Variability of DRB locus of MHC genes class II in chamois (Rupicapra spp.)

Stipoljev, Sunčica

Sveučilište u Zagrebu

FACULTY OF AGRICULTURE

Sunčica Stipoljev

VARIABILITY OF DRB LOCUS OF MHC GENES CLASS II IN CHAMOIS (*Rupicapra* spp.)

DOCTORAL THESIS

Zagreb, 2022

Sveučilište u Zagrebu

AGRONOMSKI FAKULTET

Sunčica Stipoljev

RAZNOLIKOST DRB LOKUSA MHC GENA SKUPINE II U DIVOKOZA (*Rupicapra* spp.)

DOKTORSKI RAD

Zagreb, 2022.

Sveučilište u Zagrebu

FACULTY OF AGRICULTURE

Sunčica Stipoljev

VARIABILITY OF DRB LOCUS OF MHC GENES CLASS II IN CHAMOIS (*Rupicapra* spp.)

DOCTORAL THESIS

Supervisors: Assoc. Prof. Nikica Šprem, PhD

Prof. Elena Bužan, PhD

Zagreb, 2022

Sveučilište u Zagrebu

AGRONOMSKI FAKULTET

Sunčica Stipoljev

RAZNOLIKOST DRB LOKUSA MHC GENA SKUPINE II U DIVOKOZA (*Rupicapra* spp.)

DOKTORSKI RAD

Mentori: izv. prof. dr. sc. Nikica Šprem
prof. dr. sc. Elena Bužan

Zagreb, 2022.

Bibliography data
Scientific area: Biotechnical sciences
Scientific field: Agriculture
Branch of science: Hunting
Institution: University of Zagreb, Faculty of Agriculture
Supervisors of the PhD thesis: Assoc. Prof. Nikica Šprem, PhD,
Prof. Elena Bužan, PhD
Number of pages: 84
Number of images: 21
Number of tables: 10
Number of appendices: 5
Number of references: 112
Date of oral examination: May 25, 2022
The members of the PhD defense committee:
1. Assist. Prof. Toni Safner, PhD, University of Zagreb, Faculty of Agriculture
2. Assoc. Prof. Tea Tomljanović, PhD, University of Zagreb, Faculty of Agriculture
3. Assoc. Prof. Ana Galov, PhD, University of Zagreb, Faculty of Science
Rad je pohranjen u:
Nacionalnoj i sveučilišnoj knjižnici u Zagrebu, Ulica Hrvatske bratske zajednice 4 p.p. 550, 10 000 Zagreb,
Knjižnici Sveučilišta u Zagrebu Agronomskog Fakulteta, Svetošimunska cesta 25, 10 000 Zagreb.
The doctoral thesis is stored in:
National and University Library in Zagreb, Ulica Hrvatske bratske zajednice 4 p.p. 550, 10 000 Zagreb,
Central Agricultural Library, University of Zagreb, Faculty of Agriculture, Svetošimunska cesta 25, 10 000 Zagreb.
The subject of the thesis was accepted at the session of the Faculty Council of the University of Zagreb Faculty of Agriculture, held on March 03, 2020, and approved at the session of the Senate of the University of Zagreb, held on November 10, 2020.

UNIVERSITY OF ZAGREB
FACULTY OF AGRICULUTRE

DECLARATION OF ORIGINALITY

I, **Sunčica Stipoljev**, declare that I have composed solely by myself the thesis titled:

VARIABILITY OF DRB LOCUS OF MHC GENES CLASS II IN CHAMOIS (*Rupicapra* spp.)

With my signature I confirm that:

- I am the sole author of this thesis;
- this thesis is an original report of my research, and due references have been provided on all supporting literatures and resources;
- I am familiar with the provisions of the Code of Ethics of the University of Zagreb (Article 19).

Zagreb, March 22, 2022

_____
PhD Candidate signature

Doctoral thesis grade

Doctoral thesis was defended at the University of Zagreb, Faculty of Agriculture on May 25, 2022, in front of the PhD defense committee comprised of:

1. Assist. Prof. Toni Safner, PhD _____
University of Zagreb Faculty of Agriculture

2. Assoc. Prof. Tea Tomljanović, PhD _____
University of Zagreb Faculty of Agriculture

3. Assoc. Prof. Ana Galov, PhD _____
University of Zagreb Faculty of Science

Supervisors information

Assoc. Prof. Nikica Šprem

Assoc. prof. Nikica Šprem was born in Zagreb in 1978. He graduated from the Faculty of Agriculture at the University of Zagreb in 2004 and started working at the Faculty of Fisheries, Apiculture, Wildlife management and special Zoology as a research fellow in 2005. He obtained a PhD in Hunting at the Faculty of Agriculture in Osijek, where he defended his thesis entitled "Morphological and genetic characteristics of wild boar (*Sus scrofa* L.) in the Republic of Croatia" in 2009. In 2012 he was appointed Assistant Professor and in 2017 Associate Professor.

He studied abroad at the University of Natural Resources and Applied Life Sciences, Austria; University of Gothenburg, Sweden; HAKI Institute, Hungary; Montana State University, USA; University of Kaposvar, Hungary; Instituto de Investigaciones Marinas, Spain; University of Sassari, Italy and Massey University, New Zealand. He has been the head of the faculty experiment station “Ban Josip Jelačić, Prolom" for many years, which he has significantly improved both professionally and financially.

His early career interest was applied genetics in wildlife management, especially in ungulate species (wild boar, red deer). In recent years, his research has focused on the behaviour, ecology and management of wildlife species of the subfamily Caprinae (aoudad, European mouflon, chamois) in the Mediterranean ecosystem of the Dinaric Alps. He has published 71 scientific articles which are cited in the Web of Science Core Collection (h-index 12; without single citation 701). He is co-author of four scientific books. He has participated as an active collaborator or leader in ten national and six international scientific projects, as well as in several professional projects.

He is the holder of the undergraduate module Hunting and the graduate modules Biology and Ecology of Wildlife, Hunting and Hunting Management, and in the doctoral studies of the module Research methods in wildlife management. He is also involved in several other modules (Plants in hunting area, sport and recreational fisheries). At the University of Sassari (Italy), he participates in Evolutionary Biology, the PhD Course in Life Sciences and Biotechnologies.

He has supervised 23 diploma/degree theses, and two student theses have received the Rector's / Dean's Award. In these activities, he has proved to be a good teacher who is independent in his work and his initiative motivates students towards quality research and basic study requirements, as evidenced by no less than ten original scientific papers co-authored with students, six of which are cited in the Web of Science Core Collection. To date, he has supervised five doctoral students and one postdoctoral researcher.

Assoc. Prof. Šprem has a special flair for fieldwork and has been involved in the conduct and writing of 34 technical reports (national monitoring of open waters, ichthyological studies, fisheries and hunting economics). He is also the author of 56 technical articles in domestic journals.

Prof. Elena Bužan

As a young researcher, Elena Bužan received her PhD in 2007 from the University of Ljubljana, Biotechnical faculty, in collaboration with the University of Hull, Department of Molecular Ecology, United Kingdom. Before starting her academic career, she gained six years of experience as a technologist in the field of environmental protection in the companies Droga d.o.o, Delamaris d.d. in Tomos d.o.o. During this time she completed a Master's degree in Technical Environmental Protection at the University of Maribor.

Her research areas are molecular ecology and conservation biology, i.e. the use of genetics in wildlife conservation and management. She is co-author of 53 original and peer-reviewed scientific articles, editor of two scientific monographs, she participated in more than 70 papers at international conferences and in 7 chapters in monographic publications. She is also a member of the editorial board of a prestigious international journal. Through leadership and participation in national and international projects she contributes significantly to the development of molecular ecology as a tool for conservation and sustainable management of wildlife working with policy makers and stakeholders.

She is currently leading an international Horizon 2020 project Step Change, which major aim is the inclusion of citizen science in the fields of health, energy, and conservation. For her academic and scientific achievements, she was awarded by University of Primorska and by regional authorities. She has supervised more than 30 graduate, master's and doctoral theses in the fields of conservation genetics and animal biology. She is a member of the Slovenian Biochemical and Genetic Society, International Biodiversity Consortium, Scientific Consortium for Conservation Genetics ConGen and the representative of Slovenia in ERGA (European Reference Genomic Atlas) consortium. As a guest lecturer and evaluator of study programmes, she has visited many universities in Brazil, Kazakhstan, USA, Lebanon, Czech Republic, Australia, Russia and Croatia. She was an educational mentor in four projects of the Public Fund for Scholarships and Human Resources Development funded by the European Social Fund, which linked the university with the local economy and enabled students to acquire additional skills for easier employment.

Acknowledgments

The research work in this dissertation was carried out within the project HRZZ- IP-2016-06-5751, “DNA as evidence of distribution and vitality of endangered Balkan chamois – BalkCham” funded by Croatian Science Foundation (HRZZ) in the period from March 01, 2017, to March 01, 2020, led by Assoc. Prof. Nikica Šprem.

I would like to thank my supervisors, Assoc. Prof. Nikica Šprem and Prof. Elena Bužan, for their invaluable advice, constant support and patience during the preparation of this thesis.

Special thanks to all my colleagues who supported me during this period and contributed in any way to the preparation of this thesis.

Summary

The chamois, *Rupicapra* spp, is a species in the family Bovidae and one of the most common mountain ungulates, distributed in the mountain ranges of Europe and Asia Minor, where it occurs in two species according to its morphological and behavioral characteristics: the Northern chamois *R. rupicapra* (with subspecies *cartusiana, rupicapra, tatrica, balcanica, carpatica, asiatica* and *caucasica*) and the Southern chamois *R. pyrenaica* (with subspecies *parva, pyrenaica* and *ornata*).

The major histocompatibility complex (MHC) is a family of genes that encode receptors that recognize and bind antigens to present them to T cells. They are therefore central to vertebrate adaptive immunity. The MHC region comprises some of the most variable loci in the vertebrate genome, which have been associated with various fitness traits and thus with the long-term persistence of populations. Because of their well-characterized function and exceptional diversity, they represent excellent markers for evolutionary ecology and conservation. Traditional methods have been commonly used for their genotyping, but the introduction of next-generation sequencing has enabled more accurate and reproducible genotyping of such polymorphic gene families.

In this work, the genetic diversity of the second exon of the MHC class II DRB locus was analyzed in 110 individuals from populations covering most of the range of the genus *Rupicapra* using a next-generation approach (Ion Torrent S5, Thermo Fisher). MHC DRB exon 2 encodes functionally important residues of the antigen binding groove and can therefore be used as a measure of the functional diversity of DRB alleles.

Twenty-five MHC DRB alleles were found, each translated into a unique amino acid sequence, indicating the functional importance of polymorphism between alleles. Fourteen novel DRB alleles were identified in this study, two of which were found only in *R. r. carpatica*, two in *R. r. asiatica*, three in *R. r. balcanica*, and four only in *R. r. rupicapra*. A gene duplication was not identified. The high ratio of the relative rates of non-synonymous to synonymous mutations and the presence of trans-species polymorphisms suggest that this locus was under strong balancing selection throughout the evolutionary history of this species.

Keywords: *Rupicapra*, major histocompatibility complex, next-generation sequencing, allelic polymorphism, conservation

Raznolikost DRB lokusa MHC gena skupine II u divokoza (*Rupicapra* spp.)

Divokoza (*Rupicapra* spp.) je papkar iz porodice šupljorožaca (Bovidae) rasprostranjen na planinskim lanacima diljem Europe i Bliskog Istoka, a prema morfoloških i bihevioralnim svojstvima dijeli se na dvije vrste: sjevernu divokozu *Rupicapra rupicapra* (sa podvrstama *cartusiana*, *rupicapra*, *tatrica*, *balcanica*, *carpatica*, *asiatica* i *caucasica*) i južnu divokozu *Rupicapra pyrenaica* (sa podvrstama *parva, pyrenaica* i *ornata*).

Glavni sustav tkivne podudarnosti ili MHC (eng. major histocompatibility complex) je porodica gena koji kodiraju receptore koji prepoznaju i vežu antigene kako bi ih predstavili T stanicama te su izuzetno važni za adaptivnu imunost kralježnjaka. MHC regija obuhvaća neke od najvarijabilnijih lokusa u genomu kralježnjaka, koji su povezani s različitim obilježjima fitnesa, a time i s dugotrajnom postojanošću populacija. Zbog svoje dobro okarakterizirane funkcije i iznimne raznolikosti, predstavljaju izvrsne markere u evolucijskoj ekologiji i konzervaciji. Za njihovu genotipizaciju obično se koriste tradicionalne metode, ali uvođenje sekvenciranja sljedeće generacije omogućilo je točniju i ponovljivu genotipizaciju takvih polimorfnih genskih porodica.

U ovom radu analizirana je genetička raznolikost egzona 2 MHC DRB lokusa skupine II u 110 jedinki iz populacija koje pokrivaju većinu područja rasprostranjenja roda *Rupicapra* korištenjem sekvenciranja sljedeće generacije (Ion Torrent S5, Thermo Fisher). MHC DRB egzon 2 kodira funkcionalno važne aminokiseline za vezanje antigena i stoga se može koristiti kao mjera funkcionalne raznolikosti DRB alela.

Pronađeno je dvadeset i pet MHC DRB alela, od kojih svaki translacijom daje jedinstveni aminokiselinski slijed, što ukazuje na funkcionalnu važnost polimorfizma među alelima. U ovoj studiji identificirano je četrnaest novih DRB alela, od kojih su dva pronađena samo u *R. r. carpatica*, dva u *R. r. asiatica*, tri u *R. r. balcanica*, a četiri u *R. r. rupicapra*. Duplikacija gena nije identificirana. Visok omjer relativnih stopa nesinonimnih i sinonimnih mutacija i prisutnost trans-specijskog polimorfizama sugeriraju da je ovaj lokus bio pod snažnom balansirajućom selekcijom tijekom evolucijske povijesti divokoza.

Ključne riječi: *Rupicapra*, glavni sustav tkivne podudarnosti, sekvenciranje sljedeće generacije, alelni polimorfizam, konzervacija

Table of contents

List of abbreviations

BIC – Bayesian information criterion
BEB – Bayes empirical Bayes
DAPC – discriminant analysis of principal components
FEL – fixed effects likelihood
FUBAR – fast unconstrained Bayesian approximation
HLA – human leukocyte antigen
LRT – likelihood ratio test
MEME – mixed effects model of evolution
MHC – major histocompatibility complex
NGS – next generation sequencing
PBR – peptide-binding region
PC – principal component
PCR – polymerase chain reaction
PSS – positively selected site
SLAC – single-likelihood ancestor counting
SSCP – single-strand conformation polymorphism

List of tables

posterior probability of > 95% are shown in standard font. Codon numbers correspond to the codons of ß1-domain in chamois (Schaschl et al., 2004).

Table 9. Recombinant sequences at exon 2 of the major histocompatibility complex (MHC) class II DRB locus in chamois.

Table 10. DRB exon 2 allele frequencies estimated with Single-Strand Conformation Polymorphism (SSCP)/Sanger sequencing and amplicon-based next-generation sequencing (NGS).

List of figures

location in the case of larger sample size (Appendix 1). DRB alleles in legend are ordered sequentially, from left to right, according to their frequency in the overall sample.

Figure 11. The number of nucleotide differences between the different MHC DRB alleles found in Northern chamois (top) and Southern chamois (down).

Figure 12. Alignment of the putative amino acid sequences of the exon 2 of MHC DRB alleles found in Northern (A) and Southern (B) chamois. Dots indicate identity in the amino acid sequence to the sequence of the Ruru-DRB01 allele and a yellow columns indicate codons involved in the peptide-binding region (PBR) in humans (Brown et al., 1993).

Figure 13. Scatterplot of the genetic differentiation resulting from a discriminant analysis of principal components (DAPC) for the genetic structure of Northern and Sothern chamois subspecies based on the DRB locus. Individuals are presented as separate dots with colours denoting chamois species/subspecies and inclusion of 95% inertia ellipses.

Figure 14. Scatterplot of the five MHC supertypes resulting from a DAPC. Each allele is represented as a dot and the supertypes as ellipses (ST1-ST5) (see spatial distribution in Figure 10).

Figure 15. Neighbour-net network of chamois DRB alleles. The alleles marked in yellow were detected in both species. Scale bar indicates the scale of the network i.e., 0.01 substitutions per nucleotide site.

Figure 16. Distribution of positively selected sites in exon 2 of chamois DRB genes is shown as estimated by EasyCodeML (model M2a). Red columns indicate the class of sites with a high probability of $\omega > 1$. In this model $\omega1 = 0.00$ and applies to ~58% of the codons (blue); $\omega2 = 1.00$ at ~20% of the sites (green); and $\omega3 = 18.24$ at ~22% of the sites (red). Diversity of the observed peptides is indicated by a sequence conservation logo. Asterisks indicate codons involved in peptide-binding region (PBR) residues in the human ortholog. Sites under positive and negative selection are indicated with respect to FEL, FUBAR, SLAC (for pervasive selection) and MEME (for episodic selection).

Figure 17. Variants ordered by descending per amplicon frequency (PAF) after sequencing errors have been added to them through the clustering step performed by AmpliSAS. Lines represent amplicons of 28 individuals. Horizontal line represents 14% PAF threshold for the filtering step.

Figure 18. Percentages of amplicon reads assigned to alleles before and after AmpliSAS clustering. On average, 54% of amplicon depths were assigned to alleles, and after clustering this proportion increased to 83%.

Figure 19. Representation of amplicon depths of 28 individuals, and the number of reads of true sequences and artefacts within each amplicon. True - sequences that match a real allele i.e. true variant; the following are sequencing errors: X - 1bp substitutions, I - insertions, D - deletions, H - homopolymer indels; XIDH - sequence with any combination

of at least two sequencing errors; leftovers - low frequency variants, chimeras or sequences containing many errors which could not be classified into the major clusters. All sequencing errors, except leftovers, were clustered with the true variant from which they were derived.

Figure 20. Alignment of the nucleotide sequences, only variable residues of the exon 2 of DRB alleles of chamois are shown. The codes before represent the species abbreviation and gene name, and the numbers after indicate allele numbers.

Figure 21. Median-joining network (ε = 0) of DRB alleles found in chamois. Alleles are represented by pie charts whose size is proportional to the number of individuals. The colors indicate the number of individuals that have particular alleles obtained with two genotyping methods. Number of mutations separating nodes is represented by slashes crossed with the network branches. Small black circles indicate hypothetical alleles, as predicted by the model.

List of appendices

Appendix 1. Basic data on Northern and Southern chamois individuals included in the study.

Appendix 2. A summary of the DRB alleles found in each chamois individual using single-strand conformation polymorphism (SSCP)/Sanger sequencing and amplicon-based next-generation sequencing (NGS).

Appendix 3. Amplicon depths of 28 individuals, and the number of reads of true alleles and artefacts within each amplicon. true = true alleles; X = 1 bp substitutions; D = deletions; I = insertions; H = homopolymer indels, XDIH = sequence with any combination of at least two errors; leftovers = low frequency variants, chimeras or sequences containing many errors which could not be classified into the major clusters.

Appendix 4. Percentages of true alleles and artefacts within each amplicon and averaged. true = true alleles; X = 1 bp substitutions; D = deletions; I = insertions; H = homopolymer indels, XDIH = sequence with any combination of at least two errors.

Appendix 5. Percentage of artefacts resulting from 1 bp substitution within each amplicon and averaged.

1. GENERAL INTRODUCTION

Chamois is the most common mountain-dwelling ungulate in Europe and Asia Minor, with two species currently recognized: Northern chamois, *Rupicapra rupicapra* (with seven subspecies) and Southern chamois, *Rupicapra pyrenaica* (with three subspecies) (Corlatti et al., 2011). Unfortunately, many isolated chamois populations are threatened by disease, habitat loss, poaching, and hybridization due to translocations of other subspecies. There is also growing evidence that chamois are affected by global warming (Chirichella et al., 2021), as mountain habitats are most at risk due to climate change (White et al., 2018). The ability of natural populations to adapt to new environmental conditions is critical to their survival and is largely determined by the genetic variation present in each population. In the conservation and management of species, genetic monitoring is critical to ensure that appropriate measures are taken to maintain the viability and adaptive potential of populations, especially in the case of small and declining populations (Leroy et al., 2018). Because neutral genetic variation is not sufficient to understand all the mechanisms that influence genetic variation (Eizaguirre and Baltazar-Soares, 2014), it has been suggested that adaptive genetic diversity in natural populations should also be monitored. In some cases, patterns of differentiation among populations can only be detected at functional genes under selection (Awadi et al., 2018). Studying the molecular polymorphism of adaptive genes (i.e., genes that directly influence fitness) can help understand how adaptive genetic variation is generated and maintained within populations (Funk et al., 2012). Among one of the most polymorphic genes in vertebrates are the major histocompatibility complex (MHC) genes, which are responsible for the adaptive immune response (Sommer, 2005). Polymorphism of MHC genes has been shown to be related to individual fitness and thus to long-term persistence of populations, making these genes important markers for various fitness traits, including factors important for population viability, such as resistance to parasites, survival, and reproductive success (Sommer, 2005). The most pronounced polymorphism of the MHC molecules occurs in the amino acid residues encoding the antigen binding groove (Moreno-Santillán et al., 2016). Because of the function of MHC molecules, pathogen-driven balancing selection is generally thought to be an important evolutionary force in maintaining MHC polymorphism (Spurgin and Richardson, 2010), but it is also generated by gene duplication, resulting in extensive copy number variation (O'Connor et al., 2016). Thus, the genetic diversity of the MHC is characterized not only by extreme allelic polymorphism and high nucleotide diversity, but also by the number of duplicated genes.

1.1. Research hypotheses and objectives

Hypotheses

1. There is a high degree of nucleotide and amino acid polymorphism in exon 2 of the major histocompatibility complex class II DRB locus in chamois populations.
2. Diversity and frequency of exon 2 alleles of DRB locus in a population depend on local adaptation and on the size and isolation of the population, which are influenced by natural selection and neutral evolutionary processes (migration and genetic drift).
3. Gene duplication is present, while the number of duplicated loci varies among populations, especially between populations that are spatially isolated.

Objectives

1. Investigate the genetic diversity in exon 2 of the major histocompatibility complex class II DRB locus in chamois populations.
2. Compare diversity and frequency of exon 2 alleles of DRB locus among chamois populations.
3. Investigate the duplication of the DRB locus within the genus *Rupicapra*.

2. OVERVIEW OF FORMER RESEARCH

2.1. Chamois, *Rupicapra* spp.

Chamois, *Rupicapra* spp. is a species in the family Bovidae, subfamily Caprinae, and one of the most iconic mammals distributed on most of the medium to high altitude mountain massifs in Europe and Asia Minor. Currently, classification of chamois considers two species according to its morphological and behavioural characteristics: the Northern chamois *Rupicapra rupicapra* and the Southern chamois *Rupicapra pyrenaica* (Figure 1). The Northern chamois is native to mountainous parts of central and southern Europe and Asia Minor, where it occurs as seven subspecies: *R. r. cartusiana* - Chartreuse chamois, *R. r. rupicapra* – Alpine chamois, *R. r. tatrica* – Tatra chamois, *R. r. carpatica* – Carpatian chamois, *R. r. balcanica* – Balkan chamois, *R. r. caucasica* - Caucasian chamois, and *R. r. asiatica* - Anatolian chamois. The Southern chamois is endemic to south-west Europe, where it occurs as three subspecies: *R. p. pyrenaica* - Pyrenean chamois, *R. p. parva* - Cantabrian chamois and *R. p. ornata* - Apennine chamois (Corlatti et al., 2011). Distribution ranges of *Rupicapra rupicapra* ssp. and *Rupicapra pyrenaica* ssp. are depicted in Figure 2.

Figure 1. Northern (left; photo: Denis Bertanzetti) and Southern (right; photo: Javier Ara) chamois (Corlatti et al., 2021).

According to the IUCN Red List of Threatened Species, both species of chamois have been assessed as Least Concern, conservation status of both species has been recently assessed by Herrero et al. (2020) and Anderwald et al. (2021). Northern chamois is widespread as a species and have stable population trend with a large population of nearly 500,000 individuals, the bulk of the population being Alpine subspecies *R. r. rupicapra*, while the Southern chamois have increasing population trend with population of 71,500 individuals (Corlatti et al., 2022). Both species are listed in Appendix III ("protected fauna species") of the Convention on the Conservation of European Wildlife and Natural Habitats (Bern Convention of 1979) and in Annex V ("animal and plant species of community interest whose taking in the wild and exploitation may be subject to management measures") of the European Union Directive 92/43 on the Conservation of Natural Habitats and Wild Fauna and Flora (Habitats Directive of 1992).

Although neither chamois species is threatened, most subspecies are found in restricted areas where they face various threats, including poaching, overhunting, habitat loss and degradation, human disturbance, disease, competition with livestock, hybridization due to translocations of other subspecies, and climate change (Corlatti et al., 2022). Four of the seven subspecies of Northern chamois are threatened, and three are declining in population size or range (Red List Assessment; Anderwald et al., 2021). Anatolian chamois *R. r. asiatica* is listed as Endangered, C1+2a(i). Its population size has declined catastrophically by approximately 60-70% in recent decades due to intense human impacts (Ambarlı, 2014). The total population is estimated at 500 to 750 individuals. Chartreuse chamois *R. r. cartusiana* is listed as Vulnerable, D1+2. It is restricted to a single mountain and the population size was estimated at 1,500 individuals in 2017-2018. Due to the earlier introduction of the *R. r. rupicapra*, the subspecies is threatened with genetic extinction through hybridization. Caucasian chamois *R. r. caucasica* is listed Vulnerable, C1. The population size is approximately 6,000 mature individuals. Since the 1960-1970s, the population has declined by approximately 70%, and this trend is believed to continue. Tatra chamois *R. r. tatrica* is listed as Endangered, B1+2ab. The subspecies has a restricted geographic range and a small population of 1,350 individuals. One of the major concern are low genetic variation and a predicted problem of hybridization with the introduced *R. r. rupicapra* (Zemanová et al., 2015). The Southern chamois is currently increasing in numbers and range (the red list assessment; Herrero et al., 2020). However, Apennine subspecies *R. p. ornata* is listed as Vulnerable because it lives in very small populations (estimated at 1,500 individuals) and has a limited range.

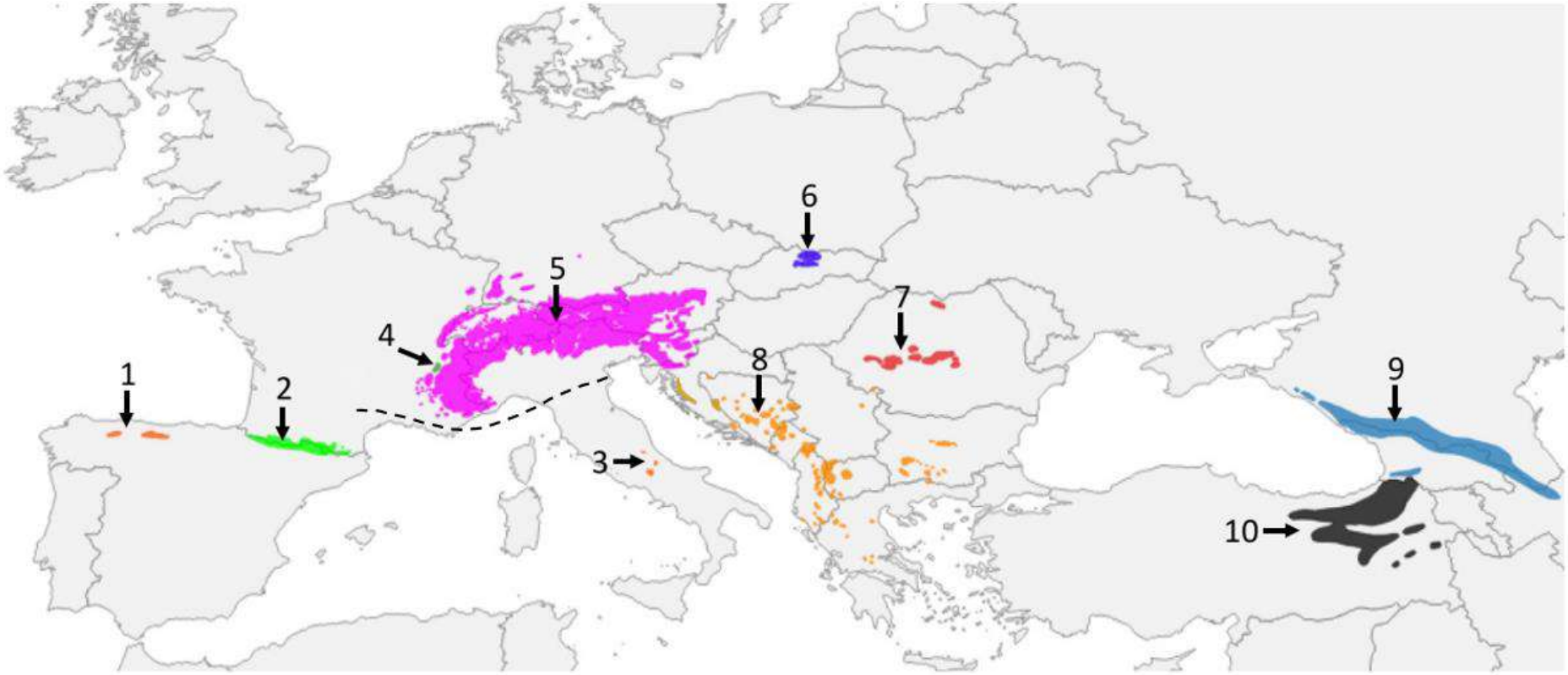

Figure 2. Distribution range of *Rupicapra* spp. The Southern chamois, *Rupicapra pyrenaica*: (1) *parva*, (2) *pyrenaica*, (3) *ornata*. The Northern chamois, *Rupicapra rupicapra*: (4) *cartusiana*, (5) *rupicapra*, (6) *tatrica*, (7) *carpatica*, (8) *balcanica*, (9) *caucasica*, (10) *asiatica*. Dashed line indicates the boundary between the two species (taken and modified from Corlatti et al., 2021).

2.2. Major histocompatibility complex

Among the most polymorphic genes in vertebrates are the major histocompatibility complex (MHC) genes, which encode cell-surface glycoproteins involved in antigen presentation to T lymphocytes and have a crucial function in the adaptive immune response against pathogens (Piertney and Oliver, 2006; Sommer, 2005). In contrast to B cell receptors which can recognize free antigen, T cell receptors recognize the antigen only if it is exposed to one of the MHC receptors, and therefore MHC molecules are crucial in triggering the immune response. These antigenic peptides are held within the binding groove of a MHC molecule. These membrane glycoproteins function as highly specialized antigen-presenting molecules with grooves that form unusually stable complexes with peptide ligands and present them on the cell surface for recognition by T cell receptors (Punt et al., 2019).

The MHC molecules has been classified in three classes: class I, class II, and class III. Members of the first two classes have a similar shape and both are responsible for displaying antigen to T cells, although they differ according to the origin of the antigens presented and in the way in which their quaternary structures are generated. MHC class I molecules presents endogenously and intracellular pathogens derived peptides to cytotoxic lymphocytes T, whereas class II molecules presents bacteria and parasite derived peptides to helper T cells (reviewed by Sommer, 2005). MHC class III region genes code for variety of circulating molecules which also have roles in the immune response, although generally do not involve direct presentation of antigen fragments to T cells.

MHC class II are heterodimers consisting of two glycoprotein chains, the light chain (α) and the heavy chain (β). The chains are connected by non-covalent interactions. Both MHC class II chains are membrane-bound glycoproteins that contain external domains, a transmembrane segment, and a cytoplasmic domain (see Figure 3, top). There are two external domains in each chain: α1- and α2-domains in one chain and β1- and β2-domains in the other. The membrane-distal α1- and β1-domains form the peptide-binding groove that can accommodate a peptide containing 13 to 18 residues (see Figure 3, middle and bottom). The groove in MHC class II molecules is formed by the association of two separate chains which is an important difference in relation to peptide-binding groove of MHC class I (Punt et al., 2019).

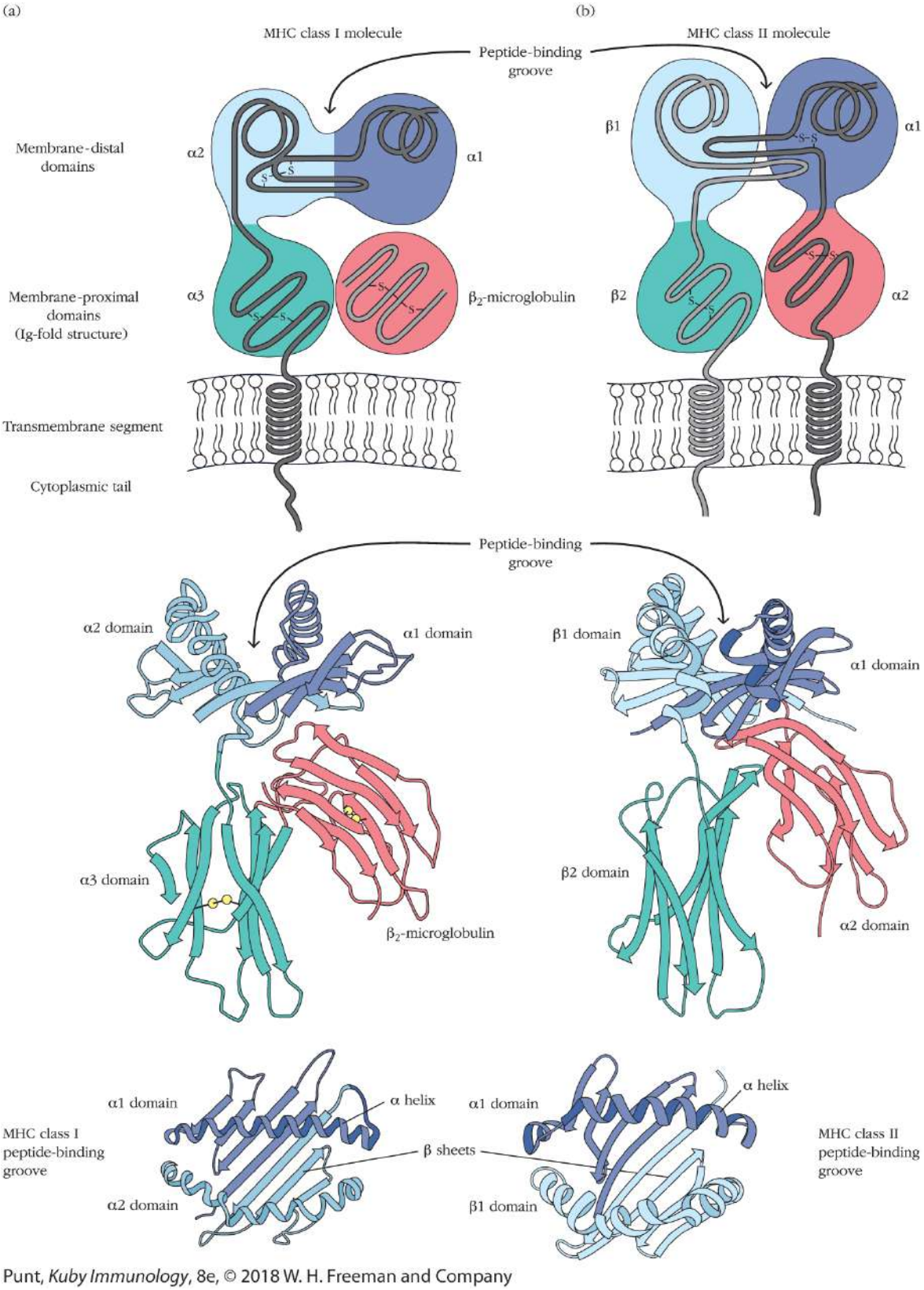

Figure 3. Schematic representations of MHC class I (a) and MHC class II (b) molecules, showing the external domains, transmembrane segments, cytoplasmic domain, and peptide-binding groove (Punt et al., 2019).

MHC genes exhibit extraordinary intraspecific and intraindividual genetic diversity due to both the large number of alleles per locus and the presence of multiple paralogous, and presumably functionally equivalent loci (Zagalska-Neubauer et al., 2010). Remarkably, polymorphism is most pronounced in MHC genes at amino acid sites encoding the peptide-binding region (PBR), as their greater allelic diversity should be associated with a response to a broader range of pathogens (Hedrick, 2002). At these sites, an excess of non-synonymous substitutions is usually observed, making MHC genes prime examples of positive selection (Piertney and Oliver, 2006). Because of their functional importance and exceptional diversity, MHC genes are therefore excellent markers to study the mechanisms that shape genetic variation in populations.

Pathogen-driven balancing selection has been proposed as one of the major evolutionary forces for the maintenance of MHC polymorphism, leading to the sharing of allelic lineages between different animal species, resulting in the pattern of trans-species polymorphism (Piertney and Oliver, 2006). Concerning balancing selection, mechanisms resulting from the Red Queen dynamics of host-pathogen coevolution are thought to maintain the high diversity of MHC genes at the population level (Ejsmond and Radwan, 2015). Evolutionary mechanisms such as overdominance (i.e. fitness advantage for heterozygotes than for both corresponding homozygotes) and negative frequency-dependent selection (i.e. rare allele advantage where individuals with rare genotypes have higher fitness) can potentially favor novel alleles that arise by mutation (Bernatchez and Landry, 2003; Sommer 2005). Although pathogen-mediated selection is important for maintaining functional MHC variation in the wild, other mechanisms, such as disassortative mating preferences, maternal-foetal interactions, recombination, and gene duplication have been proposed as alternative or complementary mechanisms for maintaining MHC diversity (Juola and Dearborn, 2011; Miller and Lambert, 2004; Spurgin and Richardson, 2010).

Genetic diversity affects population viability and its adaptive potential (Frankham, 2005; Frankham et al., 2012), and even though presumably neutral markers are often used to characterize the levels of this diversity, neutral variation is not sufficient to understand all the mechanisms that influence genetic variation (Eizaguirre and Baltazar-Soares, 2014). In many cases, differentiation patterns between populations can only be detectable at functional genes under selection (Awadi et al., 2018). Changes in the frequency of MHC alleles therefore have adaptive value, and understanding how such functional variation arises and is maintained within populations is essential for species conservation and optimal management decisions (Radwan et al., 2010; Sagonas et al., 2018).

2.3. Previous research on MHC in chamois

So far, only several studies have been published on MHC genes in chamois, despite its widespread distribution (Alvarez-Busto et al., 2007; Cavallero et al., 2012; Mona et al., 2008; Schaschl et al., 2004, 2005; Zemanová et al., 2015). Most of them were carried out on Alpine chamois *R. r. rupicapra*. To our knowledge, all studies on MHC genes in chamois carried out so far included analyses of the exon 2 of DRB locus.

However, due to generally small sample sizes analysed in individual studies, as well because many parts of chamois habitats have not been analysed, general conclusions on DRB variability in chamois populations cannot be drawn. Moreover, allelic diversity in chamois could be much higher than revealed in the up-to-date studies. Because only very small part of immune system was studied, it is also possible, however, that required immunodiversity is present on some other loci.

One of the first studies carried out by Schaschl et al. (2004) was conducted particularly on expressed genes and showed that alleles exhibited a high degree of nucleotide and amino acid diversity in comparison with alleles of other Caprinae species. Specifically, the authors found 19 alleles in a sample of 59 chamois from diverse regions of the Eastern Alps. High dN/dS ratio observed in chamois suggests strong positive selection on the DRB locus for maintaining high allelic variability. Moreover, the dN/dS ratio was higher in chamois then in other Caprinae included in the study. Rate of non-synonymous substitutions was generally lower in chamois than in the other analysed ruminants, but it was the rate of synonymous substitutions that was generally very low, which resulted in the high dN/dS ratio. The authors indicated that low dS rate that resulted in high dN/dS ration may reflect the young age of most of the chamois alleles. Similar group of authors suggested that recombination, in addition to the expected positive selection, has a key role in generating MHC diversity in a range of ungulate species including chamois (Schaschl et al., 2005).

The extremely high dN/dS ratio and the presence of trans-species polymorphisms were noted by Mona et al. (2008). The results suggest that a strong long-term balancing selection effect has operated at this locus throughout the evolutionary history of this species. The authors also examined chamois demographic processes using mitochondrial DNA polymorphism patterns and suggest that both demography and balancing selection have likely influenced the pattern of genetic variation within and among chamois populations.

Zemanova et al. (2015) were investigating conservation status of endangered Tatra chamois (*R. r. tatrica*), which lives only in the Tatra Mountains (Slovakia and southern Poland). A back-up population of the Tatra chamois was established between 1969 and 1976 in the Low Tatra Mts. to compensate for the risks associated with decreasing population size. Also, before recognition of the Tatra chamois as a unique subspecies, two populations of Alpine chamois were introduced into the adjacent areas by hunters in the early 1960s.

The authors found drastically reduced variation of the MHC DRB gene in the studied populations, with a single allele of exon 2 in the native Tatra population, which they attributed to intense demographic bottlenecks. Introgressive hybridization was also detected between the native Tatra chamois population and the introduced Alpine populations, with 19% of the genome introgressed from the Alpine chamois. Although this could lead to loss of unique genetic composition and disrupt local genetic adaptations, i.e., cause outbreeding depression, the authors also suggest that introgression could be beneficial to Tatra chamois by increasing their low genetic variation and improving their adaptability to environmental change.

Low genetic variation was found in both subspecies of *R. pyrenaica* (Alvarez-Busto et al., 2007), with seven different alleles in 98 samples of *R. p. pyrenaica* and only three alleles in 32 samples of *R. p. parva*. Because of the difference in sample size, the rarefaction method was used and the low number of alleles of *R. p. parva* was emphasized. Bottlenecks due to hunting pressure and recent parasite infections due to sarcoptic mange are likely the cause of the low allelic diversity of the subspecies *R. p. parva* (Alvarez-Busto et al., 2007).

In all the above studies, no more than one copy of the DRB locus was found. Fuselli et al. (2018) proposed an approach for genotyping complex genomic regions by combining multiple NGS methods, but could not detect sequenced multiple copies, only co-amplification of short amplicons resulting from non-specific primer binding. However, it cannot be excluded that multiple copies are present only in some unsampled individuals of the population.

2.4. SSCP and Ion Torrent sequencing

The introduction of high-throughput sequencing technologies, also known as next-generation sequencing (NGS), has enabled large-scale assessment of genetic variation at reasonable times and costs (Reuter et al., 2015). In addition, NGS methods have improved our ability to genotype highly polymorphic multigene families such as the MHC (Babik et al., 2009; Lighten et al., 2014b; Zagalska-Neubauer et al., 2010).

Great efforts have been made in evolutionary and population studies to genotype MHC loci (Grogan et al., 2016a; Huang et al., 2019). However, the task is still quite demanding and challenging due to the complex genomic organisation and high sequence variation of MHC loci, as well as the difficulties in separating true alleles from artefacts (Biedrzycka et al., 2017; Lighten et al., 2014a). MHC genes, which are often present in multiple copies, vary widely between and within species, making the identification of all alleles carried by an individual and the reconstruction of its multilocus genotype very difficult (Rekdal et al., 2018). Consequently, the issues that cause significant difficulties in MHC genotyping are (i) frequent gene duplications and variation between haplotypes in the number of loci within and between species, (ii) difficult design of locus-specific primers, (iii) varying degrees of concerted evolution, and (iv) the presence of pseudogenes that cause additional difficulties in identifying functional variants.

Various techniques such as restriction fragment length polymorphism, single-strand conformation polymorphism (SSCP), denaturing gradient gel electrophoresis, and reference strand-mediated conformational analysis in combination with cloning have been used for MHC genotyping (Babik, 2010).

The SSCP method is based on the observation that single-stranded DNA fragments with different DNA sequences will assume sequence-specific conformation when electrophoresed under non-denaturing conditions (Sunnucks et al., 2000). Since a change in a single base is sufficient to cause changes in the tertiary structure of single-stranded DNA fragments, SSCP is capable of detecting substitutions of individual bases. The SSCP method has been shown to be robust and has high sensitivity in detecting DNA sequence variations in MHC genes (Garrigan and Hedrick, 2001; Noakes et al., 2003). However, it can encounter problems when amplification of some alleles is less efficient or when certain alleles are difficult to distinguish (Babik, 2010). It becomes particularly problematic and unreliable when multiple co-amplifying copies are present in the sample (Babik, 2010).

Despite significant advances in high-throughput sequencing technologies, the genotyping of MHC systems and the ability to discriminate between true alleles and artefacts is more challenging as the number of co-amplifying genes increases with this method. Of the available high-throughput sequencing platforms, Ion Torrent and Illumina are among the most appropriate choices for MHC genotyping due to the ultra-high coverage and read lengths, offering the potential to overcome this limitation (Rekdal et al., 2018).

3. MATERIALS AND METHODS

3.1. Study area and data collection

Samples of 130 chamois from populations covering the majority of the distribution range of the genus *Rupicapra* were obtained. Samples were collected during regular hunts or after natural death or in museum collections. All study was done according to the ethical and welfare standards presented in the Official Gazette of the Republic of Croatia (OG 102/2017, Animal Protection Act) and Regulation on the Protection of Animals Used for Scientific Purposes (OG 55/13), with the approval of the Bioethical Committee for the Protection and Welfare of Animals of the University of Zagreb Faculty of Agriculture (UR.BR. 251-71-29-02/19-21-1). All the research was done in accordance with ARRIVE guidelines.

Because some of the samples were unsuccessful in laboratory procedures, this study was performed on a total of 110 chamois, 102 tissue samples, and eight bone samples. The majority of samples were Northern chamois, specifically: subspecies *R. r. rupicapra* – Alpine chamois (57), *R. r. balcanica* – Balkan chamois (31), *R. r. tatrica* – Tatra chamois (6), *R. r. carpatica* – Carpatian chamois (5), *R. r. asiatica* - Anatolian chamois (2), *R. r. caucasica* - Caucasian chamois (1). For Southern chamois we had eight samples: *R. p. pyrenaica* - Pyrenean chamois (5), *R. p. parva* – Cantabrian chamois (1) and *R. p. ornata* - Apennine chamois (2) (Figure 4). In addition, 20 published sequences of exon 2 of the MHC class II DRB gene in chamois deposited in GenBank were used (Table 1) that fully covered region of interest.

Table 1. DRB exon 2 alleles of chamois previously reported in the literature and downloaded from GenBank. Alleles in bold were also detected in this study. Alleles Rupy-DRB02 and Ruru-DRB01 are identical at DNA sequence level.

	Allele name	Accession number	Reference
1.	Rupy-DRB01	AY212149	Alvarez-Busto et al., 2007
2.	**Rupy-DRB02**	AY212150	Alvarez-Busto et al., 2007
3.	Rupy-DRB03	AY212151	Alvarez-Busto et al., 2007
4.	**Rupy-DRB04**	AY212152	Alvarez-Busto et al., 2007
5.	Rupy-DRB05	AY212153	Alvarez-Busto et al., 2007
6.	Rupy-DRB06	AY212154	Alvarez-Busto et al., 2007

	Allele name	Accession number	Reference
7.	Rupy-DRB07	AY212155	Alvarez-Busto et al., 2007
8.	Rupy-DRB08	AY212156	Alvarez-Busto et al., 2007
9.	Rupy-DRB09	AY212157	Alvarez-Busto et al., 2007
10.	Rupy-DRB10	AY898752	Schaschl et al., 2005
11.	**Rupy-DRB11**	AY898753	Schaschl et al., 2005
12.	**Rupy-DRB12**	AY898754	Schaschl et al., 2005
13.	**Rupy-DRB13**	AY898755	Schaschl et al., 2005
14.	**Ruru-DRB01**	AY368437	Schaschl et al., 2004
15.	Ruru-DRB02	AY368438	Schaschl et al., 2004
16.	Ruru-DRB03	AY368439	Schaschl et al., 2004
17.	**Ruru-DRB04**	AY368440	Schaschl et al., 2004
18.	Ruru-DRB05	AY368441	Schaschl et al., 2004
19.	Ruru-DRB06	AY368442	Schaschl et al., 2004
20.	Ruru-DRB07	AY368443	Schaschl et al., 2004
21.	Ruru-DRB08	AY368444	Schaschl et al., 2004
22.	**Ruru-DRB09**	AY368445	Schaschl et al., 2004
23.	Ruru-DRB10	AY368446	Schaschl et al., 2004
24.	Ruru-DRB11	AY368447	Schaschl et al., 2004
25.	Ruru-DRB12	AY368448	Schaschl et al., 2004
26.	Ruru-DRB13	AY368449	Schaschl et al., 2004
27.	Ruru-DRB14	AY368450	Schaschl et al., 2004
28.	**Ruru-DRB15**	AY368451	Schaschl et al., 2004
29.	Ruru-DRB16	AY368452	Schaschl et al., 2004
30.	**Ruru-DRB17**	AY368453	Schaschl et al., 2004
31.	**Ruru-DRB18**	AY368454	Schaschl et al., 2004
32.	**Ruru-DRB19**	AY368455	Schaschl et al., 2004

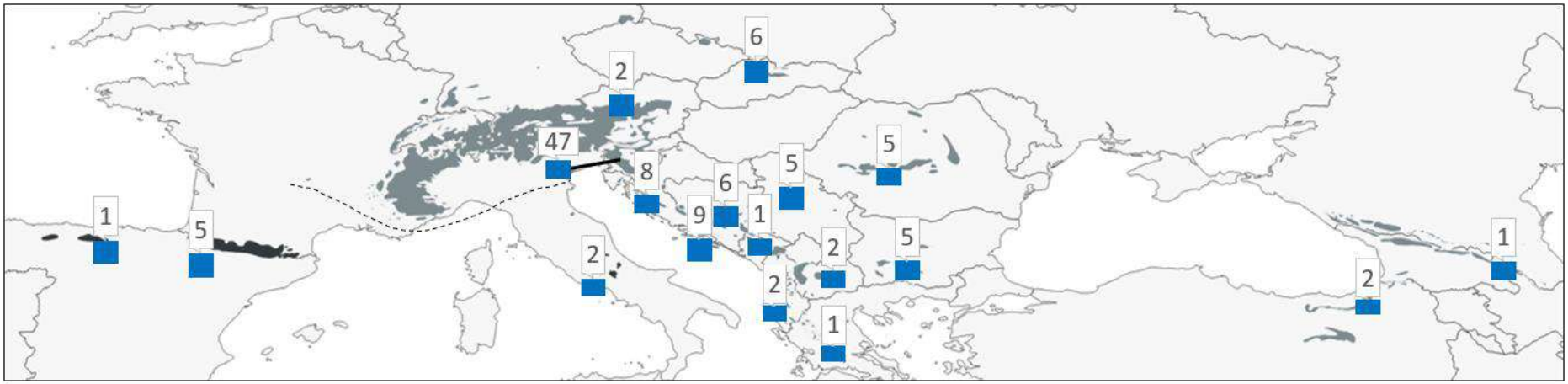

Figure 4. Geographic distribution of Northern (grey) and Southern (dark grey) chamois populations based on the IUCN Red List of Threatened Species data. The dashed line indicates the boundary between the two species. Numbers refer to the number of samples successfully analysed in this study.

3.2. Laboratory procedures

DNA was isolated from tissue samples using a peqGOLD Tissue DNA Mini Kit (VWR International) according to the manufacturer's instructions, while the bone samples were treated following the procedure described in Buzan et al. (2020). A 236 bp fragment of the second exon of the MHC class II DRB gene was amplified with primers HL030 (ATCCTCTCTGCAGCACATTTCC) and HL032 (TCGCCGCTGCACAGTGAAACTCTC) (Schaschl et al., 2004). In order to identify individuals, forward primers were designed to include the following motifs: (i) adaptor sequence required for Ion Torrent sequencing, (ii) unique 10 - 12 bp IonXpress barcode, (iii) barcode “GAT” linker, and (iv) forward primer HL030. PCR amplification was performed in triplicates in 25 µL reaction mixtures containing a 5 µL DNA template, 5X reaction buffer, 5 µL Q solution (Qiagen, Hilden, Germany), 3 µM $MgCl_2$, 0.25 mM dNTPs, 0.15 µM of each primer, and 0.08 U HotStarTaq (Qiagen). The PCR program included an initial denaturation step at 95 °C for 2 min, 40 denaturation cycles at 95 °C for 20 s, primer annealing at 60 °C for 30 s, and primer extension at 72 °C for 45 s, and a final elongation step at 72 °C for 10 min.

PCR products from triplicates were pooled and purified with Agencourt AMPure XP beads (Agencourt Bioscience Corporation, Beverly, MA, USA). The concentrations of the pooled and purified amplicons were estimated with the Qubit 3.0 Fluorometer using the Qubit dsDNA High Sensitivity Assay Kit (Thermo Fisher Scientific). Then, amplicons were normalized to 5 ng, pooled and purified again with Agencourt AMPure XP beads. Size and quality of the pooled amplicons were verified using the Agilent DNA High Sensitivity Kit on the 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA). The final library was normalized to 100 pM and sequenced with the Ion Torrent S5 on a 314 chip (Thermo Fisher Scientific). Amplicon refers to a set of reads derived from a single PCR and includes all the MHC II DRB sequences of a particular individual.

3.3. MHC genotyping

Allele calling was conducted through the pipeline based on the Amplicon Sequence Assignment (AmpliSAS) web tool designed for high-throughput genotyping of duplicated polymorphic gene families, such as the MHC (Sebastian et al., 2016). AmpliSAS is a part of AmpliSAT, a suite of next-generation amplicon sequencing analysis tools (Sebastian et al., 2016), available as a web server at: http://evobiolab.biol.amu.edu.pl/amplisat/index.php.

Initial quality and length filtering of raw data was performed with AmpliCLEAN by removing reads with a Phred quality score below 30 and all reads shorter than 279 bp and longer than 289 bp (Table 2, Figure 5).

Table 2. The relationship between sequencing error rate and sequencing base quality value. Phred score represents the base quality value, Phred = $-10\log_{10}p$; p = probability call is incorrect.

Phred score	Probability of incorrect base call	Base call accuracy
10	1/10	90%
20	1/100	99%
30	1/1000	99.90%
40	1/10000	99.99%

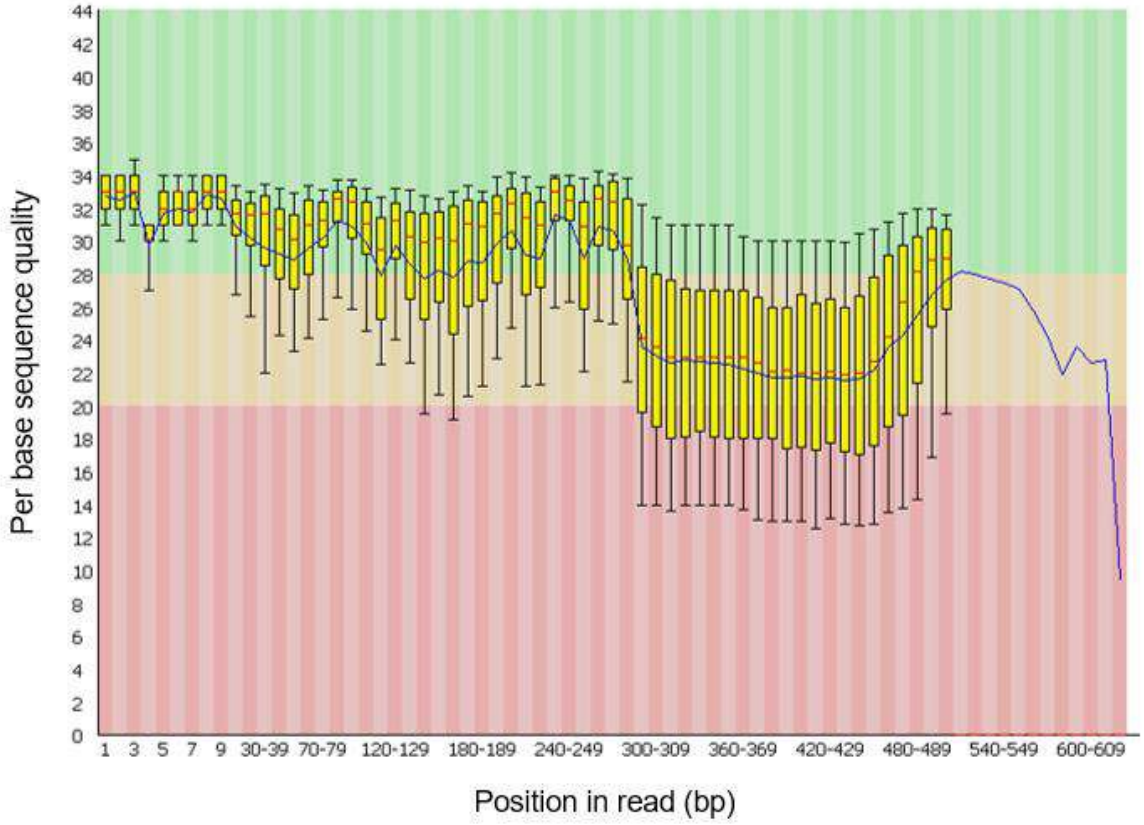

Figure 5. Overview of the range of quality values for all bases at each position in the FastQ file. The y-axis of the graph shows the quality values. The background of the graph divides the y-axis into calls of very good quality (green), calls of reasonable quality (orange) and calls of poor quality (red). The central red line is the median value; the yellow box represents the interquartile range (25-75%); the upper and lower whiskers represent the 10% and 90% points; the blue line represents the average quality. Figure was generated using FastQC (Babraham Bioinformatics, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

For preliminary examination of the data, AmpliSAT's AmpliCHECK tool was used with Ion Torrent's default parameters: 0.5% substitution error rate, 1% indel error rate, and a minimum per amplicon frequency of 1%.

AmpliSAS default parameters were also used for Ion Torrent sequencing technology: a substitution error rate of 0.5% and an indel error rate of 1%. Exact length was required for the dominant sequence within a cluster. Based on previous work in chamois (Alvarez-Busto et al., 2007; Fuselli et al., 2018; Mona et al. 2008; Schaschl et al. 2004, 2005; Zemanová et al. 2015), no more than two DRB variants per individual were expected, so the threshold for "minimum dominant frequency" of clustering was kept at a default 25%. Variants with a frequency of less than 1% within an amplicon were discarded. The maximum number of reads per amplicon that can be processed by AmpliSAS is 5,000, so for the amplicons with more than 5,000 reads, only the first 5,000 reads were used for the analysis.

AmpliSAS clusters true variants with their potential artefacts based on platform-specific error rates. This increases the coverage of true variants, i.e. alleles, as the coverage of artefacts is included in the coverage of true variants. The AmpliSAS workflow is divided into three main steps:

1. de-multiplexing reads into amplicons and unique sequences based on matching of primer and barcode sequences,
2. clustering amplicon sequences, where potential alleles with sequencing errors are clustered, increasing the read depth of the true alleles,
3. filtering sequences based on user-defined parameters and allele assignment (Figure 6).

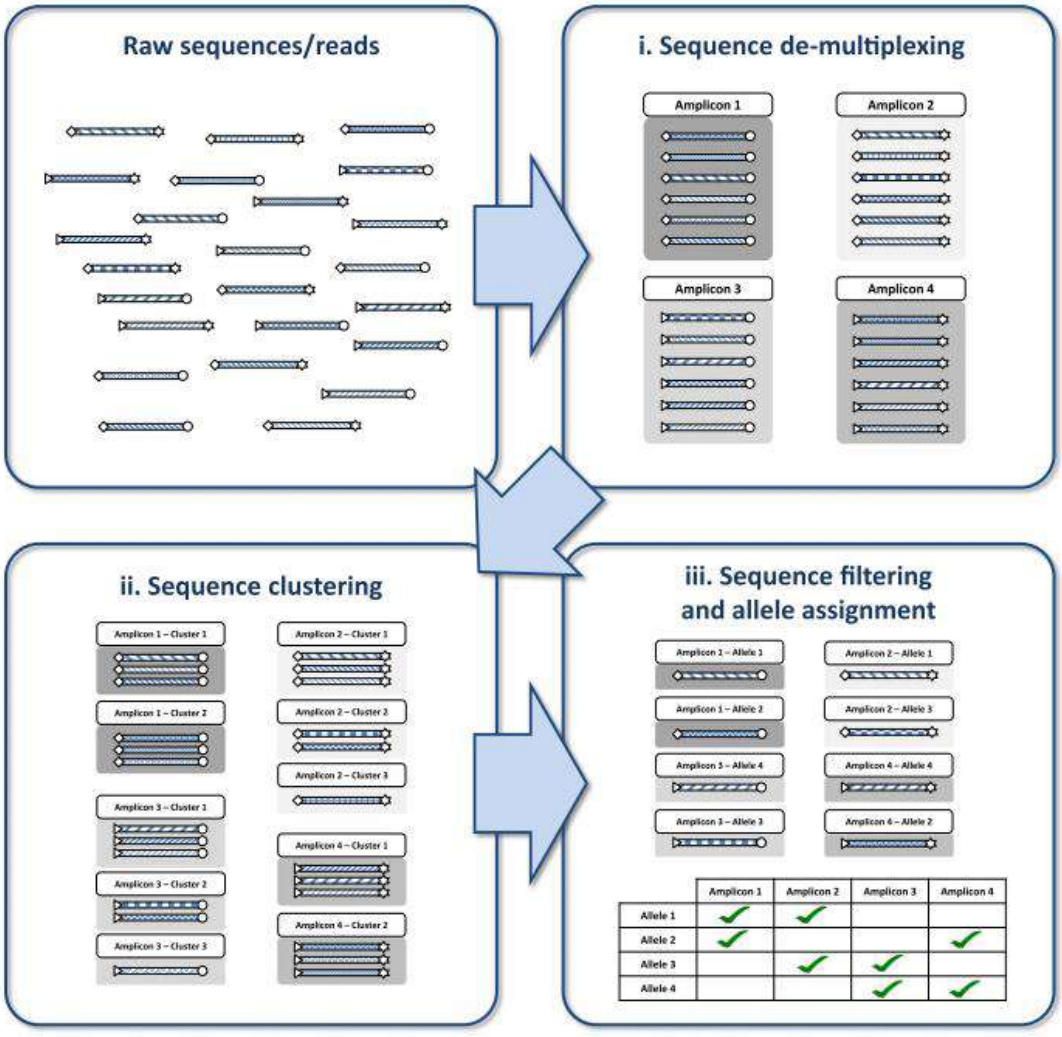

Figure 6. AMPLISAS workflow schema (Sebastian et al., 2016).

True DRB variants of the second exon were aligned and translated into protein sequences to verify whether there was evidence for pseudogenes, such as the presence of premature stop codons.

Fourteen novel DRB exon 2 sequences were identified in this study and designated as Ruru-DRB from 41 to 54, in accordance with the nomenclature previously established by Klein et al. (1990), starting from the last haplotype proposed by Mona et al. (2008). The novel alleles have been deposited in GenBank under accession numbers MT813042–MT813044 and OL421550 - OL421560.

3.4. MHC DRB allelic diversity

Unique sequences were aligned, edited, and interpreted using MEGA X v.10.0.5 (Kumar et al., 2018) with alleles downloaded from GenBank for comparison (Table 1). Sequence polymorphism measures were calculated, including the number of segregating sites (S), average number of nucleotide differences (k) and nucleotide diversity (π) for whole sequences as well as for PBR and non-PBR sites separately with DnaSP v.6.12.01 (Rozas et al., 2017). Putative PBR sites were identified based on human HLA molecules (Brown et al., 1993). Estimations of mean pairwise nucleotide distances (Jukes-Cantor model with a gamma distribution), and Poisson-corrected amino acid distances were calculated in MEGA X for all, PBR and non-PBR sites.

HP-RARE (Kalinowski, 2004) was used to estimate allelic richness (Ar), expressed as the expected number of alleles assuming the smallest sample size (i.e. eight individuals). To gain better insight into MHC gene differentiation in chamois, individuals were pooled based on geographic location or species/subspecies (in the case of a small sample size), (see Appendix 1). Pairwise values of genetic differentiation (F_{ST}) between groups were estimated and tested using Arlequin v.3.5 (Excoffier and Lischer, 2010).

To reveal genetic differentiation between Northern and Southern chamois and between Northern chamois subspecies, Discriminant Analysis of Principal Components (DAPC) was performed with “adegenet” package v.2.0.1 (Jombart, 2008; Jombart et al., 2010) in R (R Core Team, 2020).

To investigate functional MHC class II DRB diversity within genus *Rupicapra*, alleles were clustered into supertypes. 45 unique nucleotide sequences of MHC class II DRB alleles were used (Appendix 1).

The first step was to identify amino acids that reflect functional differences between alleles. Therefore, clustering based on amino acid polymorphism at the positively selected amino acid sites (PSSs) was performed. For PSSs all codons having the Bayes Empirical Bayes (BEB) posterior probability >95% in model M8 were retained (see “Signatures of selection and recombination on MHC DRB alleles” paragraph below). The PSS of each allele was numerically characterized by a set of five physicochemical descriptors for each amino acid (Doytchinova and Flower, 2005).

To cluster the alleles into supertypes, DAPC was performed with “adegenet” package in the following order. First, clusters (k) were defined using the *find.clusters()* function and all principal components (PCs) were retained. Different clustering solutions were compared based on the Bayesian information criterion (BIC) values for the increasing number of clusters. The optimal number of clusters was then chosen as the number of clusters with the lowest BIC after which the BIC decreases by a negligible amount (Jombart et al., 2010). In this case, the curve indicated five clusters should be retained (Figure 7).

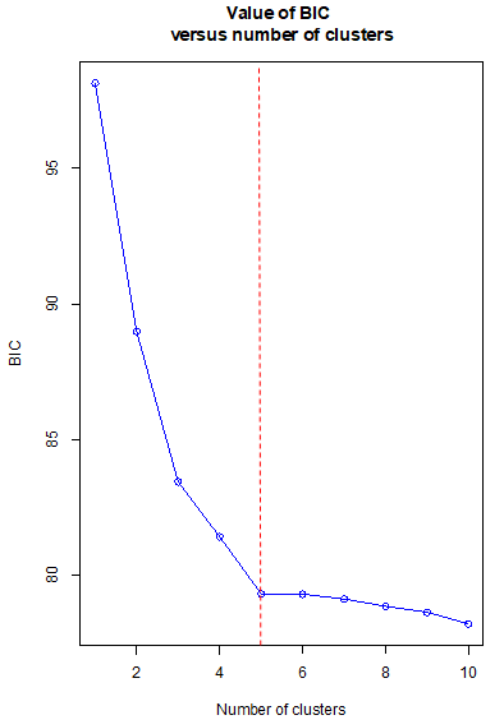

Figure 7. Choice of the number of clusters (k) for discriminant analysis of principle components (DAPC) for MHC supertypes. Dashed, vertical line indicates the chosen value of k.

After that, DAPC was applied and the optimal number of PCs was chosen to retain for discriminant function analysis using the cross-validation with *xvalDapc()* function. Finally, posterior assignments from the discriminate function analysis were used to assign MHC variants to the supertypes.

3.5. Detecting signatures of selection and recombination on MHC DRB alleles

To test whether positive selection is operating on the analysed gene, the ratio (ω) of non-synonymous (dN) and synonymous (dS) substitution rate was calculated separately on the entire MHC class II DRB exon 2 sequences and on the extracted peptide-binding region. Positive selection favours non-synonymous substitutions, so that the evolutionary distance based on non-synonymous substitutions is greater than that based on synonymous substitutions. Therefore, positive selection is indicated when $\omega > 1$, purifying (negative) when $\omega < 1$, and assuming neutral evolution, $\omega = 1$ (Nei and Gojobori, 1986). A one-tailed codon-based Z-test of selection with standard errors resulting from 10,000 bootstrap replicates in MEGA X was applied. Analysis was performed using the Nei-Gojobori model (Nei and Gojobori, 1986) with Jukes-Cantor correction.

Because positive selection is unlikely to affect an entire gene, the programme EasyCodeML (Gao et al., 2019) was used to identify codon sites affected by positive selection based on a Bayesian approach. The models implemented in the analysis were M0 (one-ratio of ω), M1a (nearly neutral, two site classes of ω: <1, =1) and M7 (ω varies among sites according to beta distribution), which do not allow for positive selection and serve as null models for M2a (positive selection, three site classes of ω: <1, =1 >1), M3 (discrete, three site classes of ω: <1,≈1 >1) and M8 (beta distribution and $\omega > 1$) respectively. The nested models (M0 vs. M3, M1a vs. M2a, and M7 vs. M8) were compared using the likelihood ratio test (LRT). Posterior probabilities for site classes in models M2a and M8 were calculated using Bayes empirical Bayes (BEB) approach.

In addition, we evaluated the impact of positive selection on individual codons using selection models: mixed effects model of evolution (MEME); Fast Unconstrained Bayesian Approximation (FUBAR); Single-Likelihood Ancestor Counting (SLAC); and Fixed Effects Likelihood (FEL). The analyses were performed on the Datamonkey web server (Weaver et al., 2018) available at: http://www.datamonkey.org.

Several methods were used to detect recombination in dataset. One method was performed using the online program GARD on the Datamonkey website.

The other methods, including RDP, BOOTSCAN, GENECONV, MAXCHI, CHIMAERA and SISCAN, are all implemented in the Recombination Detection Program (RDP4) v.4.97 (Martin and Rybicki, 2000) and were used with default settings. DnaSP was used to estimate the minimum number of recombination events.

In addition, shorter alleles described by Mona et al. (2008) were included in the dataset to further test for signals of recombination (Table 3). Compared with the alleles amplified in this study and previously by Alvarez-Busto et al. (2007) and Schaschl et al. (2004, 2005), these alleles are shorter by the 86th codon in the DRB sequence; the codon number corresponds to the codon of ß1-domain in chamois (Schaschl et al., 2004).

Table 3. DRB exon 2 alleles of chamois previously reported by Mona et al. (2008) and downloaded from GenBank for inclusion in recombination analysis.

	Allele name	**Accession number**	**Reference**
1.	Ruru-DRB20	EU887489	Mona et al., 2008
2.	Ruru-DRB21	EU887490	Mona et al., 2008
3.	Ruru-DRB22	EU887491	Mona et al., 2008
4.	Ruru-DRB23	EU887492	Mona et al., 2008
5.	Ruru-DRB24	EU887493	Mona et al., 2008
6.	Ruru-DRB25	EU887494	Mona et al., 2008
7.	Ruru-DRB26	EU887495	Mona et al., 2008
8.	Ruru-DRB27	EU887496	Mona et al., 2008
9.	Ruru-DRB28	EU887497	Mona et al., 2008
10.	Ruru-DRB29	EU887498	Mona et al., 2008
11.	Ruru-DRB30	EU887499	Mona et al., 2008
12.	Ruru-DRB31	EU887500	Mona et al., 2008
13.	Ruru-DRB32	EU887501	Mona et al., 2008
14.	Ruru-DRB33	EU887502	Mona et al., 2008
15.	Ruru-DRB34	EU887503	Mona et al., 2008
16.	Ruru-DRB35	EU887504	Mona et al., 2008
17.	Ruru-DRB36	EU887505	Mona et al., 2008
18.	Ruru-DRB37	EU887506	Mona et al., 2008
19.	Ruru-DRB38	EU887507	Mona et al., 2008
20.	Ruru-DRB39	EU887508	Mona et al., 2008
21.	Ruru-DRB40	EU887509	Mona et al., 2008

Phylogenetic relationships between MHC DRB alleles was visualized through a neighbour-net network constructed in SplitsTree v.4.14.8 (Huson and Bryant, 2006).

3.6. Comparing SSCP and Ion Torrent sequencing

Of the total samples, 28 had already determined MHC genotypes by single-strand conformation polymorphism (SSCP)/Sanger sequencing prior to this study. Therefore, these 28 samples were selected to compare the performance of the two methods, SSCP and Ion Torrent S5 sequencing, for genotyping the highly polymorphic exon 2 of the MHC class II DRB gene in chamois.

MHC Genotyping by SSCP/Sanger Sequencing was done using the same set of primers HL030 and HL032 (Schaschl et al., 2004). PCR was performed following the protocol described elsewhere (Čížková et al., 2011; Schaschl et al., 2004) and verified by electrophoresis on ethidium bromide stained agarose gel. Successful PCR products were cloned according to the protocol by Bryja et al. (2005) and, following an additional PCR step (Schaschl et al., 2004), they were evaluated by SSCP analysis and sequenced by capillary electrophoresis in an ABI 3130 analyser. Sequencing and allele identification were performed according to the protocol described in Čížková et al. (2011). The results were validated with GeneMapper v.5.0 (Thermo Fisher Scientific, Waltham, MA, USA) software. The total length of the 236 bp allele sequences was assembled using CodonCode Aligner software v.1.6.3 (CodonCode Corporation, Dedham, MA, USA) and aligned with ClustalW v.4.0, implemented in MEGA 7 (Kumar et al., 2016).

The evolutionary relationships between the alleles were analysed by a median-joining network, as implemented in the software PopART (Leigh and Bryant, 2015). The parameter ε was set to zero (default) to obtain a sparse spanning network. We compared frequencies of alleles obtained with the two genotyping approaches and estimated discrepancies in each individual's genotypes between the two methods.

4. RESEARCH RESULTS

4.1. Diversity of MHC DRB alleles

In 110 chamois tested, 25 functional alleles were found for the second exon of the MHC class II DRB gene (Figure 8) coding for 25 unique amino acid sequences; 11 (44%) were identical to alleles previously described by Alvarez-Busto et al. (2007) and Schaschl et al. (2004, 2005). None of the sequences included indels or stop codons that might indicate the presence of pseudogenes. None of the individuals had more than two alleles, i.e. there was no evidence of amplification of multiple loci, confirming previously reported data for the genus *Rupicapra* (Fuselli et al., 2018). For a summary of the allelic variants found in each individual, see Appendix 1.

The most common allele was Ruru-DRB01, which was identified in 55 individuals with a frequency of 32.7%. Twenty three alleles had a frequency <10%, while seven alleles were identified in only one individual each (Figure 8).

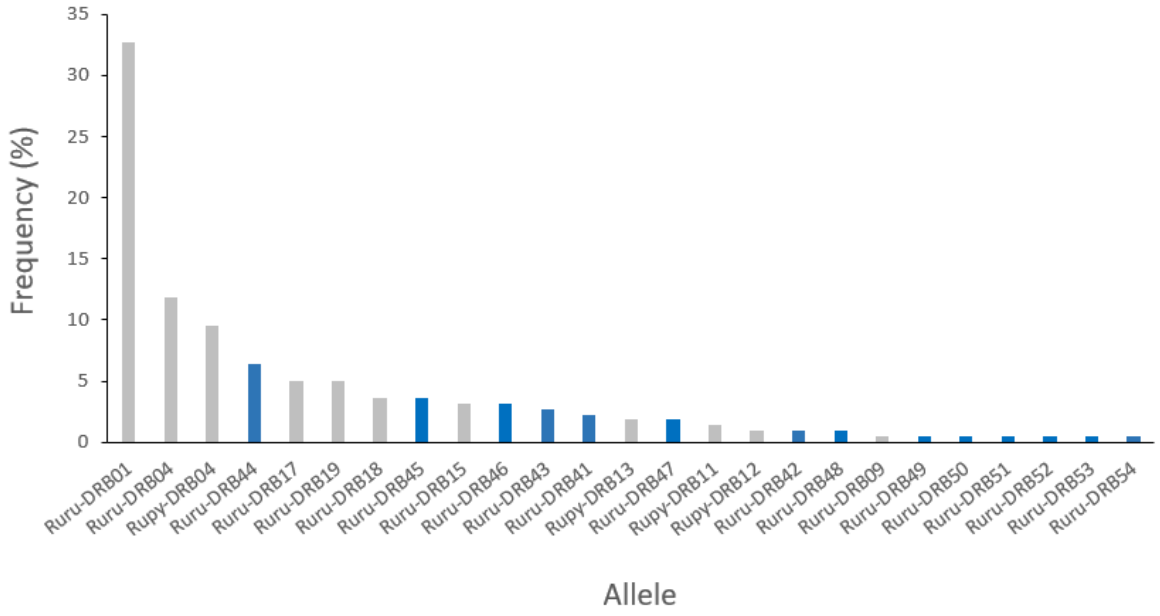

Figure 8. MHC DRB exon 2 alleles frequencies found in chamois (detected in this study). The allele Ruru-DRB01 was present in 50% and Ruru-DRB04 in 22% of the analysed individuals. Fourteen novel DRB exon 2 alleles identified in this study are marked in blue.

Thirty individuals (27.3%) were homozygous, of which most (56.7%) were homozygous for the most common allele, Ruru-DRB01. In Northern chamois, 23 unique alleles were found, with only Rupy-DRB11 and Rupy-DRB12 not found. The novel alleles Ruru-DRB53 and Ruru-DRB54 were found only in *R. r. carpatica*, Ruru-DRB51 and Ruru-DRB52 in *R. r. asiatica*, Ruru-DRB47, Ruru-DRB49 and Ruru-DRB50 in *R. r. balcanica*, and Ruru-DRB41, Ruru-DRB42, Ruru-DRB43, and Ruru-DRB48 in *R. r. rupicapra* (Figures 9; 10).

The most common allele in eight Southern chamois samples was also Ruru-DRB01 (referred to as Rupy-DRB02 by Alvarez-Busto et al., 2007). Despite the small number of samples, the number of heterozygous individuals was equal to that of homozygous individuals. The allele Rupy-DRB11 was detected only in *R. r. pyrenaica* and Rupy-DRB12 only in *R. r. ornata*.

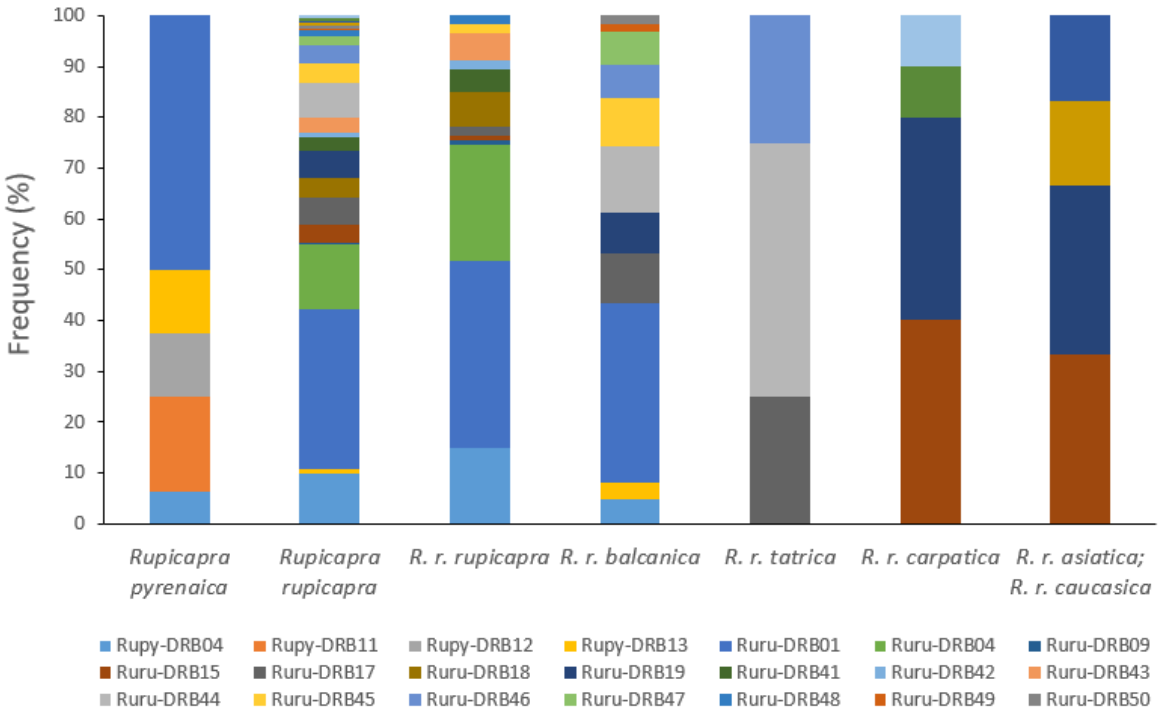

Figure 9. Plot of the frequency of MHC DRB exon 2 alleles in Southern (*Rupicapra pyrenaica*) and Northern chamois (*Rupicapra rupicapra*), and Northern chamois subspecies. The most frequent allele (Ruru-DRB01) was found with different frequencies in both species and in two of Northern chamois subspecies (*R. r. rupicapra* and *R. r. balcanica*).

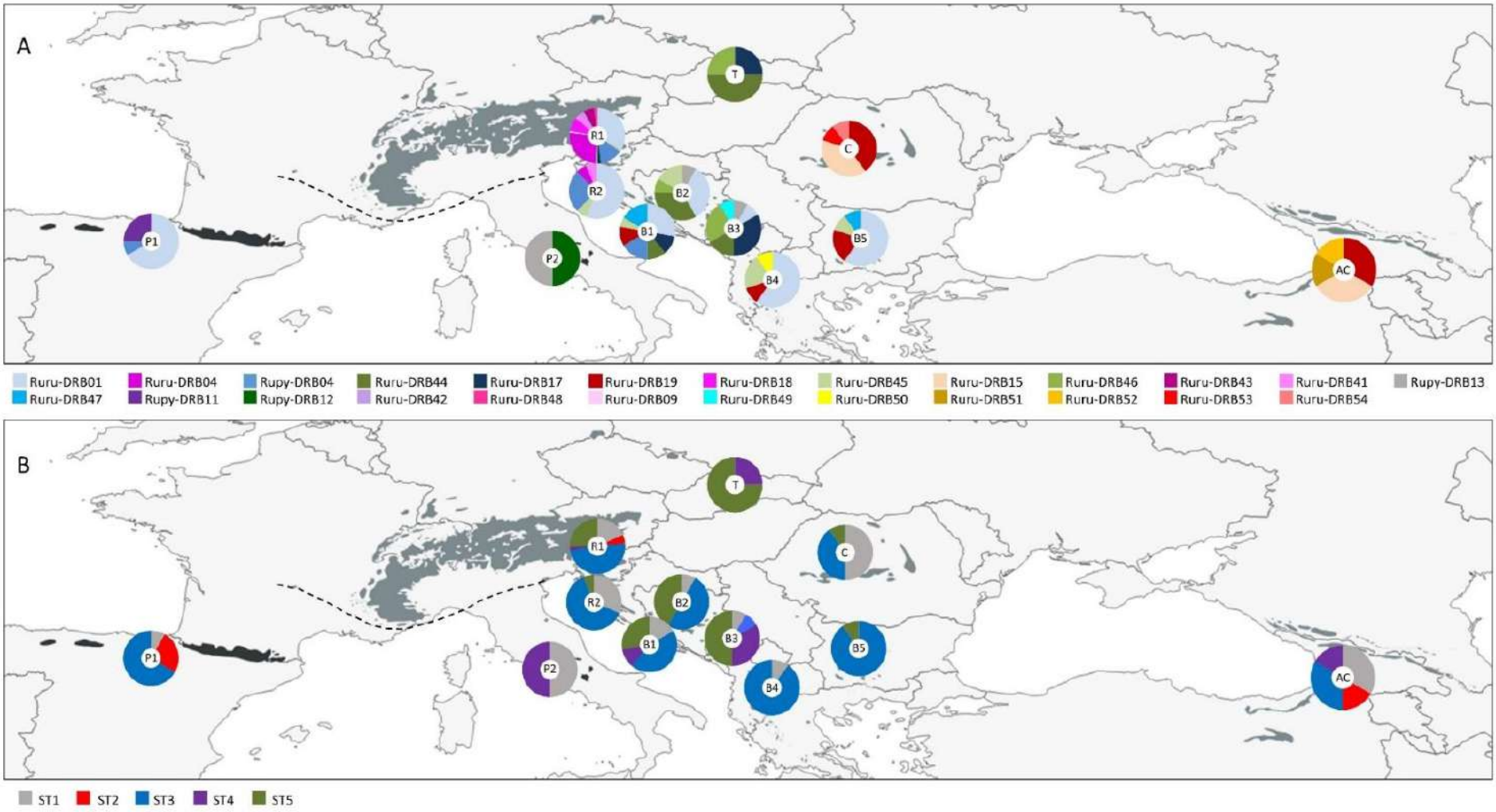

Figure 10. DRB allele (A) and supertype (B) frequencies are shown by the pie charts in groups defined by: i) subspecies (in case of a small number of samples) or ii) mountain location in the case of larger sample size (Appendix 1). DRB alleles in legend are ordered sequentially, from left to right, according to their frequency in the overall sample.

The analysis of the nucleotide alignment of 45 chamois alleles (25 detected in this study and 20 retrieved from GenBank) revealed 38 segregating (variable) nucleotide sites (Table 4), which were distributed across 78 codons.

Table 4. Sequence diversity and average nucleotide and amino acid evolutionary distances of chamois DRB exon 2 alleles, found in this study and previously published (Alvarez-Busto et al., 2007; Schaschl et al., 2004, 2005), and calculated for the complete sequences (All), peptide-binding region (PBR) and non-PBR sites. k – average number of nucleotide difference; S – the number of segregating sites; π – nucleotide diversity. Standard error (SE, 10,000 bootstrap replicates) is shown in parentheses.

				Nucleotide distance			**Amino acid distance**		
	k	**S**	**π**	**All**	**PBR**	**non-PBR**	**All**	**PBR**	**non-PBR**
Chamois	9.76	38	0.04	0.05 (0.01)	0.14 (0.04)	0.01 (0.01)	0.11 (0.03)	0.43 (0.14)	0.03 (0.02)
Southern chamois	10.87	27	0.05	0.05 (0.01)	0.17 (0.04)	0.01 (0.01)	0.12 (0.03)	0.46 (0.15)	0.03 (0.02)
Northern chamois	8.96	33	0.04	0.04 (0.01)	0.13 (0.04)	0.01 (0.01)	0.11 (0.03)	0.40 (0.14)	0.03 (0.01)

The overall nucleotide evolutionary distance, calculated using the Jukes-Cantor substitution model with a gamma distribution shape parameter, was 5%, whereas the amino acid evolutionary distance, calculated using the Poisson substitution model, was 11% (Table 4). Average nucleotide diversity was π=0.04 (0.04. in Northern and 0.05 in Southern chamois). Of the 33 segregation sites in Northern chamois, 18 were transitions and 19 were transversions, whereas in Southern chamois the number of transitions was 11 and the number of transversions was 17 in 27 segregation sites. The average number of nucleotide differences between alleles was k=9.76 and ranged from one to 19, with an average of 8.96 in Northern chamois and from two to 19, with an average of 10.87 in Southern chamois (Figure 11).

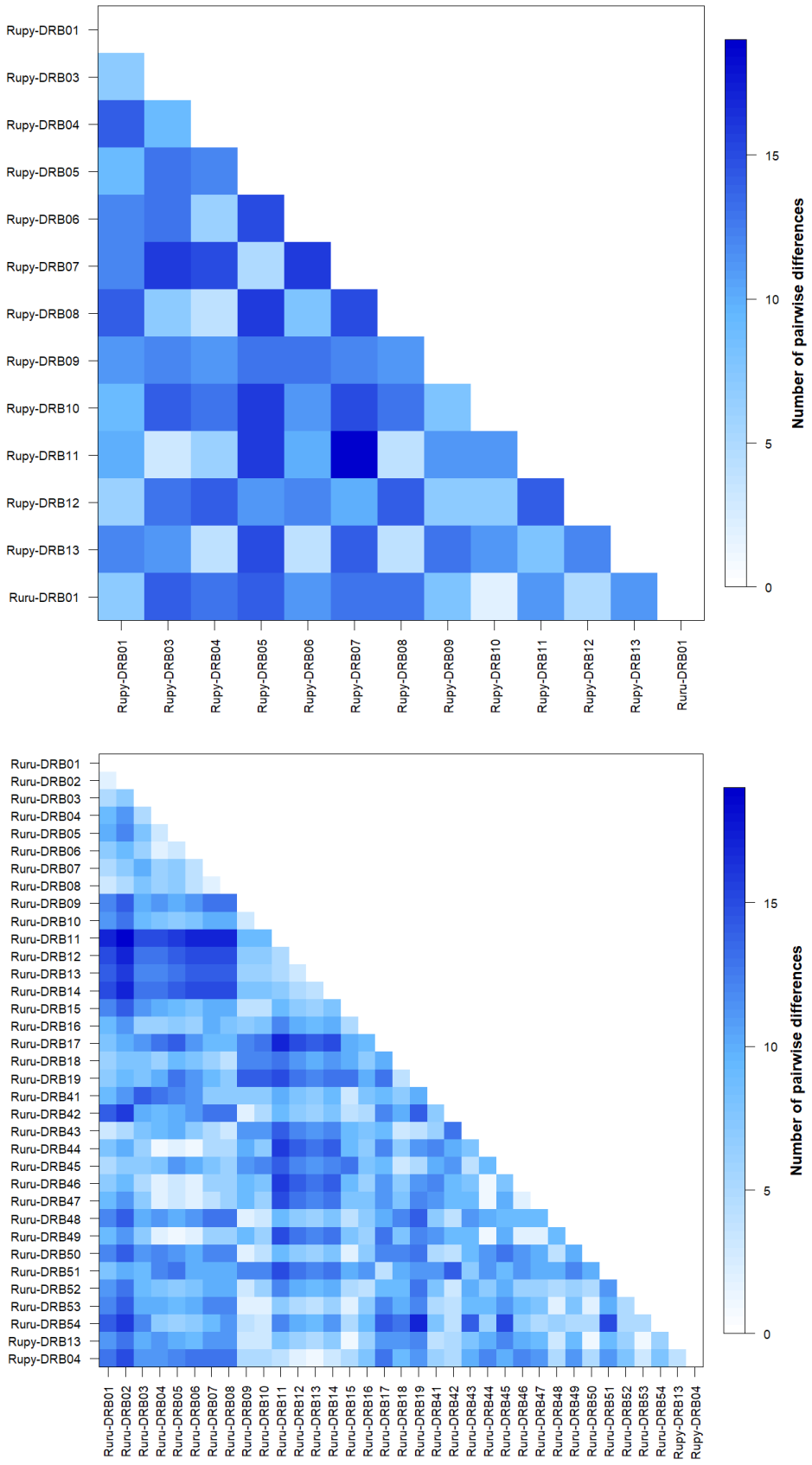

Figure 11. The number of nucleotide differences between the different MHC DRB alleles found in Southern chamois (top) and Northern chamois (down).

Following the model proposed by Brown et al. (1993) for the MHC DRB exon 2 protein structure in humans, we attributed 22 of 78 codons (28%) to the peptide-binding region (PBR). For Northern chamois, 14 of 23 variable codon positions (61%) were within the putative PBR. The higher value applies to Southern chamois, where 14 of 17 variable codon positions (82%) were within the putative PBR (Figure 12).

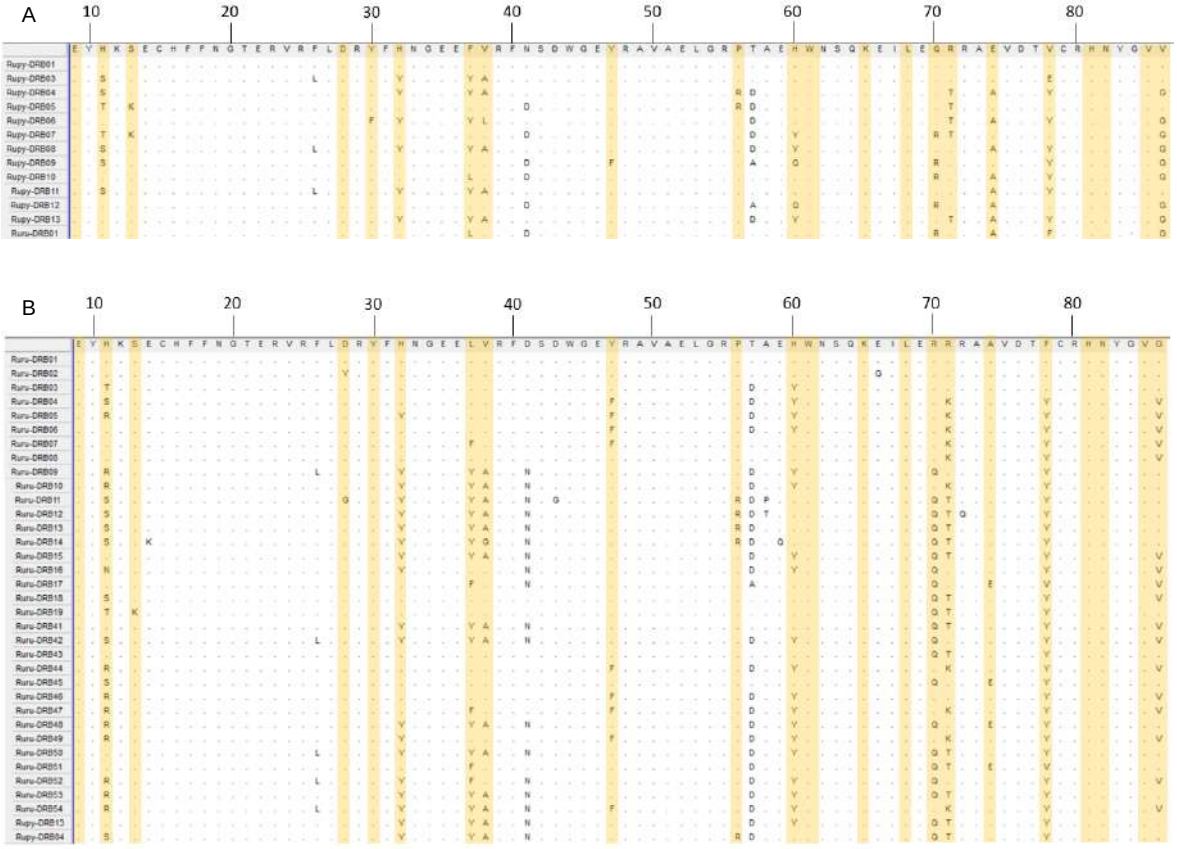

Figure 12. Alignment of the putative amino acid sequences of the exon 2 of MHC DRB alleles found in Southern (A) and Northern (B) chamois. Dots indicate identity in the amino acid sequence to the sequence of the Ruru-DRB01 allele and a yellow columns indicate codons involved in the peptide-binding region (PBR) in humans (Brown et al., 1993).

4.2. Species/subspecies genetic parameters and structure

Values of the diversity parameters for Northern and Southern chamois and for subspecies of Northern chamois for which at least six samples were available are shown in Table 5.

Table 5. DRB exon 2 genetic diversity detected in Northern and Southern chamois and its subspecies (Corlatti et al., 2011), with the number of heterozygous and homozygous individuals, the number of supertypes estimated based on 14 positively selected amino acid sites, and Tajima's D values. n – number of individuals; A – number of alleles; ST – number of supertypes; Ar – allelic richness (calculated only for sample size ≥ 8 individuals). For subspecies with sample size ≤ 6, only the number of alleles and supertypes are reported. Values in bold are significant at P < 0.05.

Species/ subspecies	n	A	ST	Ar	N° heterozygous (%)	N° homozygous (%)	Tajima's D
Southern chamois	8	5	4	5.00	4 (50)	4 (50)	1.00
Northern chamois	102	23	5	8.17	76 (75)	26 (25)	**2.38**
R. r. rupicapra	57	12	5	5.64	44 (77)	13 (23)	**2.38**
R. r. balcanica	31	11	4	7.39	21 (68)	10 (32)	1.25
R. r. tatrica	6	3	2	/	/	/	/
R. r. carpatica	5	4	3	/	/	/	/
R. r. asiatica, R. r. caucasica	3	4	4	/	/	/	/

The number of alleles (A) was five in Southern chamois and 23 in Northern chamois. At the subspecies level, alleles in Northern chamois ranged from three in *R. r. tatrica* to 12 in *R. r. rupicapra*. Allelic richness (Ar) ranged from 5.64 in *R. r. rupicapra* to 7.39 in *R. r. balcanica* and was estimated to 5.00 in Southern chamois and 8.17 in Northern chamois. The proportion of heterozygous individuals was the highest in the subspecies *R. r. rupicapra* and the lowest in the Southern chamois.

Tajima's D values were positive and significant in Northern chamois and its *R. r. rupicapra* subspecies, but non-significant within Southern chamois and *R. r. balcanica* subspecies of Northern chamois (Table 5). The average pairwise F_{ST} values between Southern and Northern chamois was positive and significant (F_{ST}=0.06, P<0.05), and it was also significant between all Northern chamois subspecies (Table 6).

Table 6. Pairwise values of genetic differentiation (F_{ST}) between Northern chamois subspecies based on DRB locus. Subspecies with small sample size (*R. r. asiatica and R. r. caucasica*) were excluded from the analysis.

R. rupicapra	*rupicapra*	*balcanica*	*tatrica*
rupicapra			
balcanica	0.12		
tatrica	0.41	0.23	
carpatica	0.33	0.24	0.49

The discriminant analysis of principal components (DAPC) separated subspecies *R. r. carpatica*, *R. r. asiatica* and *R. r. caucasica* from the rest of subspecies along the first (horizontal) axis. The differentiation between the Southern chamois species and Northern chamois subspecies was weaker and distributed mainly along the second (vertical) axis (Figure 13). To describe the relationship between the clusters, 10 PCs (four discriminant functions, 83.3% variance conserved) were retained during analyses based on their eigenvalues.

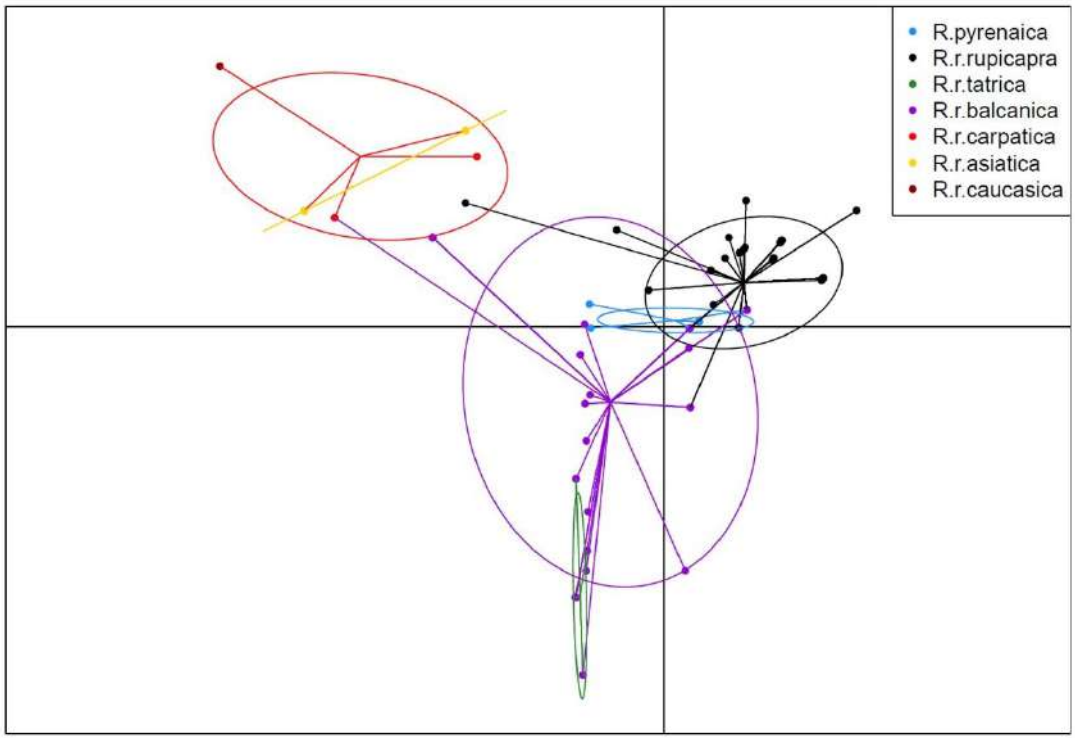

Figure 13. Scatterplot of the genetic differentiation resulting from a discriminant analysis of principal components (DAPC) for the genetic structure of Northern and Sothern chamois subspecies based on the DRB locus. Individuals are presented as separate dots with colours denoting chamois species/subspecies and inclusion of 95% inertia ellipses.

The clustering procedure revealed five supertypes in chamois based on 14 PSSs (Figure 14). To describe the relationship between the clusters, eight PCs (four discriminant functions, 86.1% variance conserved) were retained during analyses based on their eigenvalues. Five supertypes occurred in Northern chamois, and four of them in Southern chamois. The number of alleles assigned to each supertype ranged from seven to 12 (Figure 15).

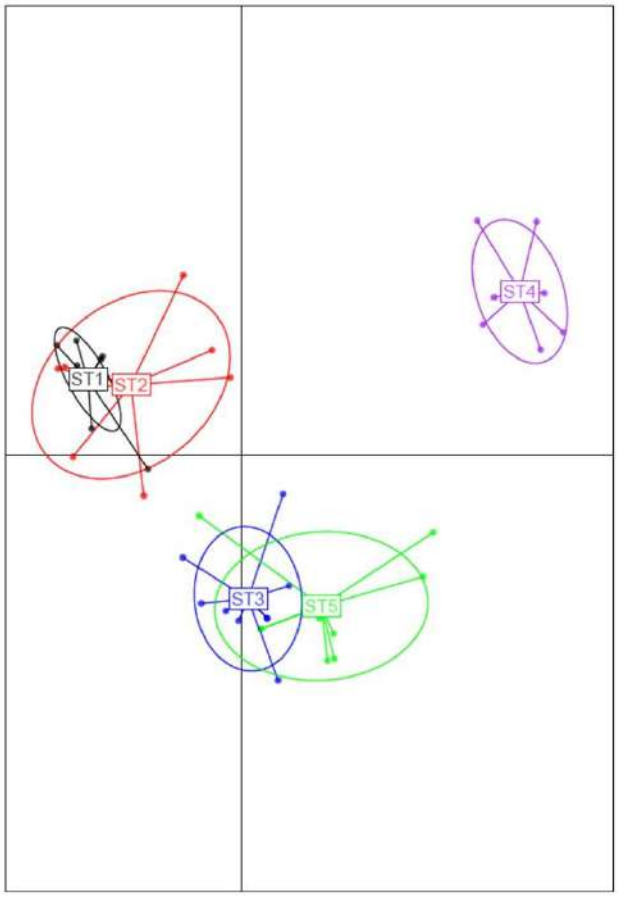

Figure 14. Scatterplot of the five MHC supertypes resulting from a DAPC. Each allele is represented as a dot and the supertypes as ellipses (ST1-ST5) (see spatial distribution in Figure 10).

The mean number of supertypes per individual was 1.50 for Southern chamois and 1.61 for Northern chamois (Appendix 1). The number of supertypes varied between two and five in different geographic regions of chamois. Despite the small sample sizes of the subspecies of Southern chamois and Northern subspecies *R. r. asiatica* and *R. r. caucasica* and the correspondingly small number of alleles detected, there was no reduction in the number of supertypes in these geographic regions, i.e., each allele corresponded to a distinct supertype (Figure 10, Table 5).

The neighbour-net network did not reveal any clear pattern of allele clustering between Northern and Southern chamois, as alleles of the two species were mainly spread across the network (Figure 15). This result is consistent with trans-species evolution of the MHC DRB locus.

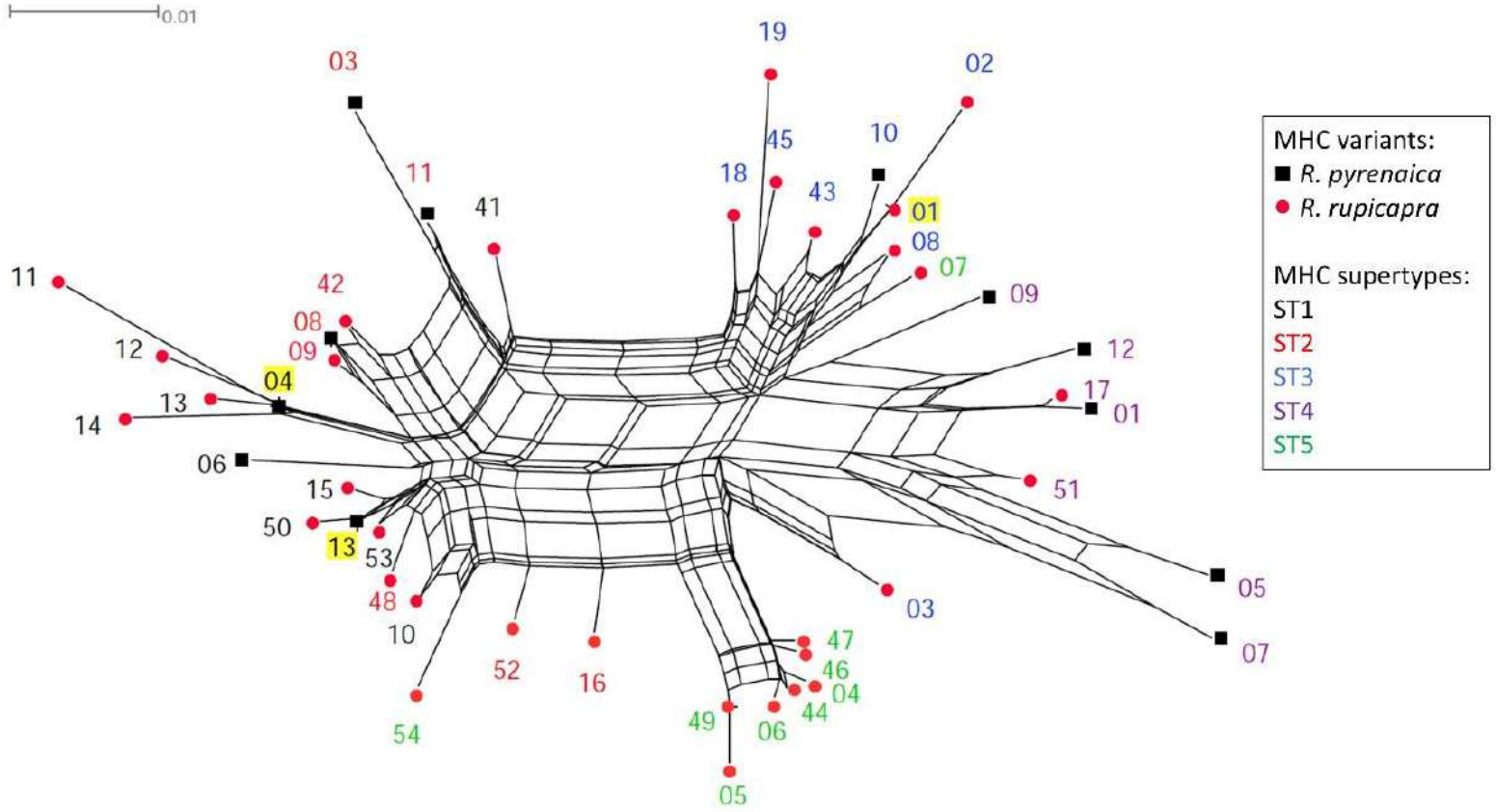

Figure 15. Neighbour-net network of chamois DRB alleles. The alleles marked in yellow were detected in both species. Scale bar indicates the scale of the network i.e., 0.01 substitutions per nucleotide site.

4.3. Signatures of selection and recombination on MHC DRB alleles

Global estimates of the ratio of relative rates of non-synonymous to synonymous mutations (dN/dS), averaged across all codon sites, indicated the presence of positive selection at the DRB locus. The non-synonymous mutation rate (dN=0.06 in Southern chamois and 0.05 in Northern chamois) exceeded the synonymous mutation rate (dS=0.01 in Southern chamois and 0.004 in Northern chamois) (Table 7).

Table 7. Relative rates of synonymous (dS) and non-synonymous (dN) substitutions, and results of one-tailed Z-test (Z) for positive selection of chamois DRB exon 2 alleles (found in this study and previously published (Alvarez-Busto et al., 2007; Schaschl et al., 2004, 2005) calculated for the complete sequences (All), peptide-binding region (PBR) and non-PBR sites. Standard error (SE, 10,000 bootstrap replicates) is shown in parentheses. Values in bold are significant at P < 0.01.

	Chamois			**Southern chamois**			**Northern chamois**		
	All	PBR	non-PBR	All	PBR	non-PBR	All	PBR	non-PBR
dS	0.01	0.01	0.004	0.01	0.03	0.004	0.004	0.01	0.003
(SE)	(0.003)	(0.01)	(0.002)	(0.01)	(0.02)	(0.004)	(0.002)	(0.01)	(0.002)
dN	0.06	0.17	0.02	0.06	0.19	0.02	0.05	0.15	0.02
(SE)	(0.01)	(0.04)	(0.01)	(0.02)	(0.05)	(0.01)	(0.01)	(0.04)	(0.01)
dN/dS	6.00	17.00	5.00	6.00	6.33	5.00	12.50	15.00	6.67
Z	**3.77**	**3.79**	1.43	**3.46**	**3.55**	1.10	**3.70**	**3.69**	1.55

The DRB locus showed evidence of strong selection pressure. Methods for calculating dN/dS values for individual codons (models M2a and M8) identified up to 14 codons (posterior probabilities > 95%) that are likely to be affected by positive selection (Table 8).

Table 8. Codon sites under positive selection as predicted by codon-based selection models M2a and M8 using the Empirical Bayes approach. The codon sites assumed to be under selection with a posterior probability of > 99% are shown in bold, and sites with a posterior probability of > 95% are shown in standard font. Codon numbers correspond to the codons of ß1-domain in chamois (Schaschl et al., 2004).

	Selection model	**Codon sites under positive selection**
Chamois	M2a	**11**, **13**, **26**, **37**, **38**, **47**, **57**, **60**, **70**, **71**, **74**, **78**, **86**
	M8	**11**, **13**, **26**, 32, **37**, **38**, **47**, **57**, **60**, **70**, **71**, **74**, **78**, **86**

The selection models yielded different levels of selection pressure at the analysed locus. The mean values of dN/dS calculated according to the M2a model for the individual codons are shown in Figure 16.

In the DRB sequences currently studied, based on the GARD and RDP approaches, no signal for recombination was found, although DnaSP detected a minimum number of 11 recombination events. When testing recombination with shorter alleles described by Mona et al. (2008) included in the dataset, the analysis revealed one recombination event (Table 9).

Table 9. Recombinant sequences at exon 2 of the major histocompatibility complex (MHC) class II DRB locus in chamois.

Recomb. event	Recombinant sequence	Major/minor parent	Consensus score	Beginning/ending breakpoint	Probability	Methods
1	Ruru-DRB19	Ruru-DRB24 /Ruru-DRB14	0.512	154/220	4.387 E-02	MaxChi, Chimaera

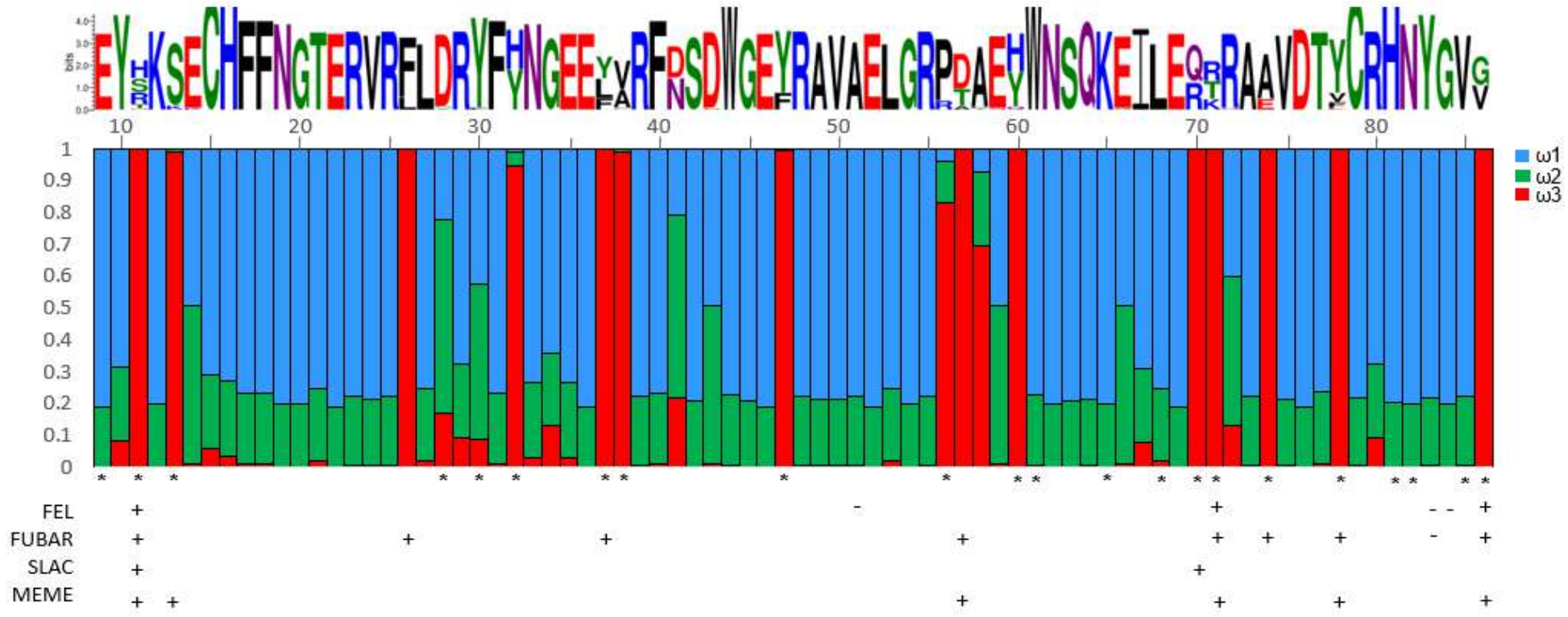

Figure 16. Distribution of positively selected sites in exon 2 of chamois DRB genes is shown as estimated by EasyCodeML (model M2a). Red columns indicate the class of sites with a high probability of ω > 1. In this model ω1 = 0.00 and applies to ~58% of the codons (blue); ω2 = 1.00 at ~20% of the sites (green); and ω3 = 18.24 at ~22% of the sites (red). Diversity of the observed peptides is indicated by a sequence conservation logo. Asterisks indicate codons involved in peptide-binding region (PBR) residues in the human ortholog. Sites under positive and negative selection are indicated with respect to FEL, FUBAR, SLAC (for pervasive selection) and MEME (for episodic selection).

4.4. Comparing SSCP and Ion Torrent sequencing

The sequence of 236 bp (without primers) of the DRB exon 2 gene was obtained for all 28 chamois. After initial filtering of the raw Ion Torrent data, the amplicon coverage ranged from 656 to the maximum of 5,000 reads allowed by AmpliSAS, with an average of 3,191 ± 1,996 (SD) reads. The coverage in this study was sufficient to separate true alleles from artefacts and obtain reliable genotypes.

When clusters within an amplicon were ordered by descending per amplicon frequency (PAF), a significant drop in frequency at the value of 3% was observed, which probably represented the boundary between true alleles and artefacts. For one individual, chimeric variant was found at a frequency of 13%. Accordingly, PAF threshold was set at 14% and the chimeric and low-frequency variants below this threshold were removed (Figure 17).

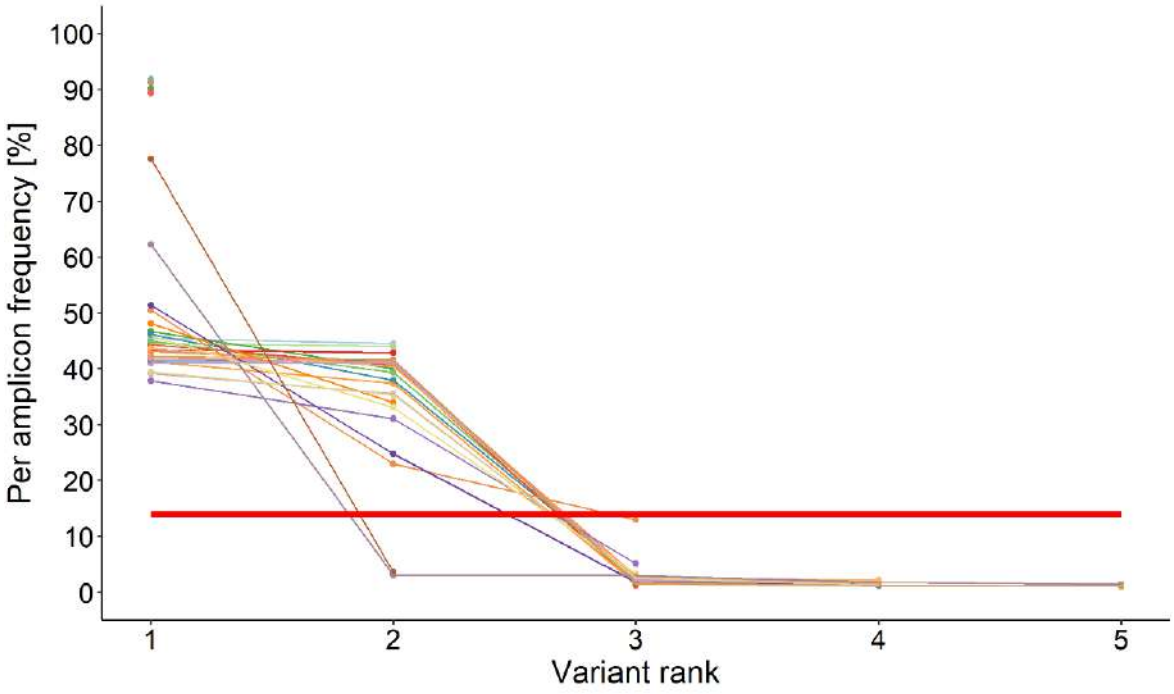

Figure 17. Variants ordered by descending per amplicon frequency (PAF) after sequencing errors have been added to them through the clustering step performed by AmpliSAS. Lines represent amplicons of 28 individuals. Horizontal line represents 14% PAF threshold for the filtering step.

Clustering sequencing errors with true variants increased the read depth of true variants, allowing to distinguish alleles more easily from low-frequency artefacts. On average, 54% of amplicon depths were assigned to alleles, and after clustering, this percentage increased to 83% (Figure 18).

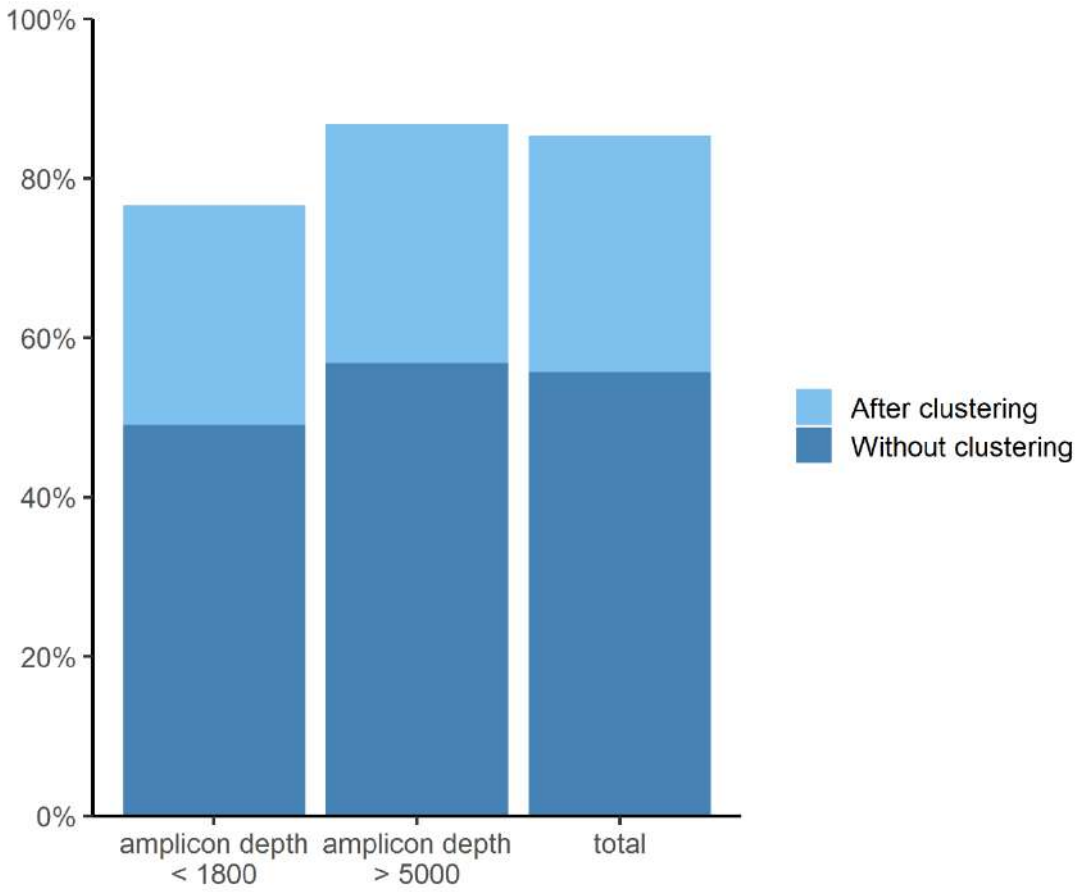

Figure 18. Percentages of amplicon reads assigned to alleles before and after AmpliSAS clustering. On average, 54% of amplicon depths were assigned to alleles, and after clustering this proportion increased to 83%.

Ion Torrent tends to produce high rates of homopolymeric indels, depending on the properties of the sequences analysed. However, this had very little effect on the results of the genotyping of chamois. Estimates of variant frequencies per amplicon after clustering were mainly influenced by 1 bp substitution (Figure 19), which accounted on average for 21% of the increase in coverage of true variants within the amplicon.

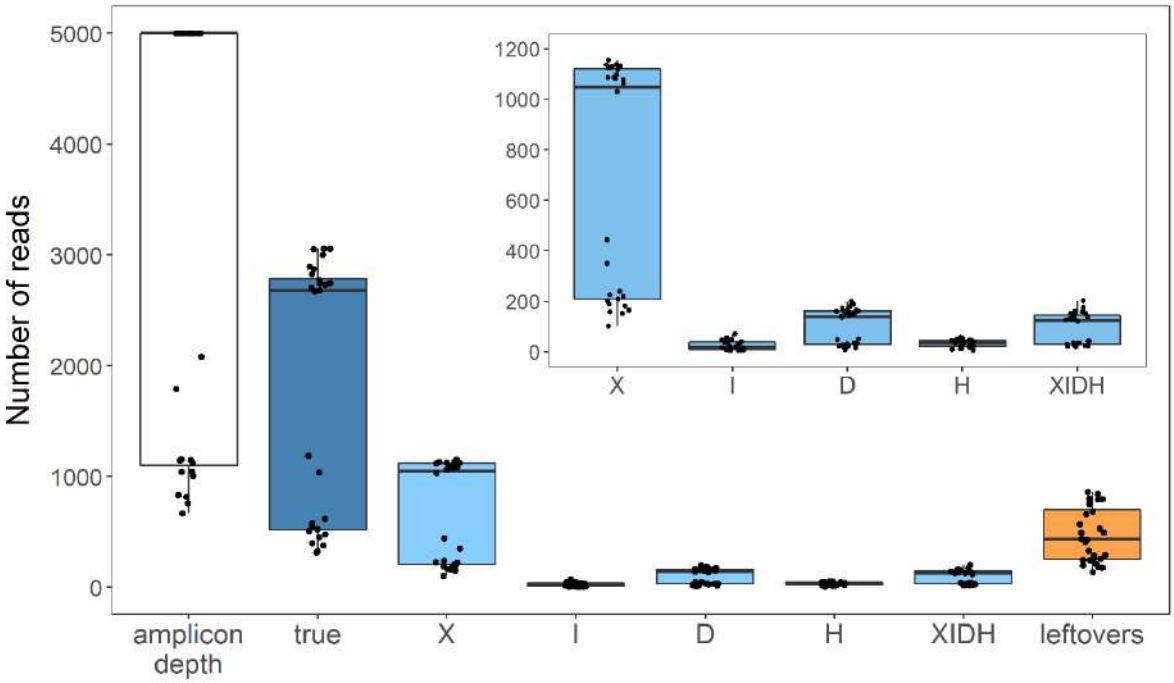

Figure 19. Representation of amplicon depths of 28 individuals, and the number of reads of true sequences and artefacts within each amplicon. True - sequences that match a real allele i.e. true variant; the following are sequencing errors: X - 1bp substitutions, I - insertions, D - deletions, H - homopolymer indels; XIDH - sequence with any combination of at least two sequencing errors; leftovers - low frequency variants, chimeras or sequences containing many errors which could not be classified into the major clusters. All sequencing errors, except leftovers, were clustered with the true variant from which they were derived.

No stop codons were found in the nucleotide sequence of all alleles identified by SSCP/Sanger sequencing and amplicon-based NGS genotyping, indicating that all alleles encode functional proteins. Allele identities were based on 17 (7%) variable nucleotide positions with the lowest number of mutations (two) between alleles Ruru-DRB09 and 42, and the highest number of 14 mutations was found between allele Ruru-DRB01 and 42 (Figures 20 and 21).

	9	10	54	72	88	89	91	99	118	145	147	148	156	187	190	211	235
*Ruru-DRB*01*	C	A	T	C	T	G	T	G	A	C	A	C	C	G	G	T	G
*Ruru-DRB*04*	A	G	.	.	.	.	.	.	T	.	G	A	T	.	A	A	T
*Ruru-DRB*09*	.	G	C	T	A	C	C	A	.	.	G	A	T	A	.	A	.
*Ruru-DRB*18*	A	G	.	.	.	.	.	.	.	.	.	.	.	A	C	A	T
*Ruru-DRB*41*	.	.	.	T	A	C	C	A	.	.	.	.	.	A	C	A	T
*Ruru-DRB*42*	A	G	C	T	A	C	C	A	.	.	G	A	T	A	.	A	T
*Ruru-DRB*43*	.	.	.	.	.	.	.	.	.	.	.	.	.	A	C	A	.
*Rupy-DRB1*04*	A	G	.	T	A	C	C	A	.	G	G	A	.	A	C	A	.

Figure 20. Alignment of the nucleotide sequences, only variable residues of the exon 2 of DRB alleles of chamois are shown. The codes before represent the species abbreviation and gene name, and the numbers after indicate allele numbers.

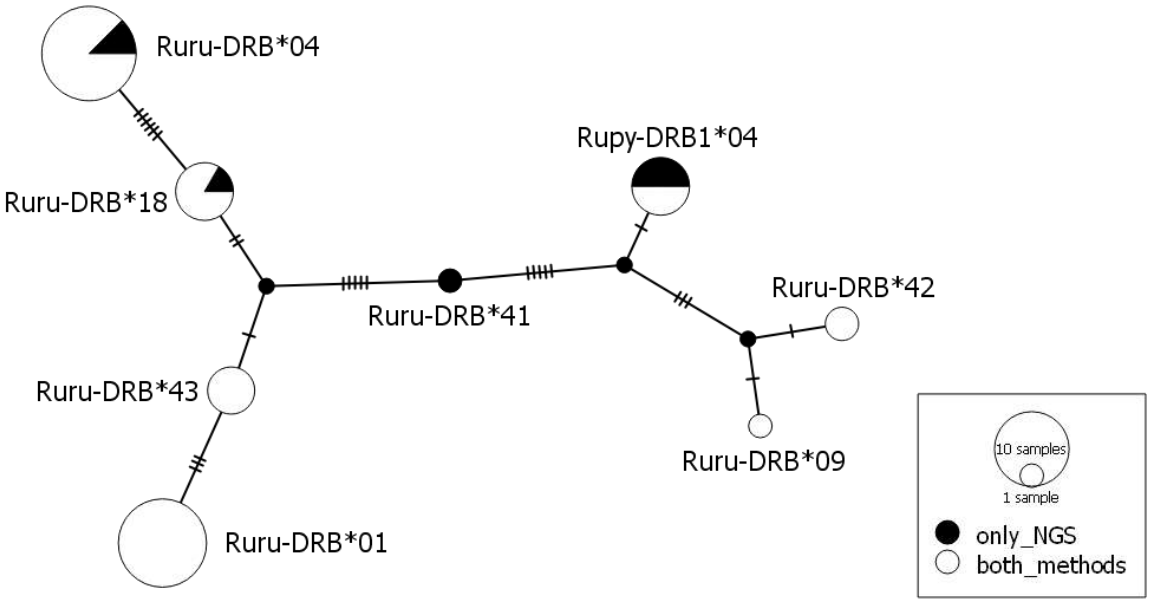

Figure 21. Median-joining network (ε = 0) of DRB alleles found in chamois. Alleles are represented by pie charts whose size is proportional to the number of individuals. The colors indicate the number of individuals that have particular alleles determined by two genotyping methods. The number of mutations separating nodes is represented by slashes crossed with network branches. Small black circles indicate hypothetical alleles as predicted by the model.

A total of eight alleles were identified in 28 individuals. Five of these had already been described (Alvarez-Busto et al., 2008; Schaschl et al., 2004), and three were new alleles not previously described. We discovered seven alleles by SSCP/Sanger genotyping, and all were also detected by NGS sequencing.

In addition, NGS allowed detection of one additional allele and expanded the genotypes of six of the 28 individuals compared with the results of SSCP/Sanger genotyping (Appendix 2).

The maximum number of alleles per individual detected by either method was two, consistent with the expectation of a single DRB locus from previous studies. Considerable variation in frequency was observed between alleles. For both methods, the lowest frequencies were for Ruru-DRB41 (0% and 2%, respectively) and Ruru-DRB09 (2%), and the highest were for Ruru-DRB01 (36% and 32%, respectively) (Table 10).

Table 10. DRB exon 2 allele frequencies estimated with Single-Strand Conformation Polymorphism (SSCP)/Sanger sequencing and amplicon-based next-generation sequencing (NGS).

Allele	SSCP/Sanger		NGS	
	Nº observations	Allele frequency	Nº observations	Allele frequency
Ruru-DRB01	20	0.357	18	0.321
Ruru-DRB04	18	0.321	17	0.304
Ruru-DRB09	1	0.018	1	0.018
Ruru-DRB18	5	0.089	6	0.107
Ruru-DRB41	0	0.000	1	0.018
Ruru-DRB42	2	0.036	2	0.036
Ruru-DRB43	6	0.107	4	0.071
Rupy-DRB04	4	0.071	7	0.125

The number of called alleles per individual agreed between the two genotyping methods in 75% of cases. Discrepancies in individual genotypes between the two methods were mainly caused by allelic dropout in SSCP/Sanger analysis. Six of 28 individuals (21%) were misclassified as homozygous in SSCP/Sanger analysis because of dropout events.

In addition to the allelic dropout, there was one false-positive allele call for an individual with SSCP/Sanger analysis that was due to weak conserved sequence. Overall, 25% of individuals had genotyping discrepancies. A summary of the DRB alleles found in each individual using both genotyping methods is provided in Appendix 2.

5. DISCUSSION

In this study, polymorphism of the MHC class II DRB locus was investigated in chamois with Ion Torrent S5 next-generation sequencing. Using the exon 2 of the MHC class II DRB gene, this study provides a characterization of the adaptive genetic diversity of chamois throughout their range and fills the gap in previously published data for Northern and Southern chamois (Alvarez-Busto et al., 2007; Schaschl et al., 2004, 2005). From twenty-five identified exon 2 DRB variants in 110 chamois individuals, eleven had already been described, and fourteen were new, undescribed alleles. Consistent with previous studies (Alvarez-Busto et al., 2007; Mona et al., 2008; Schaschl et al., 2004, 2005), this study reports low to moderate MHC diversity likely shaped by past demographic processes and balancing selection.

Although the Northern and Southern chamois are not threatened, most subspecies are found in restricted areas where they face various threats (Corlatti et al., 2022). Small or endangered populations are under serious threat from infectious diseases, as these can lead to sudden population declines and possibly to population extinction (Daszak et al., 2000). Therefore, understanding the regulation of resistance to pathogens in natural populations is central to conservation and management decisions, especially for populations threatened by disease.

There is currently no evidence that measurement of polymorphism in MHC is sufficient to predict immunological fitness. Therefor, measurement of polymorphism in MHC genes cannot be an indirect measure of immunological fitness of populations. Polymorphism in MHC genes may determine individual-level survivability and population-level adaptability to the pathogen causing the selective pressure under certain pathogen challenges, but this occurs in combination with the effects of the other genes involved in the immune response, as MHC genes definitely do not alone determine the ability to survive infection (Acevedo-Whitehouse and Cunningham, 2006). Although polymorphism at amino acid sites encoding PBR is generally considered an indicator of susceptibility or resistance to infectious disease in wildlife (O'Brien and Evermann, 1988), maintenance of variability in other immune genes may most likely account for resistance in species with otherwise uniform or low genetic MHC variation (Acevedo-Whitehouse and Cunningham, 2006).

5.1. Demography of chamois

Demography of chamois populations determined using mitochondrial and nuclear genes (Crestanello et al., 2009; Papaioannou et al., 2019; Pérez et al., 2002, 2017; Rodríguez et al., 2009, 2010; Šprem and Buzan, 2016) revealed a probable reduction in population size due to historical events (in the last 5,000 to 30,000 years) and recent habitat fragmentation and isolation (Corlatti and Herrero, 2021). Unlike many other species that experienced a demographic decline in Mediterranean refugia during the Last Glacial Maximum and a subsequent expansion as temperatures rose (Petit et al., 2003), the results of this study support the view of Mona et al. (2008) that chamois took a different path. Higher temperatures reduced the available territory for chamois and trapped them on mountain peaks, likely reducing population size. This also resulted in a positive and significant Tajima's D value for the MHC DRB locus for Northern chamois and its subspecies *R. r. rupicapra*, suggesting that Northern chamois and its subspecies *R. r. rupicapra* may have experienced a recent demographic decline. Tajima's D value was positive but not significant for the subspecies *R. r. balcanica*, likely due to its larger distribution range and the presence of multiple populations in our analysis. For Southern chamois, Tajima's D value was also positive but not significant, likely due to the small sample size.

5.2. Genetic diversity and variation between chamois species /subspecies

Mountain species often resemble archipelagos because of the irregular spatial structure of suitable habitats, so MHC diversity of mountain populations is often the result of demographic history and habitat fragmentation (Zemanova et al., 2015). Mountain habitats are inherently fragmented at various spatial scales, and therefore dispersal between habitats may be limited (Brown, 2001). This fragmentation of habitat impacts genetic structure by limiting gene flow between isolated populations and maintaining effective population size low (Ezard and Travis, 2006). Restricted gene flow between small fragmented populations leads to increased isolation by distance and strong genetic drift (Willi et al., 2007). Different molecular markers have been used to study genetic variability in chamois, e.g. mitochondrial DNA (mtDNA), microsatellites, Y-chromosome markers (Pérez et al. 2014). The Alpine chamois shows the highest values of genetic diversity for almost all markers analyzed. In contrast, the population of Appenine *R. p. ornata* shows an extremely low genetic diversity for both microsatellites and mtDNA. Most other chamois populations have intermediate values of genetic diversity.

According to the neutral genetic markers, the heterogeneity of populations is very high. Populations from the Italian Alps (Soglia et al., 2010) and Slovenia (Buzan et al., 2013) show higher genetic diversity than populations from the central Dinaric Mountains (Šprem and Buzan, 2016), which could indicate stronger genetic drift and bottleneck within these populations. A similar deficit of heterozygotes was observed in chamois populations from Bulgaria (Markov et al., 2016) and Greece (Papaioannou et al., 2019). The lowest genetic variation with a high degree of inbreeding was observed in chamois populations from Slovakia, which experienced a severe bottleneck (Zemanová et al., 2015).

Evidence of recent hybridization has only been reported between different subspecies of *Rupicapra rupicapra* (reviewed in Iacolina et al., 2019), in regions where chamois have been introduced for hunting purposes. Legal and illegal translocations of individuals were carried out neglecting genetic issues, which greatly increased the risk of losing differentiated gene pools. In recent years, detection of hybridization has been mainly performed using mtDNA, Y-chromosome and microsatellites. However, the consequences of hybridization in chamois are still largely unknown and future management interventions (translocations and reintroductions) should consider the genetic background of autochthonous and translocated populations to avoid possible genetic extinction of taxa. Furthermore, future genomic studies using other markers, such as single nucleotide polymorphisms (SNPs), could help to clarify between gene introgression and hybridization.

Discriminant function of principal components generally showed discrimination between subspecies of Northern chamois, but weaker discrimination between Northern and Southern chamois (Figure 13). The closer clustering and allele sharing between *R. r. rupicapra* and *R. r. balcanica* could be a consequence of historical gene flow, but also an influence of past translocations, which has already been confirmed by neutral markers (Šprem and Buzan, 2016). This assumption can also be confirmed by significant pairwise F_{ST} differences (Table 6). Given the low migration between chamois populations, it is reasonable to assume that some of new alleles indicate the presence of private alleles in the subspecies in which they were found, which in turn suggests diversifying selection, i.e., local MHC adaptation. In view of the above, this study suggests that the pattern of MHC genetic variation within and between regions and populations is likely due to both demographic processes, i.e. genetic drift, and balancing selection. This view is justified when the patterns of geographic variation are considered in detail (Figure 10). However, to draw a definitive conclusion, it is necessary to compare the genetic variation of the MHC with the neutral variation (Biedrzycka et al., 2020).

Uniform selection pressure across different populations could result in less differentiation at functional loci due to balancing selection (Oosterhout et al., 2006), but genetic drift in small and isolated populations may overcome balancing selection and reduce genetic diversity (Radwan et al., 2010). On the other hand, it must be taken into account that chamois populations have been subjected to the pressure of unsustainable hunting during the last two centuries, leading to the local extinction of some subspecies (*R. p. parva*, *R. p. ornata*, *R. r. balcanica*, *R. r. asiatic*a) (Shackleton, 1997). In the late 1970s, populations in the Alps experienced severe bottlenecks due to sarcoptic mange epidemics with catastrophic outbreaks in which up to 80% of local populations were lost (Buzan et al., 2013; Fuchs et al., 2000; Rossi et al., 1995), and this scenario may have occurred even earlier in the history of populations. Therefore, the observed differences in alleles within the range of chamois could be due to long-term isolation between populations/subspecies, but may also suggest possible differential pressures from pathogens that may have influenced local MHC adaptation in some mountainous regions (Arbanasić et al., 2014; Cavallero et al., 2012). In addition, differentiation of supertypes in chamois could be due to demographic processes and genetic differentiation over the history of the species, but also to differential pressure from pathogens in different mountain ranges where the species occurs (Biedrzycka et al., 2020; Rossi et al., 1995).

5.3. The role of selection

Depending on selection pressures, natural populations maintain genetic variation in functional genes (Sommer, 2005). Balancing selection is thought to counteract genetic drift and slow the rate of allele fixation (Sommer, 2005). In many cases, differentiation patterns between populations can only be detectable at functional genes under selection (Awadi et al., 2018). Changes in the frequency of MHC alleles therefore have adaptive value, and understanding how such functional variation arises and is maintained within populations is essential for species conservation and optimal management decisions (Radwan et al., 2010; Sagonas et al., 2018). The strong selection pressure on MHC genes is ultimately responsible for the high polymorphism and may lead to discrepancies between the variation patterns at MHC and neutral loci in natural populations (Alcaide, 2010). Mona et al. (2008) proposed long-term balancing selection throughout the phylogenetic history of chamois. It was supported by the data in this study, as all maximum likelihood codon-based selection models (M2a, M8) produced a “best fit” to the data compared to models without selection.

Positive selection presumably affects only a few codons (Nielsen, 2005; Yang and Swanson, 2002), so interpreting selection results based on dN/dS averaged over an entire genetic region could be misleading. As in most other taxa with well-characterised MHC loci, codons were found to be subject to positive selection and, most importantly, all codons were located at or near PBR sites involved in pathogen recognition (Piertney and Oliver, 2006). Nevertheless, the ratio of dN/dS=6.00 in Southern chamois and dN/dS=12.50 in Northern chamois is lower than in most other ungulate species in Europe (Buczek et al., 2016; Schaschl et al., 2006). Recombination has also been detected at sites at or near a PBR codons. Common features in the evolution of MHC polymorphism are duplication of genes, recombination and point mutations (Hughes and Yeager, 1998). Our results suggest that positive selection likely played an important role in the historical evolution of genes involved in pathogen recognition in chamois. Additional analysis of individual codons indicated ten codons to be under positive selection. Amino acid positions 11, 71, and 86 were estimated to be under positive selection by at least three of our tests (Figure 16).

Polymorphisms occurring prior to species divergence due to long lineage maintenance, so-called trans-species polymorphism, are considered evidence of balancing selection (Takahata, 1990; Takahata and Nei, 1990). Consequently, similar or identical alleles are shared by species that evolved from a common ancestor and are therefore more closely related than MHC alleles within a species. The Northern and Southern chamois still share some MHC alleles (Figure 10), although the divergence of species is estimated to be about 1.6 million years before present (Lalueza-Fox et al., 2005). The most common allele in both Southern and Northern chamois, Ruru-DRB01 (referred to as Rupy-DRB02 by Alvarez-Busto et al., 2007), was previously reported as a shared allele in both chamois species (Schaschl et al., 2005). Two other shared alleles were found in this study: Rupy-DRB04 and Rupy-DRB13, which were previously found in Southern chamois (Figure 10). Nevertheless, the sharing of alleles between species that are closely related may not be due to balancing selection, but the result of incomplete lineage sorting (Těšický and Vinker, 2015). Following a speciation event, shared genetic variation may gradually disappear or be maintained by balancing selection for longer than would be expected under neutral processes (genetic drift and mutation). However, trans-species polymorphism between species that have only recently diverged, such as chamois, does not necessarily indicate balancing selection, i.e., there may not have been sufficient time for divergence of MHC alleles between species (Těšický and Vinker, 2015). Another possible explanation for trans-species polymorphism was proposed to be relatively recent extensive hybridization between the two chamois species during the late glacial maximum (Schaschl et al. 2005; Rodríguez et al. 2010).

When living in sympatry, introgression of resistance alleles may be advantageous and significantly increase MHC diversity (Nadachowska-Brzyska et al. 2012).

Remarkable differences in nucleotide diversity were observed, in particular for silent substitutions, ranging from dS=0.01 for Southern chamois to dS=0.004 for Northern chamois, pointing to the species differing demographic histories, although the small sample size of Southern chamois requires cautious interpretation. In fact, rates of synonymous and non-synonymous substitutions in MHC DRB exon 2 alleles are generaly lower in chamois than in the other ruminant species studied by Schaschl et al. (2006). Rate of non-synonymous substitutions was generally lower in chamois than in the other analysed ruminants, but it was the rate of synonymous substitutions that was generally very low, which resulted in the high dN/dS ratio. Therefore, the dN/dS ratio was higher in chamois than in the other Caprinae examined in the Schaschl et al. (2006) study, reflecting differences in species demography.

The value of dN/dS ratio may provide information about the age of alleles at a particular locus (Arbanasić et al., 2013). With the accumulation of synonymous mutations over time, the dN/dS value decreases, and consequently the evidence of positive selection decreases over the course of evolution (Oosterhout et al., 2006). For this reason, higher ratio of dN/dS in Northern chamois may not indicate that positive selection is stronger at that locus, but that polymorphism is younger, i.e., it may reflect the young age of most chamois alleles (Schaschl et al., 2004). Ratio dN/dS was high across MHC DRB sequences and various approaches involving different statistical methods identified a limited number of codons affected by selection. Inferences drawn from selection tests and the putative trans-species polymorphism, suggest that balancing selection has strongly influenced the variability of MHC class II DRB locus in chamois. Moreover, selection analyses suggested that the DRB locus has been affected by different evolutionary processes in the Northern and Southern chamois. This study therefore argues that there is good evidence that selection shape sequence diversity in chamois MHC DRB genes. However, it is also emphasised that this work is limited by the discrepancy between the sample size for species/subspecies and by the small sample size for Southern chamois.

Although comparisons between species/subspecies are confounded by demographic histories and the lack of samples at different temporal and spatial scales, they can be useful to identify more general trends in MHC diversity (Cai et al., 2015). In the case of chamois, it is important to examine the relative diversity of the MHC DRB locus among *Rupicapra* subspecies that have not been analysed until this study.

Investigation of immunogenetic status, particularly of small or endangered chamois populations, would allow risk assessment for conservation management (Acevedo-Whitehouse and Cunningham 2006). Further screening of MHC class I loci is also necessary, especially given that some populations of Southern chamois were affected by severe outbreaks of a disease caused by pestivirus in the early 1990s (Arnal et al., 2004). An important question that remains to be answered is the extent to which selection shapes the adaptive genetics of chamois in its complex scenarios where other evolutionary processes such as genetic drift and introgression play important roles. This is particularly important at a time when climate change poses a major threat to chamois populations.

5.4. Comparing SSCP and Ion Torrent sequencing

The results of this study have shown that next-generation sequencing technologies perform better in sequencing highly variable regions such as the MHC and may offer real advantages compared with the SSCP/Sanger method. However, knowing the complexity of the MHC system in the species under study and, accordingly, the read depth that would be satisfactory for obtaining such genotypes is really important to exclude sequencing errors (Biedrzycka et al., 2017). In this case, at least 600 reads per amplicon was sufficient to separate true alleles from artefacts and obtain reliable genotypes.

In the conservation and management of species, genetic monitoring is essential to ensure that appropriate measures are taken to maintain the viability and adaptive potential of the population, particularly in the case of small and declining populations (Leroy et al., 2018). It is therefore important how easily genetic data can be obtained. Traditional methods were commonly used for MHC genotyping, but the introduction of next-generation sequencing facilitated more accurate and reproducible genotyping of such polymorphic gene families and allowed for a large-scale assessment of genetic variation at reasonable times and costs (Reuter et al., 2015). Nevertheless, due to the high polymorphism of the MHC genes, genotyping is inherently difficult. Here, we compared the performance of traditional and next-generation sequencing on genotyping of the MHC class II DRB locus in chamois. With conventional SSCP/Sanger sequencing, we were able to detect seven alleles, all of which were also detected with NGS. We found inconsistencies in the individual genotypes between the two methods, which were mainly caused by allelic dropout in the SSCP/Sanger method. Six out of 28 individuals were falsely classified as homozygous with SSCP/Sanger analysis. Although the MHC system in chamois is quite simple, we found genotyping discrepancies between the two methods.

Overall, discrepancies in genotyping between the two methods were found in 25% of individuals, indicating a higher detection capacity and more accurate genotyping when using NGS and, on the other hand, an underestimation of the true MHC variability when using SSCP/Sanger sequencing.

More than a decade of considerable effort has been devoted to MHC genotyping in evolutionary and population studies (Huang et al., 2018; Rekdal et al., 2018). However, the task is still quite demanding and challenging due to the complex genomic organisation and the high sequence variation of MHC loci, as well as the difficulties to separate true alleles from artefacts (Grogan et al., 2016; Lighten at al., 2014; Sebastian et al., 2014). MHC genes, often in multiple copies, vary widely between and within species, making the identification of all alleles carried by an individual and the reconstruction of multilocal genotype very challenging (Grogan et al., 2016). Consequently, the problems that cause significant difficulties in MHC genotyping and can result in ambiguous genotyping are (i) frequent gene duplications and variation between haplotypes in the number of loci within and between species, (ii) difficult design of locus-specific primers, (iii) varying degrees of concerted evolution, and (iv) the presence of pseudogens, which cause additional difficulties in identifying functional variants (Huang et al., 2019). Although there have been significant advances in high-throughput sequencing technologies, genotyping of MHC systems and the ability to distinguish between true alleles and artefacts is more challenging because of the increasing number of co-amplifying genes in this method. Of the available high-throughput sequencing platforms, Ion Torrent and Illumina are among the most suitable options for MHC genotyping due to their extremely high coverage and read lengths and offer the potential to overcome this limitation (Rekdal et al., 2018).

Recently, there have been many publications highlighting errors that can occur during the NGS sequencing process (Lighten et al., 2014a; Rekdal et al., 2018). It is important to recognise that long-term limitations can also occur in genotyping by sequencing because, similar to traditional SSCP/Sanger genotyping, NGS sequencing approaches are still based on amplification of the target DNA sequence by PCR. The number of variants corresponding to true alleles may be greatly exceeded by the number of artefacts that occur during the PCR step, and the total number of reads derived from artefacts may be higher than those from true alleles (Sommer et al., 2013). PCR may also be responsible for the allelic dropout caused by differences in amplification efficiency between alleles, especially in the SSCP/Sanger methodology (Montero et al., 2019). Allelic dropouts and false alleles can lead to significant genotyping errors.

This in turn can have a significant impact on downstream analyses based on heterozygosity, allelic diversity, and genotype composition that directly affect population genetics (Sommer et al., 2013). Stochastic variation in allele frequencies will inevitably worsen with NGS methods as the number of reads is reduced. Ion Torrent sequencing is currently one of the methods of choice for MHC genotyping, but for accurate genotyping it is important to cover a sufficient number of reads per amplicon and to apply rigorous approaches to quality control of generated data and allele calling (Grogan et al., 2016b).

Using Illumina sequencing, Biedrzycka et al. (2017) tested four genotyping approaches to characterize MHC diversity in a passerine bird where the maximum number of co-amplifying alleles exceeded 40 per individual. The four different genotyping methods tested in their study were developed based on different principles: (i) setting a threshold based on the observed frequencies of variants that could be explained as sequencing errors (Radwan et al., 2012); (ii) using two amplicon replicates to separate true alleles from artefacts (Sommer et al., 2013); (iii) establish a threshold between true alleles and artefacts based on the drop in sequencing depth between them (Lighten et al., 2014b); and (iv) use the error rate of a given high-throughput sequencing platform to cluster true alleles with their potential artefacts (Sebastian et al., 2016). Despite these differences, the authors found generally high agreement between genotyping methods and showed that high-throughput sequencing allows reliable genotyping of highly complex MHC systems when sufficient coverage per amplicon is achieved (in their case, at least 5,000 reads per amplicon). NGS technologies will most likely replace cloning and Sanger sequencing for genotyping MHC. However, conventional genotyping methods may remain an option to consider for genotyping a limited number of samples, as NGS is still expensive (Sommer et al., 2013). The results of Ion Torrent and traditional cloning/Sanger sequencing in this study are consistent with the results of Sommer et al. (2013), who showed that the results of 454 pyrosequencing and traditional cloning/Sanger sequencing are qualitatively comparable, but that more putative allelic variants can be detected with NGS. It must be emphasized that this lack of compatibility between NGS and traditional methods is probably species-specific and depends on the complexity of the MHC II DRB system under investigation. The chamois MHC II DRB system is quite simple with only one functional copy of the gene, but the difference between the methods can be much bigger in more complex systems (Biedrzycka et al., 2017). This result supports the opinion that the higher probability of allele detection with NGS is based on the higher sequencing depth, which can compensate for the limitations of SSCP/Sanger sequencing, since the difficult amplification of some alleles i.e., differences in amplification efficiency between alleles, is a real limitation of this traditional method.

Accurate characterization of the MHC diversity of an individual is the most important prerequisite for understanding the functional significance of MHC allelic diversity in evolutionary ecology, pathogen resistance, and conservation (Sommer et al., 2013). The comparison of the methodological approaches is necessary to enable researchers to evaluate in detail and understand the data generated by NGS, which will improve confidence in the approach and in subsequent analyses and applications. The major histocompatibility complex is polymorphic and polygenic in most species and therefore inherently difficult to genotype. We have shown that NGS technologies perform better in sequencing highly variable regions such as the MHC compared with the SSCP/Sanger method, but stringent allele calling pipelines are required to distinguish between true alleles and artefacts. Since SSCP/Sanger genotyping may underestimate the actual MHC variability, this method will most likely have very limited applicability. However, SSCP/Sanger analysis may remain an alternative for genotyping a limited number of samples and may be applicable to species with a less complex MHC system, such as chamois.

6. CONCLUSIONS

- Chamois has low to moderate level of genetic diversity of the second exon of the MHC class II DRB locus which are likely to be shaped by past demographic processes and balancing selection.
- A total of 25 alleles were found at the DRB locus, each of which translated into a unique amino acid sequence, indicating the functional significance of polymorphism between alleles.
- Fourteen novel DRB exon 2 alleles were identified in this study, two of which were found only in *R. r. carpatica*, two in *R. r. asiatica*, three in *R. r. balcanica*, and four only in *R. r. rupicapra*.
- Trans-species polymorphism has been confirmed and is evidence for the effect of balancing selection at this locus during the evolution of the species.
- A significant excess of non-synonymous over synonymous mutations was detected in alleles, indicating positive selection at this locus during the evolution of the species.
- A higher number of non-synonymous than synonymous mutations in the protein binding sites of the DRB locus is evidence of balancing selection.
- Most of the amino acid changes found among DRB alleles are located in the binding site of the receptor, which indicates the functionality of the polymorphism and its importance for the immunocompetence of individuals
- Amino acid evolutionary distances are greater than nucleotide distances, which is due to retention of non-synonymous mutations due to balancing selection.
- Recombination between alleles was confirmed, consistent with the important role of recombination in the formation of polymorphisms in the DRB gene.
- The existence of codons affected by positive selection was determined at the investigated locus.
- A gene duplication was not identified. The second locus was either not expressed or was not amplified under the conditions used.
- NGS technologies are better performing in sequencing highly variable regions such as the MHC compared to the SSCP/Sanger methodology, allowing more accurate genotyping of highly polymorphic genes.

7. REFERENCES

Acevedo-Whitehouse K., Cunningham A.A. (2006). Is MHC enough for understanding wildlife immunogenetics? Trends Ecol Evol 21 (8): 433–438. doi:10.1016/j.tree.2006.05.010

Alcaide M. (2010). On the relative roles of selection and genetic drift in shaping MHC variation. Mol Ecol 19 (18): 3842–3844. doi:10.1111/J.1365-294X.2010.04772.X

Alvarez-Busto J., García-Etxebarria K., Herrero J., Garin I., Jugo B.M. (2007). Diversity and evolution of the Mhc-*DRB1* gene in the two endemic Iberian subspecies of Pyrenean chamois, *Rupicapra pyrenaica*. Heredity 99 (4): 406–413. doi:10.1038/sj.hdy.6801016

Ambarlı, H. 2014. Status and management of Anatolian chamois (*Rupicapra rupicapra asiatica*): implications for conservation. In: Chamois international congress (Antonucci A., Di Domenico G., eds.), Lama dei Peligni, Majella National Park, Italy, pp. 137-138.

Anderwald P., Ambarli H., Avramov S., Ciach M., Corlatti L., Farkas A., Jovanovic M., Papaioannou H., Peters W., Sarasa M., Šprem N., Weinberg P., Willisch C. (2021). *Rupicapra rupicapra (amended version of 2020 assessment). The IUCN Red List of Threatened Species* 2021: e.T39255A195863093. Available at: https://dx.doi.org/10.2305/IUCN.UK.2021-1.RLTS.T39255A195863093.en [Accessed 11 March 2022]

Arbanasić H., Đuras M., Podnar M., Gomerčić T., Ćurković S., Galov A. (2014). Major histocompatibility complex class II variation in bottlenose dolphin from Adriatic Sea: inferences about the extent of balancing selection. Mar Biol 161 (10): 2407–2422. doi:10.1007/s00227-014-2515-6

Arbanasić H., Huber Đ., Kusak J., Gomerčić T., Hrenović J., Galov A. (2013). 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 81 (1): 19–27. doi:10.1111/tan.12029

Arnal M.C., Fernández-de-Luco D., Riba L., Maley M., Gilray J., Willoughby K., Vilcek S., Nettleton P.F. (2004). A novel pestivirus associated with deaths in Pyrenean chamois (*Rupicapra pyrenaica pyrenaica*). J Gen Virol 85 (12): 3653–3657. doi:10.1099/vir.0.80235-0

Awadi A., Ben Slimen H., Smith S., Knauer F., Makni M., Suchentrunk F. (2018). Positive selection and climatic effects on MHC class II gene diversity in hares (*Lepus capensis*) from a steep ecological gradient. Sci Rep 8 (1): 11514.

doi:10.1038/s41598-018-29657-3

Babik W. (2010). Methods for MHC genotyping in non-model vertebrates. Mol Ecol Resour 10 (2): 237–251. doi:10.1111/j.1755-0998.2009.02788.x

Babik W., Pabijan M., Radwan J. (2008). Contrasting patterns of variation in MHC loci in the Alpine newt. Mol Ecol 17 (10): 2339–2355. doi:10.1111/J.1365-294X.2008.03757.X

Babik W., Taberlet P., Ejsmond M.J., Radwan J. (2009). New generation sequencers as a tool for genotyping of highly polymorphic multilocus MHC system. Mol Ecol Resour 9 (3): 713–719. doi:10.1111/j.1755-0998.2009.02622.x

Bernatchez L., Landry C. (2003). MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? J Evol Biol 16 (3): 363–377. doi.org/10.1046/j.1420-9101.2003.00531.x

Biedrzycka A., Konopiński M., Hoffman E., Trujillo A., Zalewski A. (2020). Comparing raccoon major histocompatibility complex diversity in native and introduced ranges: Evidence for the importance of functional immune diversity for adaptation and survival in novel environments. Evol Appl 13 (4): 752–767. doi:10.1111/eva.12898

Biedrzycka A., Sebastian A., Migalska M., Westerdahl H., Radwan J. (2017). Testing genotyping strategies for ultra-deep sequencing of a co-amplifying gene family: MHC class I in a passerine bird. Mol Ecol Resour 17 (4): 642–655. doi:10.1111/1755-0998.12612

Brown J.H. (2001). Mammals on mountainsides: elevational patterns of diversity. Glob Ecol Biogeogr 10 (1): 101–109. doi:10.1046/J.1466-822X.2001.00228.X

Brown J.H., Jardetzky T.S., Gorga J.C., Stern L.J., Urban R.G., Strominger J.L., Wiley D.C. (1993). Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. Nature 364 (6432): 33–39. doi:10.1038/364033a0

Bryja J., Galan M., Charbonnel N., Cosson J.F. (2005). Analysis of major histocompatibility complex class II gene in water voles using capillary electrophoresis-single stranded conformation polymorphism. Mol Ecol Notes 5 (1): 173–176. doi:10.1111/j.1471-8286.2004.00855.x

Buczek M., Okarma H., Demiaszkiewicz A.W., Radwan J. (2016). MHC, parasites and antler development in red deer: No support for the Hamilton & Zuk hypothesis. J Evol Biol 29 (3): 617–632. doi:10.1111/jeb.12811

Buzan E.V., Bryja J., Zemanová B., Kryštufek B. (2013). Population genetics of chamois in the contact zone between the Alps and the Dinaric Mountains: Uncovering the role of habitat fragmentation and past management. Conserv Genet 14 (2): 401–412. doi:10.1007/s10592-013-0469-8

Buzan E., Potušek S., Urzi F., Pokorny B., Šprem N. (2020). Genetic characterisation of

wild ungulates: Successful isolation and analysis of DNA from widely available bones can be cheap, fast and easy. ZooKeys 965: 141–156. doi:10.3897/zookeys.965.54862

Cai R., Shafer A.B.A., Laguardia A., Lin Z., Liu S., Hu D. (2015). Recombination and selection in the major histocompatibility complex of the endangered forest musk deer (*Moschus berezovskii*). Sci Rep 5: 17285. doi:10.1038/srep17285

Cavallero S., Marco I., Lavín S., D'Amelio S., López-Olvera J.R. (2012). Polymorphisms at MHC class II DRB1 exon 2 locus in Pyrenean chamois (*Rupicapra pyrenaica pyrenaica*). Infect Genet Evol 12 (5): 1020–1026. doi:10.1016/j.meegid.2012.02.017

Charbonnel N., Pemberton J. (2005). A long-term genetic survey of an ungulate population reveals balancing selection acting on MHC through spatial and temporal fluctuations in selection. Heredity 95 (5): 377–388. doi:10.1038/sj.hdy.6800735

Chirichella R., Stephens P.A., Mason T.H.E., Apollonio M. (2021). Contrasting effects of climate change on Alpine chamois. J Wildl Manage 85 (1): 109–120. doi:10.1002/jwmg.21962

Corlatti L., Herrero J., Ferretti F., Anderwald P., García-González R., Hammer S., Nores C., Rossi L., Lovari S. (2021). Northern Chamois, *Rupicapra rupicapra* (Linnaeus, 1758) and Southern Chamois, *Rupicapra pyrenaica* Bonaparte, 1845. In: Handbook of the Mammals of Europe – Terrestrial Cetartiodactyla (Corlatti L., Zachos F., eds.), Springer Nature

Corlatti L., Iacolina L., Safner T., Apollonio M., Buzan E., Ferretti F., Hammer S., Herrero J., Rossi L., Serrano E., Arnal M.C., Brivio F., Chirichella R., Cotza A., Espunyes J., Fernández de Luco D., Gačić D., Grassi L., Grignolio S., Hauffe H.C., Kavčić K., Kinser A., Lioce F., Malagnino A., Miller C., Pokorny B., Reiner R., Rezić A., Stipoljev S., Tešija T., Yankov Y., Zwijacz-Kozica T., Šprem N. (2022). Past, present, and future of chamois science. Wildlife Biology e01025. doi:10.1002/WLB3.01025

Corlatti L., Lorenzini R., Lovari S. (2011). The conservation of the chamois *Rupicapra* spp. Mamm Rev 41 (2): 163–174. doi:10.1111/j.1365-2907.2011.00187.x

Crestanello B., Pecchioli E., Vernesi C., Mona S., Martínková N., Janiga M., Hauffe H.C., Bertorelle G. (2009). The genetic impact of translocations and habitat fragmentation in chamois (*Rupicapra*) spp. J Hered 100 (6): 691–708. doi:10.1093/jhered/esp053

Čížková D., de Bellocq J.G., Baird S.J.E., Piálek J., Bryja J. (2011). Genetic structure and contrasting selection pattern at two major histocompatibility complex genes in wild house mouse populations. Heredity 106 (5): 727–740. doi:10.1038/hdy.2010.112

Daszak P., Cunningham A.A., Hyatt A.D. (2000). Emerging infectious diseases of wildlife--threats to biodiversity and human health. Science 287 (5452): 443–449. doi:10.1126/SCIENCE.287.5452.443

Doytchinova I.A., Flower D.R. (2005). In silico identification of supertypes for class II MHCs. J Immunol 174 (11): 7085–7095. doi:10.4049/jimmunol.174.11.7085

Eizaguirre C., Baltazar-Soares M. (2014). Evolutionary conservation-evaluating the adaptive potential of species. Evol Appl 7 (9): 963–967. doi:10.1111/eva.12227

Ejsmond M.J., Radwan J. (2015). Red queen processes drive positive selection on major histocompatibility complex (MHC) genes. PLoS Comput Biol 11 (11): e1004627. doi:10.1371/journal.pcbi.1004627

Excoffier L., Lischer H. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10 (3): 564–567. doi:10.1111/J.1755-0998.2010.02847.X

Ezard T.H.G., Travis J.M.J. (2006). The impact of habitat loss and fragmentation on genetic drift and fixation time. Oikos 114 (2): 367–375. doi:10.1111/J.2006.0030-1299.14778.X

Frankham R. (2005). Genetics and extinction. Biol Conserv 126 (2): 131–140. doi:10.1016/j.biocon.2005.05.002

Frankham R., Ballou J.D., Dudash M.R., Eldridge M.D.B., Fenster C.B., Lacy R.C., Mendelson J.R., Porton I.J., Ralls K., Ryder O.A. (2012). Implications of different species concepts for conserving biodiversity. Biol Conserv 153: 25–31. doi:10.1016/j.biocon.2012.04.034

Fuchs K., Deutz A., Gressmann G. (2000). Detection of space-time clusters and epidemiological examinations of scabies in chamois. Vet Parasitol 92 (1): 63–73. doi:10.1016/S0304-4017(00)00269-7

Funk W.C., McKay J.K., Hohenlohe P.A., Allendorf F.W. (2012). Harnessing genomics for delineating conservation units. Trends Ecol Evol 27 (9): 489–496. doi:10.1016/j.tree.2012.05.012

Fuselli S., Baptista R.P., Panziera A., Magi A., Guglielmi S., Tonin R., Benazzo A., Bauzer L.G., Mazzoni C.J., Bertorelle G. (2018). A new hybrid approach for MHC genotyping: high-throughput NGS and long read MinION nanopore sequencing, with application to the non-model vertebrate Alpine chamois (*Rupicapra rupicapra*). Heredity 121 (4): 293–303. doi:10.1038/s41437-018-0070-5

Gao F., Chen C., Arab D.A., Du Z., He Y., Ho S.Y.W. (2019). EasyCodeML: A visual tool for analysis of selection using CodeML. Ecol Evol 9 (7): 3891–3898. doi:10.1002/ece3.5015

Garrigan D., Hedrick P.W. (2001). Class I MHC polymorphism and evolution in endangered California Chinook and other Pacific salmon. Immunogenetics 53 (6): 483–489. doi:10.1007/s002510100352

Garrigan D., Hedrick P.W. (2003). Perspective: detecting adaptive molecular

polymorphism: lessons from the MHC. Evolution 57 (8): 1707–1722. doi:10.1111/J.0014-3820.2003.TB00580.X

Grogan K.E., McGinnis G.J., Sauther M.L., Cuozzo F.P., Drea C.M. (2016). Next-generation genotyping of hypervariable loci in many individuals of a non-model species: technical and theoretical implications. BMC Genomics 17 (1): 204. doi:10.1186/s12864-016-2503-y

Hedrick P.W. (2002). Pathogen resistance and genetic variation at MHC loci. Evolution (N Y) 56 (10): 1902–1908. doi:10.1111/J.0014-3820.2002.TB00116.X

Herrero J., Lovari, S.Nores C., Toigo C. (2020). *Rupicapra pyrenaica. The IUCN Red List of Threatened Species* 2020: e.T19771A171131310. Available at: https://dx.doi.org/10.2305/IUCN.UK.2020-2.RLTS.T19771A171131310.en [Accessed 11 March 2022]

Huang K., Zhang P., Dunn D.W., Wang T., Mi R., Li B. (2019). Assigning alleles to different loci in amplifications of duplicated loci. Mol Ecol Resour 19 (5): 1240–1253. doi:10.1111/1755-0998.13036

Hughes A.L., Yeager M. (1998). Natural selection at major histocompatibility complex loci of vertebrates. Annu Rev Genet 32: 415–435. doi:10.1146/annurev.genet.32.1.415

Huson D.H., Bryant D. (2006). Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23 (2): 254–267. doi:10.1093/MOLBEV/MSJ030

Iacolina L., Corlatti L., Buzan E., Safner T., Šprem N. (2019). Hybridisation in European ungulates: an overview of the current status, causes, and consequences. Mamm Rev 49 (1): 45–59. doi:10.1111/mam.12140

Jombart T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. Bioinformatics 24 (11): 1403–1405. doi:10.1093/bioinformatics/btn129

Jombart T., Devillard S., Balloux F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet 11 (1): 94. doi:10.1186/1471-2156-11-94

Juola F.A., Dearborn D.C. (2011). Sequence-based evidence for major histocompatibility complex-disassortative mating in a colonial seabird. Proc R Soc B Biol Sci 279 (1726): 153–162. doi:10.1098/rspb.2011.0562

Kalinowski S.T. (2004). Counting alleles with rarefaction: private alleles and hierarchical sampling designs. Conserv Genet 5 (4): 539–543. doi:10.1023/B:COGE.0000041021.91777.1a

Klein J., Bontrop R.E., Dawkins R.L., Erlich H.A., Gyllensten U.B., Heise E.R., Jones P.P., Parham P., Wakeland E.K., Watkins D.I. (1990). Nomenclature for the major histocompatibility complexes of different species: a proposal. Immunogenetics 31 (4): 217–219. doi:10.1007/BF00204890

Kumar S., Stecher G., Li M., Knyaz C., Tamura K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35 (6): 1547–1549. doi:10.1093/molbev/msy096

Kumar S., Stecher G., Tamura K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol 33 (7): 1870–1874. doi:10.1093/molbev/msw054

Lalueza-Fox C., Castresana J., Sampietro L., Marquès-Bonet T., Alcover J.A., Bertranpetit J. (2005). Molecular dating of caprines using ancient DNA sequences of *Myotragus balearicus*, an extinct endemic Balearic mammal. BMC Evol Biol 5 (1): 70. doi:10.1186/1471-2148-5-70

Leigh J.W., Bryant D. (2015). POPART: Full-feature software for haplotype network construction. Methods Ecol Evol 6 (9): 1110–1116. doi:10.1111/2041-210X.12410

Leroy G., Carroll E.L., Bruford M.W., DeWoody J.A., Strand A., Waits L., Wang J. (2018). Next-generation metrics for monitoring genetic erosion within populations of conservation concern. Evol Appl 11 (7): 1066–1083. doi:10.1111/eva.12564

Lighten J., van Oosterhout C., Bentzen P. (2014a). Critical review of NGS analyses for de novo genotyping multigene families. Mol Ecol 23 (16): 3957–3972. doi:10.1111/mec.12843

Lighten J., van Oosterhout C., Paterson I.G., Mcmullan M., Bentzen P. (2014b). Ultra-deep Illumina sequencing accurately identifies MHC class IIb alleles and provides evidence for copy number variation in the guppy (*Poecilia reticulata*). Mol Ecol Resour 14 (4): 753–767. doi:10.1111/1755-0998.12225

Markov G., Zhelev P., Ben Slimen H., Suchentrunk F. (2016). Population genetic data pertinent to the conservation of Bulgarian chamois (*Rupicapra rupicapra balcanica*). Conserv Genet 17 (1): 155–164. doi:10.1007/s10592-015-0768-3

Martin D., Rybicki E. (2000). RDP: detection of recombination amongst aligned sequences. Bioinformatics 16 (6): 562–563. doi:10.1093/bioinformatics/16.6.562

Miller H.C., Lambert D.M. (2004). Gene duplication and gene conversion in class II MHC genes of New Zealand robins (Petroicidae). Immunogenetics 56 (3): 178–191. doi:10.1007/s00251-004-0666-1

Mona S., Crestanello B., Bankhead-Dronnet S., Pecchioli E., Ingrosso S., D'Amelio S., Rossi L., Meneguz P.G., Bertorelle G. (2008). Disentangling the effects of recombination, selection, and demography on the genetic variation at a major histocompatibility complex class II gene in the alpine chamois. Mol Ecol 17 (18): 4053–4067. doi:10.1111/j.1365-294X.2008.03892.x

Montero B.K., Refaly E., Ramanamanjato J.B., Randriatafika F., Rakotondranary S.J., Wilhelm K., Ganzhorn J.U., Sommer S. (2019). Challenges of next-generation

sequencing in conservation management: insights from long-term monitoring of corridor effects on the genetic diversity of mouse lemurs in a fragmented landscape. Evol Appl 12 (3): 425–442. doi:10.1111/eva.12723

Moreno-Santillán D.D., Lacey E.A., Gendron D., Ortega J. (2016). Genetic variation at exon 2 of the MHC class II *DQB* locus in blue whale (*Balaenoptera musculus*) from the Gulf of California. PLoS One 11 (1): e0141296. doi:10.1371/journal.pone.0141296

Nei M., Gojobori T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3 (5): 418–426. doi.org/10.1093/oxfordjournals.molbev.a040410

Nielsen R. (2005). Molecular signatures of natural selection. Annu Rev Genet 39: 197–218. doi:10.1146/annurev.genet.39.073003.112420

Noakes M.A., Reimer T., Phillips R.B. (2003). Genotypic characterization of an MHC class II locus in lake trout (*Salvelinus namaycush*) from Lake Superior by single-stranded conformational polymorphism analysis and reference strand-mediated conformational analysis. Mar Biotechnol 5 (3): 270–278. doi:10.1007/s10126-002-0079-9

O'Brien S.J., Evermann J.F. (1988). Interactive influence of infectious disease and genetic diversity in natural populations. Trends Ecol. Evol. 3: 254–259

O'Connor E.A., Strandh M., Hasselquist D., Nilsson J., Westerdahl H. (2016). The evolution of highly variable immunity genes across a passerine bird radiation. Mol Ecol 25 (4): 977–989. doi:10.1111/mec.13530

Oosterhout C., Joyce D.A., Cummings S.M., Blais J., Barson N.J., Ramnarine I.W., Mohammed R.S., Persad N., Cable J. (2006). Balancing selection, random genetic drift, and genetic variation at the major histocompatibility complex in two wild populations of guppies (*Poecilia reticulata*). Evolution 60 (12): 2562–2574. doi:10.1111/j.0014-3820.2006.tb01890.x

Papaioannou H., Fernández M., Pérez T., Domínguez A. (2019). Genetic variability and population structure of chamois in Greece (*Rupicapra rupicapra balcanica*). Conserv Genet 20 (4): 939–945. doi:10.1007/s10592-019-01177-1

Pérez T., Albornoz J., Domínguez A. (2002). Phylogeography of chamois (*Rupicapra* spp.) inferred from microsatellites. Mol Phylogenet Evol 25 (3): 524–534. doi:10.1016/S1055-7903(02)00296-8

Pérez T., Fernández M., Hammer S.E., Domínguez A. (2017). Multilocus intron trees reveal extensive male-biased homogenization of ancient populations of chamois (*Rupicapra* spp.) across Europe during Late Pleistocene. PLoS One 12 (2): e0170392. doi:10.1371/journal.pone.0170392

Pérez, T., Rodríguez, F., Essler, S., Palacios, B., Albornoz, J., Domínguez, A. (2014).

Biogeography of chamois: What different molecular markers can tell us. In: Chamois International Congress Proceedings (Antonucci A. and G. Di Domenico, eds.), Lama dei Peligni, Majella National Park, Italy.

Petit R.J., Aguinagalde I., De Beaulieu J.L., Bittkau C., Brewer S., Cheddadi R., Ennos R., Fineschi S., Grivet D., Lascoux M., Mohanty A., Müller-Starck G., Demesure-Musch B., Palmé A., Martín J.P., Rendell S., Vendramin G.G. (2003). Glacial refugia: hotspots but not melting pots of genetic diversity. Science 300 (5625): 1563–1565. doi:10.1126/science.1083264

Piertney S.B., Oliver M.K. (2006). The evolutionary ecology of the major histocompatibility complex. Heredity 96 (1): 7–21. doi:10.1038/sj.hdy.6800724

Punt J., Stranford S.A., Jones P.P., Owen J.A. (2019). Kuby Immunology. 8th edition. W. H. Freeman and Company, New York

R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.r-project.org/

Radwan J., Biedrzycka A., Babik W. (2010). Does reduced MHC diversity decrease viability of vertebrate populations? Biol Conserv 143 (3): 537–544. doi:10.1016/j.biocon.2009.07.026

Radwan J., Zagalska-Neubauer M., Cichoń M., Sendecka J., Kulma K., Gustafsson L., Babik W. (2012). MHC diversity, malaria and lifetime reproductive success in collared flycatchers. Mol Ecol 21 (10): 2469–2479. doi:10.1111/j.1365-294X.2012.05547.x

Rekdal S.L., Anmarkrud J.A., Johnsen A., Lifjeld J.T. (2018). Genotyping strategy matters when analyzing hypervariable major histocompatibility complex-Experience from a passerine bird. Ecol Evol 8 (3): 1680–1692. doi:10.1002/ece3.3757

Reuter J.A., Spacek D. V., Snyder M.P. (2015). High-throughput sequencing technologies. Mol Cell 58 (4): 586–597. doi:10.1016/j.molcel.2015.05.004

Rodríguez F., Hammer S., Pérez T., Suchentrunk F., Lorenzini R., Michallet J., Martinkova N., Albornoz J., Domínguez A. (2009). Cytochrome b phylogeography of chamois (*Rupicapra* spp.). Population contractions, expansions and hybridizations governed the diversification of the genus. J Hered 100 (1): 47–55. doi:10.1093/jhered/esn074

Rodríguez F., Pérez T., Hammer S.E., Albornoz J., Domínguez A. (2010). Integrating phylogeographic patterns of microsatellite and mtDNA divergence to infer the evolutionary history of chamois (genus *Rupicapra*). BMC Evol Biol 10 (1): 222. doi:10.1186/1471-2148-10-222

Rossi L., Meneguz P.G., Martin P. De, Rodolfi M. (1995). The epizootiology of sarcoptic mange in chamois, *Rupicapra rupicapra*, from the Italian eastern Alps.

Parassitologia 37 (2-3): 233–240

Rozas J., Ferrer-Mata A., Sanchez-DelBarrio J.C., Guirao-Rico S., Librado P., Ramos-Onsins S.E., Sanchez-Gracia A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol 34 (12): 3299–3302. doi:10.1093/molbev/msx248

Sagonas K., Runemark A., Antoniou A., Lymberakis P., Pafilis P., Valakos E.D., Poulakakis N., Hansson B., Methods S.I.S. (2019). Selection, drift, and introgression shape MHC polymorphism in lizards. Heredity 122 (4): 468–484. doi:10.1038/s41437-018-0146-2

Schaschl H., Goodman S.J., Suchentrunk F. (2004). Sequence analysis of the MHC class II DRB alleles in Alpine chamois (*Rupicapra r. rupicapra*). Dev Comp Immunol 28 (3): 265–277. doi:10.1016/j.dci.2003.08.003

Schaschl H., Suchentrunk F., Hammer S., Goodman S.J. (2005). Recombination and the origin of sequence diversity in the DRB MHC class II locus in chamois (*Rupicapra* spp.). Immunogenetics 57 (1–2): 108–115. doi:10.1007/s00251-005-0784-4

Schaschl H., Wandeler P., Suchentrunk F., Obexer-Ruff G., Goodman S.J. (2006). Selection and recombination drive the evolution of MHC class II *DRB* diversity in ungulates. Heredity 97 (6): 427–437. doi:10.1038/sj.hdy.6800892

Sebastian A., Herdegen M., Migalska M., Radwan J. (2016). AMPLISAS: A web server for multilocus genotyping using next-generation amplicon sequencing data. Mol Ecol Resour 16 (2): 498–510. doi:10.1111/1755-0998.12453

Shackleton D. (1997). Wild sheep and goats and their relatives. International Union for Conservation of Nature, Gland

Soglia D., Rossi L., Cauvin E., Citterio C., Ferroglio E., Maione S., Meneguz P.G., Spalenza V., Rasero R., Sacchi P. (2010). Population genetic structure of Alpine chamois (*Rupicapra r. rupicapra*) in the Italian Alps. Eur J Wildl Res 56 (6): 845–854. doi:10.1007/s10344-010-0382-0

Sommer S. (2005). The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Front Zool 2: 1–18. doi:10.1186/1742-9994-2-16

Sommer S., Courtiol A., Mazzoni C.J. (2013). MHC genotyping of non-model organisms using next-generation sequencing: a new methodology to deal with artefacts and allelic dropout. BMC Genomics 14 (1): 542. doi:10.1186/1471-2164-14-542

Spurgin L.G., Richardson D.S. (2010). How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. Proc R Soc B Biol Sci 277 (1684): 979–988. doi:10.1098/rspb.2009.2084

Sunnucks P., Wilson A.C.C., Beheregaray L.B., Zenger K., French J., Taylor A.C. (2000). SSCP is not so difficult: the application and utility of single-stranded conformation

polymorphism in evolutionary biology and molecular ecology. Mol Ecol 9 (11): 1699–1710. doi:10.1046/j.1365-294X.2000.01084.x

Šprem N., Buzan E. (2016). The genetic impact of chamois management in the dinarides. J Wildl Manage 80 (5): 783–793. doi:10.1002/jwmg.21081

Takahata N. (1990). A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. Proc Natl Acad Sci 87 (7): 2419–2423. doi:10.1073/PNAS.87.7.2419

Takahata N., Nei M. (1990). Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. Genetics 124 (4): 967–978. doi: 10.1093/genetics/124.4.967

Těšický M., Vinkler M. (2015). Trans-Species Polymorphism in Immune Genes: General Pattern or MHC-Restricted Phenomenon? J Immunol Res, 838035 doi:10.1155/2015/838035

Weaver S., Shank S.D., Spielman S.J., Li M., Muse S.V., Kosakovsky Pond S.L. (2018). Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. Mol Biol Evol 35 (3): 773–777. doi:10.1093/molbev/msx335

White K.S., Gregovich D.P., Levi T. (2018). Projecting the future of an alpine ungulate under climate change scenarios. Glob Chang Biol 24 (3): 1136–1149. doi:10.1111/gcb.13919

Willi Y., Van Buskirk J., Schmid B., Fischer M. (2007). Genetic isolation of fragmented populations is exacerbated by drift and selection. J Evol Biol 20 (2): 534–542. doi:10.1111/J.1420-9101.2006.01263.X

Yang Z., Swanson W.J. (2002). Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. Mol Biol Evol 19 (1): 49–57. doi:10.1093/oxfordjournals.molbev.a003981

Zagalska-Neubauer M., Babik W., Stuglik M., Gustafsson L., Cichoń M., Radwan J. (2010). 454 sequencing reveals extreme complexity of the class II Major Histocompatibility Complex in the collared flycatcher. BMC Evol Biol 10 (1): 395. doi:10.1186/1471-2148-10-395

Zemanová B., Hájková P., Hájek B., Martínková N., Mikulíček P., Zima J., Bryja J. (2015). Extremely low genetic variation in endangered Tatra chamois and evidence for hybridization with an introduced Alpine population. Conserv Genet 16 (3): 729–741. doi:10.1007/s10592-015-0696-2

8. AUTHOR'S BIOGRAPHY

Sunčica Stipoljev completed her undergraduate study in Molecular Biology at the Faculty of Science of the University of Zagreb in 2015, where she also received her Master's degree in Ecology and Nature Preservation (*magna cum laude*) in 2018 with the topic "Fine-scale genetic structure of the threatened red coral, *Corallium rubrum* (Linnaeus, 1758) inferred from microsatellite markers" carried out at the Institut de Ciències del Mar - Consejo superior de Investigaciones Científicas (CSIC) in Barcelona, Spain.

Since the same year, she has been employed at the Faculty of Agriculture in Zagreb as a PhD student on the project "DNA as evidence of distribution and vitality of endangered Balkan chamois" within the Croatian Science Foundation project "Young researchers' career development project – training of doctoral students".

She co-authored and published seven a1 scientific papers and participated as author or co-author in seven international scientific conferences. Her scientific interest is genetic research of wildlife using neutral and adaptive genetic markers.

A1 scientific publications:

1. Svetličić I., Konjević D., Bužan E., Bujanić M., Duniš L., **Stipoljev S.**, Martinčić J., Šurina M., Galov A. (2022). Performance comparison of different approaches in genotyping MHC-DRB: the contrast between single-locus and multi-locus species. doi:10.21203/RS.3.RS-1583484/V1 (submitted)

2. Corlatti, L., Iacolina, L., Safner, T., Apollonio, M., Buzan, E., Ferretti, F., Hammer, S., Herrero, J., Rossi, L., Serrano, E., Arnal, M.C., Brivio, F., Chirichella, R., Cotza, A., Espunyes, J., Fernández de Luco, D., Gačić, D., Grassi, L., Grignolio, S., Hauffe, H.C., Kavčić, K., Kinser, A., Lioce, F., Malagnino, A., Miller, C., Pokorny, B., Reiner, R., Rezić, A., **Stipoljev, S.**, Tešija, T., Yankov, Y., Zwijacz-Kozica, T., Šprem, N., 2022. Past, present, and future of chamois science. Wildlife Biology e01025. doi:10.1002/WLB3.01025

3. **Stipoljev, S.**, Safner, T., Gančević, P., Galov, A., Stuhne, T., Svetličić, I., Grignolio, S., Cassinello, J., Šprem, N., 2021. Population structure and genetic diversity of non-native aoudad populations. Sci. Reports 11, 12300. doi.org/10.1038/s41598-021-91678-2

4. Šprem, N., **Stipoljev, S.**, Ugarković, D., Buzan, E., 2021. First genetic analysis of introduced axis deer from Croatia. Mammalian Biology 101, 1121–1125. doi.org/10.1007/s42991-021-00164-9

5. **Stipoljev, S.**, Bužan, E., Rolečková, B., Iacolina, L., Šprem, N., 2020. MHC genotyping by SSCP and amplicon - based NGS approach in chamois. Animals 10, 1–9. doi.org/10.3390/ani10091694

6. Bujanić, M., Bužan, E., Galov, A., Arbanasić, H., Potušek, S., **Stipoljev, S.**, Šprem, N., Križanović, K., Konjević, D., 2020. Variability of the DRB locus of MHC genes class II in red deer (*Cervus elaphus*) from a mountain region of Croatia. Veterinarski arhiv 90, 385–392. doi.org/10.24099/vet.arhiv.0862

7. Arbanasić, H., Konjević, D., Vranković, L., Bujanić, M., **Stipoljev, S.**, Balažin, M., Šprem, N., Škorić, D., Galov, A., 2019. Evolution of MHC class II SLA-DRB1 locus in the Croatian wild boar (*Sus scrofa*) implies duplication and weak signals of positive selection. Animal Genetics 50, 33–41. doi.org/10.1111/age.12734

9. APPENDICES

Appendix 1. Basic data on Northern and Southern chamois individuals included in the study.

	Sample ID	Country	Location	Subspecies	Geographic group	DRB allele	Supertype
1.	LME472	Spain	Pyrenees	*R. p. pyrenaica*	P1	Rupy-DRB11 Ruru-DRB01	ST2 ST3
2.	LME474	Spain	Pyrenees	*R. p. pyrenaica*	P1	Rupy-DRB11 Ruru-DRB01	ST2 ST3
3.	LME475	Spain	Pyrenees	*R. p. pyrenaica*	P1	Rupy-DRB04 Ruru-DRB01	ST1 ST3
4.	LME477	Spain	Pyrenees	*R. p. pyrenaica*	P1	Rupy-DRB11 Ruru-DRB01	ST2 ST3
5.	LME1208	Spain	Pyrenees	*R. p. pyrenaica*	P1	Ruru-DRB01	ST3
6.	LME1209	Spain	Cantabrian Mountains	*R. p. parva*	P1	Ruru-DRB01	ST3
7.	LME1187	Italy	Maiella National Park	*R. p. ornata*	P2	Rupy-DRB13	ST1
8.	LME1188	Italy	Maiella National Park	*R. p. ornata*	P2	Rupy-DRB12	ST4
9.	LME551	Slovenia	Osilnica	*R. r. rupicapra*	R1	Ruru-DRB01	ST3
10.	LME616	Slovenia	Postojna	*R. r. rupicapra*	R1	Ruru-DRB01 Ruru-DRB04	ST3 ST5

	Sample ID	Country	Location	Subspecies	Geographic group	DRB allele	Supertype
11.	LME657	Slovenia	Mokrec	*R. r. rupicapra*	R1	Ruru-DRB04 Ruru-DRB41	ST5 ST1
12.	LME680	Slovenia	Triglav National Park	*R. r. rupicapra*	R1	Ruru-DRB01 Ruru-DRB45	ST3 ST3
13.	LME803	Slovenia	Pohorje	*R. r. rupicapra*	R1	Ruru-DRB04	ST5
14.	LME870	Slovenia	Psica	*R. r. rupicapra*	R1	Ruru-DRB01	ST3
15.	LME871	Slovenia	Dedec	*R. r. rupicapra*	R1	Ruru-DRB04 Ruru-DRB43	ST5 ST3
16.	LME885	Slovenia	Begunjščica	*R. r. rupicapra*	R1	Ruru-DRB41 Ruru-DRB48	ST1 ST2
17.	LME580	Slovenia	Pohorje	*R. r. rupicapra*	R1	Ruru-DRB18 Ruru-DRB43	ST3
18.	LME582	Slovenia	Pohorje	*R. r. rupicapra*	R1	Ruru-DRB09 Ruru-DRB18	ST2 ST3
19.	LME589	Slovenia	Pohorje	*R. r. rupicapra*	R1	Ruru-DRB18 Ruru-DRB42	ST3 ST2
20.	LME591	Slovenia	Pohorje	*R. r. rupicapra*	R1	Rupy-DRB04 Ruru-DRB18	ST1 ST3
21.	LME612	Slovenia	Postojna	*R. r. rupicapra*	R1	Ruru-DRB01 Ruru-DRB43	ST3
22.	LME622	Slovenia	Postojna	*R. r. rupicapra*	R1	Ruru-DRB01	ST3
23.	LME626	Slovenia	Postojna	*R. r. rupicapra*	R1	Ruru-DRB01 Ruru-DRB04	ST3 ST5

	Sample ID	Country	Location	Subspecies	Geographic group	DRB allele	Supertype
24.	LME628	Slovenia	Pohorje	*R. r. rupicapra*	R1	Rupy-DRB04 Ruru-DRB04	ST1 ST5
25.	LME632	Slovenia	Pohorje	*R. r. rupicapra*	R1	Ruru-DRB04 Ruru-DRB42	ST5 ST2
26.	LME814	Slovenia	Log pod Mangartom	*R. r. rupicapra*	R1	Rupy-DRB04 Ruru-DRB01	ST1 ST3
27.	LME817	Slovenia	Log pod Mangartom	*R. r. rupicapra*	R1	Ruru-DRB04 Ruru-DRB17	ST5 ST4
28.	LME818	Slovenia	Log pod Mangartom	*R. r. rupicapra*	R1	Rupy-DRB04 Ruru-DRB04	ST1 ST5
29.	LME822	Slovenia	Log pod Mangartom	*R. r. rupicapra*	R1	Ruru-DRB01 Ruru-DRB43	ST3
30.	LME829	Slovenia	Log pod Mangartom	*R. r. rupicapra*	R1	Ruru-DRB01	ST3
31.	LME832	Slovenia	Log pod Mangartom	*R. r. rupicapra*	R1	Ruru-DRB01 Ruru-DRB04	ST3 ST5
32.	LME834	Slovenia	Log pod Mangartom	*R. r. rupicapra*	R1	Ruru-DRB01 Ruru-DRB41	ST3 ST1
33.	LME866	Slovenia	Kozji Vrh	*R. r. rupicapra*	R1	Rupy-DRB04 Ruru-DRB01	ST1 ST3
34.	LME867	Slovenia	Medvedjek	*R. r. rupicapra*	R1	Ruru-DRB04 Ruru-DRB41	ST5 ST1
35.	LME869	Slovenia	Grebenec	*R. r. rupicapra*	R1	Ruru-DRB01 Ruru-DRB18	ST3
36.	LME876	Slovenia	Grebenec	*R. r. rupicapra*	R1	Rupy-DRB04 Ruru-DRB01	ST1 ST3

	Sample ID	Country	Location	Subspecies	Geographic group	DRB allele	Supertype
37.	LME878	Slovenia	Stegovnik	*R. r. rupicapra*	R1	Ruru-DRB04 Ruru-DRB48	ST5 ST2
38.	LME881	Slovenia	Storžič	*R. r. rupicapra*	R1	Rupy-DRB04 Ruru-DRB01	ST1 ST3
39.	LME905	Slovenia	Ig	*R. r. rupicapra*	R1	Rupy-DRB04	ST1
40.	LME579	Slovenia	Pohorje	*R. r. rupicapra*	R1	Ruru-DRB04 Ruru-DRB18	ST5 ST3
41.	LME593	Slovenia	Idrija, Jelenk	*R. r. rupicapra*	R1	Ruru-DRB01 Ruru-DRB04	ST3 ST5
42.	LME594	Slovenia	Idrija, Jelenk	*R. r. rupicapra*	R1	Ruru-DRB01 Ruru-DRB04	ST3 ST5
43.	LME596	Slovenia	Idrija, Jelenk	*R. r. rupicapra*	R1	Ruru-DRB01 Ruru-DRB04	ST3 ST5
44.	LME598	Slovenia	Idrija, Jelenk	*R. r. rupicapra*	R1	Ruru-DRB01	ST3
45.	LME603	Slovenia	Idrija 1	*R. r. rupicapra*	R1	Rupy-DRB04 Ruru-DRB04	ST1 ST5
46.	LME610	Slovenia	Idrija 1	*R. r. rupicapra*	R1	Ruru-DRB01 Ruru-DRB04	ST3 ST5
47.	LME635	Slovenia	Pohorje	*R. r. rupicapra*	R1	Ruru-DRB04 Ruru-DRB18	ST5 ST3
48.	LME637	Slovenia	Pohorje	*R. r. rupicapra*	R1	Ruru-DRB04	ST5
49.	LME649	Slovenia	LD Mokrc, Gaberje	*R. r. rupicapra*	R1	Ruru-DRB01	ST3

	Sample ID	Country	Location	Subspecies	Geographic group	DRB allele	Supertype
50.	LME734	Slovenia	Osilnica	*R. r. rupicapra*	R1	Ruru-DRB01 Ruru-DRB04	ST3 ST5
51.	LME897	Slovenia	Ig	*R. r. rupicapra*	R1	Ruru-DRB01	ST3
52.	LME898	Slovenia	Ig	*R. r. rupicapra*	R1	Rupy-DRB04 Ruru-DRB01	ST1 ST3
53.	LME899	Slovenia	Ig	*R. r. rupicapra*	R1	Rupy-DRB04 Ruru-DRB01	ST1 ST3
54.	LME900	Slovenia	Ig	*R. r. rupicapra*	R1	Ruru-DRB04 Ruru-DRB43	ST5 ST3
55.	LME903	Slovenia	Ig	*R. r. rupicapra*	R1	Ruru-DRB04 Ruru-DRB43	ST5 ST3
56.	LME434	Austria	Aussee	*R. r. rupicapra*	R1	Rupy-DRB04 Ruru-DRB18	ST1 ST3
57.	LME435	Austria	Hochschwab	*R. r. rupicapra*	R1	Ruru-DRB15 Ruru-DRB17	ST1 ST4
58.	LME452	Croatia	Gorski Kotar	*R. r. rupicapra*	R2	Ruru-DRB01 Ruru-DRB41	ST3 ST1
59.	LME541	Croatia	Gorski Kotar	*R. r. rupicapra*	R2	Rupy-DRB04 Ruru-DRB01	ST1 ST3
60.	LME382	Croatia	North Velebit	*R. r. rupicapra*	R2	Ruru-DRB01 Ruru-DRB04	ST3 ST5
61.	LME383	Croatia	North Velebit	*R. r. rupicapra*	R2	Rupy-DRB04	ST1
62.	LME384	Croatia	North Velebit	*R. r. rupicapra*	R2	Ruru-DRB01 Ruru-DRB45	ST3

	Sample ID	Country	Location	Subspecies	Geographic group	DRB allele	Supertype
63.	LME385	Croatia	North Velebit	*R. r. rupicapra*	R2	Ruru-DRB01	ST3
64.	LME386	Croatia	North Velebit	*R. r. rupicapra*	R2	Rupy-DRB04 Ruru-DRB01	ST1 ST3
65.	LME387	Croatia	North Velebit	*R. r. rupicapra*	R2	Ruru-DRB01	ST3
66.	LME367	Croatia	Biokovo	*R. r. balcanica*	B1	Ruru-DRB17 Ruru-DRB44	ST4 ST5
67.	LME373	Croatia	Biokovo	*R. r. balcanica*	B1	Ruru-DRB19	ST3
68.	LME371	Croatia	Biokovo	*R. r. balcanica*	B1	Ruru-DRB47	ST5
69.	LME374	Croatia	Biokovo	*R. r. balcanica*	B1	Ruru-DRB45 Ruru-DRB47	ST3 ST5
70.	LME379	Croatia	Biokovo	*R. r. balcanica*	B1	Ruru-DRB17 Ruru-DRB44	ST4 ST5
71.	LME380	Croatia	Biokovo	*R. r. balcanica*	B1	Ruru-DRB01	ST3
72.	LME443	Croatia	Dinara	*R. r. balcanica*	B1	Rupy-DRB04 Ruru-DRB01	ST1 ST3
73.	LME444	Croatia	Dinara	*R. r. balcanica*	B1	Rupy-DRB04 Ruru-DRB01	ST1 ST3
74.	LME563	Croatia	Dinara	*R. r. balcanica*	B1	Rupy-DRB04 Ruru-DRB01	ST1 ST3
75.	LME558	Bosnia and Herzegovina	Prenj	*R. r. balcanica*	B2	Ruru-DRB44	ST5

	Sample ID	Country	Location	Subspecies	Geographic group	DRB allele	Supertype
76.	LME1193	Bosnia and Herzegovina	Čabulja	*R. r. balcanica*	B2	Ruru-DRB01	ST3
77.	LME1194	Bosnia and Herzegovina	Čvrsinca	*R. r. balcanica*	B2	Ruru-DRB01 Ruru-DRB44	ST3 ST5
78.	LME1204	Bosnia and Herzegovina	Prenj	*R. r. balcanica*	B2	Ruru-DRB45 Ruru-DRB46	ST3 ST5
79.	LME1207	Bosnia and Herzegovina	Prenj	*R. r. balcanica*	B2	Rupy-DRB13 Ruru-DRB44	ST1 ST5
80.	LME560	Bosnia and Herzegovina	Čvrsnica	*R. r. balcanica*	B2	Ruru-DRB01 Ruru-DRB45	ST3
81.	LME481	Serbia	Kanjon Drine	*R. r. balcanica*	B3	Ruru-DRB17	ST4
82.	LME491	Serbia	Žito	*R. r. balcanica*	B3	Ruru-DRB46 Ruru-DRB49	ST5
83.	LME498	Serbia	Tara	*R. r. balcanica*	B3	Rupy-DRB13 Ruru-DRB01	ST1 ST3
84.	LME499	Serbia	Tara	*R. r. balcanica*	B3	Ruru-DRB17 Ruru-DRB44	ST4 ST5
85.	LME506	Serbia	Tara	*R. r. balcanica*	B3	Ruru-DRB44 Ruru-DRB46	ST5
86.	LME1217	Montenegro	Kanjon Tare	*R. r. balcanica*	B3	Ruru-DRB17 Ruru-DRB46	ST4 ST5
87.	LME1303	Albania	Martanesh	*R. r. balcanica*	B4	Ruru-DRB45 Ruru-DRB50	ST3 ST1
88.	LME1487	Albania	Korab	*R. r. balcanica*	B4	Ruru-DRB01 Ruru-DRB45	ST3

	Sample ID	Country	Location	Subspecies	Geographic group	DRB allele	Supertype
89.	LME1337	North Macedonia	Šar Mountains	*R. r. balcanica*	B4	Ruru-DRB01	ST3
90.	LME1321	North Macedonia	Korab	*R. r. balcanica*	B4	Ruru-DRB01 Ruru-DRB19	ST3
91.	LME1342	Greece	Olympus	*R. r. balcanica*	B4	Ruru-DRB01	ST3
92.	LME567	Bulgaria	Rhodope Mountains	*R. r. balcanica*	B5	Ruru-DRB01 Ruru-DRB19	ST3
93.	LME569	Bulgaria	Rhodope Mountains	*R. r. balcanica*	B5	Ruru-DRB01	ST3
94.	LME571	Bulgaria	Rhodope Mountains	*R. r. balcanica*	B5	Ruru-DRB01	ST3
95.	LME573	Bulgaria	Rhodope Mountains	*R. r. balcanica*	B5	Ruru-DRB45 Ruru-DRB47	ST3 ST5
96.	LME575	Bulgaria	Rhodope Mountains	*R. r. balcanica*	B5	Ruru-DRB01 Ruru-DRB19	ST3
97.	LME522	Slovakia	High Tatras (Tatra National Park)	*R. r. tatrica*	T	Ruru-DRB44 Ruru-DRB46	ST5
98.	LME524	Slovakia	High Tatras (Tatra National Park)	*R. r. tatrica*	T	Ruru-DRB17 Ruru-DRB44	ST4 ST5
99.	LME526	Slovakia	High Tatras (Tatra National Park)	*R. r. tatrica*	T	Ruru-DRB17 Ruru-DRB46	ST4 ST5
100.	LME528	Slovakia	High Tatras (Tatra National Park)	*R. r. tatrica*	T	Ruru-DRB44	ST5

	Sample ID	Country	Location	Subspecies	Geographic group	DRB allele	Supertype
101.	LME530	Slovakia	High Tatras (Tatra National Park)	*R. r. tatrica*	T	Ruru-DRB17 Ruru-DRB44	ST4 ST5
102.	LME532	Slovakia	High Tatras (Tatra National Park)	*R. r. tatrica*	T	Ruru-DRB44 Ruru-DRB46	ST5
103.	LME1396	Romania	Cumpana	*R. r. carpatica*	C	Ruru-DRB15 Ruru-DRB19	ST1 ST3
104.	LME1413	Romania	Plaisor	*R. r. carpatica*	C	Ruru-DRB19 Ruru-DRB53	ST3 ST1
105.	LME1421	Romania	Olanesti	*R. r. carpatica*	C	Ruru-DRB15	ST1
106.	LME1425	Romania	Avrig	*R. r. carpatica*	C	Ruru-DRB19	ST3
107.	LME1426	Romania	Latorita	*R. r. carpatica*	C	Ruru-DRB15 Ruru-DRB54	ST1 ST5
108.	LME1390	Turkey	İspir/Erzurum	*R. r. asiatica*	AC	Ruru-DRB15 Ruru-DRB51	ST1 ST4
109.	LME1391	Turkey	Şavşat/Artvin	*R. r. asiatica*	AC	Ruru-DRB19 Ruru-DRB52	ST3 ST2
110.	LME1341	Russia	Dagestan	*R. r. caucasica*	AC	Ruru-DRB15 Ruru-DRB19	ST1 ST3

Appendix 2. A summary of the DRB alleles found in each chamois individual using single-strand conformation polymorphism (SSCP)/Sanger sequencing and amplicon-based next-generation sequencing (NGS).

Individual	LME579		LME582		LME589		LME591		LME593		LME594		LME596		LME598		LME610	
Allele	sscp	ngs	sscp	ngs	sscp	ngs	sscp	ngs	sscp	ngs	sscp	ngs	sscp	ngs	sscp	ngs	sscp	ngs
Ruru-DRB01									+	+	+	+	+	+	+	+	+	+
Ruru-DRB04	+	+							+	+	+	+	+	+			+	+
Ruru-DRB09			+	+														
Ruru-DRB18	+	+	+	+	+	+	+	+										
Ruru-DRB41																		
Ruru-DRB42					+	+												
Ruru-DRB43																		
Rupy-DRB04							+	+										
Individual	LME612		LME616		LME622		LME626		LME628		LME632		LME635		LME637		LME649	
Allele	sscp	ngs	sscp	ngs	sscp	ngs	sscp	ngs	sscp	ngs	sscp	ngs	sscp	ngs	sscp	ngs	sscp	ngs
Ruru-DRB01	+	+	+	+	+	+	+	+									+	+
Ruru-DRB04			+	+			+	+	+	+	+	+	+	+	+	+		
Ruru-DRB09																		
Ruru-DRB18													+	+				
Ruru-DRB41																		
Ruru-DRB42											+	+						
Ruru-DRB43	+	+																
Rupy-DRB04									+	+								

Individual	LME734		LME897		LME905		LME603		LME657		LME898		LME899		LME900		LME903		LME580	
Allele	sscp	ngs	sscp	ngs	sscp	ngs	sscp	ngs	sscp	ngs	sscp	ngs	sscp	ngs	sscp	ngs	sscp	ngs	sscp	ngs
Ruru-DRB01	+	+	+	+							+	+	+	+						
Ruru-DRB04	+	+					+	+	+	+						+		+	+	
Ruru-DRB09																				
Ruru-DRB18																				+
Ruru-DRB41										+										
Ruru-DRB42																				
Ruru-DRB43															+	+	+	+	+	+
Rupy-DRB04					+	+		+				+		+						

+ false positive allele call

six individuals (21%) falsely classified as homozygous with SSCP/Sanger analysis

Appendix 3. Amplicon depths of 28 individuals, and the number of reads of true alleles and artefacts within each amplicon.
true = true alleles; X = 1 bp substitutions; D = deletions; I = insertions; H = homopolymer indels, XDIH = sequence with any combination of at least two errors; leftovers = low frequency variants, chimeras or sequences containing many errors which could not be classified into the major clusters.

	LME616	true	clustered	amplicon depth		true	X	D	I	H	XDIH		leftovers
Ruru-DRB04	415	249	166			249	109	17	3	18	19	166	
Ruru-DRB01	376	230	146			230	112	5	2	16	11	146	
				984	sum:	479	221	22	5	34	30		193
	LME657	true	clustered	amplicon depth		true	X	D	I	H	XDIH		leftovers
Ruru-DRB04	365	217	148			217	101	4	3	19	21	148	
Ruru-DRB41	258	161	97			161	63	4	4	11	15	97	
				744	sum:	378	164	8	7	30	36		121
	LME580	true	clustered	amplicon depth		true	X	D	I	H	XDIH		leftovers
Ruru-DRB18	338	208	130			208	81	21	4	4	20	130	
Ruru-DRB43	154	105	49			105	20	14	10	1	4	49	
				656	sum:	313	101	35	14	5	24		164
	LME582	true	clustered	amplicon depth		true	X	D	I	H	XDIH		leftovers
Ruru-DRB09	409	271	138			271	109	10	3	5	11	138	
Ruru-DRB18	370	251	119			251	92	6	6	5	10	119	
				1026	sum:	522	201	16	9	10	21		247

	LME589	true	clustered	amplicon depth		true	X	D	I	H	XDIH		leftovers
Ruru-DRB42	460	319	141			319	105	9	6	13	8	141	
Ruru-DRB18	460	300	160			300	120	14	4	10	12	160	
				1106	sum:	619	225	23	10	23	20		186
	LME591	true	clustered	amplicon depth		true	X	D	I	H	XDIH		leftovers
Ruru-DRB18	316	193	123			193	87	8	8	7	13	123	
Rupy-DRB04	259	133	126			133	64	25	5	4	28	126	
				811	sum:	326	151	33	13	11	41		236
	LME612	true	clustered	amplicon depth		true	X	D	I	H	XDIH		leftovers
Ruru-DRB01	537	350	187			350	127	16	3	12	29	187	
Ruru-DRB43	259	157	102			157	54	19	13	1	15	102	
				1023	sum:	507	181	35	16	13	44		227
	LME622	true	clustered	amplicon depth		true	X	D	I	H	XDIH		leftovers
Ruru-DRB01	714	453	261			453	189	16	5	25	26	261	
				1120	sum:	453	189	16	5	25	26		406
	LME626	true	clustered	amplicon depth		true	X	D	I	H	XDIH		leftovers
Ruru-DRB01	456	312	144			312	107	11	4	10	12	144	
Ruru-DRB04	409	267	142			267	102	15	4	11	10	142	
				1135	sum:	579	209	26	8	21	22		270

	LME628	true	clustered	amplicon depth		true	X	D	I	H	XDIH		leftovers
Ruru-DRB04	528	332	196			332	151	11	1	17	16	196	
Rupy-DRB04	380	213	167			213	90	38	6	6	27	167	
				1126	sum:	545	241	49	7	23	43		218
	LME632	true	clustered	amplicon depth		true	X	D	I	H	XDIH		leftovers
Ruru-DRB42	758	535	223			535	174	10	9	19	11	223	
Ruru-DRB04	746	504	242			504	176	20	3	19	24	242	
				1765	sum:	1039	350	30	12	38	35		261
	LME905	true	clustered	amplicon depth		true	X	D	I	H	XDIH		leftovers
Rupy-DRB04	632	396	236			396	159	28	6	18	25	236	
				792	sum:	396	159	28	6	18	25		160
	LME579	true	clustered	amplicon depth		true	X	D	I	H	XDIH		leftovers
Ruru-DRB18	2277	1536	741			1536	534	80	26	25	76	741	
Ruru-DRB04	2229	1521	708			1521	552	61	9	31	55	708	
				5000	sum:	3057	1086	141	35	56	131		494
	LME598	true	clustered	amplicon depth		true	X	D	I	H	XDIH		leftovers
Ruru-DRB01	4592	3055	1537			3055	1078	200	55	53	151	1537	
				5000	sum:	3055	1078	200	55	53	151		408
	LME603	true	clustered	amplicon depth		true	X	D	I	H	XDIH		leftovers

Ruru-DRB04	2227	1441	786			1441	612	79	16	22	57	786	
Rupy-DRB04	2203	1457	746			1457	543	95	22	15	71	746	
				5000	sum:	2898	1155	174	38	37	128		570
	LME635	true	clustered	amplicon depth		true	X	D	I	H	XDIH		leftovers
Ruru-DRB18	2336	1500	836			1500	601	87	20	28	100	836	
Ruru-DRB04	2001	1329	672			1329	530	59	12	19	52	672	
				5000	sum:	2829	1131	146	32	47	152		663
	LME637	true	clustered	amplicon depth		true	X	D	I	H	XDIH		leftovers
Ruru-DRB04	4507	3002	1505			3002	1120	165	33	43	144	1505	
				5000	sum:	3002	1120	165	33	43	144		493
	LME649	true	clustered	amplicon depth		true	X	D	I	H	XDIH		leftovers
Ruru-DRB01	4564	3059	1505			3059	1064	187	49	43	162	1505	
				5000	sum:	3059	1064	187	49	43	162		436
	LME897	true	clustered	amplicon depth		true	X	D	I	H	XDIH		leftovers
Ruru-DRB01	4469	2875	1594			2875	1137	161	46	47	203	1594	
				5000	sum:	2875	1137	161	46	47	203		531
	LME898	true	clustered	amplicon depth		true	X	D	I	H	XDIH		leftovers
Rupy-DRB04	2174	1374	800			1374	575	74	30	26	95	800	
Ruru-DRB01	2143	1376	767			1376	560	88	19	21	79	767	
				5000	sum:	2750	1135	162	49	47	174		683

	LME899	true	clustered	amplicon depth		true	X	D	I	H	XDIH		leftovers
Ruru-DRB01	2217	1404	813			1404	620	67	19	18	89	813	
Rupy-DRB04	2030	1327	703			1327	504	70	28	29	72	703	
				5000	sum:	2731	1124	137	47	47	161		753
	LME593	true	clustered	amplicon depth		true	X	D	I	H	XDIH		leftovers
Ruru-DRB04	898	596	302			596	244	25	6	12	15	302	
Ruru-DRB01	850	590	260			590	199	27	9	10	15	260	
				2071	sum:	1186	443	52	15	22	30		323
	LME594	true	clustered	amplicon depth		true	X	D	I	H	XDIH		leftovers
Ruru-DRB04	2071	1335	736			1335	570	79	5	20	62	736	
Ruru-DRB01	2069	1340	729			1340	556	70	10	16	77	729	
				5000	sum:	2675	1126	149	15	36	139		860
	LME596	true	clustered	amplicon depth		true	X	D	I	H	XDIH		leftovers
Ruru-DRB04	2309	1536	773			1536	570	92	11	21	79	773	
Ruru-DRB01	1898	1233	665			1233	460	77	27	22	79	665	
				5000	sum:	2769	1030	169	38	43	158		793
	LME610	true	clustered	amplicon depth		true	X	D	I	H	XDIH		leftovers
Ruru-DRB04	2251	1476	775			1476	572	98	22	29	54	775	
Ruru-DRB01	1968	1269	699			1269	510	82	18	15	74	699	
				5000	sum:	2745	1082	180	40	44	128		781

	LME734	true	clustered	amplicon depth		true	X	D	I	H	XDIH		leftovers
Ruru-DRB04	2095	1340	755			1340	588	73	9	22	63	755	
Ruru-DRB01	2060	1364	696			1364	510	83	13	23	67	696	
				5000	sum:	2704	1098	156	22	45	130		845
	LME900	true	clustered	amplicon depth		true	X	D	I	H	XDIH		leftovers
Ruru-DRB43	2156	1351	805			1351	550	78	54	27	96	805	
Ruru-DRB04	2045	1333	712			1333	537	84	18	25	48	712	
				5000	sum:	2684	1087	162	72	52	144		799
	LME903	true	clustered	amplicon depth		true	X	D	I	H	XDIH		leftovers
Ruru-DRB43	2183	1408	775			1408	586	82	12	21	74	775	
Ruru-DRB04	2021	1335	686			1335	543	69	8	20	46	686	
				5000	sum:	2743	1129	151	20	41	120		796

Appendix 4. Percentages of true alleles and artefacts within each amplicon and averaged.
true = true alleles; X = 1 bp substitutions; D = deletions; I = insertions; H = homopolymer indels, XDIH = sequence with any combination of at least two errors

individual	amplicon depth	true	%:	X	D	I	H	XDIH	sum:	%:
LME616	984	479	48.68	221	22	5	34	30	312	31.71
LME657	744	378	50.81	164	8	7	30	36	245	32.93
LME580	656	313	47.71	101	35	14	5	24	179	27.29
LME582	1026	522	50.88	201	16	9	10	21	257	25.05
LME589	1106	619	55.97	225	23	10	23	20	301	27.22
LME591	811	326	40.20	151	33	13	11	41	249	30.70
LME612	1023	507	49.56	181	35	16	13	44	289	28.25
LME622	1120	453	40.45	189	16	5	25	26	261	23.30
LME626	1135	579	51.01	209	26	8	21	22	286	25.20
LME628	1126	545	48.40	241	49	7	23	43	363	32.24
LME632	1765	1039	58.87	350	30	12	38	35	465	26.35
LME905	792	396	50.00	159	28	6	18	25	236	29.80
LME579	5000	3057	61.14	1086	141	35	56	131	1449	28.98
LME598	5000	3055	61.10	1078	200	55	53	151	1537	30.74
LME603	5000	2898	57.96	1155	174	38	37	128	1532	30.64
LME635	5000	2829	56.58	1131	146	32	47	152	1508	30.16
LME637	5000	3002	60.04	1120	165	33	43	144	1505	30.10
LME649	5000	3059	61.18	1064	187	49	43	162	1505	30.10
LME897	5000	2875	57.50	1137	161	46	47	203	1594	31.88

individual	amplicon depth	true	%:	X	D	I	H	XDIH	sum:	%:
LME898	5000	2750	55.00	1135	162	49	47	174	1567	31.34
LME899	5000	2731	54.62	1124	137	47	47	161	1516	30.32
LME593	2071	1186	57.27	443	52	15	22	30	562	27.14
LME594	5000	2675	53.50	1126	149	15	36	139	1465	29.30
LME596	5000	2769	55.38	1030	169	38	43	158	1438	28.76
LME610	5000	2745	54.90	1082	180	40	44	128	1474	29.48
LME734	5000	2704	54.08	1098	156	22	45	130	1451	29.02
LME900	5000	2684	53.68	1087	162	72	52	144	1517	30.34
LME903	5000	2743	54.86	1129	151	20	41	120	1461	29.22
average:			53.62							29.20

Appendix 5. Percentage of artefacts resulting from 1 bp substitution within each amplicon and averaged.

individual	amplicon depth	X	%:
LME616	984	221	22.46
LME657	744	164	22.04
LME580	656	101	15.40
LME582	1026	201	19.59
LME589	1106	225	20.34
LME591	811	151	18.62
LME612	1023	181	17.69
LME622	1120	189	16.88
LME626	1135	209	18.41
LME628	1126	241	21.40
LME632	1765	350	19.83
LME905	792	159	20.08
LME579	5000	1086	21.72
LME598	5000	1078	21.56
LME603	5000	1155	23.10
LME635	5000	1131	22.62
LME637	5000	1120	22.40
LME649	5000	1064	21.28
LME897	5000	1137	22.74
LME898	5000	1135	22.70
LME899	5000	1124	22.48
LME593	2071	443	21.39
LME594	5000	1126	22.52
LME596	5000	1030	20.60
LME610	5000	1082	21.64
LME734	5000	1098	21.96
LME900	5000	1087	21.74
LME903	5000	1129	22.58
average:			20.92