doi.org/10.1034/j.1399-0039.2002.590303.x.

Kennedy, L.J.; Huson, H.J.; Leonard, J.; Angles, J.M.; Fox, L.E.; Wojciechowski, J.W.; Yuncker, C.; Happ, G.M. Association of Hypothyroid Disease in Doberman Pinscher Dogs with a Rare Major Histocompatibility Complex DLA Class II Haplotype. *Tissue Antigens* **2006**, *67*, 53–56, https://doi.org/10.1111/j.1399-0039.2005.00518.x.

Kennedy, L.J.; Quarmby, S.; Fretwell, N.; Martin, A.J.; Jones, P.G.; Jones, C.A.; Ollier, W.E.R. High-Resolution Characterization of the Canine DLA-DRBI Locus Using Reference Strand-Mediated Conformational Analysis. *Journal of Heredity* **2005**, *96*, 836–842, https://doi.org/10.1093/jhered/esi112.

Kennedy, L.J.; Randall, D.A.; Knobel, D.; Brown, J.J.; Fooks, A.R.; Argaw, K.; Shiferaw, F.; Ollier, W.E.R.; Sillero-Zubiri, C.; Macdonald, D.W.; et al. Major Histocompatibility Complex Diversity in the Endangered Ethiopian Wolf (*Canis simensis*). *Tissue Antigens* **2011**, *77*, 118–125, https://doi.org/10.1111/j.1399-0039.2010.01591.x.

Kindt, T.J.; Goldsby, R.A.; Osborne, B.A.; Kuby, J. *Kuby Immunology*; 6th edn.; W.H. Freeman, New York, **2007**, ISBN: 978-1429202114

Klein, J. *Natural History of the Major Histocompatibility Complex*; 1st edn.; John Wiley & Sons: New York, **1986**;

Klein, J.; Sato, A.; Nagl, S.; O'hUigín, C. Molecular Trans-Species Polymorphism. *Annual Review of Ecology and Systematics* **1998**, *29*, 1-21, https://doi.org/10.1146/annurev.ecolsys.29.1.1

Klein, J.; Sato, A.; Nikolaidis, N. MHC, TSP, and the Origin of Species: From Immunogenetics to Evolutionary Genetics. *Annual Review of Genetics* **2007**, *41*, 281–304, https://doi.org/10.1146/annurev.genet.41.110306.130137.

Koenig, D.; Hagmann, J.; Li, R.; Bemm, F.; Slotte, T.; Neuffer, B.; Wright, S.I.; Weigel, D. Long-Term Balancing Selection Drives Evolution of Immunity Genes in *Capsella*. *eLife* **2019**, *8*, e43606, https://doi.org/10.7554/eLife.43606.

Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution* **2016**, *33*, 1870–4, https://doi.org/10.1093/molbev/msw054.

Kutal, M.; Belotti, E.; Volfová, J.; Mináriková, T.; Bufka, L.; Poledník, L.; Krojerová, J.; Bojda, M.; Váňa, M.; Kutalová, L.; et al. Occurrence of Large Carnivores – *Lynx lynx*, *Canis lupus*, and *Ursus arctos* – and of *Felis silvestris* in the Czech Republic and Western Slovakia in 2012–2016 (Carnivora). *Lynx, new series* **2017**, *48*, 93–107, https://doi.org//10.2478/lynx-2017-0006.

Letunic, I.; Bork, P. Interactive Tree Of Life (ITOL) v4: Recent Updates and New Developments. *Nucleic Acids Research* **2019**, *47*, W256–W259, https://doi.org/10.1093/nar/gkz239.

Lewontin, R.C.; Kojima, K. The Evolutionary Dynamics of Complex Polymorphisms. *Evolution* **1960**, *14*, 458–472, https://doi.org/10.2307/2405995.

Llaurens, V.; Whibley, A.; Joron, M. Genetic Architecture and Balancing Selection: The Life and Death of Differentiated Variants. *Molecular Ecology* **2017**, *26*, 2430–2448, https://doi.org/10.1111/mec.14051.

Loiseau, C.; Zoorob, R.; Garnier, S.; Birard, J.; Federici, P.; Julliard, R.; Sorci, G. Antagonistic Effects of a Mhc Class I Allele on Malaria-Infected House Sparrows. *Ecology Letters* **2008**, *11*, 258–265, https://doi.org/10.1111/j.1461-0248.2007.01141.x.

Lokki, M.-L.; Paakkanen, R. The Complexity and Diversity of Major Histocompatibility Complex Challenge Disease Association Studies. *HLA* **2019**, *93*, 3–15, https://doi.org/10.1111/tan.13429.

Maccari, G.; Robinson, J.; Ballingall, K.; Guethlein, L.A.; Grimholt, U.; Kaufman, J.; Ho, C.-S.; de Groot, N.G.; Flicek, P.; Bontrop, R.E.; et al. IPD-MHC 2.0: An Improved Inter-Species Database for the Study of the Major Histocompatibility Complex. *Nucleic Acids Research* **2017**, *45*, D860–D864, https://doi.org/10.1093/nar/gkw1050.

Masius, P.; Sprenger, J. *A Fairytale in Question: Historical Interactions between Humans and Wolves*; White Horse Press: Cambridge, **2015**; ISBN 978-1874267843.

Matzaraki, V.; Kumar, V.; Wijmenga, C.; Zhernakova, A. The MHC Locus and Genetic Susceptibility to Autoimmune and Infectious Diseases. *Genome Biology* **2017**, *18*, 76, https://doi.org/10.1186/s13059-017-1207-1.

Mérot, C.; Llaurens, V.; Normandeau, E.; Bernatchez, L.; Wellenreuther, M. Balancing Selection via Life-History Trade-Offs Maintains an Inversion Polymorphism in a Seaweed Fly. *Nature Communications* **2020**, *11*, 670, https://doi.org/10.1038/s41467-020-14479-7.

Meyer-Lucht, Y.; Otten, C.; Püttker, T.; Sommer, S. Selection, Diversity and Evolutionary Patterns of the MHC Class II DAB in Free-Ranging Neotropical Marsupials. *BMC Genetics* **2008**, *9*, 1–14, https://doi.org/10.1186/1471-2156-9-39.

Milenković, M.; Paunović, M.; Ćirović, D. *Action Plan for Wolf Canis Lupus L., 1758 Conservation in Serbia. Phase I - Strategic Plan.*; Project report, 2007;

Milinski, M. The Major Histocompatibility Complex, Sexual Selection, and Mate Choice. *Annual Review of Ecology, Evolution and Systematics* **2006**, *37*, 159–186.

Montgomery, R.A.; Tatapudi, V.S.; Leffell, M.S.; Zachary, A.A. HLA in Transplantation. *Nature Reviews Nephrology* **2018**, *14*, 558–570, https://doi.org/10.1038/s41581-018-0039-x.

Mugal, C.F.; Wolf, J.B.W.; Kaj, I. Why Time Matters: Codon Evolution and the Temporal Dynamics of dN/dS. *Molecular Biology and Evolution* **2014**, *31*, 212–231, https://doi.org/10.1093/molbev/mst192.

Murray, K.A.; Preston, N.; Allen, T.; Zambrana-Torrelio, C.; Hosseini, P.R.; Daszak, P. Global Biogeography of Human Infectious Diseases. *Proceedings of the. National Academy of Sciences USA* **2015**, *112*, 12746–12751, https://doi.org/10.1073/pnas.1507442112.

Muse, S. V; Gaut, B.S. A Likelihood Approach for Comparing Synonymous and Nonsynonymous Nucleotide Substitution Rates, with Application to the Chloroplast Genome. *Molecular Biology and Evolution* **1994**, *11*, 715–724, https://doi.org/10.1093/oxfordjournals.molbev.a040152.

Neff, B.D.; Garner, S.R.; Heath, J.W.; Heath, D.D. The MHC and Non-Random Mating in a Captive Population of Chinook Salmon. *Heredity* **2008**, *101*, 175–185, https://doi.org/10.1038/hdy.2008.43.

Nei, M.; Kumar, S. *Molecular Evolution and Phylogenetics*; Oxford University Press: New York, **2000**, ISBN 9780195135855.

Nelson, D.L.; Cox, M.M. *Lehninger Principles of Biochemistry*; 4th ed.; W.H. Freeman and Company: New York, 2005; ISBN 0-7167-4339-6.

Nielsen, R.; Yang, Z. Likelihood Models for Detecting Positively Selected Amino Acid Sites and Applications to the HIV-1 Envelope Gene. *Genetics* **1998**, *148*, 929–936.

Niskanen, A.K.; Kennedy, L.J.; Ruokonen, M.; Kojola, I.; Lohi, H.; Isomursu, M.; Jansson, E.; Pyhäjärvi, T.; Aspi, J. Balancing Selection and Heterozygote Advantage in Major Histocompatibility Complex Loci of the Bottlenecked Finnish Wolf Population. *Molecular Ecology* **2014**, 23, 875–889, https://doi.org/10.1111/mec.12647.

Ohta, T. The Nearly Neutral Theory of Molecular Evolution. *Annual Review of Ecology and Systematics* **1992**, *23*, 263–286, https://doi.org/10.1146/annurev.es.23.110192.001403.

Pedersen, N.; Liu, H.; Millon, L.; Greer, K. Dog Leukocyte Antigen Class II–Associated Genetic Risk Testing for Immune Disorders of Dogs: Simplified Approaches Using Pug Dog Necrotizing Meningoencephalitis as a Model. *Journal of Veterinary Diagnostic Investigation* **2011**, *23*, 68–76, https://doi.org/10.1177/104063871102300110.

Pond, S.L.K.; Frost, S.D.W.; Muse, S. V HyPhy: Hypothesis Testing Using Phylogenies. *Bioinformatics* **2005**, *21*, 676–679, https://doi.org/10.1093/bioinformatics/bti079.3

Promerová, M.; Alavioon, G.; Tusso, S.; Burri, R.; Immler, S. No Evidence for MHC Class II-Based Non-Random Mating at the Gametic Haplotype in Atlantic Salmon. *Heredity* **2017**, *118*, 563–567, https://doi.org/10.1038/hdy.2016.129.

QGIS Development Team. QGIS Geographic Information System, **2018**. Open Surce Geospatioal Foundation Project. http://qgis.osgeo.org

Quinnell, R.J.; Kennedy, L.J.; Barnes, A.; Courtenay, O.; Dye, C.; Garcez, L.M.; Shaw, M.-A.; Carter, S.D.; Thomson, W.; Ollier, W.E.R. Susceptibility to Visceral Leishmaniasis in the Domestic Dog Is Associated with MHC Class II Polymorphism. *Immunogenetics* **2003**, *55*, 23–28, https://doi.org/10.1007/s00251-003-0545-1.

Radwan, J.; Kawałko, A.; Wójcik, J.M.; Babik, W. MHC-DRB3 Variation in a Free-Living Population of the European Bison, *Bison bonasus*. *Molecular Ecology* **2007**, *16*, 531–540, https://doi.org/10.1111/j.1365-294X.2006.03179.x.

Rigg, R.; Skrbinšek, T.; Linell, J.D.C. Engaging Hunters and Other Stakeholders in a Pilot Study of Wolves in Slovakia Using Non-Invasive Genetic Sampling. *Final Report within the Project. "Support to the European Commission's Policy on Large Carnivores Under the Habitats Directive – phase 2"* **2014**, 1–38.

Robinson, J.; Halliwell, J.A.; Hayhurst, J.D.; Flicek, P.; Parham, P.; Marsh, S.G.E. The IPD and IMGT/HLA Database: Allele Variant Databases. *Nucleic Acids Research* **2015**, *43*, D423–D431, https://doi.org/10.1093/nar/gku1161.

Roche, P.A.; Furuta, K. The Ins and Outs of MHC Class II-Mediated Antigen Processing and Presentation. *Nature Reviews Immunology* **2015**, *15*, 203–216, https://doi.org/10.1038/nri3818.

Rymešová, D.; Králová, T.; Promerová, M.; Bryja, J.; Tomášek, O.; Svobodová, J.; Šmilauer, P.; Šálek, M.; Albrecht, T. Mate Choice for Major Histocompatibility Complex Complementarity in a Strictly Monogamous Bird, the Grey Partridge (*Perdix perdix*). *Frontiers in Zoology* **2017**, *14*, 9, https://doi.org/10.1186/s12983-017-0194-0.

Sacchi, N.; Castagnetta, M.; Miotti, V.; Garbarino, L.; Gallina, A. High-Resolution Analysis of the HLA-A, -B, -C and -DRB1 Alleles and National and Regional Haplotype Frequencies Based on 120 926 Volunteers from the Italian Bone Marrow Donor Registry. *HLA* **2019**, *94*, 285–295, https://doi.org/10.1111/tan.13613.

Safra, N.; Pedersen, N.C.; Wolf, Z.; Johnson, E.G.; Liu, H.W.; Hughes, A.M.; Young, A.; Bannasch, D.L. Expanded Dog Leukocyte Antigen (DLA) Single Nucleotide Polymorphism (SNP) Genotyping Reveals Spurious Class II Associations. *Veterinary Journal* **2011**, *189*, 220–226, https://doi.org/10.1016/j.tvjl.2011.06.023.

Saitou, N.; Nei, M. The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees. *Molecular Biology and Evolution* **1987**, *4*, 406–425, https://doi.org/10.1093/oxfordjournals.molbev.a040454.

Salvatori, V.; Okarma, H.; Ionescu, O.; Dovhanych, Y.; Find'o, S.; Boitani, L. Hunting Legislation in the Carpathian Mountains: Implications for the Conservation and Management of Large Carnivores. *Wildlife Biology* **2002**, *8*, 3–10, https://doi.org/10.2981/wlb.2002.002.

Sandberg, J.K.; Glas, R. Antigen Processing. In *Encyclopedia of Life Sciences*, **2001**, Nature Publishing Group,doi:10.1038/npg.els.0004012.

Sandhu, S.S. *Recombinant DNA Technology*; I K International Publishing House, **2010**; ISBN 9789380578446.

Seddon, J.M.; Ellegren, H. A Temporal Analysis Shows Major Histocompatibility Complex Loci in the Scandinavian Wolf Population Are Consistent with Neutral Evolution. *Proceedings of the Royal Society B: Biological Sciences* **2004**, *271*, 2283–2291, https://doi.org/10.1098/rspb.2004.2869.

Shiina, T.; Hosomichi, K.; Inoko, H.; Kulski, J.K. The HLA Genomic Loci Map: Expression, Interaction, Diversity and Disease. *Journal of Human Genetics* **2009**, *54*, 15–39, https://doi.org/10.1038/jhg.2008.5.

Slatkin, M. Linkage Disequilibrium-Understanding the Evolutionary Past and Mapping the Medical Future. *Nature Reviews Genetics* **2008**, *9*, 477–485, https://doi.org/10.1038/nrg2361.

Slatkin, M.; Excoffier, L. Testing for Linkage Disequilibrium in Genotypic Data Using the Expectation-Maximization Algorithm. *Heredity* **1996**, *76*, 377–383, https://doi.org/10.1038/hdy.1996.55.

Spurgin, L.G.; Richardson, D.S. How Pathogens Drive Genetic Diversity: MHC, Mechanisms and Misunderstandings. *Proceedings of the Royal Society B: Biological Sciences* **2010**, *277*, 979–988, https://doi.org/10.1098/rspb.2009.2084.

Strand, T. European Black Grouse: MHC Genetic Diversity and Population Structure. *PhD Dissertation, Acta Universitatis Upsaliensis*. **2011**, retrieved from: http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-160042

Tajima, F. Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genetics* **1989**, *123*, 585–595.

Takahata, N.; Nei, M. Allelic Genealogy under Overdominant and Frequency-Dependent Selection and Polymorphism of Major Histocompatibility Complex Loci. *Genetics* **1990**, *124*, 967–978.

Venkataraman, G.M.; Kennedy, L.J.; Little, M.-T.E.; Graves, S.S.; Harkey, M.A.; Torok-Storb, B.J.; Storb, R. Thirteen Novel Canine Dog Leukocyte Antigen-88 Alleles Identified by Sequence-Based Typing. *HLA* **2017**, *90*, 165–170, https://doi.org/10.1111/tan.13077.

Wagner, J.L. Molecular Organization of the Canine Major Histocompatibility Complex. *Journal of Heredity* **2003**, *94*, 23–26, https://doi.org/10.1093/jhered/esg002.

Wagner, J.L.; Burnett, R.C.; DeRose, S.A.; Storb, R. Molecular Analysis and Polymorphism of DLA-DQA Gene. *Tissue Antigens* **1996**, *48*, 199–204, https://doi.org/ 10.1111/j.1399-0039.1996.tb02629.x.

Wagner, J.L.; Burnett, R.C.; Works, J.D.; Storb, R. Molecular Analysis of DLA-DRBB1 Polymorphism. *Tissue Antigens* **1996**, *48*, 554–561, https://doi.org/10.1111/j.1399-0039.1996.tb02669.x.

Wall, J.D.; Pritchard, J.K. Haplotype Blocks and Linkage Disequilibrium in the Human Genome. *Nature Reviews Genetics* **2003**, *4*, 587–597, https://doi.org/10.1038/nrg1123.

Weinberg, W. Uber Den Nachweis Der Vererbung Beim Menschen*. *Jahreshefte des Vereins für vaterländische Naturkunde in Württemberg* **1908**, *64*, 369–382.

Xu, S.X.; Ren, W.H.; Li, S.Z.; Wei, F.W.; Zhou, K.Y.; Yang, G. Sequence Polymorphism and Evolution of Three Cetacean MHC Genes. *Journal of Molecular Evolution* **2009**, *69*, 260–275, https://doi.org/10.1007/s00239-009-9272-z.

Yamaguchi, T.; Dijkstra, J.M. Major Histocompatibility Complex (MHC) Genes and Disease Resistance in Fish. *Cells* **2019**, *8*, 378, https://doi.org/10.3390/cells8040378.

Yang, W.-C.; Chou, L.-S.; Hu, J.-M. Molecular Characterization of Major Histocompatibility Complex Class II DQB and DRB Genes in Bottlenose Dolphins (*Tursiops truncatus* and *T. aduncus*) from the Western Pacific. *Zoological Studies* **2007**, *46*, 664–679.

Yang, Z. Likelihood Ratio Tests for Detecting Positive Selection and Application to Primate Lysozyme Evolution. *Molecular Biology and Evolution* **1998**, *15*, 568–573, https://doi.org/10.1093/oxfordjournals.molbev.a025957.

Yang, Z.; Nielsen, R. Codon-Substitution Models for Detecting Molecular Adaptation at Individual Sites Along Specific Lineages. *Molecular Biology and Evolution* **2002**, *19*, 908–917, https://doi.org/10.1093/oxfordjournals.molbev.a004148.

Yang, Z.; Nielsen, R.; Goldman, N.; Pedersen, A.M. Codon-Substitution Models for Heterogeneous Selection Pressure at Amino Acid Sites. *Genetics* **2000**, *155*, 431–449.

Yang, Z.; Wong, W.S.W.; Nielsen, R. Bayes Empirical Bayes Inference of Amino Acid Sites under Positive Selection. *Molecular Biology and Evolution* **2005**, *22*, 1107–1118, https://doi.org/10.1093/molbev/msi097.

Yuhki, N.; Beck, T.; Stephens, R.; Neelam, B.; O’Brien, S.J. Comparative Genomic Structure of Human, Dog, and Cat MHC: HLA, DLA, and FLA. *Journal of Heredity* **2007**, *98*, 390–399, https://doi.org/10.1093/jhered/esm056.

Zhang, J.; Rosenberg, H.F.; Nei, M. Positive Darwinian Selection after Gene Duplication in Primate Ribonuclease Genes. *Proceedings of the National Academy of Sciences USA*. **1998**, *95*, 3708-3713, https://doi.org/10.1073/pnas.95.7.3708.

8. Appendix

8.1 Carpathian population tables

Table 61. Nucleotide sequences of DLA-DQA1 alleles found in grey wolves from the Carpathian population (dots represent the identical nucleotide as in the top sequence).

Allele name	Allele nucleotide sequence	Position
DQA1*005011	G A C C A T G T T G C C T A C T A C G G C A T A A A T G T C T A C C A G	[36]
DQA1*00601	. . . .	[36]
DQA1*012011	. . . .	[36]
DQA1*00301	. . . .	[36]
DQA1*005011	T C T T A C G G T C C C T C T G G C C A G T T C A C C C A T G A A T T T	[72]
DQA1*00601	. . . . A	[72]
DQA1*012011	. . . . A	[72]
DQA1*00301	. . . . A	[72]
DQA1*005011	G A T G G C G A T G A G G A G T T C T A C G T G G A C C T G G A G A A G	[108]
DQA1*00601	. . . .	[108]
DQA1*012011	. . . .	[108]
DQA1*00301	. . . .	[108]
DQA1*005011	A A G G A A A C T G T C T G G C G G C T G C C T G T G T T T A G C A C A	[144]
DQA1*00601	. . . .	[144]
DQA1*012011	. . . .	[144]
DQA1*00301	. . . .	[144]
DQA1*005011	T T T A C A A G T T T T G A C C C A C A G G G T G C A C T G A G A A A C	[180]
DQA1*00601	. . . . G	[180]
DQA1*012011	. . . G	[180]
DQA1*00301	. . . .	[180]
DQA1*005011	T T G G C T A T A A C A A A A C A A A A C T T G A A C A T C A T G A C T	[216]
DQA1*00601	. . . . T . . . . C	[216]
DQA1*012011	. . . . G	[216]
DQA1*00301	. . . . . C . G . G . . . . C	[216]
DQA1*005011	A A A A G G T C C A A C A A A A C T G C T G C T A C C A A T	[246]
DQA1*00601	. . . . C	[246]
DQA1*012011	. . . . C	[246]
DQA1*00301	. . . . . T . . . . . . C	[246]

Table 62. Nucleotide sequences of DLA-DQB1 alleles found in grey wolves from the Carpathian population (dots represent the identical nucleotide as in the top sequence).

Allele name	Allele nucleotide sequence	Position
DQB1*00401	G A T T T C G T G T T C C A G T T T A A G G G C G A G T G C T A T T T C A C C	[39]
DQB1*02002	. . . . A . . . . C	[39]
DQB1*03501	. . . . A . . . . T T	[39]
DQB1*03502	. . . . A . . . . T T	[39]
DQB1*03901	. . . . T T	[39]
DQB1*04101	. . . .	[39]
DQB1*04901	. . . . A . . . . C	[39]
DQB1*00401	A A C G G G A C G G A G C G G G T G C G G C T T C T G A C T A A A T A C A T C	[78]
DQB1*02002	. . . . G . G . A G	[78]
DQB1*03501	. . . . G . G . G . G	[78]
DQB1*03502	. . . . G . G . G . G	[78]
DQB1*03901	. . . . G	[78]
DQB1*04101	. . . .	[78]
DQB1*04901	. . . . G . G . A G	[78]
DQB1*00401	T A T A A C C G G G A G G A G T A C G T G C G C T T C G A C A G C G A C G T G	[117]
DQB1*02002	. . . . T	[117]
DQB1*03501	. . . . C	[117]
DQB1*03502	. . . . C	[117]
DQB1*03901	. . . . T	[117]
DQB1*04101	. . . . T	[117]
DQB1*04901	. . . . T	[117]
DQB1*00401	G G G G A G T A C C G G G C G G T C A C G G A G C T C G G G C G G C C C T G G	[156]
DQB1*02002	. . . . T . . . . G T C	[156]
DQB1*03501	. . . . G A C	[156]
DQB1*03502	. . . . G A C	[156]
DQB1*03901	. . . .	[156]
DQB1*04101	. . . .	[156]
DQB1*04901	. . . . T . . . . G T C	[156]
DQB1*00401	G C T G A G T A C T G G A A C C C G C A G A A G G A C G A G A T G G A C C G G	[195]
DQB1*02002	. . . . G G . . . . G A T C T . . . . G . . .	[195]
DQB1*03501	. . . . G G . . . . G C T C T . . . . G . A .	[195]
DQB1*03502	. . . . G G . . . . G C T C T . . . . G . A .	[195]
DQB1*03901	. . . .	[195]
DQB1*04101	. . . .	[195]
DQB1*04901	. . . . G G . . . . G A T C T . . . . G . . .	[195]
DQB1*00401	G T A C G G G C C G A G C T G G A C A C G G T G T G C A G A C A C A A C T A C	[234]
DQB1*02002	A A G . . . . C . G . . . . . . G	[234]
DQB1*03501	A G G	[234]
DQB1*03502	A G G	[234]
DQB1*03901	. . . .	[234]
DQB1*04101	. . . .	[234]
DQB1*04901	A A G . . . . G . . . . . . G	[234]
DQB1*00401	G G G T T G G A A G A G C T C T A C A C G T T G C A G C G G C G A	[267]
DQB1*02002	. . . A G . . . . A C	[267]
DQB1*03501	. . . .	[267]
DQB1*03502	. . . G	[267]
DQB1*03901	. . . G	[267]
DQB1*04101	. . . .	[267]
DQB1*04901	. . . A G . . . . A C	[267]

Table 63. Nucleotide sequences of DLA-DRB1 alleles found in grey wolves from the Carpathian population (dots represent the identical nucleotide as in the top sequence). Codon start is shaded light blue.

Allele name	Allele nucleotide sequence	Position
DRB1*03601	C A C A T T T C T T G G A G A T G T T A A A G T C C G A G T G C C A T T T C A	[39]
DRB1*04901	. . . . A . . . . . G . . . . . . T	[39]
DRB1*04902	. . . . A . . . . . G . . . . . . T	[39]
DRB1*05401	. . . . G . . G C . . . . T	[39]
DRB1*092011	. . . . T	[39]
DRB1*09901	. . . . G . . G C . . . . T	[39]
DRB1*12801	. . . . A . . . . . G . . . . . . T	[39]
DRB1*03601	C C A A C G G G A C G G A G C G G G T G C G G T T C G T G G A A A G A T A C A	[78]
DRB1*04901	. . . . G . . . . G . . .	[78]
DRB1*04902	. . . . G . . . . G . . .	[78]
DRB1*05401	. . . .	[78]
DRB1*092011	. . . .	[78]
DRB1*09901	. . . .	[78]
DRB1*12801	. . . . G . . . . . a . . . G . . .	[78]
DRB1*03601	T C C A T A A C C G G G A G G A G A A C G T G C G C T T C G A C A G C G A C G	[117]
DRB1*04901	. . T . . . . T	[117]
DRB1*04902	. . T . . . . T	[117]
DRB1*05401	. . . .	[117]
DRB1*092011	. . . . T T	[117]
DRB1*09901	. . . .	[117]
DRB1*12801	. . T . . . . T	[117]
DRB1*03601	T G G G G G A G T A C C G G G C G G T C A C G G A G C T C G G G C G G C G C G	[156]
DRB1*04901	. . . . T . . . . C . T	[156]
DRB1*04902	. . . . a . . . . C . T	[156]
DRB1*05401	. . . . C . .	[156]
DRB1*092011	. . . . C . .	[156]
DRB1*09901	. . . .	[156]
DRB1*12801	. . . . T . . . . C . T	[156]
DRB1*03601	A C G C T G A G T C C T G G A A C C G G C A G A A G G A G C T C T T G G A G C	[195]
DRB1*04901	C G . . . . A . . . . G	[195]
DRB1*04902	C G . . . . A . . . . G	[195]
DRB1*05401	. . . . G	[195]
DRB1*092011	. . . . T	[195]
DRB1*09901	. . . .	[195]
DRB1*12801	C G . . . . A . . . . G	[195]
DRB1*03601	G G A A G C G G G C C G A G G T G G A C A C C T A C T G C A G A C A C A A C T	[234]
DRB1*04901	A . . G . . . . G G T G	[234]
DRB1*04902	A . . G . . . . G G T G	[234]
DRB1*05401	A . . G . . . . C	[234]
DRB1*092011	A . . G	[234]
DRB1*09901	. . . .	[234]
DRB1*12801	A . . G . . . . G G T G	[234]
DRB1*03601	A C G G G G T G A T T G A G A G C T T C A C G G T G C A G C G G C G A G	[270]
DRB1*04901	. . . . G	[270]
DRB1*04902	. . . . G	[270]
DRB1*05401	. . . .	[270]
DRB1*092011	. . . . G G C	[270]
DRB1*09901	. . . .	[270]
DRB1*12801	. . . . G G C	[270]

Table 64. Haplotype distribution in individual wolves from the Carpathian Mountains

	Haplotype 1			Haplotype 2		
Sample name	DQA1*	DQB1*	DRB1*	DQA1*	DQB1*	DRB1*
CL 308	00301	00401	05401	005011	03901	04902
CL 309	012011	03501	03601	012011	03501	03601
CL 310	005011	03901	04902	005011	03901	04902
CL 333	00301	00401	05401	012011	03501	03601
CL 353	00301	04101	05401	012011	03501	03601
CL 364	00601	02002	092011	012011	03502	03601
CL 369	00301	00401	05401	00301	04101	05401
CL 372	00301	00401	05401	005011	03901	04902
CL 374	00301	00401	05401	005011	03901	04902
CL 375	00301	00401	05401	00301	00401	05401
CL 638	00301	00401	05401	012011	03501	03601
CL 639	00301	00401	05401	00301	00401	05401
CL 640	00301	04101	05401	00301	04101	05401
CL 642	00301	00401	05401	00301	00401	09901
CL 646	00301	00401	05401	005011	03901	04902
CL 647	00301	00401	05401	00301	00401	05401
CL 648	012011	03501	03601	012011	03501	03601
CL 649	00301	00401	05401	00601	02002	092011
CL 653	00301	00401	05401	005011	03901	04902
CL 656	00601	02002	092011	012011	03502	03601
CL 657	00301	00401	05401	012011	03501	03601
CL 658	00301	04101	05401	012011	03501	03601
CL 659	00301	00401	05401	012011	03501	03601
CL 660	00301	00401	05401	00301	00401	05401
CL 661	00301	00401	05401	005011	03901	04902
CL 662	00301	00401	05401	005011	03901	04902
CL 663	00301	04101	05401	005011	03901	04902
CL 664	012011	03501	03601	012011	03501	03601
CL 665	005011	03901	04902	012011	03501	03601
CL 666	005011	03901	04901	012011	03501	03601
CL 667	00301	00401	05401	00301	00401	05401
CL 679	00301	00401	09901	012011	03501	03601
CL 680	005011	03901	04902	005011	03901	04902
CL 685	00301	00401	05401	00301	00401	05401
CL 686	012011	03501	03601	012011	03501	03601
CL 688	00301	00401	05401	012011	03501	03601
CL 692	00601	04901	092011	012011	03502	03601
CL 693	00301	00401	05401	012011	03501	03601
CL 700	012011	03501	03601	012011	03501	03601
CL 701	00301	00401	05401	00301	04101	05401
CL 702	00301	00401	05401	00301	04101	05401
CL 703	00601	02002	092011	012011	03502	03601
CL 704	00301	00401	05401	00301	04101	05401
CL 706	012011	03501	03601	012011	03501	03601
CL 708	00301	04101	05401	00301	04101	05401
CL 709	00301	04101	05401	00601	02002	092011
CL 710	00301	00401	05401	012011	03501	03601
CL 711	00301	04101	05401	012011	03501	03601
CL 712	00301	00401	05401	012011	03501	03601
CL 713	00301	00401	05401	012011	03501	03601

Table 64. continued

Sample name	Haplotype 1 DQA1*	DQB1*	DRB1*	Haplotype 2 DQA1*	DQB1*	DRB1*
CL 714	00301	04101	05401	012011	03501	03601
CL 715	00301	00401	05401	012011	03501	03601
CL 716	00301	04101	05401	012011	03501	03601
CL 717	00301	04101	05401	012011	03501	03601
CL 718	00601	02002	092011	012011	03502	03601
CL 719	00601	02002	092011	012011	03502	03601
CL 720	00301	00401	05401	012011	03501	03601
CL 721	00301	04101	05401	012011	03501	03601
CL 723	00301	04101	05401	00301	04101	05401
CL 724	00301	04101	05401	012011	03501	03601
CL 725	00301	00401	05401	012011	03501	03601
CL 726	00301	00401	05401	012011	03501	03601
CL 727	00301	00401	05401	012011	03501	03601
CL 728	00301	04101	05401	00601	02002	092011
CL 729	00301	00401	05401	00301	04101	05401
CL 730	005011	03901	04902	012011	03501	03601
CL 731	005011	03901	04902	00601	02002	092011
CL 732	00301	00401	05401	012011	03501	03601
CL 733	00301	00401	05401	012011	03501	03601
CL 734	005011	03901	04902	012011	03501	03601
CL 735	005011	03901	04902	005011	03901	04902
CL 736	005011	03901	04901	012011	03501	03601
CL 737	005011	03901	04902	012011	03501	03601
CL 738	00301	00401	05401	00601	02002	092011
CL 739	00301	00401	05401	00601	02002	092011
CL 740	00301	00401	05401	012011	03501	03601
CL 742	005011	03901	04901	012011	03501	03601
CL 743	012011	03501	03601	012011	03501	03601
CL 744	012011	03501	03601	012011	03501	03601
CL 745	012011	03501	03601	012011	03501	03601
CL 746	00301	00401	05401	012011	03501	03601
CL 747	00301	00401	05401	00301	04101	05401
CL 748	005011	03901	04902	012011	03501	03601
CL 749	00301	00401	05401	012011	03501	03601
CL 750	00301	00401	05401	012011	03501	03601
CL 753	00301	00401	05401	00301	00401	09901
CL 754	00601	02002	092011	012011	03502	03601
CL 755	012011	03501	03601	012011	03501	03601
CL 756	005011	03901	04902	005011	03901	04902
CL 757	00301	00401	05401	00301	04101	05401
CL 759	00301	00401	05401	005011	03901	04902
CL 762	012011	03501	03601	012011	03501	03601
CL 763	012011	03501	03601	012011	03501	03601
CL 764	00301	00401	05401	00601	02002	092011
CL 765	012011	03501	03601	012011	03501	03601
CL 766	00301	00401	05401	012011	03501	03601
M.B.	005011	03901	04901	005011	03901	04901
VYS	00301	00401	05401	00601	02002	092011
X1	005011	03901	04901	005011	03901	12801

8.2 Dinaric-Balkan population tables

Table 65. Nucleotide sequences of DLA-DQA1 alleles found in grey wolves from Serbia (dots represent the identical nucleotide as in the top sequence).

Allele name	Allele nucleotide sequence	Position
DQA1*00201	G A C C A T G T T G C C T A C T A C G G C A T A A A T G T C T A C C A G	[36]
DQA1*00301	. . . .	[36]
DQA1*00401	. . . .	[36]
DQA1*00402	. . . .	[36]
DQA1*005011	. . . .	[36]
DQA1*00601	. . . .	[36]
DQA1*00701	. . . .	[36]
DQA1*012011	. . . .	[36]
DQA1*00201	T C T T A C G G T C C C T C T G G C C A G T A C A C C C A T G A A T T T	[72]
DQA1*00301	. . . .	[72]
DQA1*00401	. . . .	[72]
DQA1*00402	. . . .	[72]
DQA1*005011	. . . . T	[72]
DQA1*00601	. . . .	[72]
DQA1*00701	. . . .	[72]
DQA1*012011	. . . .	[72]
DQA1*00201	G A T G G C G A T G A G G A G T T C T A C G T G G A C C T G G A G A A G	[108]
DQA1*00301	. . . .	[108]
DQA1*00401	. . . .	[108]
DQA1*00402	. . . . T T	[108]
DQA1*005011	. . . .	[108]
DQA1*00601	. . . .	[108]
DQA1*00701	. . . .	[108]
DQA1*012011	. . . .	[108]
DQA1*00201	A A G G A A A C T G T C T G G C G G C T G C C T G T G T T T A G C A C A	[144]
DQA1*00301	. . . .	[144]
DQA1*00401	. . . .	[144]
DQA1*00402	. . . .	[144]
DQA1*005011	. . . .	[144]
DQA1*00601	. . . .	[144]
DQA1*00701	. . . .	[144]
DQA1*012011	. . . .	[144]
DQA1*00201	T T T A C A A G T T T T G A C C C A C A G G G T G C A C T G A G A A A C	[180]
DQA1*00301	. . . .	[180]
DQA1*00401	. . . .	[180]
DQA1*00402	. . . .	[180]
DQA1*005011	. . . .	[180]
DQA1*00601	. . . . G	[180]
DQA1*00701	. . . .	[180]
DQA1*012011	. . . G	[180]
DQA1*00201	T T G G C T A T A A C A A A A C A A A A C T T G A A C A T C A T G A C T	[216]
DQA1*00301	. . . . . C . G . G . . . . C	[216]
DQA1*00401	. . . . T . . . . C	[216]
DQA1*00402	. . . . T . . . . C	[216]
DQA1*005011	. . . .	[216]
DQA1*00601	. . . . T . . . . C	[216]
DQA1*00701	. . . .	[216]
DQA1*012011	. . . . G	[216]
DQA1*00201	A A A A G G T C C A A C A A A A C T G C T G C T A C C A A T	[246]
DQA1*00301	. . . . . T . . . . . . C	[246]
DQA1*00401	. . . . C	[246]
DQA1*00402	. . . . C	[246]
DQA1*005011	. . . .	[246]
DQA1*00601	. . . . C	[246]
DQA1*00701	. . . . C	[246]
DQA1*012011	. . . . C	[246]

Table 66. Nucleotide sequences of DLA-DQB1 alleles found in grey wolves from Serbia (dots represent the identical nucleotide as in the top sequence).

Allele name	Allele nucleotide sequence	Position
DQB1*00401	G A T T T C G T G T T C C A G T T T A A G G G C G A G T G C T A T T T C A C C	[39]
DQB1*02002	. . . . A . . . . C	[39]
DQB1*02301	. . . . A	[39]
DQB1*02305	. . . . A	[39]
DQB1*02901	. . . . A . . . . . G . . . . . C	[39]
DQB1*02902	. . . . A . . . . . G . . . . . C	[39]
DQB1*03501	. . . . A . . . . T T	[39]
DQB1*03801	. . . . T T	[39]
DQB1*03901	. . . . T T	[39]
DQB1*04901	. A T . . . . A . C A . . . . C . . . . C A . .	[39]
DQB1*039**	. . . . T T	[39]
DQB1*00401	A A C G G G A C G G A G C G G G T G C G G C T T C T G A C T A A A T A C A T C	[78]
DQB1*02002	. . . . G . G . A G	[78]
DQB1*02301	. . . .	[78]
DQB1*02305	. . . .	[78]
DQB1*02901	. . . . T . . . . . G	[78]
DQB1*02902	. . . . T . . . . . G	[78]
DQB1*03501	. . . . G . G . G . G	[78]
DQB1*03801	. . . .	[78]
DQB1*03901	. . . . G	[78]
DQB1*04901	. . . . G . G . A G	[78]
DQB1*039**	. . . . G	[78]
DQB1*00401	T A T A A C C G G G A G G A G T A C G T G C G C T T C G A C A G C G A C G T G	[117]
DQB1*02002	. . . . T	[117]
DQB1*02301	. . . .	[117]
DQB1*02305	. . . .	[117]
DQB1*02901	. . . . T	[117]
DQB1*02902	. . . . T	[117]
DQB1*03501	. . . . C	[117]
DQB1*03801	. . . .	[117]
DQB1*03901	. . . . T	[117]
DQB1*04901	. . . . T	[117]
DQB1*039**	. . . . T	[117]
DQB1*00401	G G G G A G T A C C G G G C G G T C A C G G A G C T C G G G C G G C C C T G G	[156]
DQB1*02002	. . . . T . . . . G T C	[156]
DQB1*02301	. . . . C .	[156]
DQB1*02305	. . . . C .	[156]
DQB1*02901	. . . . T . . . . G A C	[156]
DQB1*02902	. . . . T . . . . G A C	[156]
DQB1*03501	. . . . G A C	[156]
DQB1*03801	. . . .	[156]
DQB1*03901	. . . .	[156]
DQB1*04901	. . . . T . . . . G T C	[156]
DQB1*039**	. . . . T	[156]
DQB1*00401	G C T G A G T A C T G G A A C C C G C A G A A G G A C G A G A T G G A C C G G	[195]
DQB1*02002	. . . . G G . . . . G A T C T . . . . G . . .	[195]
DQB1*02301	. . . .	[195]
DQB1*02305	. . . .	[195]
DQB1*02901	. . . . G G . . . . G C T C T . . . . G . A .	[195]
DQB1*02902	. . . . C . . . . G G . . . . G C T C T . . . . G . A .	[195]
DQB1*03501	. . . . G G . . . . G C T C T . . . . G . A .	[195]
DQB1*03801	. . . .	[195]
DQB1*03901	. . . .	[195]
DQB1*04901	. . . . G G . . . . G A T C T . . . . G . . .	[195]
DQB1*039**	. . . .	[195]

Table 66. continued

Allele name	Allele nucleotide sequence	Position
DQB1*00401	G T A C G G G C C G A G C T G G A C A C G G T G T G C A G A C A C A A C T A C	[234]
DQB1*02002	A A G . . . . C . G . . . . . . G	[234]
DQB1*02301	. . . .	[234]
DQB1*02305	. . . .	[234]
DQB1*02901	. A G . . . . . A A C . G	[234]
DQB1*02902	. A G . . . . . A A C . G	[234]
DQB1*03501	A G G	[234]
DQB1*03801	. . . .	[234]
DQB1*03901	. . . .	[234]
DQB1*04901	A A G . . . . G . . . . . . G	[234]
DQB1*039**	. . . .	[234]
DQB1*00401	G G G T T G G A A G A G C T C T A C A C G T T G C A G C G G C G A	[267]
DQB1*02002	. . . A G . . . . A C	[267]
DQB1*02301	. . . . A C	[267]
DQB1*02305	. . . . A C . T	[267]
DQB1*02901	. . . G	[267]
DQB1*02902	. . . G	[267]
DQB1*03501	. . . .	[267]
DQB1*03801	. . . . A C	[267]
DQB1*03901	. . . G	[267]
DQB1*04901	. . . A G . . . . A C	[267]
DQB1*039**	. . . G	[267]

Table 67. Nucleotide sequences of DLA-DRB1 alleles found in grey wolves from Serbia (dots represent the identical nucleotide as in the top sequence). Codon start is shaded light blue.

Allele name	Allele nucleotide sequence	Position
DRB1*01501	C A C A T T T C T T G G A G A T G G T A A A G T T C G A G T G C C A T T T C A	[39]
DRB1*03201	. . . . G . . T . C C A . T . T . . . C C	[39]
DRB1*03202	. . . . G . . T . C C A . T . T . . . C C	[39]
DRB1*03601	. . . . T . . . . . . C	[39]
DRB1*04301	. . . . T	[39]
DRB1*04901	. . . . A	[39]
DRB1*05401	. . . . G . . . C . . . . . C . . . . T	[39]
DRB1*08001	. . . . G . . T	[39]
DRB1*090012	. . . . T	[39]
DRB1*092011	. . . . T	[39]
DRB1*092012	. . . . T	[39]
DRB1*036**	. . . . T . . . . . . C	[39]
DRB1*01501	C C A A C G G G A C G G A G C G G G T G C G G C T T C T G G T G A G A G A C A	[78]
DRB1*03201	. . . . T . C G . . . A A . . . T . . .	[78]
DRB1*03202	. . . . T . C G . . . A A . . . T . . .	[78]
DRB1*03601	. . . . T . C G . . . A A . . . T . . .	[78]
DRB1*04301	. . . . T A	[78]
DRB1*04901	. . . . T . G G . . . A A	[78]
DRB1*05401	. . . . T . C G . . . A A . . . T . . .	[78]
DRB1*08001	. . . . T . G G . . . A A . . . T . . .	[78]
DRB1*090012	. . . . T A . . . . C . . . . . C . . .	[78]
DRB1*092011	. . . . T . C G . . . A A . . . T . . .	[78]
DRB1*092012	. . . . T . C G . . . A A . . . T . . .	[78]
DRB1*036**	. . . . T . C G . . . A A . . . T . . .	[78]
DRB1*01501	T C T A T A A C C G G G A G G A G C A C G T G C G C T T C G A C A G C G A C G	[117]
DRB1*03201	. . C . . . . T T	[117]
DRB1*03202	. . C . . . . T T	[117]
DRB1*03601	. . C . . . . A	[117]
DRB1*04301	. . . .	[117]
DRB1*04901	. . . . T	[117]
DRB1*05401	. . C . . . . A	[117]
DRB1*08001	. . . . T a . . t	[117]
DRB1*090012	. . . . A T . C	[117]
DRB1*092011	. . C . . . . T T	[117]
DRB1*092012	. . C . . . . T T	[117]
DRB1*036**	. . C . . . . A	[117]
DRB1*01501	T G G G G G A G T A C C G G G C G G T C A C G G A G C T C G G G C G G C C C G	[156]
DRB1*03201	. . . .	[156]
DRB1*03202	. . . .	[156]
DRB1*03601	. . . . G . .	[156]
DRB1*04301	. . . .	[156]
DRB1*04901	. . . . T . . . . T	[156]
DRB1*05401	. . . .	[156]
DRB1*08001	. . . . G . .	[156]
DRB1*090012	. . . . T	[156]
DRB1*092011	. . . .	[156]
DRB1*092012	. . . . A .	[156]
DRB1*036**	. . . . G . .	[156]
DRB1*01501	A C G C T G A G T A C T G G A A C G G G C A G A A G G A G C T C T T G G A G C	[195]
DRB1*03201	T . . . . C . . . . A	[195]
DRB1*03202	T . . . . C . . . . A	[195]
DRB1*03601	. . . . C . . . . C	[195]
DRB1*04301	. . . . C	[195]
DRB1*04901	C G	[195]
DRB1*05401	. . . . C	[195]
DRB1*08001	. . . . C . . . . C	[195]
DRB1*090012	. . . . C . . . . C	[195]
DRB1*092011	. . . . C . . . . C . . . . T	[195]
DRB1*092012	. . . . C . . . . C . . . . T	[195]
DRB1*036**	. . . . C	[195]

Table 67. continued

Allele name	Allele nucleotide sequence	Position
DRB1*01501	A G A G G C G G G C C G A G G T G G A C A C G G T G T G C A G A C A C A A C T	[234]
DRB1*03201	. . G A . . . . . . A A C	[234]
DRB1*03202	. . G A . . . . . . A A C . . . . C T A C	[234]
DRB1*03601	G . . A . . . . C T A C	[234]
DRB1*04301	G	[234]
DRB1*04901	. . . .	[234]
DRB1*05401	. . . . C . . . . C T A C	[234]
DRB1*08001	. . . . C . . . . C T A C	[234]
DRB1*090012	. . . A . . . . C . . . . C T A C	[234]
DRB1*092011	. . . . C T A C	[234]
DRB1*092012	. . . . C T A C	[234]
DRB1*036**	G . . A . . . . C T A C	[234]
DRB1*01501	A C G G G G T G A T T G A G A G C T T C A C G G T G C A G C G G C G A G	[270]
DRB1*03201	. . . .	[270]
DRB1*03202	. . . .	[270]
DRB1*03601	. . . .	[270]
DRB1*04301	. . C . . . . . G G C	[270]
DRB1*04901	. . . . G	[270]
DRB1*05401	. . . .	[270]
DRB1*08001	. . C . . . . . G G C	[270]
DRB1*090012	. . C . . . . . G G C	[270]
DRB1*092011	. . . . G G C	[270]
DRB1*092012	. . . . G G C	[270]
DRB1*036**	. . . .	[270]

Table 68. Haplotype distribution in individual wolves from Serbia

	Haplotype 1			Haplotype 2		
Sample name	DQA1*	DQB1*	DRB1*	DQA1*	DQB1*	DRB1*
1	012011	03501	03601	012011	03501	090012
4	00601	02002	092011	012011	03501	03601
8	00301	00401	05401	00301	00401	05401
9	00301	00401	05401	00301	00401	05401
13	00601	02002	092011	012011	03501	03601
18	00601	02002	092011	00601	02002	092011
19	005011	039**	04901	012011	03501	03601
22	00301	00401	05401	00601	02002	092011
23	00601	02002	092011	00601	02002	092011
28	00201	02901	03202	00401	02002	092011
31	012011	03501	03601	012011	03501	090012
35	00601	02002	092011	012011	03501	03601
41	012011	03501	03601	012011	03501	090012
43	005011	03901	04901	012011	03501	03601
49	005011	03901	04901	00201	02901	03202
51	005011	03901	04901	00601	02002	092011
57	00601	02002	092011	012011	03501	090012
62	00301	03801	04301	012011	03501	03601
65	00301	00401	05401	012011	03501	090012
75	00201	02901	03202	00301	03801	04301
76	005011	039**	04901	00601	02002	092011
78	005011	03901	04901	005011	03901	04901
84	00601	02002	092011	012011	03501	03601
85	00301	00401	05401	00601	02002	092011
86	00601	02002	092011	00601	02002	092011
95	00601	02002	092011	00601	02002	092011
96	00301	03801	04301	00701	03801	04301
98	00301	03801	04301	00601	02002	092011
100	00301	00401	05401	012011	03501	03601
101	00601	02002	092011	012011	03501	03601
104	00601	02002	092011	012011	03501	03601
150	00301	00401	05401	012011	03501	03601
151	00201	02901	03601	00301	03801	04301
158	00301	03801	04301	00601	02002	092011
159	00601	02002	092011	012011	03501	090012
163	00601	02002	092011	00601	02002	092011
166	00601	02002	092011	012011	03501	03601
170	00201	02901	03202	00401	02002	092011
171	00201	02901	03202	012011	03501	03601
175	00601	02002	092011	012011	03501	03601
177	00301	00401	05401	012011	03501	03601
185	00201	02901	03202	00201	02901	03202
193	00301	00401	05401	00601	02002	092011
196	00601	02002	092011	012011	03501	090012
199	00201	02901	03202	00601	02002	092011
204	00201	02901	03601	00301	03801	04301

Table 68. continued

Sample name	Haplotype 1			Haplotype 2		
	DQA1*	DQB1*	DRB1*	DQA1*	DQB1*	DRB1*
205	00301	03801	04301	005011	039**	04901
206	00201	02901	03202	00601	02002	092011
208	005011	03901	04901	00601	02002	092011
210	00201	02901	03202	00301	00401	05401
264	00301	03801	04301	00301	03801	04301
265	00601	02002	092012	012011	03501	036**
268	00401	02002	092012	005011	03901	04901
269	00301	00401	05401	00601	02002	092011
271	005011	03901	04901	012011	03501	03601
275	00301	00401	05401	00601	02002	092011
279	00301	00401	05401	005011	03901	04901
280	00301	03801	04301	005011	039**	04901
281	00601	02002	092011	00601	02002	092011
295	00301	03801	04301	005011	03901	04901
299		could not resolve			could not resolve	
315	00601	02002	092012	012011	03501	036**
316	00201	02902	03201	00601	02002	092011
329	00301	00401	05401	00601	04901	092011
333	005011	039**	04901	00601	02002	092011
342	00301	00401	05401	005011	03901	04901
343	00601	02002	092011	012011	03501	03601
350	00301	00401	05401	00301	03801	04301
352	00201	02901	03202	012011	03501	090012
356	00301	00401	05401	005011	03901	04901
378	005011	03901	04901	005011	03901	04901
387	005011	03901	04901	005011	039**	04901
390	00301	00401	05401	00601	02002	092011
395	00301	00401	05401	00601	02002	092011
404	00201	02902	03201	00301	00401	05401
411	00301	00401	05401	00301	00401	05401
417	005011	03901	04901	012011	03501	03601
420	00301	03801	04301	012011	03501	03601
428	00601	02002	092011	00601	02002	092011

8. 3. List of abbreviations

DLA – dog leukocyte antigen

dN – rate of synonymous substitutions

DNA – deoxyribonucleic acid

dS – rate of non-synonymous substitutions

ER – endoplasmic reticulum

ESRI - Environmental Systems Research Institute

H_0 – null hypothesis

HA – alternative hypothesis

HLA – human leukocyte antigen

Hsp – heat shock protein

IPD – Immuno Polymorphism Database

IPTG – Isopropyl β-D-1-thiogalactopyranoside

iTOL – Interactive Tree Of Life

IUCN – International Union for Conservation of Nature

Lat – latitude

LB – lysogeny broth

Long – longitude

MEGA – Molecular Evolutionary Genetics Analysis

MHC – major histocompatibility complex

ML – maximum likelihood

PCR – polymerase chain reaction

QGIS – Quantum geographic information system

Sample ID – sample identification

SNP – single nucleotide polymorphism

TNF – tumour necrosis factor

USA – United States of America

UV – ultraviolet

Var - variance

WGS – World Geodetic System

8. Biography

Željko Pavlinec obtained his BSc in molecular biology in 2012 at the University of Zagreb, Faculty of Science, Department of Biology and his MSc in molecular biology at the same Department in 2014. He currently works as an assistant at Croatian Veterinary Institute in Zagreb, where his main tasks include development and implementation of molecular methods for diagnostic and research purposes. He is also an assistant at the University of Zagreb, Faculty of science, Depatment of Biology, where during his PhD he held practical classes in Methods in Immunology, Fundamentals of Molecular Ecology, Molecular Evolution, Biological Evolution and Evolutionary Biology. During his PhD, he was awarded a two-month scholarship from the Government of the French Republic for his professional development in bioinformatics, which was conducted at the The Institute of Evolutionary Science of Montpellier. In the same year he was awarded with the European Association of Fish Pathologists (EAFP) Student Award for his work on the genomic comparisons of serologically different isolates of *Vibrio harveyi*, presented at the EAFP conference in Porto. He has participated in three international and four national projects. His ongoing projects are: „Mediterranean Aquaculture Integrated Development“ (MedAID) – four year project financed in the scope of Horizon 2020 programme, „Enhancing Innovation and Sustainability in Adriatic Aquaculture“ (AdriAquaNet) – three year project financed in the scope of Interreg Italy – Croatia programme and „ Rotaviruses in Croatian Ecosystem: molecular epidemiology and zoonotic potential“, five year project financed by the Croatian Science Foundation (HRZZ-UIP-2017-05-8580). He has co-authored nine scientific papers, one professional paper and one diagnostic manual.

Publications:

1. Zrnčić, S.; Fioravanti, M.; Gustinelli, A.; Oraić, D.; Zupičić, I. G.; **Pavlinec**, **Ž**.; Padrós, F.; Furones, D.; Basurco, B. Survey on laboratories and consultants working in the diagnostics of European seabass and gilthead seabream diseases: preliminary results. *Bulletin of the European Association of Fish Pathologists*, **2021**, *41*(2), 81-90.
2. **Pavlinec**, **Ž**.; Zupičić, I. G.; Oraić, D.; Petani, B.; Mustać, B.; Mihaljević, Ž.; Beck, R.; Zrnčić, S. Assessment of predominant bacteria in noble pen shell (*Pinna nobilis*) collected in the Eastern Adriatic Sea. *Environmental monitoring and assessment,* **2020**, *192*, 1-10, doi:10.1007/s10661-020-08541-6.
3. Jaki Tkalec, V.; Furmeg, S.; Kiš, M.; Sokolović, J.; Benić, M.; Cvetnić, L.; **Pavlinec**, **Ž**.; Špičić, S.; Cvetnić, Ž. Detection of pathogenic bacteria in raw milk and dairy products with special regard to *Yersinia enterocolitica*. *Veterinarska stanica*, **2020**, *51*(5), 487-495.
4. Zrnčić, S.; Oraić, D.; Zupičić, I. G.; **Pavlinec**, **Ž**.; Brnić, D.; Acinger Rogić, Ž.; Sučec, I.; Steinhagen, D.; Adamek, M. Koi herpesvirus and carp edema virus threaten common carp aquaculture in Croatia. *Journal of fish diseases*, **2020**, *43*, 1-13, doi:10.1111/jfd.13163.
5. Zupičić, I.G.; **Pavlinec**, **Ž**.; Oraić, D.; Zrnčić, S. Emerging vibriosis of Mediterranean fish caused by the bacterial species *Vibrio harveyi*: an overview. *Veterinarska stanica*, **2019**, *50*(5), 455-463.
6. Špičić, S.; Duvnjak, S.; Zdelar-Tuk, M.; Laroucau, K.; Reil, I.; Velić, L.; Eterović, T.; **Pavlinec**, **Ž**.; Šegota, M.; Habrun, B.; Cvetnić, Ž. Identification and MLVA genotyping of *Chlamydia abortus* from abortion cases in small ruminants in Croatia. *Veterinarska stanica*, **2019**, *50*(4), 307-314.
7. Cvetnić, Ž.; Zdelar-Tuk, M.; Duvnjak, S.; Reil, I.; Mikulić, M.; **Pavlinec**, **Ž**.; Cvetnić, M.; Špičić, S. Tuberculous and nontuberculous mycobacteria in human and animal infection. *Veterinary Journal of Republic of Srpska*, **2018**, *18*(2), 356-369.
8. Cvetnić, Ž.; Zdelar-Tuk, M.; Duvnjak, S.; Reil, I.; Mikulić, M.; **Pavlinec**, **Ž**.; Cvetnić, M.; Špičić, S. Reservoirs of *Mycobacterium tuberculosis* complex and their significance in the infection of animals and humans. *Veterinarska stanica*, **2018**, *49*(4), 261-271.
9. Špičić, S.; Laroucau, K.; Zdelar-Tuk, M.; Duvnjak, S.; **Pavlinec**, **Ž**.; Reil, I.; Kompes, G.; Habrun, B.; Stepanić, M.; Želježić, D.; Cvetnić, Ž. Glanders – an almost forgotten zoonosis. *Veterinarska stanica*, **2018**, *49*(1), 31-36.
10. Duvnjak, S.; Špičić, S.; Kušar, D.; Papić, B.; Reil, I.; Zdelar-Tuk, M.; **Pavlinec**, **Ž**.; Đuras, M.; Gomerčić, T.; Hendriksen, S. R.; Cvetnić, Ž. Whole-Genome Sequence of the First Sequence Type 27 Brucella ceti Strain Isolated from European Waters. *Genome Announcements*, **2017**, *5*(37); e00988-17-e00988-17, doi:10.1128/genomeA.00988-17.