



**HAL**  
open science

# The role of the major histocompatibility complex in the wild: the case of the Alpine marmot (*Marmota marmota*)

Mariona Ferrandiz-Rovira

## ► To cite this version:

Mariona Ferrandiz-Rovira. The role of the major histocompatibility complex in the wild: the case of the Alpine marmot (*Marmota marmota*). Animal biology. Université Claude Bernard - Lyon I, 2015. English. NNT: 2015LYO10089 . tel-01186054

**HAL Id: tel-01186054**

**<https://theses.hal.science/tel-01186054>**

Submitted on 24 Aug 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Année 2015

# THÈSE

Présentée devant L'Université Claude Bernard - Lyon 1

pour l'obtention du  
diplôme de Doctorat

(arrêté du 7 août 2006)

Soutenue publiquement le 3 juillet 2015

par

**Mariona FERRANDIZ-ROVIRA**

---

**THE ROLE OF THE MAJOR HISTOCOMPATIBILITY  
COMPLEX IN THE WILD:  
the case of the Alpine marmot (*Marmota marmota*)**

---

Sous la direction de: Dominique ALLAINÉ

Co-encadrement de: Aurélie COHAS et Marie-Pierre CALLAIT-CARDINAL

Jury:

Marie CHARPENTIER

Gabriele SORCI

Elise HUCHARD

Pierre JOLY

Dominique ALLAINÉ

Marie-Pierre CALLAIT-CARDINAL

Aurélie COHAS



# UNIVERSITE CLAUDE BERNARD - LYON 1

**Président de l'Université**

**M. François-Noël GILLY**

Vice-président du Conseil d'Administration

M. le Professeur Hamda BEN HADID

Vice-président du Conseil des Etudes et de la Vie Universitaire

M. le Professeur Philippe LALLE

Vice-président du Conseil Scientifique

M. le Professeur Germain GILLET

Directeur Général des Services

M. Alain HELLEU

## **COMPOSANTES SANTE**

Faculté de Médecine Lyon Est – Claude Bernard

Directeur: M. le Professeur J. ETIENNE

Faculté de Médecine et de Maïeutique Lyon Sud – Charles Mérieux

Directeur: Mme la Professeure C. BURILLON

Faculté d'Odontologie

Directeur: M. le Professeur D. BOURGEOIS

Institut des Sciences Pharmaceutiques et Biologiques

Directeur: Mme la Professeure C. VINCIGUERRA

Institut des Sciences et Techniques de la Réadaptation

Directeur: M. le Professeur Y. MATILLON

Département de formation et Centre de Recherche en Biologie Humaine

Directeur: Mme. la Professeure A-M. SCHOTT

## **COMPOSANTES ET DEPARTEMENTS DE SCIENCES ET TECHNOLOGIE**

Faculté des Sciences et Technologies

Directeur: M. F. DE MARCHI

Département Biologie

Directeur: M. le Professeur F. FLEURY

Département Chimie Biochimie

Directeur: Mme Caroline FELIX

Département GEP

Directeur: M. Hassan HAMMOURI

Département Informatique

Directeur: M. le Professeur S. AKKOUCHE

Département Mathématiques

Directeur: M. le Professeur Georges TOMANOV

Département Mécanique

Directeur: M. le Professeur H. BEN HADID

Département Physique

Directeur: M. Jean-Claude PLENET

UFR Sciences et Techniques des Activités Physiques et Sportives

Directeur: M. Y.VANPOULLE

Observatoire des Sciences de l'Univers de Lyon

Directeur: M. B. GUIDERDONI

Polytech Lyon

Directeur: M. P. FOURNIER

Ecole Supérieure de Chimie Physique Electronique

Directeur: M. G. PIGNAULT

Institut Universitaire de Technologie de Lyon 1

Directeur: M. le Professeur C. VITON

Ecole Supérieure du Professorat et de l'Education

Directeur: M. le Professeur A. MOUGNIOTTE

Institut de Science Financière et d'Assurances

Directeur: M. N. LEBOISNE





## RESUME

La diversité génétique intra-spécifique constitue le potentiel adaptatif des espèces et, à ce titre, elle est donc indispensable pour l'évolution de celles-ci. Chez les vertébrés, les gènes du complexe majeur d'histocompatibilité (CMH) sont une composante essentielle de quoi permet de faire face aux parasites en initiant une réponse immunitaire. La pression de sélection exercée par les parasites et la sélection sexuelle *via* le choix du partenaire devraient donc agir sur la diversité génétique du CMH. Cependant, la distinction empirique des pressions sélectives agissant sur la diversité génétique du CMH en milieu naturel nécessite de suivre un grand nombre d'individus tout au long de leur vie et d'effectuer leur génotypage. Le premier objectif de cette thèse a donc été développer et appliquer un protocole de génotypage chez la marmotte Alpine (*Marmota marmota*), sur quatre loci du CMH décrits précédemment. Ceci permet par la suite d'étudier, dans une population de marmottes Alpines vivant en milieu naturel, si les caractéristiques génétiques du CMH influencent (1) le choix de partenaire, (2) la présence et/ou l'abondance de trois espèces de parasites intestinaux et (3) leur survie juvénile. Ce travail a fourni une méthode appropriée pour la détermination de génotypes fiables sur un grand nombre d'échantillons en utilisant des techniques de séquençage de nouvelle génération. Ensuite, nous avons constaté l'existence d'un choix de partenaire basé sur le CMH mais aussi sur les caractéristiques de l'ensemble du génome. Par la suite, nous avons mis en évidence le faible rôle du CMH sur la présence et abondance de trois espèces de parasites intestinaux. Finalement, nous avons constaté que l'association entre la survie juvénile et les caractéristiques génétiques du CMH et de l'ensemble du génome ont changé au cours des vingt-trois ans de suivi de la population. Dans l'ensemble, cette thèse présente une approche intégrée de l'étude des rôles du CMH sur une population contemporaine de marmottes Alpines.

*Mots clés:* choix du partenaire, CMH, consanguinité, défense immunitaire, diversité génétique, sélection balancé, sélection induite par des parasites, sélection naturelle, sélection sexuelle, séquençage de nouvelle génération, valeur sélective

**ABSTRACT**

Intra-specific genetic diversity represents the true potential of adaptation of species and is thus essential for evolutionary change. In vertebrates, the genes of the major histocompatibility complex (MHC) play a critical role in vertebrate disease resistance by initiating immune response. The selective pressure carried out by parasites and sexual selection *via* mate choice are supposed to maintain the extreme diversity found in the MHC. Yet, empirical differentiation of selective pressures acting on MHC in the wild requires individually based monitoring of a large number of individuals and genotyping them. The aim of this thesis was firstly to develop and apply a genotyping protocol in Alpine marmots (*Marmota marmota*) to genotype four previously described MHC loci. This allows subsequently to evaluate, in a wild population of Alpine marmots, if MHC characteristics play a role (1) on mate choice, (2) on the presence and/or abundance of three intestinal parasite species and (3) on juvenile survival. This work provided a suitable method to reliably genotype large number of individuals using next-generation sequencing techniques. Then, we found evidences for female mate choice based on MHC but also on neutral genetic characteristics. Subsequently, we evidenced the weak role of MHC characteristics on the presence and abundance of three intestinal parasites. Finally, we found evidences for a change of the effect of genetic diversity at both MHC and neutral loci on juvenile survival during the 23-year monitoring study. Overall, this thesis comprises an integrated approach for the study of the roles of MHC in a contemporaneous population of Alpine marmots.

*Keywords:* balancing selection, fitness, genetic diversity, immune defence, inbreeding, MHC, natural selection, next-generation sequencing, mate choice, parasite-driven selection, sexual selection



## CONTENTS

<b>LIST OF FIGURES</b>	11
<b>LIST OF TABLES</b>	13
<b>INTRODUCTION</b>	17
1. GENETIC DIVERSITY AND EVOLUTIONARY CHANGE	19
2. THE MAJOR HISTOCOMPATIBILITY COMPLEX	21
3. MAINTENANCE OF MHC POLYMORPHISM	24
4. AIM OF THE STUDY	29
<b>MATERIAL AND METHODS</b>	31
1. BIOLOGICAL MODEL	33
2. STUDY SITE AND POPULATION MONITORING	46
<b>CHAPTER 1: LARGE-SCALE GENOTYPING</b>	51
1. INTRODUCTION	54
2. MATERIAL AND METHODS	58
3. RESULTS	65
4. DISCUSSION	73
5. APPENDIX	78
<b>CHAPTER 2: MATE CHOICE FOR GENETIC CHARACTERISTICS</b>	91
1. INTRODUCTION	94
2. MATERIAL AND METHODS	97
3. RESULTS	106
4. DISCUSSION	113
5. APPENDIX	119
<b>CHAPTER 3: MHC CHARACTERISTICS AND PARASITES</b>	147
1. INTRODUCTION	150
2. MATERIAL AND METHODS	154
3. RESULTS	163
4. DISCUSSION	168
5. APPENDIX	174

	Contents
<b>CHAPTER 4: GENETIC CHARACTERISTICS AND SURVIVAL</b>	183
1. INTRODUCTION	186
2. MATERIAL AND METHODS	191
3. RESULTS	196
4. DISCUSSION	198
5. APPENDIX	203
<b>DISCUSSION</b>	205
1. ADITIONAL THOUGHTS ABOUT MATE CHOICE	208
2. THE DIFFICULTIES OF THE STUDY OF PARASITES	211
3. MHC-CHARACTERISTICS AND FITNESS	214
4. THE STUDY OF INTRA-SPECIFIC GENETIC DIVERSITY	217
5. FROM MHC TO THE STUDY OF OVERALL IMMUNE GENES	220
6. CONCLUSIONS	223
<b>APPENDIX A. THE DEAR ENEMY PHENOMENON</b>	225
<b>APPENDIX B. GENETIC VARIATION AND ALARM CALLS</b>	237
<b>APPENDIX C. FOLLOW THE ALPINE MARMOT TRACKS</b>	241
<b>APPENDIX D. MHC GENOTYPING IN A NON-MODEL ANURAN</b>	245
<b>APPENDIX E. PRE- AND POST-COPULATORY TRAITS</b>	249
<b>REFERENCES</b>	261
<b>ACKNOWLEDGEMENTS</b>	283

## LIST OF FIGURES

Figure 1:	Contemporaneous distribution of the fifteen described species of the genus <i>Marmota</i> . Stippling ( <i>M. flaviventris</i> ) and darkest shading ( <i>M. caligata</i> ) are used to distinguish 2 species with large, overlapping ranges in western North America	33
Figure 2:	Contemporaneous distribution of the Alpine marmot. Natural populations are represented in Yellow, reintroduced populations in light Purple and introduced populations in dark Purple	35
Figure 3:	Alpine marmot from the population of La Grande Sassièrè Nature Reserve (Alps, France)	36
Figure 4:	Alpine marmot life cycle: annual (cycle) and for the entire life (blue strip)	38
Figure 5:	Four pups of the same litter few days after their emergence from the natal burrow	39
Figure 6:	Four intestinal parasites of Alpine marmot: <i>Ctenotaenia marmotae</i> egg (a), <i>Citellina alpina</i> egg (b), <i>Ascaris laevis</i> egg and <i>Eimeria</i> sp. oocyst (c), <i>C. marmotae</i> and <i>A. laevis</i> (d) and <i>C. alpina</i> (e)	46
Figure 7:	360° view of the study site at the end of the spring in 2011	47
Figure 8:	Distribution and demarcation of different family territories studied in 2013. Orange territories are followed intensively and Yellow territories are followed since 2013	48
Figure 9:	Adult male staying in front of a trap baited with dandelion	49
Figure 10:	Distribution of the reads according to the frequency of the variant $i$ within the amplicon $j$ ( $F_{ij}$ ) they belong over all amplicons with twelve or more reads for <i>Mama-UB</i> (a), <i>Mama-UD</i> (b), <i>Mama-DRB1</i> (c), and <i>Mama-DRB2</i> (d)	69
Figure 11:	Distribution of the average base quality in each base site along the read length of all reads obtained from the 454 run with the package ShortReads	89
Figure 12:	Residual of the number of EPY in a litter as a function of relatedness between social pairs (a) and as a function of MHC class II protein dissimilarity between social pairs (b)	112



Figure 13:	Spatial distribution of the MHC allelic dissimilarity estimated over all loci (MHC diss), over MHC class I loci (MHC I diss), over MHC class II loci (MHC II diss) over <i>Mama-UD</i> and <i>Mama-DRB1</i> loci (UD+DRB1 diss), over <i>Mama-UD</i> locus (UD diss) and over <i>Mama-DRB1</i> locus (DRB1 diss)	140
Figure 14:	Dispersal distance (measured in the number of territories between the natal and the dominance territory) of the males and the females that reached dominance between 1990 and 2014 in the Nature Reserve of La Grande Sassièrè	143
Figure 15:	Spatial distribution of the relatedness estimated with the relatedness ( $R_{qg}$ ) and the MHC protein dissimilarity estimated over all loci (MHC diss), over MHC class I loci (MHC I diss) and over MHC class II loci (MHC II diss)	144
Figure 16:	Residual of the abundance of <i>Eimeria monacis</i> as a function of the presence of the <i>Mama-DRB1*02</i> protein (N = 316)	164
Figure 17:	Yearly variation ( $\pm$ 95% CI) in standardized litter size of Alpine marmots ( <i>Marmota marmota</i> ) and standardized snow cover in winter (from December to March) at the Grande Sassièrè Natural Reserve (French Alps)	189
Figure 18:	Yearly changes in the proportion of groups having helpers during winter between 1990 and 2013 at the Grande Sassièrè Natural Reserve (French Alps)	190
Figure 19:	Annual estimates ( $\pm$ 95% CI, vertical segments) of juvenile (from 0 to 1 year of age) survival rate of Alpine marmots at the Grande Sassièrè Natural Reserve from 1990 to 2012, derived from a full-time dependent capture-mark-recapture model, and the corresponding temporal trends ( $\pm$ 95% CI, dashed lines)	190
Figure 20:	Residual of the juvenile survival as a function of standardized heterozygosity during the first period (1990-2001) (a) and during the second period (2002-2012) (b)	197
Figure 21:	Residual of the juvenile survival as a function of MHC class II protein diversity during the first period (1990-2001) (a) and during the second period (2002-2012) (b)	198
Figure 22:	Dominant male (with a plastic tag on the right ear indicating its social status and sex) smelling one of two glass tubes placed during an experiment	232

## LIST OF TABLES

Table 1:	The currently recognized species and conservation status of the genus <i>Marmota</i> according to the IUCN Red List of Threatened Species	34
Table 2:	Comparison of large-scale studies of amplicon-based NGS using 454 platforms and post-processing for characterizing MHC of non-model vertebrates	56
Table 3:	Summary of the individuals and amplicons used for MHC and microsatellite sequencing	62
Table 4:	Amplicon-based next generation processing steps to minimize the impact of sequencing errors among all MHC studied loci	67
Table 5:	Allelic frequencies of 4 MHC loci for the The Grande Sassi�re Nature Reserve population	71
Table 6:	Validation of 1 363 assigned genotypes after 454 post-processing by checking mother-father-offspring triads among all MHC studied loci for the Grande Sassi�re Nature Reserve	72
Table 7:	Forward (F) and Reverse (R) individual tags for the 4 locus used in one section of PicoTiterPlate of a Roche® 454 FLX sequencing instrument	79
Table 8:	Characteristics of 16 microsatellites of Alpine marmots for the Nature Reserve of The Grande Sassi�re	86
Table 9:	Spearman's correlation coefficients between standardized heterozygosity (SH) and MHC diversity estimators - MHC protein diversity and MHC protein distance - (a) and between the relatedness ( $R_{qg}$ ) and the MHC dissimilarity estimators - MHC protein dissimilarity and MHC pair protein distance - (b)	107
Table 10:	Genetic characteristics of real social males and pairs compared to the potential candidate males and pairs residing in the same territory or in the immediate vicinity of a given female	109
Table 11:	Generalized estimating equation models showing the effects of the genetic characteristics of the real social males and pairs on both the presence and the number of EPY	111
Table 12:	Genetic characteristics of social males and pairs compared to the corresponding extra-pair males and pairs	113

Table 13:	Summary of empirical studies carried out in the wild examining female mate choice using both genome-wide and MHC genetic markers	118
Table 14:	Primer pairs used for microsatellite and MHC genotyping	120
Table 15:	Characteristics of 16 microsatellites of Alpine marmots	122
Table 16:	Tags used to barcode individuals for next generation sequencing	124
Table 17:	Characteristics of 4 MHC loci of Alpine marmots	127
Table 18:	Spearman's correlation coefficients (N = 102 dominant males) between MHC diversity estimators calculated over the 2 MHC loci and over the MHC class I or the MHC class II loci (a) and between MHC diversity estimators calculated over the MHC class I locus and over the MHC class II locus (b)	127
Table 19:	Spearman's correlation coefficients (N = 146 observed pairs) between MHC dissimilarity estimators calculated over the 2 MHC loci and over the MHC class I or the MHC class II loci (a) and between MHC dissimilarity estimators calculated over the MHC class I locus and over the MHC class II locus (a)	128
Table 20:	Spearman's correlation coefficients (N = observed dominant males) between MHC protein diversity and MHC protein distance calculated over the 2 MHC loci, over the MHC class I locus and over the MHC class II locus	128
Table 21:	Spearman's correlation coefficients (N = observed dominant pairs) between MHC protein dissimilarity and MHC pair protein distance calculated over the 2 MHC loci, over the MHC class I locus and over the MHC class II locus	128
Table 22:	Spearman's correlation coefficients (N = observed dominant males) between MHC allelic and protein diversity and between MHC individual genetic distance and protein distance calculated over the 2 MHC loci, over the MHC class I locus and over the MHC class II locus	129
Table 23:	Spearman's correlation coefficients (N = observed dominant pairs) between MHC allelic and protein dissimilarity and between MHC pair allelic distances and protein distances calculated over the 2 MHC loci, over the MHC class I locus and over the MHC class II locus	129
Table 24:	Spearman's correlation coefficients (N = 102 dominant males) between MHC genetic diversity estimators calculated over the four MHC loci and over the MHC class I or the MHC class II loci (a), over the two MHC loci presenting protein polymorphism and over the <i>Mama-UD</i> or the <i>Mama-DRB1</i> locus (b), over the MHC class I loci and over the MHC class II loci (c) and over the <i>Mama-UD</i> and the <i>Mama-DRB1</i> locus (d)	131

Table 25:	Spearman's correlation coefficients (N = 146 real pairs) between MHC genetic dissimilarity estimators calculated over the four MHC loci and over the MHC class I or the MHC class II loci (a), over the two MHC loci presenting protein polymorphism and over the <i>Mama-UD</i> or the <i>Mama-DRB1</i> locus (b), over the MHC class I loci and over the MHC class II loci (c) and over the <i>Mama-UD</i> and the <i>Mama-DRB1</i> locus (d)	132
Table 26:	Spearman's correlation coefficients (N = real dominant males) between MHC allelic diversity and MHC individual genetic distance calculated over the four MHC loci, over the MHC class I loci and over the MHC class II loci (a) and the two MHC loci presenting protein polymorphism, over the <i>Mama-UD</i> locus and over the <i>Mama-DRB1</i> locus (b)	134
Table 27:	Spearman's correlation coefficients (N = real dominant pairs) between MHC allelic dissimilarity and MHC pair genetic distance calculated over the four MHC loci, over the MHC class I loci and over the MHC class II loci (a) and the two MHC loci presenting protein polymorphism, over the <i>Mama-UD</i> locus and over the <i>Mama-DRB1</i> locus (b)	134
Table 28:	Spearman's correlation coefficients between standardized heterozygosity (SH) and MHC genetic diversity estimators - MHC allelic diversity and MHC individual genetic distance - (a, b) and between the relatedness ( $R_{qg}$ ) and the MHC genetic dissimilarity estimators - MHC allelic dissimilarity and MHC pair genetic distance - (c, d)	136
Table 29:	Genetic characteristics of real social males and pairs compared to the potential candidate males and pairs residing in the same territory or in the immediate vicinity of a given female, residing at a maximal distance of five territories away or residing in the overall studied population	138
Table 30:	Generalized estimating equation models showing the effects of the MHC allelic characteristics of the real social males and pairs on both the presence and the number of EPY	141
Table 31:	Genetic characteristics of social males and pairs compared to the corresponding extra-pair males and pairs	142
Table 32:	Genetic characteristics of real social males and pairs compared to the ones of potential candidate males and pairs depending on the spatial distances between the candidate males and a given female	145
Table 33:	Generalized linear mixed models showing the effects of the MHC protein characteristics of individuals on the abundance of <i>E. monacis</i> , the presence of <i>Eimeria sp.</i> and the presence of <i>C. marmotae</i>	166

Table 34:	Generalized linear mixed models showing the effects of the presence or absence of each allele leading to polymorphic proteins of individuals on the abundance of <i>E. monacis</i> , the presence of <i>Eimeria sp.</i> and the presence of <i>C. marmotae</i>	167
Table 35:	Baseline model (a) obtained after selection of the confounding variables for the analysis of the abundance of <i>Eimeria monacis</i> and its associated parameters (b) (N = 340)	178
Table 36:	Baseline model (a) obtained after selection of the confounding variables for the analysis of the presence of <i>Eimeria sp.</i> and its associated parameters (b) (N = 331)	179
Table 37:	Baseline model (a) obtained after selection of the confounding variables for the analysis of the presence of <i>C. marmotae</i> and its associated parameters (b) (N = 225)	179
Table 38:	Generalized linear mixed models showing the effects of the MHC allelic characteristics of individuals on the abundance of <i>E. monacis</i> , the presence of <i>Eimeria sp.</i> and the presence of <i>C. marmotae</i>	181
Table 39:	Number (N) and percentage of individuals carrying the 10 polymorphic proteins	182
Table 40:	Characteristics of juvenile individuals and whether characteristics between the two studied periods	195
Table 41:	Generalized Linear Mixed Models showing the effects of the genetic characteristics on the juvenile survival on both the first and the second period	198
Table 42:	Generalized Linear Mixed Models showing the effects of the MHC allelic characteristics on the juvenile survival on both the first and the second period	204
Table 43:	Number of subfossil samples from universities or museum collections and their corresponding location	220
Table 44:	Ethogram of behaviours and actions exhibited by Alpine marmots during the dear enemy experiments	235

# **INTRODUCTION**

---



## 1 GENETIC DIVERSITY AND EVOLUTIONARY CHANGE

The International Union for the Conservation of Nature (IUCN) has recognized genetic diversity as one of the three levels of biological diversity requiring conservation (McNeely et al. 1990). The main reason for the genetic diversity conservation is that genetic variation is essential for evolutionary change (Frankel and Soulé 1981). Among both inter and intra-specific genetic diversity, intra-specific genetic diversity represents the true potential of adaptation and evolution of species. Indeed, intra-specific genetic diversity allows populations to evolve in response to (1) biotic challenges such as new diseases, changes in parasites, changes in predators or changes in competitors and to (2) abiotic challenges such as environmental changes derived from human activity (e.g. greenhouse warming, ozone layer depletion or pollution) (Frankham 1996). Low levels of genetic diversity may limit the ability of populations to continuously respond to these challenges in both the short and long term evolutionary time (Simberloff 1988, Ellstrand and Elam 1993).

Infectious diseases caused by the great variety of extant parasites are considered as one of the most important and common environmental challenge encountered by species and individuals (Haldane 1949, Weatherall 2003). Indeed, most of the living organisms are parasites and any single host species can be infected by numerous parasite species (from viruses to worms) (Windsor 1998). Under natural conditions, individuals are constantly exposed to parasites during its whole lifetime and parasites cause injuries or diseases resulting in a decrease the host's fitness (fitness is defined here as the lifetime reproductive success, which can be both survival and reproductive success (Clutton-Brock 1988), or the individual contribution to population growth (Coulson et al. 2006)) or even in the host's death. As a consequence, parasites act as one of the major demographic and evolutionary forces (Altizer et al. 2003).



Although interactions between hosts and parasites are permanent, their relationships are far from being stable since co-evolution between hosts and parasites under the form of arms race exists. Specifically, the co-evolution between hosts and parasites, also referred as Red Queen Hypothesis (Van Valen 1973) is based on the necessity for the host to avoid or resist parasites, which leads to the evolution of alleles conferring resistance to parasites (resistance is defined here following Thomas et al. (2007) by the ability of hosts to prevent parasite infection or limit parasite load), and the necessity for the parasites to encounter and infect its hosts, which leads to the evolution of alleles conferring more virulent or more resistance capacities.

Host defence against parasites depends on the host's ability to prevent, control and clear infections caused by parasites. Host defences include natural barriers (e.g. skin, mucous membranes), immune system (i.e. biological structures and processes that provide specific and unspecific protection for individuals against pathogens encountered) or changes in behaviour (e.g. spatial avoidance, temporal avoidance, changes in diet) (Schmid-Hempel 2011). Whereas the prevention against parasites is achieved by either natural barriers, immune systems or changes in behaviour, the cleaning of infections is mainly achieved by the immune system, although behavioural defences or physiological responses not directly associated with the immune system may also contribute to clean infections (Schmid-Hempel 2011). Vertebrate immune system can be classified in two subsystems, the innate and the adaptive immune systems. Whereas the innate immune system is one of the main responsible for infection prevention and infection eradication at an early stage, the adaptive immune system is one of the main responsible for specific infection control and cleaning. A necessary component of the immune system is the genetic diversity at those genes or gene families which participate in both innate and adaptive immune system (i.e. immune genes) (Sommer 2005) such as chemokine receptors, interleukin receptors, leukocyte immunoglobulin-like receptor superfamily, major histocompatibility complex, toll-like receptor or vitamin D receptor (see Acevedo-Whitehouse and Cunningham 2006 for more examples, their functions

and known associations with infectious diseases). Among all these genes, the genes of the major histocompatibility complex (MHC) account for a relatively important resistance variance to infection (Jepson et al. 1997). Moreover, decades of research focussed on MHC have led to a better understanding of its structure, functions and evolution than that of other genes involved in the immune response.

## **2. THE MAJOR HISTOCOMPATIBILITY COMPLEX**

The MHC is a multi-gene family present in all jawed vertebrates that plays a crucial role in the immune system. MHC genes are divided into three families, named class I, II and III (Kelley et al. 2005, Acevedo-Whitehouse and Cunningham 2006). Among these three families, the MHC class I and II have been largely studied because of their relevance for the adaptive immune response. In order to control infections, which necessitate the alarm of T cells from the presence of foreign invaders, MHC transcript molecules play a key role. Specifically, MHC transcript molecules are able to bind antigen peptides from different parasites and present them to T cells. Then, if a T cell identifies the antigen peptide as a non-self, it trigger a cascade of immune reaction processes (Klein 1982) to control the infection. Both MHC class I and II are cell surface trans-membrane proteins, but have different functions. Whereas MHC class I molecules are present on the surface of all nucleated cells and bind intracellular antigen peptides, MHC class II molecules are present only on immune cells (e.g. dendritic cells, macrophages or B cells) and bind extra-cellular antigen peptides (Bjorkman and Parham 1990, Hughes and Yeager 1998).

The evolutionary pattern of MHC genes agrees with a birth-and-death process, where new genes are created by gene duplication processes. Then, these duplicated genes are either maintained during large periods, become non-

functional (i.e. pseudogenes) or are deleted (Nei et al. 1997). Although the overall structure and function of the MHC genes has been conserved during the evolution of the different vertebrate lineages, important differences regarding genomic organization exist between them, especially between birds and mammals. Whereas in birds extensive copy-number variation occurs (i.e. genes with a variable number of copies in the genotype of an individual), in mammals a maximum of two alleles *per* locus occurs.

A great number of species possesses extraordinary levels of MHC allelic diversity and heterozygosity (Edwards and Hedrick 1998, Knapp 2005) and MHC genes are indeed considered as the most variable genes of the vertebrate genome. For example, thousands of alleles have been described in human populations (<http://www.ebi.ac.uk/ipd/imgt/hla/>) or more than one hundred in other mammals (primates: Rhesus monkey (*Macaca mulatta*), rodents: four-striped grass mouse (*Rhabdomys pumilio*), root vole (*Microtus oeconomus*), or artiodactyls: red deer (*Cervus elaphus*) (reviewed in Winternitz et al. 2013). Although these extraordinary levels of MHC diversity found in some species, this diversity is highly variable between species (reviewed in Kelley et al. 2005, Winternitz et al. 2013) and even some species show low levels of MHC diversity (e.g. mammals: 4 alleles in a sample of 172 European bison (*Bison bonasus*) (Radwan et al. 2007) or 10 MHC class I and 4 MHC class II alleles in a sample of 141 cheetahs (*Acinonyx jubatus*) (Castro-Prieto et al. 2011); birds: 4 alleles in a sample of 20 South Island saddlebacks (*Philesturnus carunculatus*) (Sutton et al. 2013) or 3 alleles in a sample of 30 Galapagos penguins (*Spheniscus mendiculus*) (Bollmer et al. 2007)). Low levels of MHC diversity are generally found in species who have suffered bottleneck events (reviewed in Radwan et al. 2010), although some species with reduced effective population sizes have shown high MHC diversity levels despite showing a low genetic diversity across neutral markers (Hedrick et al. 2000, Aguilar et al. 2004, Hedrick and Hurt 2012).

The majority of MHC polymorphism is located at the codons supposedly involved in antigen binding sites (ABS) (Garrigan and Hedrick 2003), which is

the basis for a differential peptide binding (Reche and Reinherz 2003). Because the recognition of non-self is done on ABS, these sites are considered one of the key components of the MHC immune function and the focus of much of the research carried out on MHC genes. It is in ABS where at least two evidences of balancing selection could be found (Garrigan and Hedrick 2003). First, contrary to the neutral evolution, the rate of non-synonymous substitutions (i.e. amino-acid altering) on ABS is generally higher than that for synonymous substitutions, apparently because of selection have favoured amino-acid substitution (Hughes and Nei 1989). Second, in genetic systems that evolve under balancing selection, trans-species polymorphism (i.e. alleles in one species share their closest ancestry to alleles from another) can often be found (Klein et al. 1998). Accordingly, trans-species polymorphism in MHC alleles has been already documented in mammals (Radwan et al. 2007, Castro-Prieto et al. 2011).

Although each functional transcript MHC allele is available to present a large number of antigens from structurally divers parasites thanks to the promiscuity of MHC peptide binding molecules (Frank 2002, Nikolich-Žugich et al. 2004), individuals with high diversity at ABS are supposed to resist a broader array of parasites because different ABS are required to bind different antigen peptides from the wide range of parasite strains and species (Hughes and Yeager 1998). Also, individuals with specific alleles are supposed to resist specific parasites due to the specificity of transcript MHC alleles. Evidences for both MHC diversity and the possession of specific alleles have been associated with parasite resistance in a variety of vertebrates (reviewed in Spurgin and Richardson 2010, Sin et al. 2014). On the contrary, low levels of MHC diversity has been suggested to increase species vulnerability to parasites in some species such as in Tasmanian devils (*Sarcophilus harrisii*) (Siddle et al. 2007, 2010) or in lowland leopard frogs (*Lithobates yavapaiensis*) (Savage and Zamudio 2011). In Tasmanian devils, the spread of a transmissible tumour between individuals by biting is apparently facilitated by very low variation at MHC class I. Indeed, the low MHC diversity of Tasmanian devils may result in an immune system failure to recognize the tumor as “foreign” (Siddle et al. 2007). However, caution is

required before considering MHC depletion as a serious concern in conservation because any study has already reported the extinction of populations or species due to a reduction of MHC diversity (Radwan et al. 2010).

Besides the importance of MHC diversity on parasite resistance, having too many MHC genes could also be disadvantageous since it could promote negative T-cell selection. Negative T-cell selection is a process that removes T-cells that bind too strongly to self peptides (Starr et al. 2003) and would direct immune responses towards self-proteins. If this process may prevent that autoimmunity may array, the result of a negative T-cell selection is the reduction of the number of different T-cell receptors (Lawlor et al. 1990, Nowak et al. 1992) and therefore the reduction of the host capacity to bind divers antigen peptides. Nowak et al. (1992) suggested that, to counteract these selective forces, an intermediate number of MHC genes would allow to confer the maximum parasite resistance. Accordingly, associations between the possession of intermediate MHC diversity levels and a high parasite resistance have already been found in three-spined sticklebacks (*Gasterosteus aculeatus*) (Wegner et al. 2003a,b) or in bank voles (*Myodes glareolus*) (Kloch et al. 2010)).

### **3. MAINTENANCE OF MHC POLYMORPHISM**

Understanding the causes and consequences of MHC genetic diversity is a major issue in evolutionary biology, ecology, epidemiology and conservation (Hedrick et al. 2000, Siddle et al. 2010, Savage and Zamudio 2011). Two main non-mutually exclusive hypotheses have been proposed for the maintenance of MHC polymorphism: balancing selection through pathogen-mediated selection and MHC-dependent mate choice (Hedrick 1992, Apanius et al. 1997, Penn and Potts 1999, Knapp 2005, Milinski 2006, Spurgin and Richardson 2010).

Three types of balancing selection through pathogen-mediated selection are

supposed to participate in the maintenance of the MHC polymorphism: the heterozygote advantage (i.e. overdominance), the negative frequency-dependent selection (or rare allele advantage) and the fluctuating selection. The heterozygote advantage hypothesis proposes that individuals with higher MHC diversity should be favoured to recognize a greater number of parasites than individuals with low MHC diversity levels (Doherty and Zingernagel 1975, Penn et al. 2002). As a consequence MHC diverse individuals would be able to trigger more effective immune responses towards specific parasites (Carrington et al. 1999, Penn et al. 2002, Worley et al. 2010) and co-infections (Hughes and Nei 1992, McClelland et al. 2003). The negative frequency-dependent selection hypothesis proposes that specific alleles (often rare alleles) in a given population would confer greater protection to parasites because parasites are likely to be adapted to overcome the most common alleles (Bodmer 1972, Potts and Wakeland 1990, Slade and McCallum 1992). Under these circumstances, the co-evolution between hosts and parasites would drive a cyclic change in the frequency of MHC alleles. Finally, the fluctuating selection hypothesis proposes that temporal and/or geographical parasite diversity and virulence could lead to an heterogeneous selection of different MHC alleles through time and space, respectively (Hill 1991, Hedrick 2002).

Evidence for heterozygote advantage, the negative frequency-dependent selection, and the fluctuating selection has been found in a wide range of non-model vertebrates (from fish to mammals) and for a wide range of parasite species (e.g. viruses, bacteria, protozoan or helminths) (see Sin et al. 2014 for a recent review). However, how to differentiate between these hypotheses remain unclear because the results of the involved mechanisms may produce similar MHC patterns (Spurgin and Richardson 2010).

Because of the key role of MHC diversity and/or the presence of specific alleles in host's parasite resistance, MHC could potentially be a good candidate for the genetic basis of female mate choice if MHC characteristics of the chosen males contribute to enhance the quality of the offspring produced (Yamazaki et al. 1976, Hamilton and Zuk 1982). Two non-exclusive hypotheses involving

female mate choice for MHC characteristics have been proposed: the good genes hypothesis and the genetic compatibility hypothesis. According to the good genes hypothesis, mate choice is based on an absolute criterion common to all females (Neff and Pitcher 2005), and females are expected to preferentially mate with (1) heterozygous males if this choice results in MHC heterozygous offspring (good-genes-as-heterozygosity hypothesis, Brown 1997) because producing MHC heterozygous offspring is advantageous since heterozygous individuals are supposed to recognize a wider range of pathogens (Doherty and Zinkernagel 1975) and/or (2) with males possessing specific alleles (good genes hypothesis *sensu stricto*) that confer resistance to specific parasites (Takahata and Nei 1990). According to the genetic compatibility hypothesis, mate choice is based on a relative criterion specific to each female (Trivers 1972, Zeh and Zeh 1996, 1997, Neff and Pitcher 2005). According to this hypothesis, females are expected to choose MHC-dissimilar males because it is more likely to result in the production of MHC heterozygous offspring (see above for the expected selective advantages of MHC heterozygous individuals). In species at risk of inbreeding, female may also prefer to mate with MHC-dissimilar males if this allows avoiding to mate with a related partner (Potts and Wakeland 1990, Penn and Potts 1999). Although a choice for MHC dissimilarity could be advantageous for female fitness, these advantages could be counterbalanced by the production of offspring with an excess of MHC genes. In this case, a choice for intermediately MHC dissimilar males could be favoured to produce offspring with an intermediate number of MHC genes.

After 40 years of research on the links between MHC characteristics and mate choice on a wide range of animals (see specific reviews in Penn and Potts 1998, 1999, Arcaro and Eklund 1998, Jordan and Bruford 1998, Penn 2002, Ziegler et al. 2005, Milinski 2006, Yamazaki and Beauchamp 2007, Spurgin and Richardson 2010), mate choice for MHC genes has been reported in several taxa, but does not seem universal. A recent meta-analysis based on 48 previously published studies found significant support regarding a MHC-dependent mating preferences at both diversity and dissimilarity when these estimators were



characterized over several loci, whereas no clear pattern was evidenced in studies focussing on a unique locus (Kamiya et al. 2014).

Although balancing selection and mate choice are considered as the main processes driving MHC polymorphism, non-selective processes such as genetic drift (Ejsmond and Radwan 2011, Sutton et al. 2011), gene conversion (Martinsohn et al. 1999) or maternal-foetal interactions (Knapp et al. 1996, Ober et al. 1998) may also drive MHC polymorphism. Genetic drift has been found to result in an accelerated decrease of functional MHC diversity compared to the decrease of neutral genetic diversity (Ejsmond and Radwan 2011, Sutton et al. 2011). In order to explain these patterns, Ejsmond and Radwan (2011) hypothesised that after a bottleneck event, the most adaptive MHC alleles may become fixed whereas rare alleles, that in large populations are likely to be retained despite many of them may be temporally disadvantageous, would be lost due to genetic drift. On the contrary, gene conversion (i.e. a non-reciprocal gene transfer during the process of meiotic recombination resulting in inheritance pattern that does not follow Mendelian rule (Galtier et al. 2001)) at both inter and intra-locus is expected to increase the number of MHC alleles (Martinsohn et al. 1999). Evidence of gene conversion at MHC has already found in pot-bellied seahorse (*Hippocampus abdominalis*) (Bahr and Wilson 2012). Finally, maternal-foetal interactions could also have an impact on the maintenance of MHC polymorphism since foetuses that do not differ at MHC alleles from maternal alleles are more likely to be aborted than foetuses differing at MHC alleles, the last ones being histocompatible foetuses. Accordingly, significantly higher foetal loss rates have been found when parents do not differ at MHC in captive pigtailed macaques (*Macaca nemestrina*, Knapp et al. 1996) or at HLA (human leukocyte antigen, the human MHC) in humans (Ober et al. 1998). This interactions could indeed have favoured MHC heterozygous individuals and ultimately the maintenance of MHC polymorphism. However, the relative contribution of these non-selective processes on MHC polymorphism is, to date, uncertain (Castro-Prieto et al. 2011).

If the evolutionary causes of the MHC polymorphic maintenance have been



broadly studied, the evolutionary consequences of this polymorphism such as its impact on individual fitness remain less studied. In the few wild populations where the associations between MHC characteristics and survival or reproductive success were evaluated, MHC characteristics significantly affected individual fitness (Bernatchez and Landry 2003). For example, a study carried out on rhesus macaques (*Macaca mulatta*) found that MHC heterozygous males sired significantly more offspring than homozygous males (Sauerman et al. 2001). Also, a ten years of study on Seychelles warbler (*Acrocephalus sechellensis*) revealed a positive association between MHC diversity and juvenile survival (Brouwer et al. 2010). Similarly, a study carried out on great tits (*Parus major*) found an association between MHC super-types (i.e. grouped MHC transcript molecules according to peptide-binding specificity) and three components of individual fitness (survival, annual recruitment and lifetime reproductive success) (Sepil et al. 2013).

Empirical differentiation of selective pressures acting on MHC has been difficult and it has been suggested that the involved mechanisms could overlap (Bernatchez and Landry 2003, Sommer 2005, Spurgin and Richardson 2010). However, both a recent meta-analysis and a model of MHC dynamics have found evidence that sexual selection may be more important than natural selection for the maintenance of MHC polymorphism (Winternitz et al. 2013, Ejsmond et al. 2014). Long-term datasets collected on free-living individuals at various temporal and spatial scales are particularly useful to elucidate the role of MHC in parasite load, mate choice or fitness (Spurgin and Richardson 2010, Ejsmond et al. 2014). Moreover, it would be more informative to use a multiloci approach, with either including both MHC class I and class II loci or many MHC class I and class II loci (if multiple loci in a given MHC class exists for the studied species) to assess these eventual associations (Ejsmond et al. 2014, Kamiya et al. 2014, Sin et al. 2014). While collecting such large and detailed dataset from wild populations is in itself challenging, determining accurate MHC genotypes for such high number of individuals is no less a challenge.

Large-scale sequencing of MHC has long been restricted due to elevated

costs and methodological difficulties to obtain accurate genotypes. Nevertheless, amplicon-based next-generation sequencing (NGS) platforms such as 454, Illumina, SOLiD or Ion Torrent have allowed to obtain large number of DNA sequences at relatively affordable costs (see Glenn 2011, Shokralla et al. 2012). The major difficulty of the use of these techniques is that they are prone to errors such as artifacts (substitutions and chimeras) during PCR or indels during the sequencing process (Glenn 2011). Discerning true alleles *versus* sequencing errors is thus challenging, especially for highly polymorphic genes such as MHC, where alleles often differ by a single nucleotide (Babik 2010). Although a great number of challenges to obtain reliable genotypes by using efficient post-processing procedures have been overcome during the last years (Babik et al. 2009, Galan et al. 2010, Zagalska-Neubauer et al. 2010, Huchard et al. 2012, Radwan et al. 2012, Sepil et al. 2012, Sommer et al. 2013, Lighten et al. 2014a,b, Stutz and Bolnick 2014), a universal procedure is still lacking (see Huchard and Pechouskova 2014 for more details on MHC genotyping challenges).

#### **4. AIM OF THE STUDY**

Taking advantage of a long-term individually based monitoring of a wild population of Alpine marmots (*Marmota marmota*), the aim of this study was first to reliably genotype individuals at four previously described polymorphic MHC loci belonging to MHC class I (*Mama-UB* and *Mama-UD*) and class II (*Mama-DRB1* and *Mama-DRB2*) (Kuduk et al. 2012). Then, I investigated the role of MHC diversity in mate choice, intestinal parasites and juvenile survival in order to better understand the role of MHC in natural populations.

The Alpine marmot, a territorial, socially monogamous and cooperative breeding species, is a well suited biological model to study these questions for at least three reasons. First, an exceptional large and detailed dataset spanning 25

years is available (see <http://thealpinemarmotproject.org> for detailed information on the scientific project), with data collection including trapping and handling (i.e. capture-mark-recapture) of more than 1500 individuals belonging to more than 30 family groups. Second, although females are constrained in their choice of social pairs due to the competitive process for dominance acquirement, the existence of extra-pair paternities (Cohas et al. 2008) may be a way to adjust their mate choice and to copulate with males possessing peculiar MHC characteristics. Third, the intestinal parasites of monitored individuals are well characterised (Callait and Gauthier 2000), which enables the study of eventual associations between MHC characteristics and the presence or abundance of intestinal parasites.

This thesis is organised around four chapters coupled with a material and methods and a discussion section. Each chapter has been either accepted for publication or is in preparation for future publication. The first chapter has also been presented in a conference (2ème colloque de Génomique Environnementale, Rennes, France). In this chapter, I developed a suitable method to reliable genotype any kind of genes without copy-number variation (i.e. genes with a variable number of copies in the genotype of an individual) using any amplicon-based NGS platform and I applied it to obtain highly reliable individual genotypes in Alpine marmots at the four MHC polymorphic loci. In the second chapter, I investigated whether females base their social and extra-pair mate choice on neutral and/or on key functional loci using sixteen microsatellites and four MHC loci belonging to both MHC class I and II. In the third chapter, I assessed whether the individual MHC characteristics - MHC diversity and the presence of MHC specific alleles – were associated with the presence and/or the abundance of three intestinal parasites. In the fourth chapter, I analysed whether the juvenile survival depended on neutral (using sixteen microsatellites) and/or on key functional loci (using MHC loci belonging to both MHC class I and II) along the studied period.

# **MATERIAL AND METHODS**

---



## 1. BIOLOGICAL MODEL

### *The genus Marmota*

Marmots are sciurid rodents, belonging to the tribe Marmotini and the genus *Marmota*. The genus *Marmota* has a recent origin. The earliest marmot fossil is *M. minor*, dates from 10.3 million years ago and lived in North America (Steppan et al. 2011). Nowadays, fifteen marmots are currently recognized: six in North America and nine in Eurasia (Figure 1, Table 1).



**Figure 1.** Contemporaneous distribution of the fifteen described species of the genus *Marmota*. Stippling (*M. flaviventris*) and darkest shading (*M. caligata*) are used to distinguish 2 species with large, overlapping ranges in western North America (from Steppan et al. 2011).

From these fifteen marmot species, eleven are considered as least concern, Menzbier's marmot as vulnerable, Mongolian marmot as decreasing, Vancouver Island marmot as critically endangered and the remaining one, Forest-steppe

marmot, has not yet been evaluated (Table 1).

Scientific Name	Common Name	Red List Category	Geographic distribution
<i>M. broweri</i>	Alaska marmot	Least Concern	North America
<i>M. caligata</i>	Hoary marmot	Least Concern	North America
<i>M. flaviventris</i>	Yellow-bellied marmot	Least Concern	North America
<i>M. monax</i>	Woodchuck	Least Concern	North America
<i>M. olympus</i>	Olympic marmot	Least Concern	North America
<i>M. vancouverensis</i>	Vancouver Island marmot	Critically Endangered	North America
<i>M. baibacina</i>	Altai marmot	Least Concern	Eurasia
<i>M. bobak</i>	Bobak marmot	Least Concern	Eurasia
<i>M. camtschatica</i>	Black-capped marmot	Least Concern	Eurasia
<i>M. caudata</i>	Long-tailed marmot	Least Concern	Eurasia
<i>M. himalayana</i>	Himalayan marmot	Least Concern	Eurasia
<i>M. kastschenkoi</i>	Forest-stepe marmot	Not evaluated	Eurasia
<i>M. marmota</i>	Alpine marmot	Least Concern	Eurasia
<i>M. menzbieri</i>	Menzbier's marmot	Vulnerable	Eurasia
<i>M. sibirica</i>	Mongolian marmot	Decreasing	Eurasia

**Table 1.** The currently recognized species and conservation status of the genus *Marmota* according to the IUCN Red List of Threatened Species.

## *The Alpine marmot*

### *Generalities*

Alpine marmots (*M. marmota*) were widespread in Europe until the early Holocene (around 11,700 years BP) (Couturier 1955, Besson 1971). However, several drastic reduction of their population size, possibly due to environmental variations (Preleuthner and Pinsker 1993, Rassmann et al. 1994), occurred during the early Holocene and reduced the distribution of this species to the Alpine arc and the Carpathian mountains, its actual natural distribution (Figure 2). Nowadays, two subspecies of Alpine marmots are currently recognized: *M. marmota marmota*, who live in the Alpine and Pyrenean arc, and *M. marmota latirostris*, who live in the Carpathian mountains.



**Figure 2.** Contemporaneous distribution of the Alpine marmot. Natural populations are represented in Yellow, reintroduced populations in light Purple and introduced populations in dark Purple. Image from <http://www.iucnredlist.org>

The meta-population of Alpine marmots situated in the Alpine arc (Figure 2) is the biggest for this species and was estimated to 100,000 individuals in 1994 (Le Berre 1994).

Between 1948 and 1988, around 400 Alpine marmots were introduced into the French Pyrenees from individuals of the subspecies *M. marmota marmota* captured in the French Alps (López et al. 2010). The introduced population grew rapidly and nowadays occupy almost all the Pyrenean arc (Figure 2). In 2000, the estimated population in the Pyrenees was around 10,000 individuals (Herrero et al. 2000), although in 2010 López et al. (2010) suggested that this population was probably higher because it was still occupying new areas. The subspecies *M. marmota marmota* was also introduced in France (Massif Central, Jura and Vosges), in Germany (Black Forest) and Italy (Appennines) (Figure 2). In the Appennines, in addition to the introduction of the subspecies *M. marmota marmota*, the subspecies *M. marmota latirostris* was also introduced.

The subspecies *M. marmota latirostris* has a restricted range and a small population, with a population estimated to 1,500 individuals. According to the IUCN Red List of Threatened Species, this subspecies should be monitored and protected (<http://www.iucnredlist.org/details/12835/0>). This subspecies was reintroduced in Romania in 1973 (Figure 2).

At the European level, Alpine marmot is a protected species (Annex III Bern Convention; Annex III Convention on the Conservation of EY.R.O. Wildlife and



Natural Habitats, 82/72/CEE). In France it is considered as non domestic species, belonging to the national biological heritage. The French rural Code forbid hawking, sale, selling or buying dead specimens (R. 211.1 to R. 211.3), digging up, trapping (R. 227.5), and transport of live specimens (L. 224.8). The capture and the transport of marmots for scientific or re-population purposes (R. 224.14) as well as reintroductions (L. 211.3) are regulated.

The Alpine marmot's habitat is usually Alpine and sub-Alpine grasslands between 1000 and 3000 m asl. in altitude (López et al. 2010). Slope affects where marmots establish their territory since when they have a choice they prefer low or medium slopes (Allainé et al. 1994, López et al. 2010).

Alpine marmots are herbivorous and their diet is very diversified. They eat different plant species according to their phenological development and the nutritional requirements of the individuals (Rudatis and De Battisti 2005). The main predators of Alpine marmots are the Golden eagles (*Aquila chrysaetos*) and the Red foxes (*Vulpes vulpes*).

In this species there is a lack of sexual dimorphism. Alpine marmots possess a brown coat, with orange markings on the back and with beige to orange on the belly. They have a brown muzzle, with a white band, between the nose and the eyes, and small ears (Figure 3). The extremities of Alpine marmots are short. Finally, its tail is brown, bushy and black in the extreme.



**Figure 3.** Alpine marmot from the population of La Grande Sassièrè Nature Reserve (Alps, France). Photography: Marie-Léa Travert.

An adult Alpine marmot reaches between 45 and 68 cm in length (not including the tail, which measures between 13 to 16 cm). Their body mass significantly changes throughout the year. The minimal adult weight (2.2 Kg) is achieved in mid-April and coincides with the end of hibernation, whereas the maximal adult weight (6.5 Kg) is achieved in mid-October and coincides with the entry into hibernation (Körtner and Heldmaier 1995).

### *Life cycle*

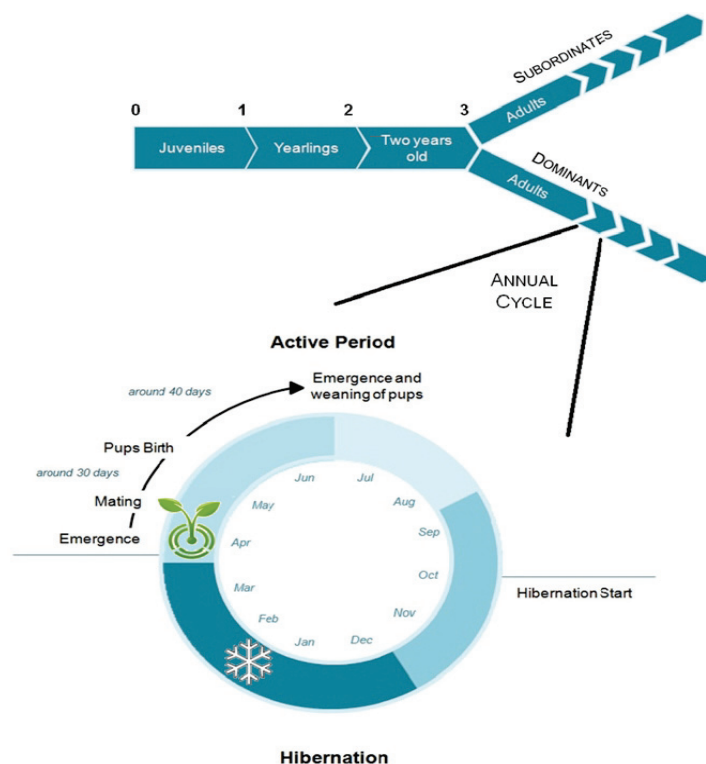
Alpine marmots reach sexual maturity at two-years old. From this moment, they can reach dominance or stay as subordinates in their natal territory (Figure 4).

Alpine marmot average lifespan is around 7 years, but the oldest individual observed to date in the studied population was a female aged of 16 years. Although being a monomorphic species, senescence differs between sexes. Whereas males show evidence of both chronological senescence (i.e. a decrease in body mass scaled to absolute age) and terminal decline (i.e. a decrease in body mass scaled to individual age at death), females do not show any detectable senescence in body mass (Tafani et al. 2013a).

Regarding the reproductive performance with age, females showed a senescence of litter size (Berger et al. 2015). Despite the decrease of litter size found in aged females, females showed an increase of offspring size with age (Berger et al. 2015). As a consequence, female reproductive performance remains constant throughout the female's lifetime (Berger et al. 2015).

Hibernation lasts for approximately 200 days (Figure 4) but may vary depending on the altitude. All members of the family group hibernate together in the same chamber of the burrow previously covered by hay and blocked with soil, gravels and faeces (Körtner and Heldmaier 1995). During hibernation, Alpine marmots undergo a substantial reduction of metabolism in order to preserve their vital functions and increase survival probabilities during the winter. They alter periods of torpor (characterised by a decreased physiological activity and body temperature) and euthermia (characterised by wakefulness and an increase of body temperature) (Arnold 1988). Although euthermia periods

represent less than 10% of the hibernation length, they cause 85% of the energetic expenditure during the hibernation period (Arnold 1988). During euthermia periods, individuals dissipate heat, which increase the burrow temperature. Social conditions during hibernation appear to play a crucial role since male subordinates (named helpers) are supposed to undergo higher and/or longer euthermia periods than other individuals, bringing more heat to the pups and therefore contributing to the increase of pups' survival (Arnold 1990, Allainé and Theuriau 2004).



**Figure 4.** Alpine marmot life cycle: annual (cycle) and for the entire life (blue strip). Image from Tafani et al. 2013b.

Individuals spend the favourable season (April to September) accumulating enough fat reserves to survive hibernation and initiate reproduction the following year (Körtner and Heldmaier 1995).

Reproductive individuals gave birth to a litter from 1 to 7 pups (median = 4) once a year. Mating occurs shortly after the end of hibernation, from mid-April to early May, and gestation lasts for 30 days (Figure 4). Once born, pups stay in the natal burrow for 40 days, where mothers lactate them (Figure 4). Once weaned,

pups emerge from the burrow between late June and mid-July, and start eating and discovering the territory (Figure 4, 5). A bias in sex-ratio at the emergence from the burrow toward males is observed in the studied population (Allainé 2004).



**Figure 5.** Four pups of the same litter few days after their emergence from the natal burrow. Photography: Carole et Denis Favre-Bonvin.

Snow depth in winter is the main environmental factor shaping litter size of Alpine marmots and the general thinning of winter snow cover that has been repeatedly reported to occur in the Alps during the last decades, have lead to a decrease in litter size (Tafani et al. 2013b).

### *Socio-spatial organization*

Alpine marmots live in family groups from 2 to 20 individuals (Allainé 2000). Family groups are composed of a dominant couple; sexually mature (2 years and older) and immature (younger than 2 years old) subordinates of both sexes; and pups born that year (Allainé et al. 2000). Individuals of both sexes may stay dominant for several years until being evicted by another individual or dying (Lardy et al. 2011).

Interactions between individuals are frequent (Perrin et al. 1993a,b). Friendly interactions can be observed between all individuals of the same social group,

although friendly interactions with the dominant male are less frequent.

Alpine marmots of the same family group share a common multi-purpose territory (range between 0.9 ha and 2.8 ha) comprising a main and secondary burrows (Perrin et al. 1993b). In their territory, individuals (1) draw their food source; (2) use the main burrow to hibernate, to sleep during the night and to give birth (Perrin et al. 1993a); and (3) use their secondary burrows to escape from eventual predators.

Although subordinate individuals are involved in territorial surveillance, they rarely defend their territory, activity relegated to the dominant couple (Arnold and Dittami 1997).

### *Dominance*

Mature subordinates (2 years and older) of both sexes can reach dominance by replacing his father or his mother. Otherwise, mature subordinates of both sexes start dispersing to reach dominance by replacing a dominant from another territory or by creating a new territory, a more unusual phenomenon (only five cases observed between 1990 and 2014 in the studied population). Female subordinates reach dominance at  $3.21 \pm 0.11$  years old (Lardy 2012) and male subordinates reach dominance at  $3.50 \pm 0.11$  years old (Lardy et al. 2012). Between 1990 and 2014, 20.2% (131 over 649) of the females and 18.6% (147 over 791) of the males born in our population became dominants in our study site. The probability to acquire dominance at 2 years old (at sexual maturity) is low in both sexes (the probability of becoming dominant the following year was  $0.04 \pm 0.01$  for yearlings, is  $0.30 \pm 0.04$  for 2 years old and  $0.45 \pm 0.06$  for adult subordinates, Ferrandiz-Rovira unpublished data), suggesting that there is a common delay of one or more years in the dispersal.

The main dispersal period is just after hibernation and early summer, where aggressive encounters between dispersal and dominant individuals from the same sex are common. The consequences of aggressive encounters are injuries inflicted by other individuals, which may cause severe infections as well as the eventual loss of a territory for the dominants. Landless individuals are highly

subjected to fitness costs since dispersing individuals which fail to reach dominance within the dispersing year have high mortality rates (Stephens et al. 2002). Dominant individuals which lose their territory have also high mortality rates (Lardy et al. 2011).

### *Mating system*

The Alpine marmot is a socially monogamous species. The dominant couple mates for one to several consecutive reproductive years (mean: 2.47, range: 1-8 between 1990 and 2014), and stays paired throughout the year.

The dominant pair monopolizes reproduction by inhibiting reproduction of same-sex subordinates through aggressive behaviour (Arnold and Dittami 1997, Hackländer et al. 2003), resulting in a high level of corticosteroids that limit testes maturation and spermatogenesis in subordinate males (Arnold and Dittami 1997); and resulting in the failure of embryo implantation and development in subordinate females (Hackländer et al. 2003).

If dominant females generally successfully monopolize all reproduction, extra-pair paternity (EPP) occurs (Goossens et al. 1998, Cohas et al. 2008). EPP was found to increase with the number of sexually mature subordinate males present in the family group (Cohas et al. 2006, 2008, Lardy et al. 2012). Moreover, the presence and the number of extra-pair young (EPY) was found to increase when females are paired with too similar or too genetically dissimilar males (measured at sixteen *a priori* neutral microsatellite loci) with respect to them (Cohas et al. 2006, 2008). Lardy et al. (2012) suggested that, in Alpine marmot, the control of many subordinates is energetically costly for dominant males, leading them to lose the control of reproduction (i.e. occurrence of EPP) and/or the dominance tenure. It has been suggested that obtaining genetic benefits is the main evolutionary force driving extra-pair paternity in Alpine marmots (Cohas et al. 2007). Indeed, the occurrence of EPP has some consequences in terms of fitness of the produced offspring. Specifically, extra-pair young were found to have a better survival rate as well as a higher probability of dominance access than within-pair young (Cohas et al. 2007).



### *Communication*

Communication is the transfer of information from one animal to another, either from conspecifics or from other species. Alpine marmots use two main types of communication, the chemical communication (scent) and the vocal communication.

#### Chemical communication

Alpine marmots, as all marmot species, have anal glands and two facial glands that may be used to communicate: jugal glands (also named as orbital or cheek glands) are located bilaterally between the eye and the ear, and buccal glands are located at both oral angles (Rausch and Bridgens 1989, Blumstein and Henderson 1996).

Cheek rubbing by jugal glands is the main territory marking mode in Alpine marmots (Bel et al. 1995, 1999). Jugal gland is an exocrine gland located in the dermis and its secretive activity is age-dependent, with no secretion observed in juvenile marmots (Bel et al. 1999). Previous studies on Alpine marmots showed that (1) scent-marking mainly occurs between the breeding period and the emergence of offspring (Bel et al. 1995, Lenti Boero 1995); (2) scent-marking is spatially distributed and mainly involves the territory boundaries and the main burrows (Bel et al. 1995, Lenti Boero 1995); (3) although all mature individuals of a family group scent-mark, this behaviour is mainly done by dominant individuals (Bel et al. 1995); and (4) foreign scent marks are removed or replaced (Bel et al. 1995). Given that scent-marking behaviour with jugal glands is the main territory marking mode and that foreign individuals represent an important threat for dominants (see Appendix A for detailed information on this phenomenon), it is likely that the dear enemy phenomenon occurs in this species. If the dear enemy phenomenon occurs in Alpine marmots, jugal glands are supposed to play a key role on this behaviour.

When same-sex adults are in conflict both individuals moves their tail from up to down, suggesting that both individuals may expose their anal glands to the other individuals. A similar behaviour has been observed in yellow-bellied

marmots, behaviour named as tail-flag (Armitage 2014).

### Vocal communication

When adult individuals discover a predator, they often emit loud vocalizations (i.e. alarm calls). These alarm calls may create confusion and may increase the caller chances to escape as well as the conspecifics chances to escape, because it also function to warn conspecifics about the presence of a predator (Shelley and Blumstein 2005). Individual variations related to the propensity to alarm are not negligible, variations that may be linked to more or less cooperative individuals (see Appendix B for more details).

### *Genetic characteristics*

Alpine marmots, as other 7 marmot species, have 38 chromosomes, whereas the number of chromosomes in the remaining marmot species range between 36 and 42 (Armitage 2014).

In Alpine marmots, low levels of genetic variability at both at neutral and functional markers have been found using allozymes (Preleuthner and Pinsker 1993), minisatellites (Kruckenhauser et al. 1997, Rassmann et al. 1994), microsatellites (Cohas et al. 2009, but see Goossens et al. 2001) and MHC (Kuduk et al. 2012, Chapter 1), and this in different populations spreading the whole Alpine arc. Demographic processes, and especially bottlenecks, have been hypothesized to have originated a huge decrease of genetic variability in this species (see discussion section and Appendix C for more details).

### Neutral markers

Sixteen microsatellites have yet been characterized for the Alpine marmot: SS-Bibl1, SS-Bibl18, SS-Bibl20, SS-Bibl31, SS-Bibl4 (Klinkicht 1993); MS41, MS45, MS47, MS53, MS56, MS6, ST10 (Hanslik and Kruckenhauser 2000); Ma002, Ma018, Ma066, Ma091 (Da Silva et al. 2003). Among these sixteen microsatellites, allelic polymorphism ranged between two and nine alleles.

Since microsatellites are non-coding regions of the genome, they are



supposed to reflect neutral evolutionary process (Jarne and Lagoda 1996). Thus microsatellites have been used to infer genome-wide heterozygosity of individuals (Slate and Pemberton 2002, Coltman and Slate 2003; but see Väli et al. 2008, Ljungqvist et al. 2010). However, caution is required to consider a small number of microsatellites as a good proxy of genome-wide heterozygosity since a recent study found that genome-wide heterozygosity is better measured from around 75 microsatellites (Miller et al. 2014).

In Alpine marmots, individual heterozygosity have been found to impact on individual fitness. Specifically, a positive correlation between heterozygosity (calculated over the sixteen microsatellites) and juvenile survival was observed (Cohas et al. 2009). The potential mechanism behind this correlation seems to be explained by the general effect (Cohas et al. 2009).

#### Functional markers

Six MHC loci have yet been characterized for the Alpine marmot: four MHC class I loci (*Mama-UA*, *Mama-UB*, *Mama-UC* and *Mama-UD*) and two MHC class II DRB loci (*Mama-DRB1* and *Mama-DRB2*) (Kuduk et al. 2012). Among these loci, two MHC class I loci (*Mama-UB* and *Mama-UD*) and two MHC class II loci (*Mama-DRB1*, and *Mama-DRB2*) were polymorphic (2 alleles for *Mama-UB*, 3 for *Mama-UD*, 8 for *Mama-DRB1*, and 3 for *Mama-DRB2*), whereas polymorphism was not detected at *Mama-UA* and *Mama-UC* (Kuduk et al. 2012). Although four loci had allelic polymorphism, only one MHC class I (*Mama-UD*) and one MHC class II loci (*Mama-DRB1*) led to polymorphic proteins (2 and 7 proteins) (Kuduk et al. 2012).

In Alpine marmots, the two MHC classes showed different patterns of evolution. Whereas no signal indicating positive selection was found (i.e. the non-synonymous and synonymous substitution ratio ( $d_N/d_S$ ) was less than one for both ABS and non-ABS sites) in MHC class I alleles, clear evidence of historical positive selection for amino acid replacements (i.e.  $d_N$  significantly exceeded  $d_S$  in putative ABS sites but not in non-ABS sites) was found in MHC class II alleles (Kuduk et al. 2012). These divergent patterns of allelic evolution may

reflect differences in selection pressures acting on class I and class II genes.

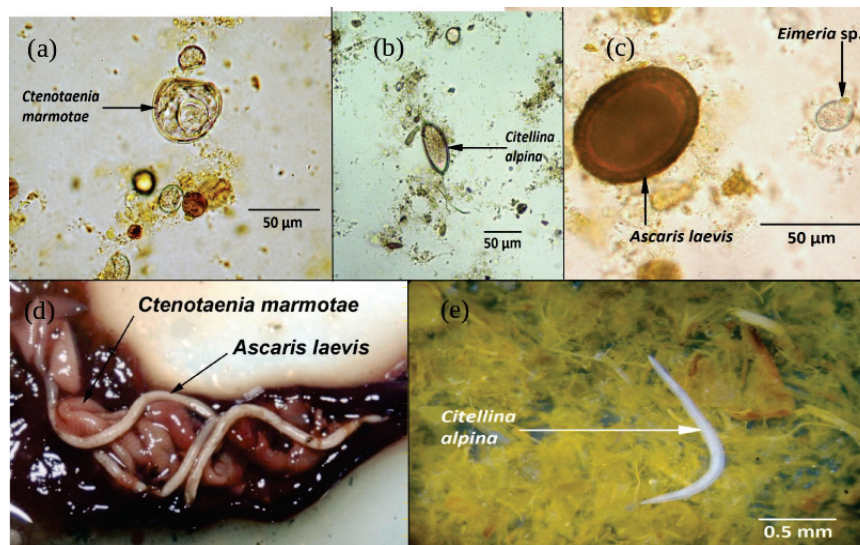
### *Alpine marmots as hosts*

The most frequently pathologies occurring in the genus *Marmota* are derived from ectoparasites and parasites located in the digestive tract (Bassano 1996). The bacterial and viral diseases are rarely described (Bassano 1996).

In a study carried out in the National Park of Berchtesgaden (Germany), a single important ectoparasite was found, the mite *Echinonyssus blanchar* (Arnold and Lichtenstein 1993). *E. blanchar* is an obligate hematophage present in all Eurasian marmot species that lives in its burrows (Arnold and Lichtenstein 1993). In this study, Arnold and Lichtenstein (1993) found that (1) the number of mites negatively correlated with body mass, (2) the number of mites increased juvenile mortality, and (3) females with a high number of mites during the preceding winter delayed the date of juvenile weaning.

Alpine marmots living in the Alpine arc are regularly parasitized by four intestinal parasites (Bassano et al. 1992, Callait and Gauthier 2000). Three out of four parasites are Helminth species (extra-cellular parasites): the Cestode species *Ctenotaenia marmotae* (Figure 6a,6d), and the two Nematoda species, *Citellina alpina* (Figure 6b,6e) and *Ascaris laevis* (Figure 6c,6d). The fourth parasite is a protozoan species (intra-cellular parasite), the Coccidia species *Eimeria monacis* (Figure 6c). Although Alpine marmots have a low specific richness, the intestinal parasite load is high. During hibernation their intestines are atrophied (Bassano et al. 1992) and this may be the cause of the large observed decrease of parasite load during this period (Callait and Gauthier 2000). Besides to consider hibernation as a strategy enabling marmots to survive during the harsh winter conditions, hibernation can also enable rid of a large number of parasites. However, each year, the parasite population dynamics allow an infestation of the most hosts of the population (Callait 1999). The consequences of the presence or abundance of intestinal parasites are difficult to identify because they do not cause any apparent pathology (Callait 1999). However, whether intestinal parasites affect or not Alpine marmots survival and/or reproductive success

remains unknown.



**Figure 6.** Four intestinal parasites of Alpine marmot: *Ctenotaenia marmotae* egg (a), *Citellina alpina* egg (b), *Ascaris laevis* egg and *Eimeria* sp. oocyst (c), *C. marmotae* and *A. laevis* (d) and *C. alpina* (e). Photography: Marie-Pierre Callait-Cardinal.

## 2. STUDY SITE AND POPULATION MONITORING

### *The Nature Reserve of the Grande Sassièr*

In 1990, the population of Alpine marmots living in the nature reserve of the Grande Sassièr (French Alps) became the object of study for a large number of researchers and students (Figure 7). Under the name of the Alpine marmot project (see <http://thealpinemarmotproject.org> for detailed information of the scientific project), a capture mark and recapture program coupled with an observation protocol were developed. Annually, there is a field season (lasts between forty to sixty days depending on the year) carried out during the active period of Alpine marmots. The field season continuously involves between four and eight people (researchers, post-docs, PhD students or bachelor students).

The nature reserve of the Grande Sassièr is managed by the National Park of

Vanoise and it is located in the Tarentaise Valley (Tignes, France). It has an altitude ranging from 1850 to 3757 m asl. and a typical Alpine climate, with low temperatures, strong daily and annual variation in temperature and marked precipitations. A large body of evidence shows us the general effects of climate change in the study site. Specifically, the average snow depth from December to March decreased by  $1.87 \pm 0.68$  (mean  $\pm$  SE) cm per year from 1991 to 2011; while the average air temperature in April increased by  $0.16 \pm 0.04$  °C per year from 1990 to 2011 (Tafari et al. 2013b).



**Figure 7.** 360° view of the study site at the end of the spring in 2011. Photography: Maxime Gay.

The flora is characteristic of Alpine and sub-Alpine grassland and forms a dense carpet of vegetation (Gensac and Rothé 1974). It is composed of a multitude of grass types and hosts a profusion of flowers when summertime arrives.

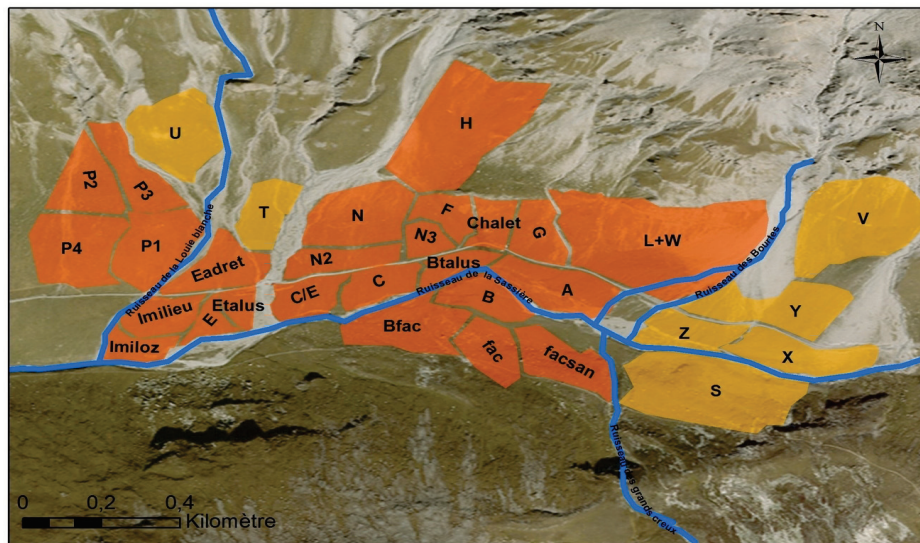
The fauna includes many emblematic species coexisting with Alpine marmots such as the chamois (*Rupicapra rupicapra*), the ibex (*ibex ibex*), the bearded vulture (*Gypaetus barbatus*), the Golden eagle and the Red fox.

### *Marmot territories*

Studied family marmots are situated in the low part of the reserve, at 2350 m asl (45°29'N, 65°90'E). All territories are roughly at the same altitude, in an area of approximately 1.5 Km long and 500 Km wide, are distributed of each side of a tourist trail that traverses the valley floor, and are exposed to North or to South (Figure 8). In North facing slopes, snows lasts longer and plants growth is delayed, thus limiting the food resources during the early active period of

marmots. In Yellow-bellied marmots, this limitation of food affected marmots' life history traits (Van Vuren and Armitage 1991).

Cartography of marmot territories is defined and actualized each year (Figure 8). During all these twenty-five years, a total of forty-one marmot families have been studied for more than one year, although twenty-five of them have been followed intensively (Figure 8).



**Figure 8.** Distribution and demarcation of different family territories studied in 2013. Orange territories are followed intensively and Yellow territories are followed since 2013.

### *Marmot trapping*

During the pups emergence period (from the 15<sup>th</sup> of June to the 15<sup>th</sup> July), each territory is carefully observed to identify the emergence of the pups. Then juveniles are counted and trapped either by small two-door live traps or by hand. Almost all of the pups are captured within the three days following their first emergence from the burrows.

Over 80% of the marmots from one year old are captured at least once every year, between mid-April to mid-July, using two-door live traps baited with dandelions (*Taraxacum densleonis*). Between one and four traps are placed in front of the entrance of the main burrows of each territory (Figure 9), which allow us to assign with certainty each trapped individual to its territory. To identify and evaluate the presence and the abundance of intestinal parasite



species, a faecal sample is collected (if available) from the trap immediately after the capture.



**Figure 9.** Adult male staying in front of a trap baited with dandelion.

### *Marmot handling*

Once captured, individuals are tranquillized with Zolétil 100 ( $0.1 \text{ mL Kg}^{-1}$ ) by intramuscular injection. Once anaesthetized, individuals are sexed, aged from their size (up to 3 years) and weighed. Morphological characteristics (presence of visible testis for males, developed teats for females, development of scent glands for both sexes) are used to confirm dominance status determined from observation (see behavioural observations section below). Captured marmots are individually marked with a transponder chip (model ID100, 0.9 cm long,  $<0.1 \text{ cm}$  in diameter, Trovan Ltd, [www.Trovan.com](http://www.Trovan.com), Identifikationssysteme, Metternicher Straße 4, 53919 Weilerswist, Germany) injected under the skin of the neck for permanent individual recognition, and with one metal tags ( $1 \text{ cm} \times 3 \text{ mm}$ ) at their ear for visual identification of sex (right side for females and left side for males). Additionally, dominants are also tagged with a coloured plastic tag ( $<1 \text{ cm}^2$ , right side for males and left side for females).

Different morphological measurements are done: jaw length, fore-foot length, ulna length, hind-foot length, tibia length, body length without tail, zygomatic width and pelvis width.

For genetic analyses, skin biopsies and hair samples are collected. For skin biopsies, a piece of skin ( $<1 \text{ mm}^3$ ) is removed with a biopsy punch (Alcyon, Lyon, France) without causing any bleeding. For chemical analysis, scent

samples are collected from jugal, buccal and anal glands when secretion is available. Finally, blood samples are also collected.

Handling lasts a maximum of 30 minutes and individuals are absent from their territory for a maximum of 90 minutes (time elapsed between the capture and the release, once the animal has completely awakened from the anaesthesia). Any individual has observed to be excluded from its territory following capture.

### *Behavioural observations*

During the field season each family group is observed to determine group composition. For each group, the number of adults, two years old, yearlings and pups of each sex are counted thanks to their ear marks and the use of 10x50 binoculars and 20 x 60 telescopes. Moreover, scent-marking behaviour and aggressive interactions allow us to categorize individuals as subordinates or dominants (Bel et al. 1999). To avoid disturbing marmots during observation, the distance between marmots and the observer varies between 80 and 200 m and observations are always done outside from the focal marmot territory. Observation sessions last one hour and are randomly distributed during the period of activity from 8:00 a.m. to 12:00 p.m. and from 15:00 p.m. to 21:00 p.m. During the field season, each family is observed a minimum of 30 hours.

Two additional behavioural observations have been carried out between 2012 and 2014 field seasons in order to (1) test the presence of the dear enemy phenomenon (see Appendix A for more details on this project) and (2) to study both the individual variations related to the propensity to alarm (see Appendix B for more details on this project) and the population variation of alarm calls. Both projects were conceived and designed as adjacent works of this thesis.

# CHAPTER 1: LARGE-SCALE GENOTYPING

---

**Large-scale genotyping of highly polymorphic loci by next generation sequencing: how to overcome the challenges to reliably genotype individuals?**

M Ferrandiz-Rovira, T Bigot, D Allainé, M-P Callait-Cardinal and A Cohas  
*Heredity (2015), 114: 485-493*





## Abstract

Studying the different roles of adaptive genes is still a challenge in evolutionary ecology and requires reliable genotyping of large numbers of individuals. Next generation sequencing (NGS) techniques enable such large-scale sequencing, but stringent data processing is required. Here, we develop an easy to use methodology to process amplicon-based NGS data and we apply this methodology to reliably genotype four MHC loci belonging to MHC class I and II of Alpine marmots (*Marmota marmota*). Our post-processing methodology allowed us to increase the number of retained reads. The quality of genotype assignment was further assessed using three independent validation procedures. A total of 3 069 high quality MHC genotypes were obtained at four MHC loci for 863 Alpine marmots with a genotype assignment error rate estimated as 0.21%. The proposed methodology could be applied to any genetic system and any organism, except when extensive copy-number variation occurs (i.e. genes with a variable number of copies in the genotype of an individual). Our results highlight the potential of amplicon-based NGS techniques combined with adequate post-processing to obtain the large-scale highly reliable genotypes needed to understand the evolution of highly polymorphic functional genes.

**Keywords:** Alpine marmot; high throughput sequencing; *Marmota marmota*; non-model species; parentage relationships; post-processing

## 1. INTRODUCTION

Understanding the evolution of highly polymorphic functional genes is a major issue in evolutionary biology, ecology and medical science (e.g. genes involved in immunity in insects (Ellis et al. 2012), in plants (Jones 2001) and in vertebrates (Hedrick 1994, Akira et al. 2001); genes regulating self-incompatibility in plants (Charlesworth and Awadalla 1998) or in fungi (Wu et al. 1998)). Reaching this goal requires the comprehension of the role of the different selective pressures acting on functional genes. For this purpose, information regarding the genetic makeup of free-living individuals at various temporal and spatial scales is needed (Vasemägi and Primmer 2005).

Amplicon-based next generation sequencing (NGS) techniques allow the sequencing of large numbers of individuals at affordable costs (the term amplicon refers here to a pool of sequenced reads obtained for a given individual, locus and PCR). Moreover, amplicon-based NGS techniques are highly suitable to genotype genes with high allelic diversity and provide a viable method to obtain an accurate genotyping of all copies for genes characterized by a variable number of copies (i.e. with copy-number variation) (Babik 2010, Promerová et al. 2012). Indeed, amplicon-based NGS techniques have been widely used to genotype individuals at a wide range of highly polymorphic functional loci such as the major histocompatibility complex (MHC) in rodents (Babik et al. 2009, Galan et al. 2010, Sommer et al. 2013), primates (Huchard et al. 2012), carnivores (Oomen et al. 2013), birds (Radwan et al. 2012, Sepil et al. 2012, Dunn et al. 2013) and fish (Herdegen et al. 2014, Lighten et al. 2014a); the CYP7B1 and SPG7 genes involved in the hereditary spastic paraplegia in humans (Schlipf et al. 2011); Toll-Like Receptor (TLR) genes in New Zealand robin (*Petroica australis*) (Grueber et al. 2012); and the S locus receptor kinase (*SRK*) gene

involved in plant self-incompatibility in rock cress (*Arabidopsis sp.*) (Jørgensen et al. 2012).

Owing to their wide range of applications, amplicon-based NGS techniques have undergone an extremely fast development (see Shokralla et al. 2012 for details on development of NGS technologies) leading to the emergence of several NGS platforms such as 454, Illumina, SOLiD or Ion Torrent. While these NGS platforms may differ in terms of the number of obtained reads per run, the number of base pairs per read, the run time or the cost per run (see Glenn 2011, Shokralla et al. 2012 for reviews of available NGS techniques), all amplicon-based NGS techniques generate multiple reads for a given amplicon and are prone to errors such as artefacts (substitutions and chimeras) during PCR or indels during the sequencing process (Glenn 2011). Discerning true alleles versus sequencing errors is thus challenging especially where several functionally distinct alleles may differ by a single nucleotide (Babik 2010), and post-processing procedures are then needed. How to deal with these errors is a fundamental challenge for large-scale sequencing studies and two approaches have been proposed (Lighten et al. 2014b) (Table 2).

Taxon	Species	MHC class	loci	Number of loci	Number of alleles/locus	Number of amplicons	Number of genotyped amplicons	% assigned genotypes	mean $\pm$ SD coverage per amplicon (max)	Cost per amplicon <sup>4</sup> (€)	Cost per genotyped amplicon <sup>5</sup> (€)	References
Mammals	<i>Marmota marmota</i>	I	UB	4	2	1 152	834	72%	134 $\pm$ 148 (1 647)	1.58	2.37	Present study
			UD		3	1 152	790	69%	100 $\pm$ 109 (850)			
	II	DRB1		8	1 152	719	63%	69 $\pm$ 85 (707)				
		DRB2		3	1 152	726	63%	99 $\pm$ 124 (1 008)				
<i>Myodes glareolus</i>	II	DRB	>8	<9.12	96	79	82%	94 $\pm$ 44	76.03	92.39	Babik et al. 2009	
	II	DRB1/DRB2	2	2/8	10	10	100%	5 199 $\pm$ 1 586	91.01	91.01	Oomen et al. 2013	
11 rodent Genera	II	I <sup>1</sup>	15	NA	1566 (22-650) <sup>3</sup>	1407 (4-602) <sup>3</sup>	90%	74 $\pm$ NA	4.66	5.19	5.19	Galan et al. 2010
<i>Microcebus murinus</i>	II	DQB	2	61	672	643	96%	394 $\pm$ 429	5.07	5.63	Huchard et al. 2012	
		DRB		60	768	654	85%	394 $\pm$ 477				
Fish	<i>Gasterosteus aculeatus</i>	II	I Ib	NA	244 <sup>6</sup>	364	295	81%	442 $\pm$ NA (4687)	20	24.68	Stutz and Bolnick 2014
Birds	<i>Ficedula albicollis</i>	II	I <sup>2</sup>	2	11.9 $\pm$ 3.2	249	222	89%	1 244 $\pm$ 474 (2 666)	29.24	32.80	Radwan et al. 2012

**Table 2.** Comparison of large-scale studies of amplicon-based NGS using 454 platforms and post-processing for characterizing MHC of non-model vertebrates. <sup>1</sup>: see in Galan et al. (2010) for more details on the sequenced loci and species; <sup>2</sup>: birds have an overall locus organization differing from that of mammals and should not be compared with mammals, see Radwan et al. (2012) and Sepil et al. (2012) for more details on how to name bird's loci; <sup>3</sup>: number of samples depending on the studied Genera. <sup>4</sup>: overall cost of 454 sequencing divided by the number of amplicons excluding DNA extraction and PCR cost. <sup>5</sup>: overall cost of 454 sequencing divided by the number of genotyped amplicons excluding DNA extraction and PCR cost. <sup>6</sup>: number of alleles at the overall MHC genotyped loci.

The first approach aims to classify each variant as an allele or an artefact based on two assumptions: (1) that variants corresponding to alleles will be over-represented compared to artefacts and (2) that artefacts will be more similar to alleles than alleles will be from each other. Protocols then filter out obvious artefacts (too short or too long reads), filter variants based on their frequency of occurrence within an entire run and/or within an amplicon and apply criteria based on sequence similarity to differentiate artefacts from alleles (e.g. Galan et al. 2010, Babik et al. 2009, Radwan et al. 2012, Huchard et al. 2012). The second and very recent approach (Lighten et al. 2014a, Stutz and Bolnick 2014), rather than applying frequency or similarity criteria to determine whether a given variant is an allele or not, groups variants into clusters based on the similarity of the corresponding sequences and then identify alleles on a cluster by cluster basis rather than on a variant by variant basis. Both these approaches suffer from validation over a very limited number of samples using cloning or duplicate PCRs of the same samples (generally less than 30 samples but see Sommer et al. 2013 for a counter example).

Building on these approaches, we propose a genotyping protocol suitable for any genes which does not present extensive copy-number variations between individuals of a given species and for data generated by amplicon-based NGS techniques. Compared to previously published protocols, this protocol (1) can be applied to amplicons with very low coverage, (2) can be readily implemented thanks to a web interface, (3) has been extensively validated (1 450 validated genotypes), and (4) is highly reliable (error rate = 0.21%). This protocol allows reliable genotyping of amplicons with low coverage thanks to the assignation of reads based only on a few base pairs of the primers instead of the whole sequence and to clustering reads containing sequencing errors such as indels. Given that amplicon-based NGS techniques remain expensive (Glenn 2011, Shokralla et al. 2012), the proposed protocol contributes to the reduction of the sequencing costs.

The proposed protocol has been applied to data generated by 454 sequencing (suitable since the targeted MHC sequences range from 175-bp to 230-bp) on four MHC loci (*Mama-UB*, *Mama-UD*, *Mama-DRB1* and *Mama-*

*DRB2*) (Kuduk et al. 2012) of 1096 wild Alpine marmots (*Marmota marmota*). The MHC is a multi-gene family present in all jawed vertebrates and the MHC genes are the most polymorphic loci known in the vertebrate genome (reviewed in Kelley et al. 2005). These genes, divided into three main families named class I, II and III (Kelley et al. 2005), encode for proteins involved in antigen presentation and play a critical role in vertebrate disease resistance by initiating the immune response (Hedrick 1994). The quality of the assigned genotypes was assessed through (1) checking intra-individual repeatability of amplicon-based NGS sequencing; (2) comparing genotypes of a given individual obtained by amplicon-based NGS and Sanger sequencing; and (3) assessing the consistency of the obtained genotypes using the parentage relationships previously established based on a panel of sixteen microsatellites.

## 2. MATERIAL AND METHODS

### *Amplicon-based NGS data processing*

NGS data processing to obtain individual genotypes can be divided into the four following steps:

- Step 1. Assignment of reads to loci and individuals, elimination of singletons and elimination of reads with inappropriate sizes;
- Step 2. Elimination of amplicons with insufficient coverage;
- Step 3. Determination of alleles;
- Step 4. Determination of homozygous and heterozygous amplicons.

All steps are fully described below. These steps are carried out using two custom and heavily commented Python scripts which are freely available at <https://github.com/tbigot/alFinder>, accompanied by example files and dataset. A user friendly web interface ([http://pbil.univ-lyon1.fr/software/alFinder/config\\_generator/alFinder\\_config\\_step1](http://pbil.univ-lyon1.fr/software/alFinder/config_generator/alFinder_config_step1)) is available

to apply the procedure.

In the case where alleles have been previously described, the user can process directly through all the steps described below. In the case where alleles have not been previously described, the user should conduct a pre-treatment to identify alleles' sequences. To this purpose, the user should follow the steps described below but retaining only reads equal to the expected allele length in step 1.3 and skipping step 3. The post-processing procedure can then be conducted (following all steps described below) to maximize the number of retained reads and amplicons thanks to indels' assignment (see step 3).

*Step 1. Assignment of reads to loci and individuals, elimination of singletons, elimination of reads with inappropriate sizes*

Only reads containing perfect analogous individual tags (forward and reverse) and the minimal number of bp allowing to distinguish with no ambiguity between the forward and reverse primers (e.g. the first x-bp of the different primers used or any x-bp within the primer sequence) instead of complete primers are retained from the FASTA files produced as a result of the sequencing run (Step1.1). The use of x-bp of the primers instead of the complete primers was designed to maximize the number of retained reads per individual. The number and position of bp used is flexible and depends on the design of the primers and on the degree of stringency needed. After cutting off tags and primers, the library file is compressed by removing singletons (i.e. variants represented by a single read) (Step1.2) as well as reads with less than 95% or more than 105% of the expected allele length (Step1.3) to decrease the opportunity for inclusion of PCR chimeras and to maximize the number of reads with indels retained.

*Step 2. Elimination of amplicons with insufficient coverage*

Given that a minimal number of reads is required to obtain a reliable genotype and that several identical reads must be present within a given amplicon to obtain a reliable allele (Galan et al. 2010), the minimal number of reads per amplicon needs to be calculated to ensure a negligible probability of missing alleles. The model proposed by Galan et al. (2010) was used to assess this number per amplicon. In this model, the confidence level ( $f$ ) to determine a correct genotype



depends on three components: (1)  $r$ , the minimum required number of copies of the given allelic variant within an amplicon; (2)  $n$ , the total number of reads for a given amplicon; and (3)  $m$ , the maximum number of alleles within an amplicon (i.e.  $m$  is fixed to 2 for one locus in a diploid species). The program “Negative Multinomial” (implemented by Galan et al. (2010) and freely available online) is used to determine the minimum value of  $n$  for a confidence level ( $f$ ) of at least 95%. Thus, amplicons with less reads than the total number of reads for a given amplicon ( $n$ ) are discarded at this stage.

### *Step 3. Determination of alleles*

Variants corresponding to previously described alleles are identified and named following the original names of the alleles. Remaining variants are classified in two groups: (1) variants with a length corresponding to the one of the previously described alleles (i.e. correct length variants) and (2) variants with a length different than the one corresponding to the previously described alleles (i.e. incorrect length variants).

To increase the number of assigned reads thanks to the use of sequencing errors, all variants are aligned using the progressive alignment (Feng and Doolittle 1987) with the default aligning parameters of the CLC Sequence Viewer software free trial version 6.7.1. Since the primary artefacts generated by amplicon-based NGS techniques differ among platforms used (e.g. 454 and Ion Torrent primary generate indels, SOLiD generates A-T bias and Illumina generates substitutions (Glenn 2011)), this step is more efficient in retaining additional reads for 454 and Ion Torrent instruments than for other technologies. Studies using SOLiD or Illumina instruments could thus skip this step until further implementations may allow the exploitation of artefacts generated by these techniques.

First, among correct length variants, those presenting indels that lead to a change in all amino-acids following the indels (i.e. variants with one insertion and one deletion resulting from sequencing errors) are assigned manually to a previously identified alleles or to another correct length variant. Other correct length variants are not assigned since they could be either artefacts (i.e. indels) or

true alleles produced by mutation-selection effects. Assigning such variants could lead to the underestimation of allelic diversity. Second, incorrect length variants presenting indels are assigned manually to previously identified alleles or to correct length variants. Assignment of variants with indels allows to increase the evidence that the most frequent variant present in a given amplicon is a true allele (Stutz and Bolnick 2014). This procedure could be especially useful when true alleles are represented by few reads due to a lower sequencing efficiency relative to other alleles (Sommer et al. 2013).

*Step 4. Determination of homozygous and heterozygous amplicons*

To determine homozygous and heterozygous amplicons, the model proposed by Hohenlohe et al. (2010), and later used by Etter et al. (2011), is used to calculate, given the sequencing error rate, the likelihood of each possible genotype given all the retained variants of an amplicon. An homozygous or an heterozygous genotype is assigned to each amplicon based on a likelihood ratio test between the most likely homozygous and the most likely heterozygous genotypes with one degree of freedom. If the likelihood ratio test is not significant, no genotype is assigned.

*Field methods and sample collection*

DNA samples were collected in three wild populations of Alpine marmots located in the Nature Reserve of the Grande Sassi re (at 2,340 m a.s.l., French Alps, 45°29'N, 65°90'E), the Gran Paradiso National Park (at 2,190 m a.s.l., Italian Alps, 45°35'N, 7°11'E) and the Llosa valley (at 1,900 m a.s.l., Catalan Pyrenees, 42° 26'N, 1°42'E).

In the Grande Sassi re Nature Reserve, Alpine marmot samples have been collected since 1990. Individuals from 26 family groups were captured, tranquillized with Zol til 100 (0.1 mL Kg<sup>-1</sup>), sexed, and their age was determined from their size (up to 3 years). Their social status was determined according to scrotal development for males and teat development for females, and hair samples and skin biopsies were collected for molecular analysis. The

composition of family groups was assessed from both capture–recapture data and intensive observations (see Cochas et al. 2006 for details on the observation protocol). Similar protocols have been used since 2008 in the Gran Paradiso and the Llosa valley populations.

From the captured individuals, 1 036, 30 and 30 genetic samples were obtained from the Grande Sassi re, the Gran Paradiso and the Llosa valley populations, respectively (Table 3).

	Number of captured individuals	Number of individuals with microsatellite genotypes	Number of individuals to be sequenced at 4 MHC loci	Number of amplicons to be sequenced per MHC locus	Number of amplicons to be sequenced for all 4 MHC loci	Number of amplicons successfully sequenced at all 4 MHC loci
Total	1 096	1 036	1 096	1 152 <sup>1</sup>	4 608 <sup>1</sup>	3 069 <sup>1</sup>
The Sassi�re	1 036	1 036	1 036	1 092 <sup>1</sup>	4 368 <sup>1</sup>	2 894 <sup>1</sup>
Grand Paradiso	30	0	30	30	120	91
The Llosa	30	0	30	30	120	84

**Table 3.** Summary of the individuals and amplicons used for MHC and microsatellite sequencing. <sup>1</sup>Includes replicated individuals.

### *DNA extraction*

For these 1 096 individuals (Table 3), genomic DNA was extracted from 15 to 30 hairs or skin biopsies by incubation at 66 °C for 80 min for hairs and at 56°C for 120 min for skin biopsies in 50 µL lysis buffer (20 mM Tris-HCl, 1.5 mM MgCl<sub>2</sub>, 25 mM KCl, 0.5% Tween 20, and 0.1 mg/ml proteinase K), followed by 20 min of proteinase K inactivation at 96°C.

### *MHC characterization*

Six MHC loci have yet been characterized for the Alpine marmot: four MHC class I loci (*Mama-UA*, *Mama-UB*, *Mama-UC* and *Mama-UD*) and two MHC class II DRB loci (*Mama-DRB1* and *Mama-DRB2*) (Kuduk et al. 2012). Among these loci, *Mama-UB*, *Mama-UD*, *Mama-DRB1*, and *Mama-DRB2* were polymorphic (2 alleles for *Mama-UB*, 3 for *Mama-UD*, 8 for *Mama-DRB1*, and 3

for *Mama-DRB2*), whereas polymorphism was not detected at *Mama-UA* and *Mama-UC*. Thus, only the four polymorphic loci were sequenced in this study.

### *Amplicon-based NGS*

Amplification of MHC loci used specific primers for Alpine marmots designed to separate each locus and to avoid sequencing several duplicated genes with a single pair of primers (see Kuduk *et al.* 2012 for more details). Amplification of MHC class I used the primers MarmMF1/MarmR2 for *Mama-UB*, and MarmMF1/MarmR4 for *Mama-UD* (Kuduk *et al.* 2012). Amplification of MHC class II DRB loci used the primers MM\_DRB\_F1/MM\_DRB\_R3 for *Mama-DRB1* and MM\_DRB\_F2a/MM\_DRB\_R2 for *Mama-DRB2* (Kuduk *et al.* 2012). MM\_DRB\_F2a (5'-ACGATTCCTGCAGCAGATGA-3') was used instead of MM\_DRB\_F2 (5'-GAGTGTCATTTTTTCAATrGGA-3') (Kuduk *et al.* 2012), because the high number of homopolymers in the MM\_DRB\_F2 primer region interferes with the extraction of reads. A GS FLX Titanium Primer Adaptor (5'-CGTATCGCCTCCCTCGCGCCATCAG-3' for forward primer and 5'-CTATGCGCCTTGCCAGCCCGCTCAG-3 for reverse primer) necessary for the 454 sequencing were added to these locus-specific primers. A 6-bp tag used for barcoding individuals was also added between the adapters and the locus-specific primers. Sixteen different tags in the forward primers and twelve different tags in the reverse primers (Table 7 in Appendix 1, Chapter 1) with a minimum of 3-bp differences between tags were used to ensure a very low probability for a read to be assigned to the wrong individual due to a typing mistake in the individual tag and to allow us to pool 144 amplicons/locus.

PCR was performed in a Mastercycler (Eppendorf, Hamburg, Germany) in 10  $\mu\text{L}$  of reaction mixture containing 5  $\mu\text{L}$  HotStarTaq Polymerase Master Mix (QIAGEN, Hilden, Germany), 0.2  $\mu\text{L}$  of both primers at 100  $\mu\text{M}$ , 3.6  $\mu\text{L}$  water and 1  $\mu\text{L}$  DNA at a concentration of 30  $\text{ng}\cdot\mu\text{L}^{-1}$ . The cycling scheme was 95°C for 15 min, followed by 34 cycles at 95°C for 30 s, primer-specific annealing temperature (50°C for *Mama-UD* and *Mama-DRB2* and 55°C for *Mama-UB* and

*Mama-DRB1*) and for 30 s, 72°C for 60 s and a final extension step at 72°C for 10 min. The concentration of the PCR products was measured by fluorometry using the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen). Equimolar amounts of 144 amplicons were pooled for a given locus, and purified using MinElute PCR Purification Kit (QIAGEN). The amplicons for each of the four loci were subsequently pooled and sequenced on an eighth of a PicoTiterPlate of a Roche® 454 FLX sequencing instrument (576 amplicons/eighth). The total of the 4 608 amplicons were thus sequenced on a full PicoTiterPlate (Table 3).

### *Validation procedures*

Three validation procedures were conducted on the 1 036 sequenced individuals from the Grande Sassièrè only (Table 3): validation through NGS repeatability, through Sanger sequencing and through parentage relationships.

For validation through NGS repeatability, thirty individuals were amplified, for each locus, in 2 to 5 independent PCRs (mean number of PCR/individual/loci  $\pm$  SD:  $2.87 \pm 0.82$ ) using different tag combinations, yielding a total of 86 amplicons/locus. Consequently, from the 1096 individuals for whom DNA was extracted, 1152 amplicons at each locus were included in the sequencing process (Table 3).

For validation through Sanger sequencing, DNA extractions and PCR amplifications were carried out for 9 individuals at four MHC loci (36 genotypes) following the same protocols as above. 30  $\mu$ L of PCR products were purified using the Axygen® Cleanup Kit for PCR. Purified PCR products were single-strand sequenced using Dideoxynucleotide Terminator (Dyeterminator, kit BigDye® v.3.1 provided by LifeTechnologies). Sanger sequences are prone to sequencing errors in the 5' region. So, for the polymorphic regions to be on the 3' of the Sanger sequences, we sequenced with the reverse primer, when the polymorphism was at the end of the sequence of interest, and with the forward primer, when the polymorphism was at the beginning of the sequence of interest (MarmR2 (forward) for *Mama-UB*, MarmR4 (forward) for *Mama-UD*,

MM\_DRB\_R3 (forward) for *Mama-DRB1* and MM\_DRB\_F2 (reverse) for *Mama-DRB2*). Sequenced products were then purified using AxyPrep Mag Dye Clean (Axygen, following manufacturer's instructions). The DNA sequencing reactions were then analyzed on an ABI3730XL 96 caps DNA Analyzer (LifeTechnologies). Obtained reads were aligned using the same procedure as stated earlier. Genotypes were considered as homozygous when they perfectly matched with a unique described allele, whereas they were considered as heterozygous when ambiguous positions in the DNA-sequence were found and perfectly matched with two described alleles.

For validation through parentage relationships, 1 036 individuals (Table 3) were genotyped at sixteen microsatellites: SS-Bib11, SS-Bib18, SS-Bib20, SS-Bib31, SS-Bib4 (Klinkicht 1993); MS41, MS45, MS47, MS53, MS56, MS6, ST10 (Hanslik and Kruckenhauser 2000); Ma002, Ma018, Ma066, Ma091 (Da Silva et al. 2003) (see Appendix 2 in Chapter 1 for details on protocol and microsatellite characteristics). Parentage analyses (see Cohas et al. 2008 and Appendix 2 in Chapter 1) enabled us to confirm maternity and paternity for 663 pups. These 663 mother-father-offspring triads were then used to check for consistency of the obtained MHC genotypes.

### **3. RESULTS**

#### *454 sequencing output*

The 454 sequencing run yielded a total of 584 745 reads. The 6-bp total length of the individual tags (forward and reverse) and the first 3-bp of the locus-specific forward and reverse primers were used to assign reads to a given individual and locus. Among the obtained reads, 459 143 reads perfectly matched the individual tags and the 3-bp primer markers. Between 966 and 1 069 amplicons were obtained by retaining 3-bp primers, whereas between 922 and 1 056 amplicons

would have been obtained by retaining reads with complete primers. The mean coverage per amplicon varied across amplicons (mean coverage per amplicon ranged between 69 and 134, see Table 2 for details on each locus) and was lower than expected (expected coverage per amplicon: 217 reads).

After removing singletons and reads of less than 95% or more than 105% of the correct length variants (Table 4), 245 178 reads were retained and the coverage was similar for all loci with respect to the 1 152 genotype calls for each locus (948 retained amplicons (82%) for *Mama-UB*; 944 (82%) for *Mama-UD*; 903 (78%) for *Mama-DRB1* and 901 (78%) for *Mama-DRB2*).

The minimum number of reads per amplicon ( $n$ ) was found to be twelve (corresponding to a confidence level of 96.1%) after fixing  $r$  to 3 following Galan et al. (2010) and  $m$  to 2 (the Alpine marmot being a diploid species and each locus being amplified separately). A total of 242 067 reads were retained after discarding amplicons for which the total number of reads per locus was lower than twelve (Table 4). Again, similar numbers of retained amplicons were found at each locus with respect to the 1 152 genotype calls for each locus (835 retained amplicons (72%) for *Mama-UB*; 793 (69%) for *Mama-UD*; 720 (63%) for *Mama-DRB1*; and 726 (63%) for *Mama-DRB2*).

At this stage, 578 variants were identified. Sixteen out of these variants corresponded to previously described alleles (Kuduk et al. 2012), 55 variants had correct length (all belonging to *Mama-UD* locus) and the remaining 507 had incorrect length (77 for *Mama-UB*, 2 for *Mama-UD*, 233 for *Mama-DRB1* and 195 for *Mama-DRB2*). Among the 55 correct length variants, 44 *Mama-UD* variants contained indels and were assigned to previously identified alleles. Among the 507 incorrect length variants, 318 contained indels and were assigned to previously identified alleles (53 for *Mama-UB*, 104 for *Mama-DRB1*, and 117 for *Mama-DRB2*). The remaining 11 correct length variants and the remaining 233 incorrect length variants could either be PCR artefacts/chimeras or new alleles differing from those described in Kuduk et al. (2012). After the determination of homozygous and heterozygous amplicons, all these 244 remaining variants were removed.

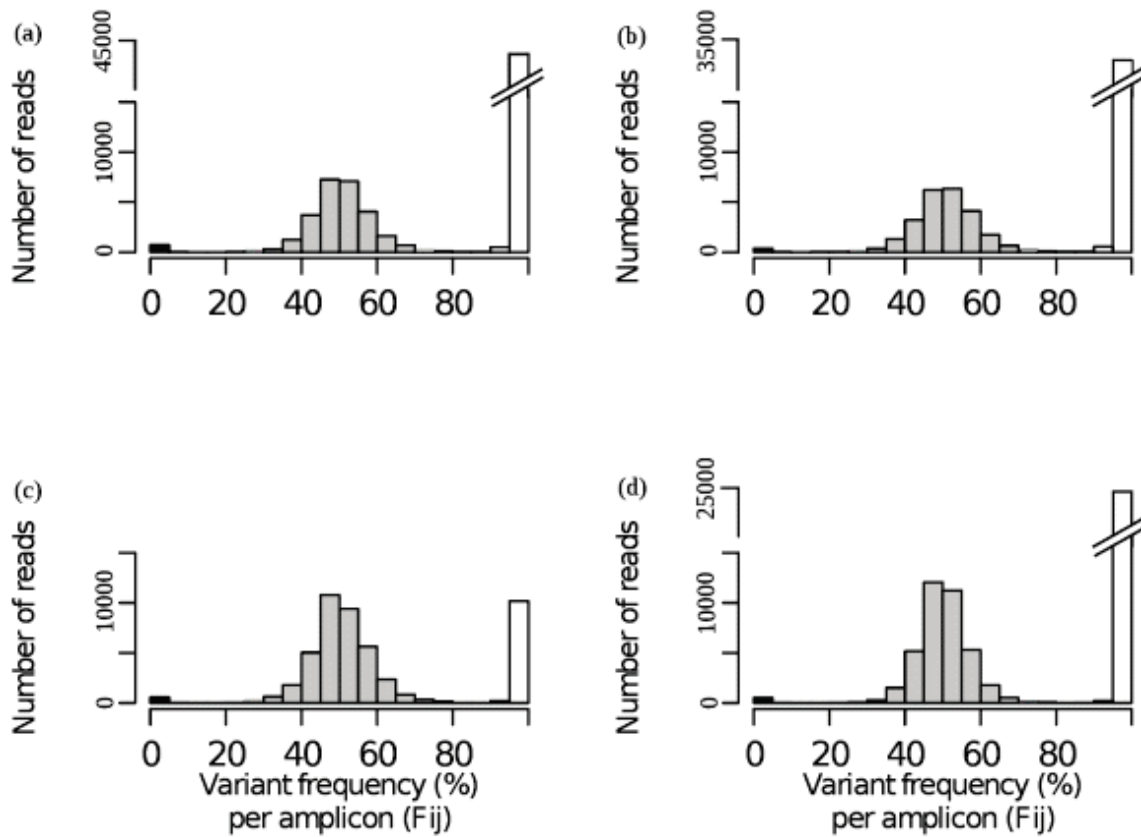
Processing description	Total number of reads (N. amp. <sup>1</sup> )	Mama-UB reads (N. amp. <sup>1</sup> )	Mama-UD reads (N. amp. <sup>1</sup> )	Mama-DRB1 reads (N. amp. <sup>1</sup> )	Mama-DRB2 reads (N. amp. <sup>1</sup> )
Number of reads (all mixed loci)	584 745	-	-	-	-
Elimination of reads without both forward and reverse individual tags and 3-bp locus-specific primers (Step1.1)	459 143 (4 103)	152 960 (1067)	113 862 (1069)	79 114 (966)	113 207 (1001)
Elimination of singletons (Step1.2) and reads with < 95% or > 105% of correct length variants (Step1.3)	245 178 (3 696)	72 214 (948)	59 642 (944)	49 212 (903)	64 110 (901)
Elimination of amplicons with < 12 reads for a given locus (Step 2)	242 067 (3 074)	71 611 (835)	58 911 (793)	48 291 (720)	63 254 (726)
Elimination of variants after determining the most likely genotype for each amplicon (Step 3 and 4)	239 879 (3 069)	70 817 (834)	58 424 (790)	47 251 (719)	63 387 (726)

**Table 4.** Amplicon-based next generation processing steps to minimize the impact of sequencing errors among all MHC studied loci. <sup>1</sup>N. amp.: number of sequenced amplicons.



The assignment of indels to previously described alleles allowed us to increase the number of retained reads by 9.56% (194 996 reads were obtained by assigning indels, whereas 215 601 would have been obtained without assigning indels) and to increase the total number of retained amplicons (3 069 amplicons were obtained by assigning indels, whereas 2 866 would have been obtained without assigning indels).

The distribution of variant frequencies per sample at each locus (Figure 10), indicated two main peaks corresponding to frequencies of 90-100% and 25-75%. This distribution is consistent with the expected distribution of homozygous and heterozygous amplicons. The sequencing error rate for 454 FLX sequencing instrument has been estimated to range between 0.18% to 0.03% (Huse et al. 2007). Sequencing error rate was assumed to be constant at each locus (but see Hohenlohe et al. 2010, Etter et al. 2011). Indeed, at this stage of the post-processing most of the sequencing errors should have been eliminated thanks to the previous post-processing steps (e.g. the elimination of singletons and reads of < 95% and > 105% correct length variants or the use of indels). We thus fixed the sequencing error rate to 0.03%. After determining the most likely genotype for each amplicon, a total of 239 879 reads were retained (Table 4). Similar numbers of retained amplicons were found at each locus with respect to the 1 152 genotype calls for each locus (834 retained amplicons (72%) for *Mama-UB*; 790 (69%) for *Mama-UD*; 719 (63%) for *Mama-DRB1*; and 726 (63%) for *Mama-DRB2*). Thus, from a total of 1 152 genotype calls at each locus (4 608 at the four MHC loci), between 719 and 834 genotypes were retained for each locus (3069 genotypes at the four MHC loci, see Table 4). At the final stage, no new variants were found suggesting that no new alleles were present in the studied population and that PCR artefacts/chimeras had been successfully removed during the post-processing.



**Figure 10.** Distribution of the reads according to the frequency of the variant  $i$  within the amplicon  $j$  ( $F_{ij}$ ) they belong over all amplicons with twelve or more reads for *Mama-UB* (a), *Mama-UD* (b), *Mama-DRB1* (c), and *Mama-DRB2* (d). White bars likely correspond to homozygous amplicons ( $F_{ij}$  of 90-100%), light gray bars likely correspond to heterozygous amplicons ( $F_{ij}$  of 25-75%) and black bars likely correspond to sequencing errors ( $F_{ij}$  of 0-5%).

Details on MHC allelic frequencies for the Grande Sassièrè population are summarized in Table 5. The number of genotypes obtained for each individual was found to significantly differ ( $\chi^2 = 5\,330$ ; d.f. = 4;  $P = <0.0001$ ) from a binomial distribution with a probability of being genotyped at a given locus of 0.67 (i.e. 3\,069 obtained genotypes divided by 4\,608 genotype calls) due to an inflated number of individuals for which no genotype was obtained (257 observed vs 14 expected individuals with no genotype).

Conducting the proposed procedure without using *a priori* knowledge of the alleles, leads to retain 2\,866 amplicons and to identify the 16 variants corresponding to the previously described alleles (Kuduk et al. 2012) as potential alleles. The use of the pre-treatment does not affect the obtained results since no differences were found between these results and the ones obtained when using *a*

*priori* knowledge of the allele.

Investigation of different sequencing error rates revealed that when the error rate was increased to 0.18, 156 additional genotypes were not assigned and 111 genotypes were found homozygous instead of heterozygous. Increasing the sequencing error rate further increased the genotyping error rate to 0.95% (13 identified erroneous genotypes) instead of 0.21% (see below for genotype validation results).

<i>Mama-UB</i> alleles	<i>Mama-UB</i> allelic freq.	<i>Mama-UD</i> alleles	<i>Mama-UD</i> allelic freq.	<i>Mama-DRB1</i> alleles	<i>Mama-DRB1</i> allelic freq.	<i>Mama-DRB2</i> alleles	<i>Mama-DRB2</i> allelic freq.
*01	0.74	*01	0.74	*01	0.39	*01	0.45
*02	0.26	*02	0.08	*02	0.23	*02	0.39
		*03	0.19	*03	0.13	*03	0.16
				*04	0.02		
				*05	0.03		
				*06	0.07		
				*07	0.09		
				*08	0.04		
Nind	754		712		664		665

**Table 5.** Allelic frequencies of 4 MHC loci for the The Grande Sassièrre Nature Reserve population. Nind: number of individuals.

### *Validation of MHC genotype assignment*

A total of 1 450 out of 2 894 assigned genotypes from the Nature Reserve of the Grande Sassièrre population (Table 6) were used to validate the above procedure. Regarding intra-individual repeatability of 454 sequencing, no difference was found among the 30 replicated individuals (16 replicates for *Mama-UB*, 13 for *Mama-UD*, 12 for *Mama-DRB1*, and 10 for *Mama-DRB2*, all validated). Concerning genotypes obtained through 454 sequencing and Sanger sequencing, no difference was found among the 36 duplicated genotypes (9 for each studied locus). In the last validation, 1 363 assigned genotypes (1016 from offspring, 161 from mothers and 186 from fathers) were used to check for genotype consistency with the parentage relationships established with microsatellites (Table 6). From a total of 1 363 checked genotypes, only 3 mismatches (0.21% of all checked genotypes) within mother-father-offspring triads (Table 6) were found. Given the three validation procedures, genotypes obtained after 454 post-processing were found to be subject to a very small error rate (0.21%).

Locus	Number of checked triads		Total offspring genotypes	Total mother genotypes	Total father genotypes
	Homozygous parents, same alleles ♀ AA and ♂ AA	Homozygous parents, different alleles ♀ AA and ♂ BB			
	Offspring AA	Offspring <sup>1</sup> AB	Offspring <sup>1</sup> AC/AD/BC/BD		
<i>Mama-UB</i>	106	10	195 <sup>2(1)</sup>	48	55
<i>Mama-UD</i>	79	0	207 <sup>2(2)</sup>	44	53
<i>Mama-DRB1</i>	2	0	210	33	37
<i>Mama-DRB2</i>	13	8	186	36	41
TOTAL	200	18	798	161	186

**Table 6.** Validation of 1 363 assigned genotypes after 454 post-processing by checking mother-father-offspring triads among all MHC studied loci for the Grande Sassièrre Nature Reserve. ♀ : e.g. of mother's genotype; ♂ : e.g. of father's genotype; <sup>1</sup>: e.g. of pup's possible genotype; <sup>2(N)</sup>: Number of non-matching genotypes.

### *MHC sequencing cost*

The expected cost of sequencing a single locus per individual was lower for 454 than for Sanger sequencing. DNA extraction costs 1.50 € (2.02 \$) and PCR costs 4.50 € (6.07 \$) which is the same whatever the sequencing method used. Although using numerous tagged primers necessary for amplicon-based NGS was expected to be costly, the extra-cost compared to using a single pair of primers necessary for Sanger sequencing was negligible when genotyping such a large number of individuals.

The cost of sequencing was 1.58 € (2.15 \$) and 4.50 € (6.07 \$) respectively for 454 and Sanger. However, after sequencing, 33% of genotype calls could not be retrieved with 454 sequencing, which represents a real cost of 2.37 € (3.20 \$) per genotype obtained without taking into account DNA extraction and PCR cost (see Table 2 to compare expected and obtained costs for similar studies).

## **4. DISCUSSION**

The proposed post-processing protocol allowed us to obtain highly reliable genotypes at two MHC class I and two MHC class II loci for 863 Alpine marmots, the largest set of individuals genotyped at MHC loci to date for a non-model species (see Table 2 and Lighten et al. 2014a).

Our protocol maximizes the number of retained reads and amplicons due to the use of the three first bp instead of the complete primer sequences to identify the loci and to the assignment of reads with indels to either previously known alleles or correct length variants. Moreover, the genotype assignment error rate of 0.21% attests the reliability of our method. None of the previous studies have conducted such an extensive validation of an amplicon-based NGS procedure. Consequently, the expected coverage per amplicon can be reduced to improve cost-efficiency without lowering the reliability of the obtained genotypes.

To date, no previous studies have used only part of the primer sequence and only two recent studies (Lighten et al. 2014a, Stutz and Bolnick 2014) used indels or repeatable base-mismatch errors to optimize genotyping procedures. Interestingly, identifying a given locus based solely on the first three base pairs instead of on the complete sequence of the primer as well as the assignment of indels allowed us to increase the number of retained amplicons. This protocol could be especially useful in future studies where users will choose a low mean coverage to optimize time and funding.

During post-processing, the differentiation of artefacts and PCR chimeras from true alleles has proven to be challenging (Babik et al. 2009, Babik 2010, Zagalska-Neubauer et al. 2010, Huchard et al. 2012, Sommer et al. 2013). The absence of artefacts and PCR chimeras after removing singletons in our study could be the result of multiples causes. First, the whole set of reads obtained during the 454 sequencing process was found to be of high quality (see Appendix 3 in Chapter 1 for a quality diagnostic), such that substitutions were likely very rare. Second, because we amplified each locus separately, the opportunity for production of PCR chimeras was largely reduced (i.e. there could have been at most two allelic variants/amplicon, and only one variant in homozygous amplicons). However, avoiding production of PCR chimeras is not always possible as, for instance, in the case of recently duplicated loci where the same primers amplify several loci and the loci cannot be separated (i.e. Babik et al. 2009, Castillo et al. 2010, Cammen et al. 2011, Oomen et al. 2013). Additionally, the moderate number of PCR cycles we used should have further reduced production of PCR chimeras (Lenz and Becker 2008). Finally, the low coverage per amplicon found in this study might have decreased the probability of finding both artefacts and chimeras after post-processing.

The extensive validation we conducted demonstrated the accuracy (genotype assignment error rate of 0.21%) of the proposed protocol for amplicon-based next generation post-processing. According to previous studies using amplicon-based NGS (Babik et al. 2009, Galan et al. 2010, Huchard et al. 2012, Oomen et al. 2013), the use of a stringent post-processing procedure has demonstrated its

usefulness in disentangling sequencing artefacts and chimeras from true alleles. Moreover, the validation of genotype assignment by confrontation with known parentage relationships has proven to be useful since, among the 47% of assigned genotypes that could be confronted with parentage relationships, 99.79% were validated. Consequently, such validation should be recommended whenever parentage relationships are available.

High percentages of genotype assignment were obtained in most studies using amplicon-based NGS techniques for MHC typing (Table 2), with the exception of the study of Sepil et al. (2012) (56% of genotype assignment) and our study (between 63% and 72% of genotype assignment). Our study was specifically designed to improve its cost-efficiency, and a cost to pay was the reduced expected coverage per amplicon. This low expected coverage combined with the high variability of the coverage per amplicon may have contributed to a higher loss of assigned genotypes compared to previous studies (Table 2). Even though care was taken to equalize the amount of DNA among amplicons within pools, and the mean coverage per amplicon seemed adequate to allow genotyping most amplicons (it exceeded the minimum requirement of twelve reads per amplicon by over an order of magnitude in this study), a relatively large proportion of amplicons could not be genotyped because they did not reach the minimal number of reads.

Variability of coverage per amplicon is a well-known issue when using NGS and can result from the use of sub-optimal primers (Sommer et al. 2013) or from the presence of homopolymer regions in the tags (Huse et al. 2007). However, our marmot-specific primers were carefully designed (Kuduk et al. 2012) and the tags we used (provided in Table 7, Appendix 2 in Chapter 1) were designed to minimize this problem, which was confirmed after sequencing (the number of homopolymers in the tags was not found to affect the obtained coverage per amplicon).

The variability of genomic DNA quality among samples is more likely to have been a key factor for the high loss of assigned genotypes in our study. Indeed, the inflated number of individuals for which no genotype was assigned



suggests a poor genomic DNA quality of some samples. Poor DNA quality may reduce the coverage per amplicon as well as the quality of the reads, resulting in the failure to obtain genotypes for some individuals. The existence of poor DNA quality samples may be due to differences in sample collection and/or sample preservation, which are likely to occur in studies conducted in the field over extended period of time. In the present study, the wide period of time elapsed between sample collection and sample sequencing (ranging from 1 to 21 years) may have resulted in a poor genomic DNA quality of some samples and therefore to the high number of obtained individuals with no assigned genotype. If genotyping a whole set of loci for a given individual is needed, efforts to maximize DNA quality (collection, storage, and DNA extraction) should be done.

If genotyping most of the individuals is a priority, one should consider a higher coverage per amplicon (e.g. Zagalska-Neubauer et al. 2010, Oomen et al. 2013), despite the associated increase in the cost of genotyping. Fortunately, other sequencing platforms such as Illumina now allow for sequencing the full length of functionally important MHC exons (eg. MHC class IIb exon 2 in guppies (*Poecilia reticulata*) (Lighten et al. 2014a)) at much lower cost than 454 sequencing. Similarly, higher minimum coverage per amplicon should be used to sequence recently duplicated genes (i.e. Castillo et al. 2010, Cammen et al. 2011, Oomen et al. 2013).

When several loci are amplified with the same set of primers but the number of loci is not known in advance, the post-processing procedure could be adapted to firstly obtain the number of sequenced loci. The minimal number of sequenced loci can be determined using the visualization of the distribution of the  $F_{ij}$  (see Results section and Figure 10 for details) and exploration of the  $F_j$  (distribution of the variants for a given amplicon) (Babik et al. 2009). Then, the parameters of the model proposed by Galan et al. (2010) could be changed according to the determined number of sequenced loci or using an extension of this model proposed by Sommer et al. (2013), that allows to take into account variation in amplification efficiency, a common pattern when amplifying several loci with a unique pair of primers. Additionally, a modification of the maximum-

likelihood framework proposed by Hohenlohe et al. (2010) and Etter et al. (2011) is required to determine homozygous and heterozygous amplicons for recently duplicated genes.

The success of our genotyping protocol stems not only from a low number of PCR and sequencing artefacts, but also from careful optimization of the primers we used (see Kuduk et al. 2012 for more details). Sub-optimal primers can result in low coverage of some true allelic variants (Sommer et al. 2013), and this would make it impossible to assign genotypes based on the likelihood ratio test between the most likely homozygous and the most likely heterozygous genotypes. The fact that we found perfect, or nearly perfect, congruence in all three of our validation procedures confirms that poorly fitting primers were not an issue in our system. While investment in the careful design of primers may often be substantial, it is a well spent effort that is more than compensated during large-scale genotyping given that methods similar to ours can be applied to any system for which a given pair of primers amplifies a given number of loci. However, these methods may not be suitable for species with extensive copy-number variation (i.e. Babik et al. 2009, Radwan et al. 2012).

No new alleles were found for any of the four MHC loci despite our extensive sampling effort, indicating that the sample of 38 Alpine marmots studied by Kuduk et al. (2012) was enough to capture all MHC alleles in these populations (MHC class I loci: two alleles for *Mama-UB* and three alleles for *Mama-UD*; MHC class II loci: eight alleles for *Mama-DRB1* and three alleles for *Mama-DRB2*). Although similar levels of allelic MHC diversity were found in wolverines (two and eight alleles were found in two MHC class II loci (Oomen et al. 2013), the low allelic MHC diversity in Alpine marmots contrasts with higher allelic diversity generally found in mammals. For example, twenty alleles at the MHC class I loci were found in woodchuck (*Marmota monax*) (Yang et al. 2000, Zhou et al. 2003) and seven in the European bison (*Bison bonasus*) (Babik et al. 2012). Concerning MHC class II loci, between seven and thirty-eight alleles were found in ten rodent species (Goüy de Bellocq et al. 2008), and sixty and sixty-one in two loci in Grey mouse lemur (*Microcebus murinus*) (Huchard et al.

2012). In addition to poor MHC diversity, low levels of genetic variability in Alpine marmots have been previously found using allozymes (Preleuthner and Pinsker 1993), minisatellites (Kruckenhauser et al. 1997, Rassmann et al. 1994) and microsatellites (Cohas et al. 2009, but see Goossens et al. 2001) and this in different populations spread across the whole Alpine arc. Alpine marmots were widespread in Europe until the early Holocene (Couturier 1955, Besson 1971). However, several drastic reductions of their population size possibly due to environmental variations (Preleuthner and Pinsker 1993, Rassmann et al. 1994) occurred during the early Holocene and reduced the distribution of this species to the Alpine arc and the Carpathian Mountains, its actual natural distribution. Such a low observed genetic diversity found at several genetic markers (MHC, allozymes, minisatellites and microsatellites) can be the result of bottlenecks occurring during Holocene. Moreover, small population size may have favored inbreeding that in turn has reduced genetic variability.

Our results emphasize the efficiency of amplicon-based NGS combined with adequate post-processing and validation procedures to obtain large numbers of accurately assigned genotypes from any species and genes in the absence of extensive copy-number variation. In conclusion, amplicon-based NGS and post-processing open new possibilities for research on the evolution of MHC and other highly polymorphic functional genes which require accurate large-scale assigned genotypes at the individual level.

## **5. APPENDIX**

### *Appendix 1*

**Table 7.** Forward (F) and Reverse (R) individual tags for the 4 locus used in one section of PicoTiterPlate of a Roche® 454 FLX sequencing instrument.

ID	<i>Mama-UB</i>				<i>Mama-UD</i>			
	MarmM1_F	TAG_F	Marm2_R	TAG_R	MarmM1_F	TAG_F	Marm4_R	TAG_R
1	MarmM1_F01	AACCGA	Marm2_R01	TTGAGT	MarmM1_F01	AACCGA	Marm4_R01	TTGAGT
2	MarmM1_F02	AAGTGT	Marm2_R01	TTGAGT	MarmM1_F02	AAGTGT	Marm4_R01	TTGAGT
3	MarmM1_F03	AGTGTT	Marm2_R01	TTGAGT	MarmM1_F03	AGTGTT	Marm4_R01	TTGAGT
4	MarmM1_F04	CCGCTG	Marm2_R01	TTGAGT	MarmM1_F04	CCGCTG	Marm4_R01	TTGAGT
5	MarmM1_F05	AACGCG	Marm2_R01	TTGAGT	MarmM1_F05	AACGCG	Marm4_R01	TTGAGT
6	MarmM1_F06	GGCTAC	Marm2_R01	TTGAGT	MarmM1_F06	GGCTAC	Marm4_R01	TTGAGT
7	MarmM1_F07	TTCTCG	Marm2_R01	TTGAGT	MarmM1_F07	TTCTCG	Marm4_R01	TTGAGT
8	MarmM1_F08	TCACTC	Marm2_R01	TTGAGT	MarmM1_F08	TCACTC	Marm4_R01	TTGAGT
9	MarmM1_F01	AACCGA	Marm2_R02	AAGCAG	MarmM1_F01	AACCGA	Marm4_R02	AAGCAG
10	MarmM1_F02	AAGTGT	Marm2_R02	AAGCAG	MarmM1_F02	AAGTGT	Marm4_R02	AAGCAG
11	MarmM1_F03	AGTGTT	Marm2_R02	AAGCAG	MarmM1_F03	AGTGTT	Marm4_R02	AAGCAG
12	MarmM1_F04	CCGCTG	Marm2_R02	AAGCAG	MarmM1_F04	CCGCTG	Marm4_R02	AAGCAG
13	MarmM1_F05	AACGCG	Marm2_R02	AAGCAG	MarmM1_F05	AACGCG	Marm4_R02	AAGCAG
14	MarmM1_F06	GGCTAC	Marm2_R02	AAGCAG	MarmM1_F06	GGCTAC	Marm4_R02	AAGCAG
15	MarmM1_F07	TTCTCG	Marm2_R02	AAGCAG	MarmM1_F07	TTCTCG	Marm4_R02	AAGCAG
16	MarmM1_F08	TCACTC	Marm2_R02	AAGCAG	MarmM1_F08	TCACTC	Marm4_R02	AAGCAG
17	MarmM1_F01	AACCGA	Marm2_R03	TTGCAA	MarmM1_F01	AACCGA	Marm4_R03	TTGCAA
18	MarmM1_F02	AAGTGT	Marm2_R03	TTGCAA	MarmM1_F02	AAGTGT	Marm4_R03	TTGCAA
19	MarmM1_F03	AGTGTT	Marm2_R03	TTGCAA	MarmM1_F03	AGTGTT	Marm4_R03	TTGCAA
20	MarmM1_F04	CCGCTG	Marm2_R03	TTGCAA	MarmM1_F04	CCGCTG	Marm4_R03	TTGCAA
21	MarmM1_F05	AACGCG	Marm2_R03	TTGCAA	MarmM1_F05	AACGCG	Marm4_R03	TTGCAA
22	MarmM1_F06	GGCTAC	Marm2_R03	TTGCAA	MarmM1_F06	GGCTAC	Marm4_R03	TTGCAA
23	MarmM1_F07	TTCTCG	Marm2_R03	TTGCAA	MarmM1_F07	TTCTCG	Marm4_R03	TTGCAA
24	MarmM1_F08	TCACTC	Marm2_R03	TTGCAA	MarmM1_F08	TCACTC	Marm4_R03	TTGCAA
25	MarmM1_F01	AACCGA	Marm2_R04	CACGTA	MarmM1_F01	AACCGA	Marm4_R04	CACGTA
26	MarmM1_F02	AAGTGT	Marm2_R04	CACGTA	MarmM1_F02	AAGTGT	Marm4_R04	CACGTA
27	MarmM1_F03	AGTGTT	Marm2_R04	CACGTA	MarmM1_F03	AGTGTT	Marm4_R04	CACGTA
28	MarmM1_F04	CCGCTG	Marm2_R04	CACGTA	MarmM1_F04	CCGCTG	Marm4_R04	CACGTA
29	MarmM1_F05	AACGCG	Marm2_R04	CACGTA	MarmM1_F05	AACGCG	Marm4_R04	CACGTA
30	MarmM1_F06	GGCTAC	Marm2_R04	CACGTA	MarmM1_F06	GGCTAC	Marm4_R04	CACGTA
31	MarmM1_F07	TTCTCG	Marm2_R04	CACGTA	MarmM1_F07	TTCTCG	Marm4_R04	CACGTA
32	MarmM1_F08	TCACTC	Marm2_R04	CACGTA	MarmM1_F08	TCACTC	Marm4_R04	CACGTA
33	MarmM1_F01	AACCGA	Marm2_R05	TAACAT	MarmM1_F01	AACCGA	Marm4_R05	TAACAT
34	MarmM1_F02	AAGTGT	Marm2_R05	TAACAT	MarmM1_F02	AAGTGT	Marm4_R05	TAACAT
35	MarmM1_F03	AGTGTT	Marm2_R05	TAACAT	MarmM1_F03	AGTGTT	Marm4_R05	TAACAT
36	MarmM1_F04	CCGCTG	Marm2_R05	TAACAT	MarmM1_F04	CCGCTG	Marm4_R05	TAACAT
37	MarmM1_F05	AACGCG	Marm2_R05	TAACAT	MarmM1_F05	AACGCG	Marm4_R05	TAACAT
38	MarmM1_F06	GGCTAC	Marm2_R05	TAACAT	MarmM1_F06	GGCTAC	Marm4_R05	TAACAT
39	MarmM1_F07	TTCTCG	Marm2_R05	TAACAT	MarmM1_F07	TTCTCG	Marm4_R05	TAACAT
40	MarmM1_F08	TCACTC	Marm2_R05	TAACAT	MarmM1_F08	TCACTC	Marm4_R05	TAACAT
41	MarmM1_F01	AACCGA	Marm2_R06	GGTCGA	MarmM1_F01	AACCGA	Marm4_R06	GGTCGA
42	MarmM1_F02	AAGTGT	Marm2_R06	GGTCGA	MarmM1_F02	AAGTGT	Marm4_R06	GGTCGA
43	MarmM1_F03	AGTGTT	Marm2_R06	GGTCGA	MarmM1_F03	AGTGTT	Marm4_R06	GGTCGA
44	MarmM1_F04	CCGCTG	Marm2_R06	GGTCGA	MarmM1_F04	CCGCTG	Marm4_R06	GGTCGA
45	MarmM1_F05	AACGCG	Marm2_R06	GGTCGA	MarmM1_F05	AACGCG	Marm4_R06	GGTCGA
46	MarmM1_F06	GGCTAC	Marm2_R06	GGTCGA	MarmM1_F06	GGCTAC	Marm4_R06	GGTCGA
47	MarmM1_F07	TTCTCG	Marm2_R06	GGTCGA	MarmM1_F07	TTCTCG	Marm4_R06	GGTCGA
48	MarmM1_F08	TCACTC	Marm2_R06	GGTCGA	MarmM1_F08	TCACTC	Marm4_R06	GGTCGA



## Chapter 1: Large-scale genotyping

ID	<i>Mama-UB</i>				<i>Mama-UD</i>			
	MarmM1_F	TAG_F	Marm2_R	TAG_R	MarmM1_F	TAG_F	Marm4_R	TAG_R
49	MarmM1_F01	AACCGA	Marm2_R07	CTTGGT	MarmM1_F01	AACCGA	Marm4_R07	CTTGGT
50	MarmM1_F02	AAGTGT	Marm2_R07	CTTGGT	MarmM1_F02	AAGTGT	Marm4_R07	CTTGGT
51	MarmM1_F03	AGTGTT	Marm2_R07	CTTGGT	MarmM1_F03	AGTGTT	Marm4_R07	CTTGGT
52	MarmM1_F04	CCGCTG	Marm2_R07	CTTGGT	MarmM1_F04	CCGCTG	Marm4_R07	CTTGGT
53	MarmM1_F05	AACGCG	Marm2_R07	CTTGGT	MarmM1_F05	AACGCG	Marm4_R07	CTTGGT
54	MarmM1_F06	GGCTAC	Marm2_R07	CTTGGT	MarmM1_F06	GGCTAC	Marm4_R07	CTTGGT
55	MarmM1_F07	TTCTCG	Marm2_R07	CTTGGT	MarmM1_F07	TTCTCG	Marm4_R07	CTTGGT
56	MarmM1_F08	TCACTC	Marm2_R07	CTTGGT	MarmM1_F08	TCACTC	Marm4_R07	CTTGGT
57	MarmM1_F01	AACCGA	Marm2_R08	TCCAGC	MarmM1_F01	AACCGA	Marm4_R08	TCCAGC
58	MarmM1_F02	AAGTGT	Marm2_R08	TCCAGC	MarmM1_F02	AAGTGT	Marm4_R08	TCCAGC
59	MarmM1_F03	AGTGTT	Marm2_R08	TCCAGC	MarmM1_F03	AGTGTT	Marm4_R08	TCCAGC
60	MarmM1_F04	CCGCTG	Marm2_R08	TCCAGC	MarmM1_F04	CCGCTG	Marm4_R08	TCCAGC
61	MarmM1_F05	AACGCG	Marm2_R08	TCCAGC	MarmM1_F05	AACGCG	Marm4_R08	TCCAGC
62	MarmM1_F06	GGCTAC	Marm2_R08	TCCAGC	MarmM1_F06	GGCTAC	Marm4_R08	TCCAGC
63	MarmM1_F07	TTCTCG	Marm2_R08	TCCAGC	MarmM1_F07	TTCTCG	Marm4_R08	TCCAGC
64	MarmM1_F08	TCACTC	Marm2_R08	TCCAGC	MarmM1_F08	TCACTC	Marm4_R08	TCCAGC
65	MarmM1_F01	AACCGA	Marm2_R09	ACTTCA	MarmM1_F01	AACCGA	Marm4_R09	ACTTCA
66	MarmM1_F02	AAGTGT	Marm2_R09	ACTTCA	MarmM1_F02	AAGTGT	Marm4_R09	ACTTCA
67	MarmM1_F03	AGTGTT	Marm2_R09	ACTTCA	MarmM1_F03	AGTGTT	Marm4_R09	ACTTCA
68	MarmM1_F04	CCGCTG	Marm2_R09	ACTTCA	MarmM1_F04	CCGCTG	Marm4_R09	ACTTCA
69	MarmM1_F05	AACGCG	Marm2_R09	ACTTCA	MarmM1_F05	AACGCG	Marm4_R09	ACTTCA
70	MarmM1_F06	GGCTAC	Marm2_R09	ACTTCA	MarmM1_F06	GGCTAC	Marm4_R09	ACTTCA
71	MarmM1_F07	TTCTCG	Marm2_R09	ACTTCA	MarmM1_F07	TTCTCG	Marm4_R09	ACTTCA
72	MarmM1_F08	TCACTC	Marm2_R09	ACTTCA	MarmM1_F08	TCACTC	Marm4_R09	ACTTCA
73	MarmM1_F01	AACCGA	Marm2_R10	GCGAGA	MarmM1_F01	AACCGA	Marm4_R10	GCGAGA
74	MarmM1_F02	AAGTGT	Marm2_R10	GCGAGA	MarmM1_F02	AAGTGT	Marm4_R10	GCGAGA
75	MarmM1_F03	AGTGTT	Marm2_R10	GCGAGA	MarmM1_F03	AGTGTT	Marm4_R10	GCGAGA
76	MarmM1_F04	CCGCTG	Marm2_R10	GCGAGA	MarmM1_F04	CCGCTG	Marm4_R10	GCGAGA
77	MarmM1_F05	AACGCG	Marm2_R10	GCGAGA	MarmM1_F05	AACGCG	Marm4_R10	GCGAGA
78	MarmM1_F06	GGCTAC	Marm2_R10	GCGAGA	MarmM1_F06	GGCTAC	Marm4_R10	GCGAGA
79	MarmM1_F07	TTCTCG	Marm2_R10	GCGAGA	MarmM1_F07	TTCTCG	Marm4_R10	GCGAGA
80	MarmM1_F08	TCACTC	Marm2_R10	GCGAGA	MarmM1_F08	TCACTC	Marm4_R10	GCGAGA
81	MarmM1_F01	AACCGA	Marm2_R11	TGGAAC	MarmM1_F01	AACCGA	Marm4_R11	TGGAAC
82	MarmM1_F02	AAGTGT	Marm2_R11	TGGAAC	MarmM1_F02	AAGTGT	Marm4_R11	TGGAAC
83	MarmM1_F03	AGTGTT	Marm2_R11	TGGAAC	MarmM1_F03	AGTGTT	Marm4_R11	TGGAAC
84	MarmM1_F04	CCGCTG	Marm2_R11	TGGAAC	MarmM1_F04	CCGCTG	Marm4_R11	TGGAAC
85	MarmM1_F05	AACGCG	Marm2_R11	TGGAAC	MarmM1_F05	AACGCG	Marm4_R11	TGGAAC
86	MarmM1_F06	GGCTAC	Marm2_R11	TGGAAC	MarmM1_F06	GGCTAC	Marm4_R11	TGGAAC
87	MarmM1_F07	TTCTCG	Marm2_R11	TGGAAC	MarmM1_F07	TTCTCG	Marm4_R11	TGGAAC
88	MarmM1_F08	TCACTC	Marm2_R11	TGGAAC	MarmM1_F08	TCACTC	Marm4_R11	TGGAAC
89	MarmM1_F01	AACCGA	Marm2_R12	CGAATC	MarmM1_F01	AACCGA	Marm4_R12	CGAATC
90	MarmM1_F02	AAGTGT	Marm2_R12	CGAATC	MarmM1_F02	AAGTGT	Marm4_R12	CGAATC
91	MarmM1_F03	AGTGTT	Marm2_R12	CGAATC	MarmM1_F03	AGTGTT	Marm4_R12	CGAATC
92	MarmM1_F04	CCGCTG	Marm2_R12	CGAATC	MarmM1_F04	CCGCTG	Marm4_R12	CGAATC
93	MarmM1_F05	AACGCG	Marm2_R12	CGAATC	MarmM1_F05	AACGCG	Marm4_R12	CGAATC
94	MarmM1_F06	GGCTAC	Marm2_R12	CGAATC	MarmM1_F06	GGCTAC	Marm4_R12	CGAATC
95	MarmM1_F07	TTCTCG	Marm2_R12	CGAATC	MarmM1_F07	TTCTCG	Marm4_R12	CGAATC
96	MarmM1_F08	TCACTC	Marm2_R12	CGAATC	MarmM1_F08	TCACTC	Marm4_R12	CGAATC

Chapter 1: Large-scale genotyping

ID	Mama-UB				Mama-UD			
	MarmM1_F	TAG_F	Marm2_R	TAG_R	MarmM1_F	TAG_F	Marm4_R	TAG_R
97	MarmM1_F09	GAACTA	Marm2_R01	TTGAGT	MarmM1_F09	GAACTA	Marm4_R01	TTGAGT
98	MarmM1_F10	CACAGT	Marm2_R01	TTGAGT	MarmM1_F10	CACAGT	Marm4_R01	TTGAGT
99	MarmM1_F11	CAATCG	Marm2_R01	TTGAGT	MarmM1_F11	CAATCG	Marm4_R01	TTGAGT
100	MarmM1_F12	CCGTCC	Marm2_R01	TTGAGT	MarmM1_F12	CCGTCC	Marm4_R01	TTGAGT
101	MarmM1_F13	AAGACA	Marm2_R01	TTGAGT	MarmM1_F13	AAGACA	Marm4_R01	TTGAGT
102	MarmM1_F14	GGTAAG	Marm2_R01	TTGAGT	MarmM1_F14	GGTAAG	Marm4_R01	TTGAGT
103	MarmM1_F15	ATAAIT	Marm2_R01	TTGAGT	MarmM1_F15	ATAAIT	Marm4_R01	TTGAGT
104	MarmM1_F16	CGTCAC	Marm2_R01	TTGAGT	MarmM1_F16	CGTCAC	Marm4_R01	TTGAGT
105	MarmM1_F09	GAACTA	Marm2_R02	AAGCAG	MarmM1_F09	GAACTA	Marm4_R02	AAGCAG
106	MarmM1_F10	CACAGT	Marm2_R02	AAGCAG	MarmM1_F10	CACAGT	Marm4_R02	AAGCAG
107	MarmM1_F11	CAATCG	Marm2_R02	AAGCAG	MarmM1_F11	CAATCG	Marm4_R02	AAGCAG
108	MarmM1_F12	CCGTCC	Marm2_R02	AAGCAG	MarmM1_F12	CCGTCC	Marm4_R02	AAGCAG
109	MarmM1_F13	AAGACA	Marm2_R02	AAGCAG	MarmM1_F13	AAGACA	Marm4_R02	AAGCAG
110	MarmM1_F14	GGTAAG	Marm2_R02	AAGCAG	MarmM1_F14	GGTAAG	Marm4_R02	AAGCAG
111	MarmM1_F15	ATAAIT	Marm2_R02	AAGCAG	MarmM1_F15	ATAAIT	Marm4_R02	AAGCAG
112	MarmM1_F16	CGTCAC	Marm2_R02	AAGCAG	MarmM1_F16	CGTCAC	Marm4_R02	AAGCAG
113	MarmM1_F09	GAACTA	Marm2_R03	TTGCAA	MarmM1_F09	GAACTA	Marm4_R03	TTGCAA
114	MarmM1_F10	CACAGT	Marm2_R03	TTGCAA	MarmM1_F10	CACAGT	Marm4_R03	TTGCAA
115	MarmM1_F11	CAATCG	Marm2_R03	TTGCAA	MarmM1_F11	CAATCG	Marm4_R03	TTGCAA
116	MarmM1_F12	CCGTCC	Marm2_R03	TTGCAA	MarmM1_F12	CCGTCC	Marm4_R03	TTGCAA
117	MarmM1_F13	AAGACA	Marm2_R03	TTGCAA	MarmM1_F13	AAGACA	Marm4_R03	TTGCAA
118	MarmM1_F14	GGTAAG	Marm2_R03	TTGCAA	MarmM1_F14	GGTAAG	Marm4_R03	TTGCAA
119	MarmM1_F15	ATAAIT	Marm2_R03	TTGCAA	MarmM1_F15	ATAAIT	Marm4_R03	TTGCAA
120	MarmM1_F16	CGTCAC	Marm2_R03	TTGCAA	MarmM1_F16	CGTCAC	Marm4_R03	TTGCAA
121	MarmM1_F09	GAACTA	Marm2_R04	CACGTA	MarmM1_F09	GAACTA	Marm4_R04	CACGTA
122	MarmM1_F10	CACAGT	Marm2_R04	CACGTA	MarmM1_F10	CACAGT	Marm4_R04	CACGTA
123	MarmM1_F11	CAATCG	Marm2_R04	CACGTA	MarmM1_F11	CAATCG	Marm4_R04	CACGTA
124	MarmM1_F12	CCGTCC	Marm2_R04	CACGTA	MarmM1_F12	CCGTCC	Marm4_R04	CACGTA
125	MarmM1_F13	AAGACA	Marm2_R04	CACGTA	MarmM1_F13	AAGACA	Marm4_R04	CACGTA
126	MarmM1_F14	GGTAAG	Marm2_R04	CACGTA	MarmM1_F14	GGTAAG	Marm4_R04	CACGTA
127	MarmM1_F15	ATAAIT	Marm2_R04	CACGTA	MarmM1_F15	ATAAIT	Marm4_R04	CACGTA
128	MarmM1_F16	CGTCAC	Marm2_R04	CACGTA	MarmM1_F16	CGTCAC	Marm4_R04	CACGTA
129	MarmM1_F09	GAACTA	Marm2_R05	TAACAT	MarmM1_F09	GAACTA	Marm4_R05	TAACAT
130	MarmM1_F10	CACAGT	Marm2_R05	TAACAT	MarmM1_F10	CACAGT	Marm4_R05	TAACAT
131	MarmM1_F11	CAATCG	Marm2_R05	TAACAT	MarmM1_F11	CAATCG	Marm4_R05	TAACAT
132	MarmM1_F12	CCGTCC	Marm2_R05	TAACAT	MarmM1_F12	CCGTCC	Marm4_R05	TAACAT
133	MarmM1_F13	AAGACA	Marm2_R05	TAACAT	MarmM1_F13	AAGACA	Marm4_R05	TAACAT
134	MarmM1_F14	GGTAAG	Marm2_R05	TAACAT	MarmM1_F14	GGTAAG	Marm4_R05	TAACAT
135	MarmM1_F15	ATAAIT	Marm2_R05	TAACAT	MarmM1_F15	ATAAIT	Marm4_R05	TAACAT
136	MarmM1_F16	CGTCAC	Marm2_R05	TAACAT	MarmM1_F16	CGTCAC	Marm4_R05	TAACAT
137	MarmM1_F09	GAACTA	Marm2_R06	GGTCGA	MarmM1_F09	GAACTA	Marm4_R06	GGTCGA
138	MarmM1_F10	CACAGT	Marm2_R06	GGTCGA	MarmM1_F10	CACAGT	Marm4_R06	GGTCGA
139	MarmM1_F11	CAATCG	Marm2_R06	GGTCGA	MarmM1_F11	CAATCG	Marm4_R06	GGTCGA
140	MarmM1_F12	CCGTCC	Marm2_R06	GGTCGA	MarmM1_F12	CCGTCC	Marm4_R06	GGTCGA
141	MarmM1_F13	AAGACA	Marm2_R06	GGTCGA	MarmM1_F13	AAGACA	Marm4_R06	GGTCGA
142	MarmM1_F14	GGTAAG	Marm2_R06	GGTCGA	MarmM1_F14	GGTAAG	Marm4_R06	GGTCGA
143	MarmM1_F15	ATAAIT	Marm2_R06	GGTCGA	MarmM1_F15	ATAAIT	Marm4_R06	GGTCGA
144	MarmM1_F16	CGTCAC	Marm2_R06	GGTCGA	MarmM1_F16	CGTCAC	Marm4_R06	GGTCGA



## Chapter 1: Large-scale genotyping

ID	<i>Mama-DRB1</i>				<i>Mama-DRB2</i>			
	MMDRB1_F	TAG_F	MMDRB1_R	TAG_R	MMDRB2_F	TAG_F	MMDRB2_R	TAG_R
1	MMDRB1_F01	AACCGA	MMDRB1_R01	TTGAGT	MMDRB2_F01	AACCGA	MMDRB2_R01	AAGCAG
2	MMDRB1_F02	CCGGAA	MMDRB1_R01	TTGAGT	MMDRB2_F02	AGTGTT	MMDRB2_R01	AAGCAG
3	MMDRB1_F03	AGTGTT	MMDRB1_R01	TTGAGT	MMDRB2_F03	CCGCTG	MMDRB2_R01	AAGCAG
4	MMDRB1_F04	CCGCTG	MMDRB1_R01	TTGAGT	MMDRB2_F04	AACGCG	MMDRB2_R01	AAGCAG
5	MMDRB1_F05	AACGCG	MMDRB1_R01	TTGAGT	MMDRB2_F05	GGCTAC	MMDRB2_R01	AAGCAG
6	MMDRB1_F06	GGCTAC	MMDRB1_R01	TTGAGT	MMDRB2_F06	TTCTCG	MMDRB2_R01	AAGCAG
7	MMDRB1_F07	TTCTCG	MMDRB1_R01	TTGAGT	MMDRB2_F07	TCACTC	MMDRB2_R01	AAGCAG
8	MMDRB1_F08	TCACTC	MMDRB1_R01	TTGAGT	MMDRB2_F08	GAACTA	MMDRB2_R01	AAGCAG
9	MMDRB1_F01	AACCGA	MMDRB1_R02	AAGCAG	MMDRB2_F01	AACCGA	MMDRB2_R02	TTGCAA
10	MMDRB1_F02	CCGGAA	MMDRB1_R02	AAGCAG	MMDRB2_F02	AGTGTT	MMDRB2_R02	TTGCAA
11	MMDRB1_F03	AGTGTT	MMDRB1_R02	AAGCAG	MMDRB2_F03	CCGCTG	MMDRB2_R02	TTGCAA
12	MMDRB1_F04	CCGCTG	MMDRB1_R02	AAGCAG	MMDRB2_F04	AACGCG	MMDRB2_R02	TTGCAA
13	MMDRB1_F05	AACGCG	MMDRB1_R02	AAGCAG	MMDRB2_F05	GGCTAC	MMDRB2_R02	TTGCAA
14	MMDRB1_F06	GGCTAC	MMDRB1_R02	AAGCAG	MMDRB2_F06	TTCTCG	MMDRB2_R02	TTGCAA
15	MMDRB1_F07	TTCTCG	MMDRB1_R02	AAGCAG	MMDRB2_F07	TCACTC	MMDRB2_R02	TTGCAA
16	MMDRB1_F08	TCACTC	MMDRB1_R02	AAGCAG	MMDRB2_F08	GAACTA	MMDRB2_R02	TTGCAA
17	MMDRB1_F01	AACCGA	MMDRB1_R03	TTGCAA	MMDRB2_F01	AACCGA	MMDRB2_R03	CACGTA
18	MMDRB1_F02	CCGGAA	MMDRB1_R03	TTGCAA	MMDRB2_F02	AGTGTT	MMDRB2_R03	CACGTA
19	MMDRB1_F03	AGTGTT	MMDRB1_R03	TTGCAA	MMDRB2_F03	CCGCTG	MMDRB2_R03	CACGTA
20	MMDRB1_F04	CCGCTG	MMDRB1_R03	TTGCAA	MMDRB2_F04	AACGCG	MMDRB2_R03	CACGTA
21	MMDRB1_F05	AACGCG	MMDRB1_R03	TTGCAA	MMDRB2_F05	GGCTAC	MMDRB2_R03	CACGTA
22	MMDRB1_F06	GGCTAC	MMDRB1_R03	TTGCAA	MMDRB2_F06	TTCTCG	MMDRB2_R03	CACGTA
23	MMDRB1_F07	TTCTCG	MMDRB1_R03	TTGCAA	MMDRB2_F07	TCACTC	MMDRB2_R03	CACGTA
24	MMDRB1_F08	TCACTC	MMDRB1_R03	TTGCAA	MMDRB2_F08	GAACTA	MMDRB2_R03	CACGTA
25	MMDRB1_F01	AACCGA	MMDRB1_R04	CACGTA	MMDRB2_F01	AACCGA	MMDRB2_R04	GGTCGA
26	MMDRB1_F02	CCGGAA	MMDRB1_R04	CACGTA	MMDRB2_F02	AGTGTT	MMDRB2_R04	GGTCGA
27	MMDRB1_F03	AGTGTT	MMDRB1_R04	CACGTA	MMDRB2_F03	CCGCTG	MMDRB2_R04	GGTCGA
28	MMDRB1_F04	CCGCTG	MMDRB1_R04	CACGTA	MMDRB2_F04	AACGCG	MMDRB2_R04	GGTCGA
29	MMDRB1_F05	AACGCG	MMDRB1_R04	CACGTA	MMDRB2_F05	GGCTAC	MMDRB2_R04	GGTCGA
30	MMDRB1_F06	GGCTAC	MMDRB1_R04	CACGTA	MMDRB2_F06	TTCTCG	MMDRB2_R04	GGTCGA
31	MMDRB1_F07	TTCTCG	MMDRB1_R04	CACGTA	MMDRB2_F07	TCACTC	MMDRB2_R04	GGTCGA
32	MMDRB1_F08	TCACTC	MMDRB1_R04	CACGTA	MMDRB2_F08	GAACTA	MMDRB2_R04	GGTCGA
33	MMDRB1_F01	AACCGA	MMDRB1_R05	TAACAT	MMDRB2_F01	AACCGA	MMDRB2_R05	TCCAGC
34	MMDRB1_F02	CCGGAA	MMDRB1_R05	TAACAT	MMDRB2_F02	AGTGTT	MMDRB2_R05	TCCAGC
35	MMDRB1_F03	AGTGTT	MMDRB1_R05	TAACAT	MMDRB2_F03	CCGCTG	MMDRB2_R05	TCCAGC
36	MMDRB1_F04	CCGCTG	MMDRB1_R05	TAACAT	MMDRB2_F04	AACGCG	MMDRB2_R05	TCCAGC
37	MMDRB1_F05	AACGCG	MMDRB1_R05	TAACAT	MMDRB2_F05	GGCTAC	MMDRB2_R05	TCCAGC
38	MMDRB1_F06	GGCTAC	MMDRB1_R05	TAACAT	MMDRB2_F06	TTCTCG	MMDRB2_R05	TCCAGC
39	MMDRB1_F07	TTCTCG	MMDRB1_R05	TAACAT	MMDRB2_F07	TCACTC	MMDRB2_R05	TCCAGC
40	MMDRB1_F08	TCACTC	MMDRB1_R05	TAACAT	MMDRB2_F08	GAACTA	MMDRB2_R05	TCCAGC
41	MMDRB1_F01	AACCGA	MMDRB1_R06	GGTCGA	MMDRB2_F01	AACCGA	MMDRB2_R06	ACTTCA
42	MMDRB1_F02	CCGGAA	MMDRB1_R06	GGTCGA	MMDRB2_F02	AGTGTT	MMDRB2_R06	ACTTCA
43	MMDRB1_F03	AGTGTT	MMDRB1_R06	GGTCGA	MMDRB2_F03	CCGCTG	MMDRB2_R06	ACTTCA
44	MMDRB1_F04	CCGCTG	MMDRB1_R06	GGTCGA	MMDRB2_F04	AACGCG	MMDRB2_R06	ACTTCA
45	MMDRB1_F05	AACGCG	MMDRB1_R06	GGTCGA	MMDRB2_F05	GGCTAC	MMDRB2_R06	ACTTCA
46	MMDRB1_F06	GGCTAC	MMDRB1_R06	GGTCGA	MMDRB2_F06	TTCTCG	MMDRB2_R06	ACTTCA
47	MMDRB1_F07	TTCTCG	MMDRB1_R06	GGTCGA	MMDRB2_F07	TCACTC	MMDRB2_R06	ACTTCA
48	MMDRB1_F08	TCACTC	MMDRB1_R06	GGTCGA	MMDRB2_F08	GAACTA	MMDRB2_R06	ACTTCA



Chapter 1: Large-scale genotyping

ID	<i>Mama-DRB1</i>				<i>Mama-DRB2</i>			
	MMDRB1_F	TAG_F	MMDRB1_R	TAG_R	MMDRB2_F	TAG_F	MMDRB2_R	TAG_R
49	MMDRB1_F01	AACCGA	MMDRB1_R07	CTTGGT	MMDRB2_F01	AACCGA	MMDRB2_R07	GCGAGA
50	MMDRB1_F02	CCGGAA	MMDRB1_R07	CTTGGT	MMDRB2_F02	AGTGTT	MMDRB2_R07	GCGAGA
51	MMDRB1_F03	AGTGTT	MMDRB1_R07	CTTGGT	MMDRB2_F03	CCGCTG	MMDRB2_R07	GCGAGA
52	MMDRB1_F04	CCGCTG	MMDRB1_R07	CTTGGT	MMDRB2_F04	AACGCG	MMDRB2_R07	GCGAGA
53	MMDRB1_F05	AACGCG	MMDRB1_R07	CTTGGT	MMDRB2_F05	GGCTAC	MMDRB2_R07	GCGAGA
54	MMDRB1_F06	GGCTAC	MMDRB1_R07	CTTGGT	MMDRB2_F06	TTCTCG	MMDRB2_R07	GCGAGA
55	MMDRB1_F07	TTCTCG	MMDRB1_R07	CTTGGT	MMDRB2_F07	TCACTC	MMDRB2_R07	GCGAGA
56	MMDRB1_F08	TCACTC	MMDRB1_R07	CTTGGT	MMDRB2_F08	GAACTA	MMDRB2_R07	GCGAGA
57	MMDRB1_F01	AACCGA	MMDRB1_R08	TCCAGC	MMDRB2_F01	AACCGA	MMDRB2_R08	TGGAAC
58	MMDRB1_F02	CCGGAA	MMDRB1_R08	TCCAGC	MMDRB2_F02	AGTGTT	MMDRB2_R08	TGGAAC
59	MMDRB1_F03	AGTGTT	MMDRB1_R08	TCCAGC	MMDRB2_F03	CCGCTG	MMDRB2_R08	TGGAAC
60	MMDRB1_F04	CCGCTG	MMDRB1_R08	TCCAGC	MMDRB2_F04	AACGCG	MMDRB2_R08	TGGAAC
61	MMDRB1_F05	AACGCG	MMDRB1_R08	TCCAGC	MMDRB2_F05	GGCTAC	MMDRB2_R08	TGGAAC
62	MMDRB1_F06	GGCTAC	MMDRB1_R08	TCCAGC	MMDRB2_F06	TTCTCG	MMDRB2_R08	TGGAAC
63	MMDRB1_F07	TTCTCG	MMDRB1_R08	TCCAGC	MMDRB2_F07	TCACTC	MMDRB2_R08	TGGAAC
64	MMDRB1_F08	TCACTC	MMDRB1_R08	TCCAGC	MMDRB2_F08	GAACTA	MMDRB2_R08	TGGAAC
65	MMDRB1_F01	AACCGA	MMDRB1_R09	ACTTCA	MMDRB2_F01	AACCGA	MMDRB2_R09	CGAATC
66	MMDRB1_F02	CCGGAA	MMDRB1_R09	ACTTCA	MMDRB2_F02	AGTGTT	MMDRB2_R09	CGAATC
67	MMDRB1_F03	AGTGTT	MMDRB1_R09	ACTTCA	MMDRB2_F03	CCGCTG	MMDRB2_R09	CGAATC
68	MMDRB1_F04	CCGCTG	MMDRB1_R09	ACTTCA	MMDRB2_F04	AACGCG	MMDRB2_R09	CGAATC
69	MMDRB1_F05	AACGCG	MMDRB1_R09	ACTTCA	MMDRB2_F05	GGCTAC	MMDRB2_R09	CGAATC
70	MMDRB1_F06	GGCTAC	MMDRB1_R09	ACTTCA	MMDRB2_F06	TTCTCG	MMDRB2_R09	CGAATC
71	MMDRB1_F07	TTCTCG	MMDRB1_R09	ACTTCA	MMDRB2_F07	TCACTC	MMDRB2_R09	CGAATC
72	MMDRB1_F08	TCACTC	MMDRB1_R09	ACTTCA	MMDRB2_F08	GAACTA	MMDRB2_R09	CGAATC
73	MMDRB1_F01	AACCGA	MMDRB1_R10	GCGAGA	MMDRB2_F01	AACCGA	MMDRB2_R10	AAGGTC
74	MMDRB1_F02	CCGGAA	MMDRB1_R10	GCGAGA	MMDRB2_F02	AGTGTT	MMDRB2_R10	AAGGTC
75	MMDRB1_F03	AGTGTT	MMDRB1_R10	GCGAGA	MMDRB2_F03	CCGCTG	MMDRB2_R10	AAGGTC
76	MMDRB1_F04	CCGCTG	MMDRB1_R10	GCGAGA	MMDRB2_F04	AACGCG	MMDRB2_R10	AAGGTC
77	MMDRB1_F05	AACGCG	MMDRB1_R10	GCGAGA	MMDRB2_F05	GGCTAC	MMDRB2_R10	AAGGTC
78	MMDRB1_F06	GGCTAC	MMDRB1_R10	GCGAGA	MMDRB2_F06	TTCTCG	MMDRB2_R10	AAGGTC
79	MMDRB1_F07	TTCTCG	MMDRB1_R10	GCGAGA	MMDRB2_F07	TCACTC	MMDRB2_R10	AAGGTC
80	MMDRB1_F08	TCACTC	MMDRB1_R10	GCGAGA	MMDRB2_F08	GAACTA	MMDRB2_R10	AAGGTC
81	MMDRB1_F01	AACCGA	MMDRB1_R11	TGGAAC	MMDRB2_F01	AACCGA	MMDRB2_R11	GCGGCA
82	MMDRB1_F02	CCGGAA	MMDRB1_R11	TGGAAC	MMDRB2_F02	AGTGTT	MMDRB2_R11	GCGGCA
83	MMDRB1_F03	AGTGTT	MMDRB1_R11	TGGAAC	MMDRB2_F03	CCGCTG	MMDRB2_R11	GCGGCA
84	MMDRB1_F04	CCGCTG	MMDRB1_R11	TGGAAC	MMDRB2_F04	AACGCG	MMDRB2_R11	GCGGCA
85	MMDRB1_F05	AACGCG	MMDRB1_R11	TGGAAC	MMDRB2_F05	GGCTAC	MMDRB2_R11	GCGGCA
86	MMDRB1_F06	GGCTAC	MMDRB1_R11	TGGAAC	MMDRB2_F06	TTCTCG	MMDRB2_R11	GCGGCA
87	MMDRB1_F07	TTCTCG	MMDRB1_R11	TGGAAC	MMDRB2_F07	TCACTC	MMDRB2_R11	GCGGCA
88	MMDRB1_F08	TCACTC	MMDRB1_R11	TGGAAC	MMDRB2_F08	GAACTA	MMDRB2_R11	GCGGCA
89	MMDRB1_F01	AACCGA	MMDRB1_R12	CGAATC	MMDRB2_F01	AACCGA	MMDRB2_R12	AGAGAC
90	MMDRB1_F02	CCGGAA	MMDRB1_R12	CGAATC	MMDRB2_F02	AGTGTT	MMDRB2_R12	AGAGAC
91	MMDRB1_F03	AGTGTT	MMDRB1_R12	CGAATC	MMDRB2_F03	CCGCTG	MMDRB2_R12	AGAGAC
92	MMDRB1_F04	CCGCTG	MMDRB1_R12	CGAATC	MMDRB2_F04	AACGCG	MMDRB2_R12	AGAGAC
93	MMDRB1_F05	AACGCG	MMDRB1_R12	CGAATC	MMDRB2_F05	GGCTAC	MMDRB2_R12	AGAGAC
94	MMDRB1_F06	GGCTAC	MMDRB1_R12	CGAATC	MMDRB2_F06	TTCTCG	MMDRB2_R12	AGAGAC
95	MMDRB1_F07	TTCTCG	MMDRB1_R12	CGAATC	MMDRB2_F07	TCACTC	MMDRB2_R12	AGAGAC
96	MMDRB1_F08	TCACTC	MMDRB1_R12	CGAATC	MMDRB2_F08	GAACTA	MMDRB2_R12	AGAGAC



## Chapter 1: Large-scale genotyping

ID	<i>Mama-DRB1</i>				<i>Mama-DRB2</i>			
	MMDRB1_F	TAG_F	MMDRB1_R	TAG_R	MMDRB2_F	TAG_F	MMDRB2_R	TAG_R
97	MMDRB1_F09	GAACTA	MMDRB1_R01	TTGAGT	MMDRB2_F09	CACAGT	MMDRB2_R01	AAGCAG
98	MMDRB1_F10	CACAGT	MMDRB1_R01	TTGAGT	MMDRB2_F10	CAATCG	MMDRB2_R01	AAGCAG
99	MMDRB1_F11	CAATCG	MMDRB1_R01	TTGAGT	MMDRB2_F11	CCGTCC	MMDRB2_R01	AAGCAG
100	MMDRB1_F12	CCGTCC	MMDRB1_R01	TTGAGT	MMDRB2_F12	AAGACA	MMDRB2_R01	AAGCAG
101	MMDRB1_F13	AAGACA	MMDRB1_R01	TTGAGT	MMDRB2_F13	GGTAAG	MMDRB2_R01	AAGCAG
102	MMDRB1_F14	GGTAAG	MMDRB1_R01	TTGAGT	MMDRB2_F14	ATAATT	MMDRB2_R01	AAGCAG
103	MMDRB1_F15	ATAATT	MMDRB1_R01	TTGAGT	MMDRB2_F15	CGTCAC	MMDRB2_R01	AAGCAG
104	MMDRB1_F16	CGTCAC	MMDRB1_R01	TTGAGT	MMDRB2_F16	AAGTGT	MMDRB2_R01	AAGCAG
105	MMDRB1_F09	GAACTA	MMDRB1_R02	AAGCAG	MMDRB2_F09	CACAGT	MMDRB2_R02	TTGCAA
106	MMDRB1_F10	CACAGT	MMDRB1_R02	AAGCAG	MMDRB2_F10	CAATCG	MMDRB2_R02	TTGCAA
107	MMDRB1_F11	CAATCG	MMDRB1_R02	AAGCAG	MMDRB2_F11	CCGTCC	MMDRB2_R02	TTGCAA
108	MMDRB1_F12	CCGTCC	MMDRB1_R02	AAGCAG	MMDRB2_F12	AAGACA	MMDRB2_R02	TTGCAA
109	MMDRB1_F13	AAGACA	MMDRB1_R02	AAGCAG	MMDRB2_F13	GGTAAG	MMDRB2_R02	TTGCAA
110	MMDRB1_F14	GGTAAG	MMDRB1_R02	AAGCAG	MMDRB2_F14	ATAATT	MMDRB2_R02	TTGCAA
111	MMDRB1_F15	ATAATT	MMDRB1_R02	AAGCAG	MMDRB2_F15	CGTCAC	MMDRB2_R02	TTGCAA
112	MMDRB1_F16	CGTCAC	MMDRB1_R02	AAGCAG	MMDRB2_F16	AAGTGT	MMDRB2_R02	TTGCAA
113	MMDRB1_F09	GAACTA	MMDRB1_R03	TTGCAA	MMDRB2_F09	CACAGT	MMDRB2_R03	CACGTA
114	MMDRB1_F10	CACAGT	MMDRB1_R03	TTGCAA	MMDRB2_F10	CAATCG	MMDRB2_R03	CACGTA
115	MMDRB1_F11	CAATCG	MMDRB1_R03	TTGCAA	MMDRB2_F11	CCGTCC	MMDRB2_R03	CACGTA
116	MMDRB1_F12	CCGTCC	MMDRB1_R03	TTGCAA	MMDRB2_F12	AAGACA	MMDRB2_R03	CACGTA
117	MMDRB1_F13	AAGACA	MMDRB1_R03	TTGCAA	MMDRB2_F13	GGTAAG	MMDRB2_R03	CACGTA
118	MMDRB1_F14	GGTAAG	MMDRB1_R03	TTGCAA	MMDRB2_F14	ATAATT	MMDRB2_R03	CACGTA
119	MMDRB1_F15	ATAATT	MMDRB1_R03	TTGCAA	MMDRB2_F15	CGTCAC	MMDRB2_R03	CACGTA
120	MMDRB1_F16	CGTCAC	MMDRB1_R03	TTGCAA	MMDRB2_F16	AAGTGT	MMDRB2_R03	CACGTA
121	MMDRB1_F09	GAACTA	MMDRB1_R04	CACGTA	MMDRB2_F09	CACAGT	MMDRB2_R04	GGTCGA
122	MMDRB1_F10	CACAGT	MMDRB1_R04	CACGTA	MMDRB2_F10	CAATCG	MMDRB2_R04	GGTCGA
123	MMDRB1_F11	CAATCG	MMDRB1_R04	CACGTA	MMDRB2_F11	CCGTCC	MMDRB2_R04	GGTCGA
124	MMDRB1_F12	CCGTCC	MMDRB1_R04	CACGTA	MMDRB2_F12	AAGACA	MMDRB2_R04	GGTCGA
125	MMDRB1_F13	AAGACA	MMDRB1_R04	CACGTA	MMDRB2_F13	GGTAAG	MMDRB2_R04	GGTCGA
126	MMDRB1_F14	GGTAAG	MMDRB1_R04	CACGTA	MMDRB2_F14	ATAATT	MMDRB2_R04	GGTCGA
127	MMDRB1_F15	ATAATT	MMDRB1_R04	CACGTA	MMDRB2_F15	CGTCAC	MMDRB2_R04	GGTCGA
128	MMDRB1_F16	CGTCAC	MMDRB1_R04	CACGTA	MMDRB2_F16	AAGTGT	MMDRB2_R04	GGTCGA
129	MMDRB1_F09	GAACTA	MMDRB1_R05	TAACAT	MMDRB2_F09	CACAGT	MMDRB2_R05	TCCAGC
130	MMDRB1_F10	CACAGT	MMDRB1_R05	TAACAT	MMDRB2_F10	CAATCG	MMDRB2_R05	TCCAGC
131	MMDRB1_F11	CAATCG	MMDRB1_R05	TAACAT	MMDRB2_F11	CCGTCC	MMDRB2_R05	TCCAGC
132	MMDRB1_F12	CCGTCC	MMDRB1_R05	TAACAT	MMDRB2_F12	AAGACA	MMDRB2_R05	TCCAGC
133	MMDRB1_F13	AAGACA	MMDRB1_R05	TAACAT	MMDRB2_F13	GGTAAG	MMDRB2_R05	TCCAGC
134	MMDRB1_F14	GGTAAG	MMDRB1_R05	TAACAT	MMDRB2_F14	ATAATT	MMDRB2_R05	TCCAGC
135	MMDRB1_F15	ATAATT	MMDRB1_R05	TAACAT	MMDRB2_F15	CGTCAC	MMDRB2_R05	TCCAGC
136	MMDRB1_F16	CGTCAC	MMDRB1_R05	TAACAT	MMDRB2_F16	AAGTGT	MMDRB2_R05	TCCAGC
137	MMDRB1_F09	GAACTA	MMDRB1_R06	GGTCGA	MMDRB2_F09	CACAGT	MMDRB2_R06	ACTTCA
138	MMDRB1_F10	CACAGT	MMDRB1_R06	GGTCGA	MMDRB2_F10	CAATCG	MMDRB2_R06	ACTTCA
139	MMDRB1_F11	CAATCG	MMDRB1_R06	GGTCGA	MMDRB2_F11	CCGTCC	MMDRB2_R06	ACTTCA
140	MMDRB1_F12	CCGTCC	MMDRB1_R06	GGTCGA	MMDRB2_F12	AAGACA	MMDRB2_R06	ACTTCA
141	MMDRB1_F13	AAGACA	MMDRB1_R06	GGTCGA	MMDRB2_F13	GGTAAG	MMDRB2_R06	ACTTCA
142	MMDRB1_F14	GGTAAG	MMDRB1_R06	GGTCGA	MMDRB2_F14	ATAATT	MMDRB2_R06	ACTTCA
143	MMDRB1_F15	ATAATT	MMDRB1_R06	GGTCGA	MMDRB2_F15	CGTCAC	MMDRB2_R06	ACTTCA
144	MMDRB1_F16	CGTCAC	MMDRB1_R06	GGTCGA	MMDRB2_F16	AAGTGT	MMDRB2_R06	ACTTCA

*Appendix 2. Methods for microsatellite genotyping, microsatellite characteristics and paternity analysis of Alpine marmots for the Nature Reserve of the Grande Sassièrè*

*Microsatellites genotyping*

Individuals were typed at 16 microsatellite loci: SSBib11, SS-Bib18, SS-Bib20, SS-Bib31, SS-Bib4 (Klinkicht 1993); MS41, MS45, MS47, MS53, MS56, MS6, ST10 (Hanslik and Kruckenhauser 2000); Ma002, Ma018, Ma066, Ma091 (Da Silva et al. 2003). To enable polymerase chain reaction (PCR) multiplexing and subsequent assessment of the allele sizes for all loci, we used primers labeled with FAM, PET, NED and VIC fluorescent dyes (FAM for SS-Bib11, MS45, Ma066 and Ma091, PET for SS-Bib18, SS-Bib20, SS-Bib4, Ma002 and Ma018, NED for SS-Bib31, MS47, MS53 and Ma002, and VIC for MS41, MS56, MS6 and ST10). PCR was carried out in three 10 µl reactions (Mix1, Mix2 and Mix3) containing 5 µl of Kit PCR (QIAGEN, Hilden, Germany) and 1 µl of DNA extract with a DNA concentration of 25-100 ng/ml. In addition, Mix1 contained 0.03 µl of each primer for MS45, 0.1 µl of each primer for SS-Bib31, MS41 and ST10, 0.2 µl of each primer for SS-Bib18 and SS-Bib4 and 0.3 µl of each primer of Ma002. Mix2 contained 0.05 µl of each primer for MS56 and MS6, 0.1 µl of each primer for MS53 and Ma091, 0.14 µl of each primer for SS-Bib11, and 0.2 µl of each primer for MS47, Ma018 and Ma066. Mix3 contained 0.2 µl of each primer for SS-Bib20.

Bib11	Bib118	Bib120	Bib131	Bib14	Ma002	Ma018	Ma066	Ma091	MS41	MS45	MS47	MS53	MS56	MS6	ST10																
Alleles	Alleles	Alleles	Alleles	Alleles	Alleles	Alleles	Alleles	Alleles	Alleles	Alleles	Alleles	Alleles	Alleles	Alleles	Alleles																
Freq	Freq	Freq	Freq	Freq	Freq	Freq	Freq	Freq	Freq	Freq	Freq	Freq	Freq	Freq	Freq																
95	0.16	132	<0.01	206	0.01	157	0.50	175	0.13	184	0.17	107	0.39	176	0.04	132	0.14	104	0.02	142	0.05	116	0.14								
97	0.22	137	0.01	208	0.20	159	0.27	188	0.17	271	0.19	298	0.75	233	0.02	167	0.09	186	0.83	109	0.50	180	0.26	140	0.44	106	0.29	158	0.88	118	0.27
101	0.43	143	0.38	216	0.38	161	0.17	190	0.66	279	0.48	241	0.36	169	0.04	111	0.11	182	0.16	142	0.42	108	0.68	160	0.07	120	0.21	120	0.21		
107	0.14	145	0.12	218	0.32	163	0.06	192	0.04	281	0.31	171	0.01	184	0.18	144	<0.01	110	0.01	130	0.05										
109	0.05	147	0.39	220	0.08	173	0.17	186	0.33	148	0.01	132	0.15	132	0.15																
	149	0.11	222	<0.01	188	0.02	190	0.01	136	0.03																					
Nind	1,032	1,031	999	1,033	1,031	1,007	1,016	1,026	1,029	1,026	1,024	1,029	1,030	1,026	1,026	1,026	1,026	1,030	1,030	1,024	1,029	1,029	1,030	1,026	1,026	1,026	1,026	1,026			

**Table 8.** Characteristics of 16 microsatellites of Alpine marmots for the Nature Reserve of The Grande Sassièrè.

Amplifications were carried out in a Mastercycler (Eppendorf, Hambourg, Germany) thermo-cycler with the following cycling conditions: 15 min at 95°C, then 28 cycles for Mix1 and Mix2 and 35 cycles for Mix3 composed of 30 s denaturing at 94°C, 90 s annealing at 57°C, 60 s extension at 72°C, and finally 30 min at 60°C to ensure complete extension. We then added 1.5 µl of Mix1 and 1.5 µl of Mix2 plus 1.5 µl of Mix3 to 0.15 µl of size standard ROX 60-415 and 10 µl of formamide. Electrophoresis was run for 3 h on an automated sequencer ABI 3130 (Applied Biosystems Inc., Foster City, CA, U.S.A.) to determine allele sizes. Microsatellite patterns were examined with Genemapper 4.0 (Applied Biosystems).

#### *Paternity analysis*

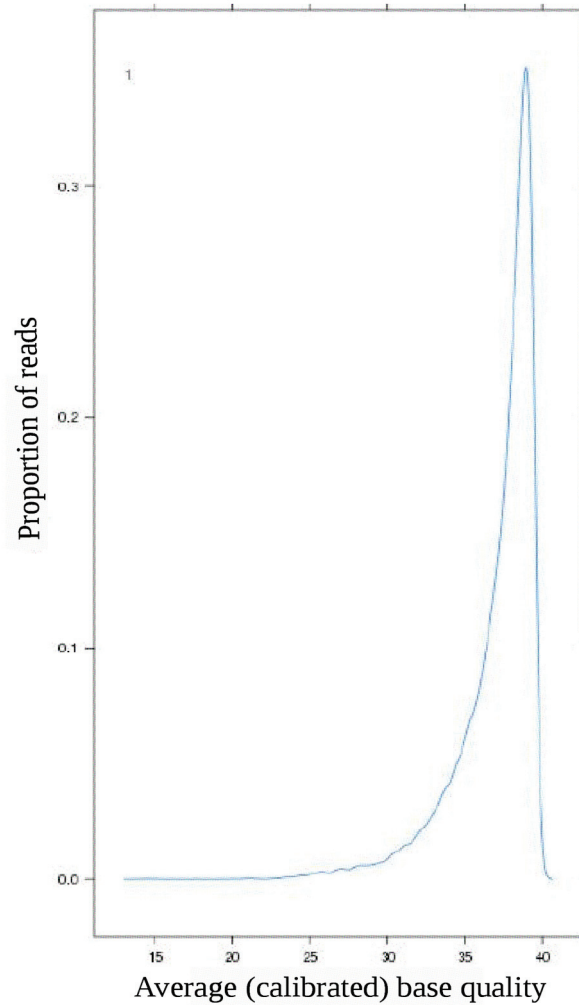
Parental relationships were confirmed thanks to genetic exclusion and parentage analyses.

Parentage analyses were performed in two ways. First, the genotypes of each pup and of the dominant pair were compared to check maternity. For 16x663 mother-pup comparisons, only 1 mismatch at one locus (SS-Bibl20) between the putative mother and one of its pups was found. We then defined pup as fathered by the dominant male if no mismatch was observed with the dominant male genotype (624 of 663 pups) and as not fathered by the dominant male if at least one mismatch was observed with the dominant male genotype (39 pups: one to seven mismatches). For thirteen pups, exclusions of paternity were based on only one mismatch with the dominant male. We nevertheless consider unlikely that these pups were fathered by the dominant male, because (1) genotyping error rate was low (probability of finding an error for one allele should not exceed 0.0003, for details see Cohas et al. 2008), (2) all these pups and their parents were retyped and their genotypes confirmed, (3) the average mutation rate for microsatellites is low ( $1.67 \times 10^{-4}$  per generation in *M. marmota*, Rassmann et al. 1994) and (4) only one mismatch with the putative mother has been found (see above). The genotypes of pups not fathered by the dominant male were then compared to the genotypes of all known sexually mature males in the family

group. The genotypes of pups not fathered by the dominant male were then compared to those of all known sexually mature males present in their family. 14 pups had genotypes compatible with that of a subordinate male in their family while 25 pups had a genotype incompatible with all subordinate males of their family.

Second, parentage analysis was repeated using the software CERVUS 3.0.3 (Kalinowski et al. 2007), with the following settings: 20 candidate fathers per offspring, 95% of candidate parents sampled, error rate of 1% to allow for mistyping and for mutations or null alleles, and assignment at the 95% confidence level. We then ran the parentage analyses with the mother identity known and all the sexually mature males present in a given year in the population as putative fathers. The previous results were confirmed except for 31 pups where paternity could be assigned to both the dominant and a subordinate male. MHC markers confirmed the dominant male to be the father of thirteen and a subordinate male to be the father of another one. In the last case, even when considering these additional markers, 17 pups could still be assigned to both the dominant and a subordinate male. However, the sexual organs of the putative subordinate father showed no sign of development at capture, and all the other pups of the litter were assigned with no ambiguity to the dominant male. Thus, we parsimoniously considered this last pup to be fathered by the dominant male. Among the 25 pups that were neither fathered by the dominant male nor by a subordinate of their family, 13 were found to be fathered by an individual born in our study population in dispersal while the other 12 were fathered by unknown males.

Appendix 3. Quality diagnostic of the 454 sequencing run



**Figure 11.** Distribution of the average base quality in each base site along the read length of all reads obtained from the 454 run with the package ShortReads (Morgan et al. 2009).





## **CHAPTER 2: MATE CHOICE FOR GENETIC CHARACTERISTICS**

---

**Genome-wide and MHC compatibility based mate choice in Alpine marmots: different targets in different contexts?**

M Ferrandiz-Rovira, D Allainé, M-P Callait-Cardinal and A Cohas

*Draft*



## Chapter 2: Mate choice for genetic characteristics

## Abstract

Sexual selection through female mate choice for genetic characteristics has been suggested to be an important evolutionary force maintaining genetic variation in animal populations. However, the genetic targets of female mate choice are not clearly identified and whether female mate choice is based on genome-wide characteristics or on specific loci remains an open question. Here, we investigated the genetic targets of female mate choice in Alpine marmots (*Marmota marmota*), a socially monogamous mammal where extra-pair paternity (EPP) occurs. We used 16 neutral microsatellites to describe genome-wide characteristics and 4 MHC loci belonging to MHC class I and II as specific genetic characteristics. Our results reveal that (1) genome-wide and specific characteristics convey different information in this species (2) social pairs show a higher MHC protein dissimilarity than expected under random choice, and (3) the occurrence of EPP increases when social pairs present a high genome-wide similarity or dissimilarity but also when they present low MHC protein dissimilarity. Thus, female mate choice is based on both genome-wide and MHC characteristics, and the genetic characteristics targeted seems to be context dependent. We emphasize the need for empirical studies of mate choice in the wild using both genome-wide and MHC characteristics because whether genome-wide and MHC characteristics convey similar information is not ubiquitous.

*Keywords:* extra-pair paternity, inbreeding avoidance, major histocompatibility complex, *Marmota marmota*, mate choice, sexual selection

## 1. INTRODUCTION

Sexual selection through female mate choice for genetic characteristics has been suggested to be an important evolutionary force maintaining or regulating genetic variation in animal populations (Jennions 1997, Tregenza and Wedell 2000, Mays and Hill 2004). Males are then supposed to be chosen based on genetic characteristics that contribute to enhance the quality of the offspring produced (Kokko et al. 2003, Mead and Arnold 2004). Two non-exclusive hypotheses involving female mate choice for genetic characteristics have been proposed: the good genes hypothesis and the genetic compatibility hypothesis. According to the good genes hypothesis, mate choice is based on an absolute criterion common to all females (Neff and Pitcher 2005), and females are expected to preferentially mate with males possessing specific alleles (good genes hypothesis *sensu stricto*) and/or with heterozygous males (good-genes-as-heterozygosity hypothesis, Brown 1997). According to the genetic compatibility hypothesis, mate choice is based on a relative criterion specific to each female (Trivers 1972, Zeh and Zeh 1996, 1997, Neff and Pitcher 2005), and females are expected to avoid mating with related or genetically similar mates. However, despite more than two decades of interest, the targets of female mate choice are not clearly identified. In particular, it is still unclear whether female mate choice is based on genome-wide characteristics or on specific loci.

Mate choice based on genome-wide characteristics can occur when female preferences translate into the production of genome-wide heterozygous offspring, because more heterozygous offspring often present a higher fitness (Crnokrak and Roff 1999, Keller and Waller 2002, Coltman and Slate 2003, Oh and Badyaev 2006, but see Kokko and Ots 2006). Females are then expected to avoid mating with homozygous (assuming this choice results in homozygous offspring)

Chapter 2: Mate choice for genetic characteristics (good genes hypothesis) and/or related partners (compatibility hypothesis), especially in those species where strong inbreeding depression occurs (Tregenza and Wedell 2000). Although a choice for dissimilar males could be advantageous for female fitness, this advantage could be, in certain cases, counterbalanced by outbreeding depression (Bateson 1983, Thornhill 1993) and a choice for intermediately dissimilar males could be favoured.

If female mate choice is based on characteristics at specific loci, the major histocompatibility complex (MHC) appears as a likely target. Indeed, the MHC is a multi-gene family present in all jawed vertebrates (Kelley et al. 2005) playing a critical role in vertebrate disease resistance by initiating immune response (Hedrick 1994). Specifically, transcript molecules from MHC class I and II genes typically recognize intracellular and extracellular pathogens respectively. Females may choose their mates to produce offspring possessing MHC alleles conferring resistance to pathogens (Takahata and Nei 1990). Associations between the presence of a specific MHC-allele and the resistance to a pathogen have been highlighted (Harf and Sommer 2005, Kloch et al. 2010, Oppelt et al. 2010, Schwensow et al. 2010, Cutrera et al. 2011). Females are then expected to prefer males possessing these specific MHC alleles (good genes hypothesis *sensu stricto*). Alternatively, females may choose their mates to produce heterozygous offspring at MHC genes. Producing MHC heterozygous offspring is advantageous since heterozygous individuals are able to recognize a wider range of pathogens (Doherty and Zinkernagel 1975). Accordingly, a negative association between heterozygosity at MHC loci and parasite load has been reported in different species (Penn et al. 2002, Froeschke and Sommer 2005, Lenz et al. 2009). Females are then expected to prefer heterozygous males at MHC genes (good-genes-as-heterozygosity) or MHC-dissimilar males (genetic compatibility hypothesis). If highly heterozygous MHC loci could be advantageous, such genetic makeup could also promote negative T-cell selection (i.e. a process removing T-cells that bind too strongly to self peptides) (Starr et al. 2003). If this process may prevent autoimmunity from occurring, negative T-cell selection could reduce the diversity of T-cell receptors (Lawlor et al. 1990,

Nowak et al. 1992) and so reduce the resistance to pathogens. To avoid such deleterious effects, a female choice for intermediately MHC-dissimilar males could be favoured. However, female preference for MHC-dissimilarity may also arise to avoid mating with related partners (Potts and Wakeland 1990, Penn and Potts 1999). A crucial assumption is then that the MHC is a marker of relatedness, an assumption thought to be true because MHC loci are among the most polymorphic loci in vertebrates, individuals that share MHC alleles are thus likely to be related (Bernatchez and Landry 2003).

Assuming that genetic characteristics at neutral markers are an indicator of genome-wide characteristics (but see Csilléry et al. 2006, Miller et al. 2014), we investigated, using microsatellite and MHC markers, whether females base their mate choice on genome-wide and/or on specific genetic characteristics in the Alpine marmot (*Marmota marmota*). Mate acquisition in Alpine marmots is a competitive process because reproduction is mostly monopolized by the dominant individuals within the family group, although extra-pair paternities (EPP) occur (Cohas et al. 2008). Whereas social mate choice might be constrained by the necessity to acquire a dominant position before reproduction, extra-pair mate choice may be a way for female Alpine marmots to adjust their mate choice and to copulate with males possessing peculiar genetic characteristics. We first investigated whether genome-wide and MHC loci convey similar information by testing if microsatellites and MHC loci characteristics were correlated. Secondly, we tested whether social and extra-pair mate choice depended on microsatellites (genome-wide) and/or MHC (specific) characteristics. For that, we investigated (i) whether microsatellites and/or MHC heterozygosity of social males and genetic similarity of social pairs differed from those expected under random mate choice, (ii) whether EPP occurrence depended on the social male's heterozygosity and/or on genetic similarity to the female measured at microsatellites and/or on MHC loci and (iii) whether extra-pair males were more heterozygous and dissimilar to the female at microsatellites and/or MHC loci than cuckolded males.

## 2. MATERIAL AND METHODS

### *Studied species*

The Alpine marmot is an hibernating ground-dwelling squirrel, living in family groups from two to 20 individuals, composed of a dominant pair, sexually mature (2 years and older) and immature (yearlings) subordinates of both sexes, and pups born that year (Allainé 2000). Within family groups, reproduction is monopolized by the dominant pair and the dominant female gives birth to a litter of 1 to 7 pups (median = 4) once a year. Dominants generally inhibit reproduction of same-sex subordinates through aggressive behaviour (Arnold and Diatami 1997, Hackländer et al. 2003), resulting in a high level of corticosteroids that limits testes maturation and spermatogenesis in subordinate males (Arnold and Diatami 1997) and results in the failure of embryo implantation and development in subordinate females (Hackländer et al. 2003). If dominant females generally successfully monopolize all reproduction, EPP occurs (Cohas et al. 2008). Lardy et al. (2012) suggested that the control of many subordinates is energetically costly for dominant males leading them to lose some reproduction and/or dominance. As a result, EPP indeed increases with the number of sexually mature subordinate males present in the family group (Cohas et al. 2006, Lardy et al. 2012). Sexually mature subordinates of both sexes can reach dominance by replacing their fathers or mothers. Otherwise, mature subordinates disperse to reach dominance mainly by replacing a dominant from another territory or rarely by establishing a new territory (only five cases observed between 1990 and 2014 in the studied population).

### *Field methods and sample collection*

Data were collected from 1990 to 2010 in a wild population of Alpine marmots located in the Nature Reserve of La Grande Sassi re (at 2,340 m a.s.l., French Alps, 45°29'N, 65°90'E). Marmots belonging to 25 family groups have been captured every year from mid-April to mid-July using two-door live traps baited with dandelions (*Taraxacum densleonis*) placed near the entrances of the main burrows in order to assign each captured individual to its family group.

Once captured, individuals were anaesthetised with Zol til 100 (0.1 mL Kg<sup>-1</sup>), sexed, aged from their size (up to 3 years), and their social status was determined according to scrotal development for males and teat development for females. All individuals were marked using a transponder and a numbered metal ear-tag placed on the right ear of females and on the left ear of males. An additional coloured plastic ear tag was placed on the opposite ear of dominant individuals. In addition, both hair samples and skin biopsies were collected on all trapped individuals for genetic analysis.

The composition of family groups was assessed from both capture-recapture data and intensive observations (see Cohan et al. 2008 for details on observations protocol). The number of subordinates of each sex and age class (pup, yearling, two-year-old and adult) was assessed for each family group and scent marking behaviour was used to confirm the identity of the dominant pair (Bel et al. 1999). Moreover, the date of pup emergence from their natal burrows and the litter size were assessed through daily observations. Virtually all emerged offspring were trapped either with smaller two-door live traps or by hand within three days of emergence.

### *Microsatellite genotyping*

1045 individuals were genotyped at sixteen microsatellites: SS-Bib1, SS-Bib18, SS-Bib20, SS-Bib31, SS-Bib4 (Klinkicht 1993); MS41, MS45, MS47, MS53, MS56, MS6, ST10 (Hanslik and Kruckenhauser 2000); Ma002, Ma018, Ma066, Ma091 (Da Silva et al. 2003) (see Appendix 1 in Chapter 2 for details on

Chapter 2: Mate choice for genetic characteristics  
genotyping methods and microsatellite characteristics).

### *MHC genotyping*

1025 individuals were genotyped at four MHC polymorphic loci - two from MHC class I exon 2 (*Mama-UB* and *Mama-UD*) and two from MHC class II DRB loci (*Mama-DRB1* and *Mama-DRB2*) - using next generation sequencing, Sanger sequencing or were deduced based on mother-father-offspring triads (see Chapter 1, Appendix 1 in Chapter 2 for details on genotyping methods and MHC characteristics).

### *Parentage analysis*

Parentage analyses were performed in two ways. First, the genotypes of each pup and of the dominant female were compared to check maternity. For 16x859 mother-pup comparisons, only 1 mismatch at one locus (SS-Bib120) between the putative mother and one of its pups was found. This mismatch likely results from a mutation in the pup since retyping of both the mother and the pup consistently yield the same genotypes and none of the two subordinate females present in the family group had a genotype compatible with this pup. We then defined a pup as a within-pair young (WPY) if no mismatch was observed with the dominant male genotype (798 of 859 pups) and as an extra-pair young (EPY) if at least one mismatch was observed with the dominant male genotype (61 of 859 pups: one to nine mismatches). For twenty-seven pups, exclusions of paternity were based on only one mismatch with the dominant male. We considered it unlikely that pups with one mismatch were fathered by the dominant male, because (1) genotyping error rate was low (0.0003, for details see Cohas et al. 2008), (2) all these offspring and their parents were retyped and their genotypes confirmed, (3) the average mutation rate for microsatellites is low ( $1.67 \times 10^{-4}$  per generation in *M. marmota*, Rassmann et al. 1994), and (4) only one mismatch with the putative mother has been found. The genotypes of the 61 pups not fathered by the dominant male were then compared to those of all known sexually mature males



Chapter 2: Mate choice for genetic characteristics present in their family. Twenty-five pups had genotypes compatible with that of a subordinate male in their family while 36 pups had a genotype incompatible with all subordinate males of their family. Among the 36 pups that were neither fathered by the dominant male nor by a subordinate of their family, 14 were found to be fathered by a dispersing individual born in our study population while the other 22 were fathered by unknown males.

Second, parentage analysis was repeated using the software CERVUS 3.0.3 (Kalinowski et al. 2007), with the following settings: 20 candidate fathers per offspring, 95% of candidate parents sampled, error rate of 1% to allow for mistyping and for mutations or null alleles, and assignment at the 95% confidence level. We then ran the parentage analyses with the mother identity known and all the sexually mature males present in a given year in the population as putative fathers. The parentage relationships of the 859 pups established by genetic exclusion were confirmed by parentage analyses with no ambiguity except for 37 pups, for which paternity could be assigned to both the dominant and a subordinate male. The comparison between MHC genotypes obtained for dominant and subordinate males (see Chapter 1, Appendix 1 in Chapter 2 for details on MHC genotyping) confirmed that 14 pups were fathered by the dominant male and another one was fathered by a subordinate male. The 22 remaining pups could still be assigned to both the dominant and a subordinate male when considering MHC markers. Even if the dominant male had a higher likelihood of being the father than the subordinate male for 11 out of these 22 pups only, the sexual organs of the putative subordinate father showed no sign of development at capture. We consequently parsimoniously considered these 22 pups to be fathered by the dominant male.

### *Genetic characteristics estimators*

Genome-wide estimators (genome-wide heterozygosity and relatedness) were calculated over the 16 microsatellite loci while MHC estimators (MHC diversity and MHC dissimilarity) were calculated over the two MHC loci, *Mama-UD* and

*Mama-DRB1*, that led to polymorphic proteins (three and eight alleles resulting in two and seven proteins respectively; *Mama-UB* and *Mama-DRB2* were not considered here since they are genetically polymorphic but do not led to polymorphic proteins). Moreover, *Mama-UD* and *Mama-DRB1* were also considered separately because in Alpine marmots MHC class I and II genes are under different selective pressures (Kuduk et al. 2012). MHC class I and II estimators were indeed not correlated except for a correlation between MHC class I and II protein dissimilarity (Appendix 2 in Chapter 2 - Table 18 and 19), suggesting that both classes could convey different information. Only individuals successfully genotyped at both MHC loci, solely at the MHC class I locus or solely at the MHC class II locus were used to calculate the MHC, MHC class I and MHC class II estimators respectively.

To test for the good genes hypothesis, both the genome-wide heterozygosity and the MHC diversity (MHC protein diversity and MHC protein distance) were estimated for each individual. The genome-wide heterozygosity was estimated as the standardized heterozygosity (SH, Coltman et al. 1999) calculated with the R function “GENHET v3.1” (Coulon 2010). Additionally, the genome-wide individual heterozygosity was estimated by the internal relatedness (IR, Amos et al. 2001) and the homozygosity by locus (HL, Aparicio et al. 2006). Since the three genome-wide individual heterozygosity estimators were highly correlated (SH vs. IR:  $\rho = -0.95$ ,  $P = < 0.001$ ; SH vs. HL:  $\rho = -0.97$ ,  $P = < 0.001$ ; IR vs. HL:  $\rho = 0.94$ ,  $P = < 0.001$ ,  $N = 95$  dominant males) and the results found were independent of the estimator used, we only present the results using the standardized heterozygosity.

Two indexes were used to estimate the MHC diversity: the MHC protein diversity and the MHC protein distance. The MHC protein diversity was estimated as the number of different MHC proteins an individual produces once the protein translated by the allele *Mama-DRB1\*08* was discarded since it has a stop codon and therefore the translated protein is non-functional (Kuduk et al. 2012). The MHC protein distance was calculated using the software MEGA version 6.06 (Tamura et al. 2013) as the sum of the p-distances (i.e. the

proportion of amino acid sites that differs between two alleles) among the considered MHC loci except for the p-distance between a given protein and the one translated from *Mama-DRB1\*08* that was set to 0.

To test for the compatible genes hypothesis, the relatedness and the MHC dissimilarity (MHC protein dissimilarity and MHC pair protein distance) were estimated for each pair of individuals. The relatedness was estimated with the Queller & Goodnight's relatedness ( $R_{qg}$ , Queller and Goodnight 1989) calculated over the 16 microsatellites using custom made scripts written in R software version 3.0.3 (R Development Core Team, 2013). Additionally, the relatedness between individuals was estimated by Lynch & Ritland's estimator ( $R_{lr}$ , Lynch and Ritland 1999) and Identity (I, Belkhir et al. 2002). Since the three relatedness estimators were highly correlated ( $R_{qg}$  vs.  $R_{lr}$ :  $\rho = 0.89$ ,  $P = < 0.001$ ;  $R_{qg}$  vs. I:  $\rho = 0.90$ ,  $P = < 0.001$ ;  $R_{lr}$  vs. I:  $\rho = 0.78$ ,  $P = < 0.001$ ;  $N = 146$  observed dominant pairs) and the results found were independent of the estimator used, we only present the results using the Queller & Goodnight's estimator as a measure of relatedness.

Finally, two indexes were used to estimate the MHC dissimilarity: the MHC protein dissimilarity and the MHC pair protein distance. The MHC protein dissimilarity was calculated as the number of proteins a male produces that differed from the ones produced by the female once the protein translated by the allele *Mama-DRB1\*08* was discarded. The MHC pair protein distance was calculated as the sum of the p-distances (i.e. the proportion of amino acid sites that differs between two alleles) averaged over all the pairwise comparisons of the proteins of a locus for a given pair. Heterozygous individuals possessing *Mama-DRB1\*08* were considered as homozygous for the remaining allele.

All MHC diversity and dissimilarity estimators were further calculated considering alleles instead of proteins (see Appendix 3 in Chapter 2 for details on the calculation of the MHC allelic diversity and dissimilarity estimators).

The different indexes used to estimate the MHC diversity and dissimilarity were found to be highly correlated, except for MHC class I allelic and protein diversity (Appendix 2, Chapter 2 - Table 20 to Table 23 and Appendix 3, Chapter

2 - Table 26 to Table 27), suggesting that these different estimators partly convey similar information.

### *Statistical analyses*

#### *Are genetic characteristics correlated?*

To test whether genome-wide and MHC estimators convey similar information, Spearman's rank correlation coefficients were calculated between standardized heterozygosity and the MHC diversity estimators of the dominant males as well as between the Queller & Goodnight's relatedness and the MHC dissimilarity estimators of the social pairs.

#### *Are social males and social pairs genetically different than expected from random mate choice?*

If there is a choice for genetic characteristics, one should expect genetic characteristics to differ between observed and candidate mates. Hence, for each reproductive event of each female (from 337 to 356 reproductive events depending on the genetic estimators considered), the genetic characteristics of observed social males (from 97 to 104 males) and of observed social pairs (from 146 to 158 pairs) were compared to the genetic characteristics of candidate males and pairs using bootstrap tests. A male was included in the pool of candidate mates for a given reproductive event of a given female if he met three criteria: (1) he was present in the population (i.e. he was observed and/or captured a given year); (2) he was at least three-year old (two-years old males rarely reach dominance: 17 out of 116 males) and (3) he resided in the same territory or in the immediate vicinity of the female (see Appendix 4 in Chapter 2 for more details on males' dispersal distances). The distributions of the difference between genetic estimators of observed *versus* candidate males and pairs were generated under the null hypothesis of random mate choice by randomly allocating, for each reproductive event of each female, a given male as the social male and one of the other males as the candidate male. This bootstrap procedure was repeated 1000

Chapter 2: Mate choice for genetic characteristics

times using the R package “boot” (Canty and Ripley 2014). In order to test whether genetic characteristics estimators of social males and social pairs differed from those expected under the hypotheses of random mate choice, the mean difference between genetic estimators of the truly observed social males and the truly observed social pairs were compared to the ones obtained from the random distributions. The exact two-tailed P-value was computed as the proportion of simulations displaying a greater mean than the observed mean value plus the proportion of cases displaying a lower mean than the symmetrical (relative to the simulated mean) of the observed mean value.

Following Huchard et al. (2013), the effect of spatial distance on the pool of candidate mates was investigated. Analyses were also conducted expanding candidate mates (1) to males residing at a maximal distance of five territories away, the maximal spatial distance recorded between the natal territory and the territory where a male established as dominant and (2) to all males residing in the studied population (see Appendix 4 in Chapter 2 for details on males' dispersal distances).

#### *Does EPP depend on genetic characteristics of social males and social pairs?*

To determine whether the EPP depended on the social males' and social pairs' genetic characteristics, the presence and the number of EPY within a litter were established for the 145 litters (produced by 83 different social pairs, 39 being present between 2 and 5 years), for which the parentage relationships were established for all pups and the number of mature male subordinates present in the family group was known.

Generalized estimating equations (GEE) (Liang and Zeger 1986) were used to account for repeated measures of social pairs. GEE were chosen since they make broader hypotheses about data structure and are better adapted to departure from normality of random effects and small sample sizes within clusters than generalized mixed models (Carlin et al. 2001, Zuur et al. 2009). An exchangeable correlation matrix was chosen to specify the same correlation between all observations of a given pair (this is analogous to the correlation structure derived

from specifying the identity of the pair as a random factor in a mixed model (Horton and Lipsitz 1999)).

Two series of GEE models were constructed using the R package “geepack” (Halekoh et al. 2006) with the presence of EPP and the number of EPY within litters as the dependent variables respectively. In the models with the presence of EPP, GEE were used with a logit link and a variance given by a binomial distribution whereas, in the models with the number of EPY, a logarithm link and a variance given by a Poisson distribution were used. In all models, litter size and the number of sexually mature male subordinates present in a given family were included as fixed effects to control for a potential confounding effect of litter size on the number of EPY in a litter and to control for the fact that EPP increases with the number of sexually mature males in Alpine marmots (Cohas et al. 2006, Lardy et al. 2012). Queller & Goodnight’s relatedness between social partners was further included as a fixed effect in all models since it affects the presence of EPY in Alpine marmots (Cohas et al. 2008). For each model, genetic estimators were then added separately. A linear effect was considered for all genetic diversity estimators while both a linear and a quadratic effects were considered for all genetic dissimilarity estimators to test for a choice for intermediately dissimilarity. Since genetic estimators were not available for all males and pairs, sample sizes vary from 134 to 140 litters in the statistical analyses.

*Are social and extra-pair males or pairs genetically different?*

If there is a choice for genetic characteristics, one could expect social males and pairs to be less diverse and/or less dissimilar or intermediately dissimilar than the respective extra-pair males and pairs. Since there are clear expectations regarding the sense of the expected differences, social males' and social pairs' genetic characteristics were compared to the ones of the respective extra-pair males and pairs using a one-tailed binomial test with the null hypothesis that the probability of success is equal to 0.5.

All statistical analyses were conducted using the R software version 3.0.3 (R

Development Core Team, 2013). All tests were two-tailed unless otherwise stated and the level of significance was set to 0.05. Analyses conducted with the MHC diversity and dissimilarity estimators calculated considering alleles instead of proteins gave qualitatively similar results, therefore only the results obtained with proteins are presented (see Appendix 3 in Chapter 2, Table 29 to Table 31 for results obtained considering alleles).

### **3. RESULTS**

#### *Correlation between genetic characteristics*

Genome-wide heterozygosity and MHC diversity were not correlated regardless of the MHC loci considered (Table 9a).

Queller & Goodnight's relatedness and the MHC dissimilarity estimators were not correlated (Table 9b), suggesting that genome-wide and MHC loci convey different information, except for a slight negative correlation (i.e. individuals more similar at microsatellites loci are more similar at MHC proteins) evidenced between the Queller & Goodnight's relatedness and the MHC protein dissimilarity calculated over the two functionally polymorphic MHC loci and over the MHC class II locus (Table 9b).

(a)	MHC protein diversity (N = 102)	MHC class I protein diversity (N = 104)	MHC class II protein diversity (N = 105)	MHC protein distance (N = 102)	MHC class I protein genetic distance (N = 104)	MHC class II protein genetic distance (N = 105)
SH	0.10 ( $P = 0.31$ )	0.03 ( $P = 0.78$ )	0.13 ( $P = 0.20$ )	0.03 ( $P = 0.78$ )	0.12 ( $P = 0.21$ )	0.10 ( $P = 0.34$ )
(b)	MHC protein dissimilarity (N = 146)	MHC class I protein dissimilarity (N = 148)	MHC class II protein dissimilarity (N = 152)	MHC pair protein distance (N = 146)	MHC class I pair protein distance (N = 148)	MHC class II pair protein distance (N = 152)
R <sub>qg</sub>	-0.19**	-0.05 ( $P = 0.56$ )	-0.19**	-0.07 ( $P = 0.42$ )	-0.04 ( $P = 0.66$ )	-0.06 ( $P = 0.44$ )

**Table 9** Spearman's correlation coefficients between standardized heterozygosity (SH) and MHC diversity estimators - MHC protein diversity and MHC protein distance - (a) and between the Queller & Goodnight's relatedness (R<sub>qg</sub>) and the MHC dissimilarity estimators - MHC protein dissimilarity and MHC pair protein distance - (b). MHC protein diversity and protein dissimilarity estimators are calculated over the two MHC loci, over the MHC class I locus and over the MHC class II locus (N = number of dominant males or pairs).



*Are social males and social pairs genetically different than expected from random mate choice?*

Social males were no more genetically diverse than expected under random mate choice, and this is true of whichever the loci considered – microsatellites or MHC –, the diversity estimators (Table 10) or the pool of candidate males considered (Appendix 4 in Chapter 2 - Table 32).

Social pairs were not less related nor more intermediately related than expected under random mate choice among the males residing in the same territory or in the immediate vicinity of a given female (Table 10). However, since Queller & Goodnight's relatedness decreased among individuals residing farther apart (Appendix 4 in Chapter 2 - Figure 15), social pairs were more related and less intermediately related than expected under random choice among the males residing farther away (Appendix 4 in Chapter 2, Figure 15 - Table 32).

Social pairs showed a higher protein dissimilarity and a higher pair protein distance at the two functionally polymorphic MHC loci and at the MHC class II locus than expected under random choice among the males residing in the same territory or in the immediate vicinity of a given female (Table 10). However, given that dissimilarity increased among individuals residing farther apart (Appendix 4 in Chapter 2 - Figure 15), social pairs showed a MHC dissimilarity consistent with random choice among the males residing farther away (Appendix 4 in Chapter 2 - Table 32).

Genetic characteristics	Observed mean	Observed difference	Simulated difference [CI 95%]	p-value
SH	0.95	-0.02	-0.0002 [-0.02 , 0.02]	0.09
MHC protein diversity	2.91	0.01	0.003 [-0.06 , 0.06]	0.62
MHC class I protein diversity	1.15	0.01	0.01 [-0.03 , 0.04]	0.75
MHC class II protein diversity	1.75	0.01	-0.003 [-0.05 , 0.04]	0.86
MHC protein distance	0.05	-0.0004	-0.001 [-0.01 , 0.003]	0.99
MHC class I protein distance	0.003	<0.0001	0.0002 [-0.0004 , 0.001]	0.75
MHC class II protein distance	0.05	-0.001	-0.001 [-0.01 , 0.003]	0.99
$R_{qg}$	0.10	0.01	0.01 [-0.02 , 0.03]	0.98
$R_{qg}-R_{qg\_intermediate}$	0.20	-0.001	0.003 [-0.02 , 0.03]	0.99
MHC protein dissimilarity	1.50	<b>0.18</b>	<b>-0.03 [-0.05 , 0.11]</b>	<b>&lt;0.0001</b>
MHC class I protein dissimilarity	0.20	0.01	0.01 [-0.02 , 0.05]	0.50
MHC class II protein dissimilarity	1.31	<b>0.17</b>	<b>&lt;0.0001 [-0.06 , 0.07]</b>	<b>&lt;0.0001</b>
MHC pair protein distance	0.02	<b>0.001</b>	<b>&lt;0.0001 [-0.001 , 0.001]</b>	<b>0.02</b>
MHC class I pair protein distance	0.002	0.0002	0.0002 [-0.0001 , 0.0004]	0.49
MHC class II pair protein distance	0.04	<b>0.002</b>	<b>-0.0004 [-0.002 , 0.001]</b>	<b>0.04</b>

**Table 10.** Genetic characteristics of real social males and pairs compared to the potential candidate males and pairs residing in the same territory or in the immediate vicinity of a given female. MHC diversity and dissimilarity estimators were calculated over the two MHC loci, over the MHC class I locus and over the MHC class II locus. SH: standardized heterozygosity;  $R_{qg}$ : Queller & Goodnight's relatedness.  $R_{qg}-R_{qg\_intermediate}$ : absolute difference between the observed value of the relatedness and the intermediate value of the Queller & Goodnight's relatedness (0.17) obtained from the generalized estimating equation model of the number of EPY within litters. P-values <0.05 are indicated in bold.

### *EPP pattern*

A total of 522 pups from 145 different litters were retained from families where the number of sexually mature male subordinates (mean  $\pm$  SD =  $0.83 \pm 1.23$ ; range: 0-6) and all pups of the year were known. The mean number of pups within a litter was  $3.60 \pm 1.23$  ( $\pm$  SD) (range: 1 to 7). From a total of 522 pups, 38 (7.28%) were found to be EPY and from a total of 145 examined litters, 20 (13.79%) contained at least one EPY. From these 20 litters, 9 (45%), 6 (30%), 4 (20%) and 1 (5%) contained one, two, three and five EPY respectively. Four out of 20 litters were sired only by the extra-pair male whereas 16 were mixed litters (pups sired by both the social and an extra-pair male). The mean proportion of

EPY within the mixed litters was  $0.42 \pm 0.16$  ( $\pm$  SD) (range: 0.20 to 0.75). From the total of 20 litters that contained at least one EPY, 45% (9 litters) were fathered by a subordinate male of their family, 15% (3 litters) were fathered by a dispersing individual born in our study population and the remaining 40% (8 litters) were fathered by unknown males.

### *Does EPP depend on genetic characteristics of social males and social pairs?*

As expected, the presence and the number of EPY within litters increased with the number of sexually mature male subordinates present in the family group (presence of EPY: estimate  $\pm$  SE =  $0.47 \pm 0.14$ ,  $P = 0.001$ ; number of EPY: estimate  $\pm$  SE =  $0.27 \pm 0.13$ ,  $P = 0.05$ ,  $N = 140$ ) but did not depend on litter size (presence of EPY: estimate  $\pm$  SE =  $0.08 \pm 0.17$ ,  $P = 0.65$ ; number of EPY: estimate  $\pm$  SE =  $0.12 \pm 0.14$ ,  $P = 0.37$ ,  $N = 140$ ).

Neither the presence nor the number of EPY within litters depended on the social male's diversity, whatever the loci considered – microsatellites or MHC – or the diversity estimators considered (Table 11).

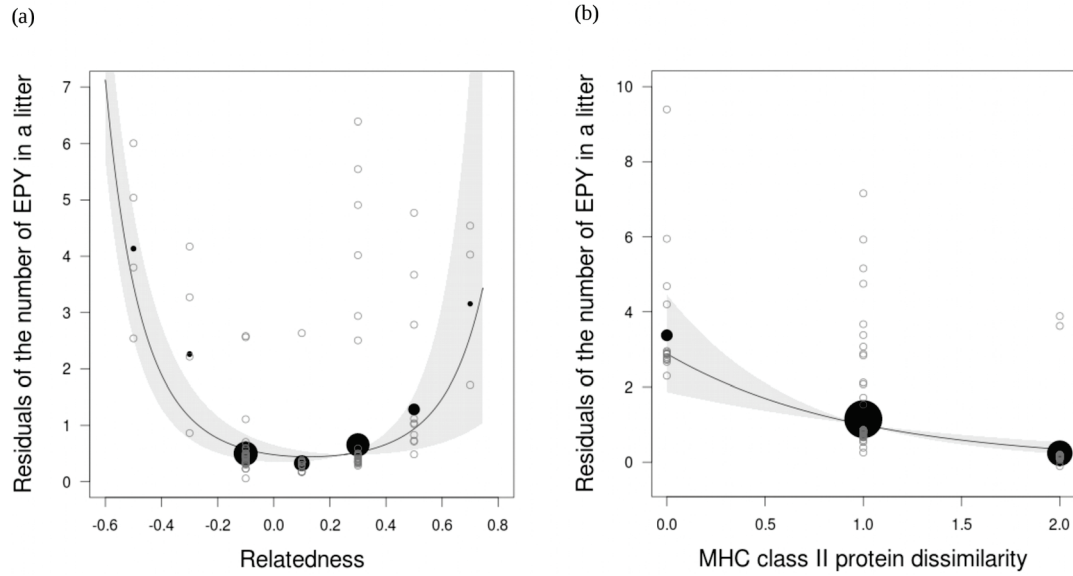
The presence and the number of EPY within litters depended on a quadratic relationship with the social pair Queller & Goodnight's relatedness (Table 11, Figure 12a), both being lower for an intermediate Queller & Goodnight's relatedness of respectively 0.14 and 0.17. In addition, the number of EPY within litters decreased linearly when the social pair MHC protein dissimilarity, measured at the two functionally polymorphic MHC loci and at the functionally polymorphic MHC class II locus only, increased (Table 11, Figure 12b). Interestingly, when the MHC dissimilarity was further calculated considering alleles instead of proteins, the same pattern was evidenced with the MHC allelic dissimilarity, but only when the MHC allelic dissimilarity was estimated on the two MHC loci leading to polymorphic proteins (*Mama-UD* and *Mama-DRB1*) or on the *Mama-DRB1* locus (Appendix 3 in Chapter 2 – Table 30b). However, neither the presence nor the number of EPY within litters depended on the social

Chapter 2: Mate choice for genetic characteristics

pair's MHC pair protein distance, regardless of the MHC loci considered (Table 11).

Independent variable	Presence of EPY		Number of EPY		N
	estimate $\pm$ SE	p-value	estimate $\pm$ SE	p-value	
SH	1.47 $\pm$ 1.49	0.32	1.11 $\pm$ 1.09	0.30	140
MHC protein diversity	0.50 $\pm$ 0.63	0.42	0.75 $\pm$ 0.48	0.12	134
MHC class I protein diversity	0.44 $\pm$ 0.92	0.63	0.64 $\pm$ 0.75	0.39	135
MHC class II protein diversity	0.65 $\pm$ 0.79	0.41	1.02 $\pm$ 0.67	0.13	139
MHC protein distance	13.32 $\pm$ 9.90	0.18	14.43 $\pm$ 7.62	0.06	134
MHC class I protein distance	26.08 $\pm$ 54.35	0.63	37.93 $\pm$ 44.09	0.39	135
MHC class II protein distance	8.94 $\pm$ 8.71	0.30	4.92 $\pm$ 8.57	0.57	139
$R_{qg}$	-1.73 $\pm$ 0.95	0.07	-1.29 $\pm$ 0.83	0.12	140
$(R_{qg})^2$	<b>8.40 <math>\pm</math> 2.74</b>	<b>0.002</b>	<b>5.30 <math>\pm</math> 1.64</b>	<b>0.001</b>	140
MHC protein dissimilarity	-0.53 $\pm$ 0.34	0.12	<b>-0.56 <math>\pm</math> 0.22</b>	<b>0.01</b>	134
MHC class I protein dissimilarity	-0.19 $\pm$ 0.89	0.83	-0.69 $\pm$ 0.85	0.41	135
MHC class II protein dissimilarity	-0.86 $\pm$ 0.46	0.06	<b>-1.05 <math>\pm</math> 0.46</b>	<b>0.02</b>	138
MHC pair protein genetic distance	-12.82 $\pm$ 20.53	0.53	-9.94 $\pm$ 17.54	0.57	134
MHC class I pair protein genetic distance	-9.68 $\pm$ 93.06	0.92	-35.27 $\pm$ 88.93	0.69	135
MHC class II pair protein genetic distance	-10.97 $\pm$ 10.98	0.32	-16.53 $\pm$ 14.76	0.26	138

**Table 11.** Generalized estimating equation models showing the effects of the genetic characteristics of the real social males and pairs on both the presence and the number of EPY. The MHC diversity and dissimilarity estimators were calculated over the two MHC loci, over the MHC class I locus and over the MHC class II locus. All models include the number of sexually mature male subordinates present in a given family, the litter size, the social pair Queller & Goodnight's relatedness and its associated quadratic term. P-values  $<0.05$  are indicated in bold.



**Figure 12.** Residual of the number of EPY in a litter as a function of Queller & Goodnight's relatedness between social pairs (a) and as a function of MHC class II protein dissimilarity between social pairs (b). Residuals are corrected for confounding factors (the number of sexually mature male subordinates present in a given family at the figure 12a and for the number of sexually mature male subordinates present in a given family, the social pair Queller & Goodnight's relatedness and its associated quadratic term at the figure 12b). Open circles represent the observed residual number of EPY in a litter. Thick lines represent predictions of the model and the grey surface represents standard errors of the fitted.

*Are social and extra-pair males or pairs genetically different?*

Social males were no more diverse than extra-pair males, no more dissimilar or no more intermediately dissimilar than extra-pair pairs regardless of the loci considered (Table 12).

Genetic characteristics	Nbr. social < extra-pair	p-value	N
SH	11	0.17	17
MHC protein diversity	1	1.00	14
MHC class I protein diversity	0	1.00	14
MHC class II protein diversity	1	1.00	15
MHC protein distance	2	0.99	14
MHC class I protein distance	0	1.00	14
MHC class II protein distance	3	0.99	15
$R_{qg}$	7*	0.83	17
$R_{qg}-R_{qg\_intermediate}$	7*	0.83	17
MHC protein dissimilarity	1	1.00	14
MHC class I protein dissimilarity	0	1.00	14
MHC class II protein dissimilarity	2	0.99	14
MHC pair protein distance	1	1.00	14
MHC class I pair protein distance	0	1.00	14
MHC class II pair protein distance	2	0.99	14

**Table 12.** Genetic characteristics of social males and pairs compared to the corresponding extra-pair males and pairs. The MHC diversity and dissimilarity estimators were calculated over the two MHC loci, over the MHC class I locus and over the MHC class II locus. SH: standardized heterozygosity;  $R_{qg}$ : Queller & Goodnight's relatedness;  $R_{qg}-R_{qg\_intermediate}$ : absolute difference between the observed and the intermediate value of Queller & Goodnight's relatedness (0.17) obtained from the generalized estimating equation model of the number of EPY within litters. Nbr. (number) of social pairs with Queller & Goodnight's relatedness superior to the one of the corresponding extra-pair.

#### 4. DISCUSSION

While a choice based on genome-wide characteristics have now been extensively

studied in wild populations, the existence of a choice based on MHC loci have still been seldom studied under natural conditions. In Alpine marmots, the relatedness of social pairs conformed with random mate choice but their MHC protein dissimilarity was higher than expected under random choice. EPPs were rare for intermediate relatedness and for high MHC protein dissimilarity between social pairs' members. However, no difference between the genetic characteristics of social and extra-pair males was evidenced. Our results thus provide the first evidence that female mate choice in the Alpine marmot is based on both genome-wide and MHC characteristics. These results are in agreement with the genetic compatibility hypothesis rather than with the good genes hypothesis. Moreover, in this species, genome-wide characteristics and MHC characteristics convey different information and mate choice for these genetic characteristics seems to be context dependent.

We found no evidence of a choice for good genes, mate choice being based neither on male's genome-wide heterozygosity nor on male's MHC diversity. Instead, we found a choice for compatible genes. Specifically, mate choice was based on both genome-wide and MHC compatibility. Previous studies based either on genome-wide characteristics (reviewed in Mays et al. 2008 for birds) or on MHC characteristics (e.g. birds: Ekblom et al. 2004, Westerdahl 2004, Richardson et al. 2005, Bollmer et al. 2012; primates: reviewed in Setchell and Huchard 2010) have found evidence for a choice for good genes, compatible genes, both or none. These contrasted empirical results suggest that the species' ecological and biological characteristics might play a significant role on the targeted genes (Setchell and Huchard 2010). In particular, a choice for compatible genes has been proposed to be favoured in inbred populations (Bernatchez and Landry 2003, Setchell and Huchard 2010, but see Bichet et al. 2014). Alpine marmots are characterized by a low genome-wide variability as estimated with allozymes (Preleuthner and Pinsker 1993), minisatellites (Rassmann et al. 1994, Kruckenhauser et al. 1997) and microsatellites (Cohas et al. 2009, but see Goossens et al. 2001) and by a low MHC variability (Kuduk et al. 2012, Chapter 2). The social organization and the dispersal patterns of this



species increase the risk for a female to mate with a related male (see Appendix 4 in Chapter 2 for detailed information regarding dispersal patterns in the studied population) and the costs of such mating are expected to be high since juvenile survival decreases with offspring homozygosity (Cohas et al. 2009). Therefore, choosing a mate with compatible genetic characteristics could be a way for female Alpine marmots to avoid such costs.

Social mate choice was based on MHC characteristics (protein dissimilarity) but not on genome-wide characteristics (relatedness). The population genetic structure is strongly influenced by dispersal patterns, females of species with low dispersal ranges are more prone to encounter related mates than females of species showing longer dispersal ranges (Perrin and Mazalov 2000). In Alpine marmots, more than 48% of individuals from both sexes acquire dominance in their natal territory or in the immediate vicinity resulting in most of the available social males being genetically similar to females (see Appendix 4 in Chapter 2 for detailed information regarding dispersal patterns – Figure 14 - and spatial distribution of genetic characteristics – Figure 15 - in the studied population). Thus, female Alpine marmots have reduced opportunity to avoid inbreeding when choosing a social partner. On the contrary, the occurrence of EPP depended on both genome-wide and MHC protein dissimilarity of the social pair, suggesting that female Alpine marmots choose to cuckold their social males to promote optimal outbreeding and offspring MHC-heterozygosity. If true, the genetic characteristics of the pools of social males and of potential extra-pair males are expected to differ, with the genetic similarity of the potential extra-pair males to the female being lower. Females may thus have more opportunities to choose extra-pair mates according to both genome-wide and MHC characteristics. The absence of statistical difference found between social and extra-pair mates found in this study is probably due to a bias regarding the identified extra-pair males. Indeed, among the seventeen identified extra-pair males, twelve of them were subordinates of the social-pair family and five were dispersing individuals born in another family group originating from our study area. But, half of extra-pair males were not identified and thus probably came

from outside the study area.

Our results indicated that females based both their social mate choice and their decision to cuckold their mates on MHC characteristics. Parasite pressures that individuals face could modulate the fitness pay-offs of mate choice. Females may then choose MHC compatible mates to produce MHC-heterozygous offspring (Takahata and Nei 1990). Producing MHC-heterozygous offspring is advantageous since heterozygous individuals are able to recognize a wider range of pathogens (Doherty and Zinkernagel 1975). In accordance, MHC diversity has been found to be negatively associated with parasite load (Penn et al. 2002, Froeschke and Sommer 2005, Lenz et al. 2009). The strong positive selection detected at the antigen binding sites of the MHC class II but not of the MHC class I alleles of Alpine marmots suggests that strong selective pressures are exerted by extracellular parasites (Kuduk et al. 2012). In agreement, we found evidence of female mate choice based on MHC class II but not for MHC class I genes. Similarly, the low parasite pressures encountered in insular populations has been hypothesized to be at the origin of the absence of mate choice for MHC genes in an insular population of house sparrows (*Passer domesticus*) (Bichet et al. 2014).

We found that genome-wide and MHC characteristics were not correlated indicating that genome-wide and MHC loci convey different information. Therefore, when Alpine marmots choose for both genome-wide and MHC characteristics they may target different genetic characteristics in order to achieve different genetic benefits in terms of fitness of the produced offspring. Alpine marmots also have the possibility to target distinct characteristics in different contexts, and they do so when choosing social mates and cuckolding them. Indeed, female Alpine marmots choose their social males for different genetic characteristics than they cuckolded them. Similarly, in two primate species where genome-wide and MHC loci were found not to be correlated, the occurrence of EPP depended on MHC characteristics but not on genome-wide characteristics in fat-tailed dwarf lemurs (*Cheirogaleus medius*) (Schwensow et al. 2008a), whereas evidence that social males are chosen based on both genome-wide and

MHC genetic characteristics was found in grey mouse lemurs (*Microcebus murinus*) (Schwensow et al. 2008b). Also, in scarlet rosefinches (*Carpodacus erythrinus*) genome-wide and MHC loci were not correlated and extra-pair males are chosen based genome-wide characteristics but not on MHC characteristics (Winternitz et al. 2015). However, the possibility to choose separately for genome-wide or MHC characteristics (or for both) may not exist in all species. A correlation between genome-wide and MHC loci leads to females apparently targeting both genetic characteristics (such a correlation is not ubiquitous, see Table 5). However, in such cases one cannot exclude the apparent choice for a genetic characteristic to be a by-product of a choice for the other one (Piertney and Oliver 2006, Kempenaers 2007, Huchard and Pechouskova 2014). For example, in mandrills (*Mandrillus sphinx*), females mate according to both genome-wide and MHC characteristics (Setchell et al. 2010). However, because genome-wide and MHC characteristics are correlated in this species, one cannot exclude that evidence for MHC-dependent female mate choice is a by-product of a genome-wide inbreeding avoidance. Indeed, MHC loci being among the most polymorphic loci in vertebrates, individuals that share MHC alleles are likely to be related (Bernatchez and Landry 2003) and it has been argued that females may avoid inbreeding through MHC-based female choice (Potts and Wakeland 1990, Penn and Potts 1999). To better understand the evolutionary causes and consequences of female mate choice, these considerations advocate for the need to not only combine both genome-wide and MHC characteristics (and not to consider a single target) but also to examine the correlation between them.

Taxa	Scientific name	Type of choice	Genome-wide markers		MHC markers		Correlation between genetic markers	References
			Good genes	Compatible genes	Good genes	Compatible genes		
Reptile	<i>Sphenodon punctatus</i>	social vs. random	na	±	na	±	na	Miller et al. 2009
		mated vs. non mated	=	na	=	na		
Bird	<i>Acrocephalus sechellensis</i>	social vs. random	na	na	=	=		Richardson et al. 2005
		occurrence of EPP	na	=	+	=	yes	
		social vs. extra-pair	na	na	+	=		
		social vs. random	na	=	±*	=		
	<i>Geothlypis trichas</i>	occurrence of EPP	na	=	=	±§	na	Bollmer et al. 2012
		social vs. extra-pair	na	=	=	=		
		social vs. random	na	=	na	=		
		extra-pair vs. random	na	=	na	=		
	<i>Passer domesticus</i>	occurrence of EPP	na	=	na	na		Bichet et al. 2014
		social vs. extra-pair	na	+	na	+		
		social vs. random	=	=	=	=		
		genetic vs. random	+	=	=	=	no	
	<i>Carpodacus erythrinus</i>	occurrence of EPP	+	=	=	=		Winternitz et al. 2015
		social vs. extra-pair	=	=	=	=		
		social vs. random	+	=	+	+	no	
		social vs. random	na	+	=	±#	yes	
Mammal	<i>Microcebus murinus</i>	social vs. random	+	=	+	+	no	Schwensow et al. 2008b
		social vs. random	na	=	=	±#	Huchard et al. 2013	
	<i>Cheirogaleus medius</i>	social vs. random	+	=	+	+		Schwensow et al. 2008a
		occurrence of EPP	=	=	=	=	no	
		social vs. extra-pair	=	=	=	=		
		reproductor vs. random	±	±	+	+	yes	
	<i>Mandrillus sphinx</i>	social vs. random	=	=	=	±§		Setchell et al. 2010
		social vs. random	=	=	=	+		
	<i>Marmota marmota</i>	occurrence of EPP	=	+	=	±§	no	Present study
		social vs. extra-pair	=	=	=	=		

**Table 13.** Summary of empirical studies carried out in the wild examining female mate choice using both genome-wide and MHC genetic markers. na: not applicable or not studied; +: Evidence of a female mate choice for the studied genetic marker; =: No evidence of a female mate choice for the studied genetic marker; ±: Evidence of a female mate choice mixed; \*: Evidence of a female mate choice for MHC class I alleles and no evidence for MHC class II alleles; #: Evidence of a female mate choice for MHC class II alleles and no evidence for MHC class I alleles; §: Evidence of a female mate choice for MHC DRB locus and no evidence for MHC DQB locus.

### *Conclusions*

Our results suggest that female Alpine marmots base their mate choice on both genome-wide and MHC loci. This choice seems to be context-dependent as indicated by the fact that the genes involved in social mate choice and in the occurrence of EPP differ. Population genetic structure, dispersal patterns and parasite pressures may play a key role in female mating decisions. Whether females actively choose for genetic characteristics or whether this choice is a by-product of the geographic origins of males remains an open question. In the case of an assessment of genetic characteristics, olfactory cues are likely to be involved. For example, characteristics of chemical compounds have been proven to depend on the genetic characteristics of wild or semi-free-ranging individuals (e.g. genome-wide loci: in ring-tailed lemurs (*Lemur catta*) (Charpentier et al. 2008) and in black-legged kittiwakes (*Rissa tridactyla*) (Leclaire et al. 2012); or MHC loci: in mandrills (Setchell et al. 2011)). Exploration of the association between genetic characteristics at either genome-wide or MHC loci, smell characteristics, and mating preferences in wild populations is a promising avenue of research that awaits further investigation.

## **5. APPENDIX**

### *Appendix 1. Molecular analyses and genetic markers characteristics*

#### *DNA extraction*

For 1045 individuals, genomic DNA was extracted from 15 to 30 hairs or skin biopsies by incubation at 66 °C for 80 min for hairs and at 56°C for 120 min for tissue in 50 µL lysis buffer (20 mM Tris-HCl, 1.5 mM MgCl<sub>2</sub>, 25 mM KCl, 0.5% Tween 20, and 0.1 mg/ml proteinase K), followed by 20 min of proteinase K inactivation at 96°C.

*Microsatellites genotyping*

Individuals were typed at 16 microsatellite loci: SS-Bib1, SS-Bib18, SS-Bib20, SS-SS-Bib31, SS-Bib4 (Klinkicht 1993); MS41, MS45, MS47, MS53, MS56, MS6, ST10 (Hanslik and Kruckenhauser 2000); Ma002, Ma018, Ma066, Ma091 (Da Silva et al. 2003) (see Table 14 for primers design).

Locus name	Forward primer (sequence 5' – 3')	Reverse primer (sequence 5' – 3')
SS-Bib1	CTGAAGCAGCCATCCAGTA	TGGTGTGCCATTGTTCT
SS-Bib18	ATGGTCATGGAAGGGAAG	GGCATCTTCACAGTTGATC
SS-Bib20	ATTCTCTAGTCGTTAACAAGAATC	CACCAGTGAAACTACATACAGTG
SS-Bib31	TTACACCTTCTCTGGCTCC	TCTGAGCGGATTGTCTTTAT
SS-Bib4	CCTAGGTTTCAGTCTTCAACACA	TGGTGTGCCATTGTTCT
Ma002	CATTTAGACGCACATTTTG	GGGATGGAGAATGAGGAAG
Ma018	ATCCGTCCAATAAAGAAATTC	GTTTCTTGTGGCTCAGTGGTCAGATG
Ma066	AATATGTTAAGGCAGTTCTAGC	GTTTCTTCCTGATATGGAAAGATGATGT
Ma091	CCTGTGTGAGTCCTGGAGTC	AGCCATTTAGGTTACATCTGC
MS41	GGTGTATATGGGAATAGGGGG	GCCTTCAAATCAAAGCAGGTTG
MS45	CTGTCTCTTTGTCCCTGCC	CTCCTTACCATCATCTTTCCG
MS47	CCTGATGTAGTCAGTCAG	TGTGGGAAATGGCACATC
MS53	ATTGAGGAGCAGCATCTAGG	TCAGGGAAAGGCAGACCTG
MS56	CAGACTCCCACCAGTGACC	CCTGATCTATGTAGGTTCCAT
MS6	CTGATGGGGTTAAGATTGCC	CCCCACTGACCCACCTCC
ST10	TTGTGATCCTCCAGGGAGTT	GTGATTTCCAAACCCCATTC
<i>Mama-UB</i>	AyCTCCGTGTCCCGGCC	GCAGGTTGCTCAGGCCCACTT
<i>Mama-UD</i>	AyCTCCGTGTCCCGGCC	GCGCAGGGTGTTC AAGCACAT
<i>Mama-DRB1</i>	GAGTGTCAATTTCTCCAACsrGA	TyAmCTCTCCkCTCCACAGTGAA
<i>Mama-DRB2</i>	ACGATTCCTGCAGCAGATGA	TTCGGCTTAACTCTCCGCTTTGA

**Table 14.** Primer pairs used for microsatellite and MHC genotyping.

To enable polymerase chain reaction (PCR) multiplexing and subsequent assessment of the allele sizes for all loci, we used primers labeled with FAM, PET, NED and VIC fluorescent dyes (FAM for SS-Bib1, MS45, Ma066 and Ma091, PET for SS-Bib18, SS-Bib20, SS-Bib4, Ma002 and Ma018, NED for SS-Bib31, MS47, MS53 and Ma002, and VIC for MS41, MS56, MS6 and ST10). PCR was carried out in three 10 µl reactions (Mix1, Mix2 and Mix3)

## Chapter 2: Mate choice for genetic characteristics

containing 5  $\mu$ l of Kit PCR (QIAGEN, Hilden, Germany) and 1  $\mu$ l of DNA extract with a DNA concentration of 25-100 ng/ml. In addition, Mix1 contained 0.03  $\mu$ l of each primer for MS45, 0.1  $\mu$ l of each primer for SS-Bib131, MS41 and ST10, 0.2  $\mu$ l of each primer for SS-Bib118 and SS-Bib14 and 0.3  $\mu$ l of each primer of Ma002. Mix2 contained 0.05  $\mu$ l of each primer for MS56 and MS6, 0.1  $\mu$ l of each primer for MS53 and Ma091, 0.14  $\mu$ l of each primer for SS-Bib11, and 0.2  $\mu$ l of each primer for MS47, Ma018 and Ma066. Mix3 contained 0.2  $\mu$ l of each primer for SS-Bib120. Amplifications were carried out in a Mastercycler (Eppendorf, Hambourg, Germany) thermo-cycler with the following cycling conditions: 15 min at 95°C, then 28 cycles for Mix1 and Mix2 and 35 cycles for Mix3 composed of 30 s denaturing at 94°C, 90 s annealing at 57°C, 60 s extension at 72°C, and finally 30 min at 60°C to ensure complete extension. We then added 1.5  $\mu$ l of Mix1 and 1.5  $\mu$ l of Mix2 plus 1.5  $\mu$ l of Mix3 to 0.15  $\mu$ l of size standard ROX 60-415 and 10  $\mu$ l of formamide. Electrophoresis was run for 3 h on an automated sequencer ABI 3130 (Applied Biosystems Inc., Foster City, CA, U.S.A.) to determine allele sizes. Microsatellite patterns were examined with Genemapper 4.0 (Applied Biosystems). Depending on the locus, 97-100% of individuals were genotyped and details on microsatellites characteristics are summarized in Table 15.



SS-Bib11		SS-Bib18		SS-Bib20		SS-Bib31		SS-Bib14		Ma002		Ma018		Ma066		Ma091		MS41		MS45		MS47		MS53		MS56		MS6		ST10							
Alleles	Freq	Alleles	Freq	Alleles	Freq	Alleles	Freq	Alleles	Freq	Alleles	Freq	Alleles	Freq	Alleles	Freq	Alleles	Freq	Alleles	Freq	Alleles	Freq	Alleles	Freq	Alleles	Freq	Alleles	Freq	Alleles	Freq	Alleles	Freq						
95	0.15	132	<0.01	206	0.01	157	0.50	175	0.13	265	<0.01	296	0.25	231	0.63	159	0.13	184	0.17	107	0.40	176	0.03	132	0.13	104	0.02	142	0.05	116	0.15						
97	0.22	137	0.01	208	0.19	159	0.27	178	<0.01	271	0.19	298	0.75	233	0.02	167	0.09	186	0.83	109	0.49	180	0.26	140	0.44	106	0.30	158	0.88	118	0.26						
101	0.44	143	0.36	216	0.39	161	0.18	188	0.17	279	0.50			241	0.35	169	0.04			111	0.11	182	0.17	142	0.42	108	0.68	160	0.07	120	0.22						
103	<0.01	145	0.13	218	0.33	163	0.05	190	0.66	281	0.31					171	0.01					184	0.18	144	<0.01	110	<0.01	130	0.05								
107	0.15	147	0.39	220	0.08			192	0.04	283	<0.01					173	0.17					186	0.33	148	0.01			132	0.15								
109	0.04	149	0.11	222	<0.01											175	0.44					188	0.02					134	0.14								
																177	0.02					190	0.01					136	0.03								
																179	0.09																				
																188	<0.01																				
Nind	1045	1042	1042	1010	1043	1041	1015	1025	1035	1038	1032	1040	1033	1037	1038	1033	1033	1033	1038	1037	1033	1033	1033	1038	1037	1033	1033	1033	1033	1031	1031	1031	1031	1031	1031	1031	1031

**Table 15.** Characteristics of 16 microsatellites of Alpine marmots. Nind: number of individuals.

Hardy-Weinberg equilibrium was tested using “HWE.test” implemented on the library “genetics” (Warnes 2012) of the R version 3.1.1 (R Development Core Team 2013). These tests were performed on 200 randomized individuals to avoid potential bias caused by family structure and on all cohorts pooled to ensure sufficient sample size. None of the loci showed deviation from Hardy–Weinberg equilibrium after Bonferroni correction ( $P < 0.05/16 = 0.003$ ), except for Ma002 ( $\chi^2 = 33.69$ ,  $P < 0.0001$ , 10000 replicates).

### *MHC genotyping*

Individuals were genotyped at four MHC loci, two from MHC class I exon 2 (*Mama-UB* and *Mama-UD*) and two from MHC class II DRB loci (*Mama-DRB1* and *Mama-DRB2*). Three different methods were used to genotype individuals: next generation sequencing, Sanger sequencing or were deduced based on mother-father-offspring triads.

### Next generation sequencing

Two runs of a Roche® 454 FLX sequencing instrument were done. Methods used to genotype individuals at four MHC loci have been described in details elsewhere (Chapter 1) and are briefly summarized here.

Amplification of MHC loci used primers specifically designed to separate each locus and to avoid sequencing several duplicated genes of Alpine marmots (Table 14). A 6-bp tag was added between the adapters and the locus-specific reverse and forward primers to barcode individuals. A minimum of 3-bp differences between tags were used to ensure a very low probability for a read to be assigned to the wrong individual due to a typing mistake in the individual tag (Table 16 for details on tag library). PCR was performed in a Mastercycler (Eppendorf, Hamburg, Germany) in 10  $\mu\text{L}$  of reaction mixture containing 5  $\mu\text{L}$  HotStarTaq Polymerase Master Mix (QIAGEN, Hilden, Germany), 0.2  $\mu\text{L}$  of both primers at 100  $\mu\text{M}$ , 3.6  $\mu\text{L}$  water and 1  $\mu\text{L}$  DNA at a concentration of 30  $\text{ng}\cdot\mu\text{L}^{-1}$ . The cycling scheme was 95°C for 15 min, followed by 34 cycles at 95°C for 30 s, primer-specific annealing temperature (50°C for *Mama-UD* and *Mama-*

*DRB2* and 55°C for *Mama-UB* and *Mama-DRB1*) and for 30 s, 72°C for 60 s and a final extension step at 72°C for 10 min. The concentration of the PCR products was measured by fluorometry using the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen). Equimolar amounts of amplicons were pooled for a given locus and purified using MinElute PCR Purification Kit (QIAGEN) and the four loci were subsequently pooled and sequenced on an eight of a PicoTiterPlate of a Roche® 454 FLX sequencing instrument (576 amplicons/eight).

Group	Forward tag	Reverse tag	
1	AACCGA	TTGAGT	
	AAGTGT	AAGCAG	
	AGTGTT	TTGCAA	
	CCGCTG	CACGTA	
	AACGCG	TAACAT	
	GGCTAC	GGTCGA	
	TTCTCG	CTTGGT	
	TCACTC	TCCAGC	
		ACTTCA	
		GCGAGA	
		TGGAAC	
		CGAATC	
	2	GAACTA	TTGAGT
		CACAGT	AAGCAG
CAATCG		TTGCAA	
CCGTCC		CACGTA	
AAGACA		TAACAT	
GGTAAG		GGTCGA	
ATAATT			
CGTCAC			

**Table 16.** Tags used to barcode individuals for next generation sequencing. All combinations of a forward and a reverse tags within each of the two groups were used.

Although next generation sequencing is a suitable method for large-scale genotyping and generates multiple reads for a given amplicon, it is also prone to errors (Glenn 2011). Discerning true alleles versus sequencing errors is thus

Chapter 2: Mate choice for genetic characteristics challenging (Babik 2010). An accurate post-processing of obtained sequences was conducted in four steps in order to ensure the reliability of assigned MHC genotypes: (1) assignment of reads to loci and individuals, elimination of singletons and elimination of reads with inappropriate sizes; (2) elimination of reads with insufficient coverage; (3) determination of alleles; and (4) determination of homozygous and heterozygous amplicons (see Chapter 1 for all details on post-processing). Finally, 3422 genotypes (686 *Mama-UB*, 898 *Mama-UD*, 915 *Mama-DRB1* and 923 *Mama-DRB2*) were obtained. The reliability of obtained genotypes was assessed through three independent methods: (1) intra-individual next generation sequencing repeatability; (2) comparison of obtained genotypes with Sanger and next generation sequencing; and (3) comparison of the consistency of mother-father-offspring triad genotypes, with an error rate estimated to 0.3% (see Chapter 1 for all details on methods to assess genotype reliability).

### Sanger sequencing

For Sanger sequencing, PCR amplifications were carried out following the same protocol used for next generation sequencing (with the exception that no individual barcoding or adaptator were added to the primers). 30  $\mu$ L of PCR products were purified using Axygen® Cleanup Kit for PCR following the manufacturer's instructions. Purified PCR products were single strand sequenced using Dideoxynucleotide Terminator (Dyeterminator, kit BigDye® v.3.1 provided by LifeTechnologies). Sanger sequences are prone to sequencing errors in the 5' region. So, for the polymorphic regions to be on the 3' of the Sanger sequences, we sequenced with the reverse primer, when the polymorphism was at the end of the sequence of interest, and with the forward primer, when the polymorphism was at the beginning of the sequence of interest (MarmR2 (forward) for *Mama-UB*, MarmR4 (forward) for *Mama-UD*, MM\_DRB\_R3 (forward) for *Mama-DRB1* and MM\_DRB\_F2 (reverse) for *Mama-DRB2*). Sequenced products were then purified using AxyPrep Mag Dye Clean (Axygen, following manufacturer's instructions). The DNA sequencing reactions were then

Chapter 2: Mate choice for genetic characteristics analyzed on an ABI3730XL 96 caps DNA Analyzer (LifeTechnologies) (see Chapter 1 for more details on the protocol). Obtained reads were aligned using the progressive alignment (Feng and Doolittle 1987) with the default aligning parameters of the CLC Sequence Viewer software free trial version 6.7.1. to assign 358 genotypes (186 *Mama-UB*, 64 *Mama-UD*, 56 *Mama-DRB1* and 52 *Mama-DRB2*).

#### Deduction from parentage relationships

Thanks to established parentage relationships, when both parents were homozygous at a given MHC loci (obtained either by next generation sequencing or Sanger sequencing) a total of 139 MHC genotypes of their pups could be deduced (106 deduced genotypes at *Mama-UB*, 14 at *Mama-UD*, 1 at *Mama-DRB1* and 18 at *Mama-DRB2*). Genotypes were only deduced when, for a given individual, no more genetic material was available or after the failure of next generation sequencing and/or Sanger sequencing. However, we considered as reliable all genotypes inferred using information on parental genotypes since (1) parentage relationships were highly reliable (see methods regarding reliability of parentage analysis) and (2) only 0.3% of MHC genotyping errors were found after an extended validation procedure of obtained genotypes (see Chapter 1 for more details). The bias in the number of deduced genotypes among different loci (106 at *Mama-UB*, 14 at *Mama-UD*, 1 at *Mama-DRB1* and 18 at *Mama-DRB2*) is inherent to the high variation in allelic diversity observed among the studied loci (Table 17). For instance, the higher assignment rate of genotypes at *Mama-UB* is a consequence of the existence of only two alleles at this locus, which increases the likelihood of homozygous pairs.

Finally, a total of 3919 genotypes from 1025 individuals were obtained among the four MHC loci (3422 using next generation sequencing, 358 using Sanger sequencing and 139 deduced on mother-father-offspring triads. Details on MHC loci characteristics are summarized in Table 17.

Chapter 2: Mate choice for genetic characteristics

<i>Mama-UB</i>	<i>Mama-UB</i>	<i>Mama-UD</i>	<i>Mama-UD</i>	<i>Mama-DRB1</i>	<i>Mama-DRB1</i>	<i>Mama-DRB2</i>	<i>Mama-DRB2</i>
alleles	allelic freq.	alleles	allelic freq.	alleles	allelic freq.	alleles	allelic freq.
*01	0.73	*01	0.73	*01	0.39	*01	0.45
*02	0.27	*02	0.08	*02	0.23	*02	0.40
		*03	0.19	*03	0.13	*03	0.15
				*04	0.02		
				*05	0.03		
				*06	0.07		
				*07	0.09		
				*08	0.04		
Nind	978		976		972		993

**Table 17.** Characteristics of 4 MHC loci of Alpine marmots. Nind: number of individuals.

*Appendix 2. Correlations between the genetic characteristics estimators*

(a)	MHC class I protein diversity	MHC class II protein diversity	MHC class I protein distance	MHC class II protein distance
MHC protein diversity	0.67***	0.78***	-	-
MHC protein distance	-	-	0.24**	0.98***
(b)	MHC class II protein diversity		MHC class II protein distance	
MHC class I protein diversity	0.05 ( $P=0.60$ )		-	
MHC class I protein distance	-		0.02 ( $P=0.80$ )	

**Table 18** Spearman's correlation coefficients (N = 102 dominant males) between MHC diversity estimators calculated over the 2 MHC loci and over the MHC class I or the MHC class II loci (a) and between MHC diversity estimators calculated over the MHC class I locus and over the MHC class II locus (b). \*:  $0.10 < P < 0.05$ ; \*\*:  $0.05 < P < 0.0001$ ; \*\*\*:  $P < 0.0001$ .

Chapter 2: Mate choice for genetic characteristics

(a)	MHC class I protein dissimilarity	MHC class II protein dissimilarity	MHC class I pair protein distance	MHC class II pair protein distance
MHC protein dissimilarity	0.65***	0.88***	-	-
MHC pair protein distance	-	-	0.35***	0.96***

(b)	MHC class II protein dissimilarity	MHC class II pair protein distance
MHC class I protein dissimilarity	0.24***	-
MHC class I pair protein distance	-	0.10 ( $P=0.24$ )

**Table 19.** Spearman's correlation coefficients ( $N = 146$  observed pairs) between MHC dissimilarity estimators calculated over the 2 MHC loci and over the MHC class I or the MHC class II loci (a) and between MHC dissimilarity estimators calculated over the MHC class I locus and over the MHC class II locus (a). \*:  $0.10 < P > 0.05$ ; \*\*:  $0.05 < P > 0.0001$ ; \*\*\*:  $P < 0.0001$ .

	MHC protein distance ( $N = 102$ )	MHC class I protein distance ( $N = 104$ )	MHC class II protein distance ( $N = 105$ )
MHC protein diversity	0.68***	-	-
MHC class I protein diversity	-	1.00***	-
MHC class II protein diversity	-	-	0.74***

**Table 20.** Spearman's correlation coefficients ( $N =$  observed dominant males) between MHC protein diversity and MHC protein distance calculated over the 2 MHC loci, over the MHC class I locus and over the MHC class II locus. \*:  $0.10 < P > 0.05$ ; \*\*:  $0.05 < P > 0.0001$ ; \*\*\*:  $P < 0.0001$ .

	MHC pair protein distance ( $N = 146$ )	MHC class I pair protein distance ( $N = 148$ )	MHC class II pair protein distance ( $N = 152$ )
MHC protein dissimilarity	0.62***	-	-
MHC class I protein dissimilarity	-	0.85***	-
MHC class II protein dissimilarity	-	-	0.58***

**Table 21.** Spearman's correlation coefficients ( $N =$  observed dominant pairs) between MHC protein dissimilarity and MHC pair protein distance calculated over the 2 MHC loci, over the MHC class I locus and over the MHC class II locus. \*:  $0.10 < P > 0.05$ ; \*\*:  $0.05 < P > 0.0001$ ; \*\*\*:  $P < 0.0001$ .



Allelic	Protein	MHC diversity (N=102)	MHC class I diversity (N=104)	MHC class II diversity (N=105)	MHC protein distance (N=102)	MHC class I protein distance (N=104)	MHC class II protein distance (N=105)
MHC diversity		0.39***	-	-	-	-	-
MHC class I diversity		-	0.16( $P = 0.10$ )	-	-	-	-
MHC class II diversity		-	-	0.76***	-	-	-
MHC genetic distance		-	-	-	0.84***	-	-
MHC class I genetic distance		-	-	-	-	0.17*	-
MHC class II genetic distance		-	-	-	-	-	0.84***

**Table 22.** Spearman's correlation coefficients (N = observed dominant males) between MHC allelic and protein diversity and between MHC individual genetic distance and protein distance calculated over the 2 MHC loci, over the MHC class I locus and over the MHC class II locus. \*:  $0.10 < P > 0.05$ ; \*\*:  $0.05 < P > 0.0001$ ; \*\*\*:  $P < 0.0001$ .

Allelic	Protein	MHC dissimilarity (N=146)	MHC class I dissimilarity (N=148)	MHC class II dissimilarity (N=152)	MHC pair protein distance (N=146)	MHC class I pair protein distance (N=148)	MHC class II pair protein distance (N=152)
MHC dissimilarity		0.66**	-	-	-	-	-
MHC class I dissimilarity		-	0.29***	-	-	-	-
MHC class II dissimilarity		-	-	0.88***	-	-	-
MHC pair genetic distance		-	-	-	0.82**	-	-
MHC class I pair genetic distance		-	-	-	-	0.25***	-
MHC class II pair genetic distance		-	-	-	-	-	0.88***

**Table 23.** Spearman's correlation coefficients (N = observed dominant pairs) between MHC allelic and protein dissimilarity and between MHC pair allelic distances and protein distances calculated over the 2 MHC loci, over the MHC class I locus and over the MHC class II locus. \*:  $0.10 < P > 0.05$ ; \*\*:  $0.05 < P > 0.0001$ ; \*\*\*:  $P < 0.0001$ .

*Appendix 3. MHC genetic diversity and dissimilarity estimators and results obtained with these estimators*

*MHC genetic diversity and MHC genetic dissimilarity estimators*

All MHC estimators were calculated over the four MHC allelic polymorphic loci as well as over the only two MHC loci leading to a polymorphic proteins (*Mama-UD* and *Mama-DRB1*). Both MHC estimators were calculated over MHC class I and class II loci separately because MHC class I and II genes are under different selective pressures leading to MHC class I and II estimators not to be correlated except for a correlation between MHC class I and II allelic dissimilarity (Table 24 and 25) and suggesting that both classes could convey different information. Only individuals successfully genotyped at the four MHC loci, at the two MHC class I loci or at the two MHC class II loci were used to calculate the MHC, MHC class I and MHC class II estimators respectively.

(a)	MHC class I allelic diversity	0.75***	MHC class II allelic diversity	0.64***	MHC class I genetic distance	-	MHC class II genetic distance	-
	MHC genetic distance	-	-	0.20*	-	0.97***	-	-
(b)	UD allelic diversity	0.81***	DRB1 allelic diversity	0.54***	UD genetic distance	0.06 ( $P=0.54$ )	DRB1 genetic distance	0.60***
	UD and DRB1 allelic diversity	0.81***	0.54***	-	-	-	-	-
	UD and DRB1 genetic distance	-	-	-	0.06 ( $P=0.54$ )	0.60***	-	-
(c)	MHC class I allelic diversity	0.02 ( $P=0.87$ )	MHC class II allelic diversity	0.02 ( $P=0.87$ )	MHC class I genetic distance	-	MHC class II genetic distance	-
	MHC class I genetic distance	-	-	-	-0.02 ( $P=0.82$ )	-	-	-
(d)	UD allelic diversity	-0.05 ( $P=0.65$ )	DRB1 allelic diversity	-0.05 ( $P=0.65$ )	UD genetic distance	-	DRB1 genetic distance	-
	UD genetic distance	-	-	-	-0.12 ( $P=0.24$ )	-	-	-

**Table 24.** Spearman's correlation coefficients ( $N = 102$  dominant males) between MHC genetic diversity estimators calculated over the four MHC loci and over the MHC class I or the MHC class II loci (a), over the two MHC loci presenting protein polymorphism and over the *Mama-UD* or the *Mama-DRB1* locus (b), over the MHC class I loci and over the MHC class II loci (c) and over the *Mama-UD* and the *Mama-DRB1* locus (d). \*:  $0.10 < P > 0.05$ ; \*\*:  $0.05 < P > 0.0001$ ; \*\*\*:  $P < 0.0001$ .

(a)	MHC class I allelic dissimilarity	MHC class II allelic dissimilarity	MHC class I pair genetic distance	MHC class II pair genetic distance
MHC allelic dissimilarity	0.80***	0.81***	-	-
MHC pair genetic distance	-	-	0.21**	0.95***
(b)	UD allelic dissimilarity	DRB1 allelic dissimilarity	UD pair genetic distance	DRB1 pair genetic distance
UD and DRB1 allelic dissimilarity	0.74***	0.76***	-	-
UD and DRB1 pair genetic distance	-	-	0.14*	0.97***
(c)	MHC class I allelic dissimilarity	MHC class II allelic dissimilarity	MHC class I pair genetic distance	MHC class II pair genetic distance
MHC class I allelic dissimilarity	0.32**	-	-	-
MHC class I pair genetic distance	-	-	-0.02 ( $P=0.82$ )	-
(d)	UD allelic dissimilarity	DRB1 allelic dissimilarity	UD pair genetic distance	DRB1 pair genetic distance
UD allelic dissimilarity	0.14*	-	-	-
UD pair genetic distance	-	-	-0.05 ( $P=0.51$ )	-

**Table 25.** Spearman's correlation coefficients ( $N = 146$  real pairs) between MHC genetic dissimilarity estimators calculated over the four MHC loci and over the MHC class I or the MHC class II loci (a), over the two MHC loci presenting protein polymorphism and over the *Mama-UD* or the *Mama-DRB1* locus (b), over the MHC class I loci and over the MHC class II loci (c) and over the *Mama-UD* and the *Mama-DRB1* locus (d). \*:  $0.10 < P > 0.05$ ; \*\*:  $0.05 < P > 0.0001$ ; \*\*\*:  $P < 0.0001$ .

Two indexes were used to estimate the MHC genetic diversity: the MHC allelic diversity and the MHC individual genetic distance. The MHC allelic diversity was calculated as the number of different MHC alleles an individual possesses. The individual genetic distance was calculated using the software MEGA version 6.06 (Tamura et al. 2013) as the sum of the p-distances (i.e. the proportion of nucleotide sites that differs between two alleles) among the considered MHC loci.

Two indexes were used to estimate the MHC genetic dissimilarity: the MHC allelic dissimilarity and the MHC pair genetic distance. The MHC allelic dissimilarity was calculated as the number of alleles a male possesses but the female does not. The MHC pair genetic distance was calculated as the sum of the p-distances (i.e. the proportion of nucleotide sites that differs between two alleles) averaged over all the pairwise comparisons of the alleles of a locus for a given pair (Bollmer et al. 2012).

The different index used to calculate MHC genetic diversity – MHC allelic diversity and MHC individual genetic distance – and MHC genetic dissimilarity – MHC allelic dissimilarity and the MHC pair genetic distance – estimators were found to be highly correlated (see Table 26 and 27), suggesting that different index convey the same information.

Chapter 2: Mate choice for genetic characteristics

(a)	MHC genetic distance (N = 102)	MHC class I genetic distance (N = 104)	MHC class II genetic distance (N = 105)
MHC allelic diversity	0.51***	-	-
MHC class I allelic diversity	-	1.00***	-
MHC class II allelic diversity	-	-	0.58***

(b)	UD and DRB1 genetic distance (N = 102)	UD genetic distance (N = 104)	DRB1 genetic distance (N = 105)
UD and DRB1 allelic diversity	0.40***	-	-
UD allelic diversity	-	1.00***	-
DRB1 allelic diversity	-	-	0.63***

**Table 26.** Spearman's correlation coefficients (N = real dominant males) between MHC allelic diversity and MHC individual genetic distance calculated over the four MHC loci, over the MHC class I loci and over the MHC class II loci (a) and the two MHC loci presenting protein polymorphism, over the *Mama-UD* locus and over the *Mama-DRB1* locus (b). \*: 0.10 < P > 0.05; \*\*: 0.05 < P > 0.0001; \*\*\*: P < 0.0001.

(a)	MHC pair genetic distance (N = 146)	MHC class I pair genetic distance (N = 148)	MHC class II pair genetic distance (N = 152)
MHC allelic dissimilarity	0.43***	-	-
MHC class I allelic dissimilarity	-	0.75***	-
MHC class II allelic dissimilarity	-	-	0.52***

(b)	UD and DRB1 pair genetic distance (N = 146)	UD pair genetic distance (N = 148)	DRB1 pair genetic distance (N = 152)
UD and DRB1 allelic dissimilarity	0.47***	-	-
UD allelic dissimilarity	-	0.77***	-
DRB1 allelic dissimilarity	-	-	0.58***

**Table 27.** Spearman's correlation coefficients (N = real dominant pairs) between MHC allelic dissimilarity and MHC pair genetic distance calculated over the four MHC loci, over the MHC class I loci and over the MHC class II loci (a) and the two MHC loci presenting protein polymorphism, over the *Mama-UD* locus and over the *Mama-DRB1* locus (b). \*: 0.10 < P > 0.05; \*\*: 0.05 < P > 0.0001; \*\*\*: P < 0.0001.

*Results*

Correlation between genetic characteristics

Genome-wide heterozygosity and MHC genetic diversity were not correlated (Table 28a,b). Although a negative correlation (i.e. individuals more similar at genome-wide loci are more similar at MHC loci and *vice versa*) was evidenced between relatedness and the MHC allelic dissimilarity calculated over all the MHC loci, the MHC class II loci, the two MHC loci presenting protein polymorphism and the *Mama-DRB1* locus (Table 28c,d), relatedness and all the other MHC genetic dissimilarity estimators were not correlated (Table 28c,d).



(a)	MHC allelic diversity (N = 102)	MHC class I allelic diversity (N = 104)	MHC class II allelic diversity (N = 105)	MHC genetic distance (N = 102)	MHC class I genetic distance (N = 104)	MHC class II genetic distance (N = 105)
SH	0.02 ( $P = 0.82$ )	0.01 ( $P = 0.89$ )	0.09 ( $P = 0.39$ )	0.06 ( $P = 0.58$ )	0.01 ( $P = 0.94$ )	0.06 ( $P = 0.52$ )
(b)	UD and DRB1 allelic diversity (N = 102)	UD allelic diversity (N = 104)	DRB1 allelic diversity (N = 105)	UD and DRB1 genetic distance (N = 102)	UD genetic distance (N = 104)	DRB1 genetic distance (N = 105)
SH	0.04 ( $P = 0.72$ )	0.04 ( $P = 0.66$ )	0.04 ( $P = 0.72$ )	0.05 ( $P = 0.62$ )	0.03 ( $P = 0.76$ )	0.10 ( $P = 0.32$ )
(c)	MHC allelic dissimilarity (N = 146)	MHC class I allelic dissimilarity (N = 148)	MHC class II allelic dissimilarity (N = 152)	MHC pair genetic distance (N = 146)	MHC class I pair genetic distance (N = 148)	MHC class II pair genetic distance (N = 152)
$R_{ig}$	-0.27**	-0.13 ( $P = 0.13$ )	-0.25**	-0.11 ( $P = 0.18$ )	-0.14 ( $P = 0.10$ )	-0.07 ( $P = 0.36$ )
(d)	UD and DRB1 allelic dissimilarity (N = 146)	UD allelic dissimilarity (N = 148)	DRB1 allelic dissimilarity (N = 152)	UD and DRB1 pair genetic distance (N = 146)	UD pair genetic distance (N = 148)	DRB1 pair genetic distance (N = 152)
$R_{ig}$	-0.22**	-0.05 ( $P = 0.54$ )	-0.24**	-0.10 ( $P = 0.24$ )	-0.10 ( $P = 0.23$ )	-0.07 ( $P = 0.36$ )

**Table 28.** Spearman's correlation coefficients between standardized heterozygosity (SH) and MHC genetic diversity estimators - MHC allelic diversity and MHC individual genetic distance - (a, b) and between the relatedness ( $R_{ig}$ ) and the MHC genetic dissimilarity estimators - MHC allelic dissimilarity and MHC pair genetic distance - (c, d). MHC diversity and dissimilarity estimators were calculated over the four MHC, the two MHC class I and the two MHC class II loci (a, c) or over the two MHC loci presenting protein polymorphism, the *Mama-UD* locus and the *Mama-DRB1* locus (b, d) (N = number of real dominant males or pairs).

Are social males and social pairs genetically different than expected from random mate choice?

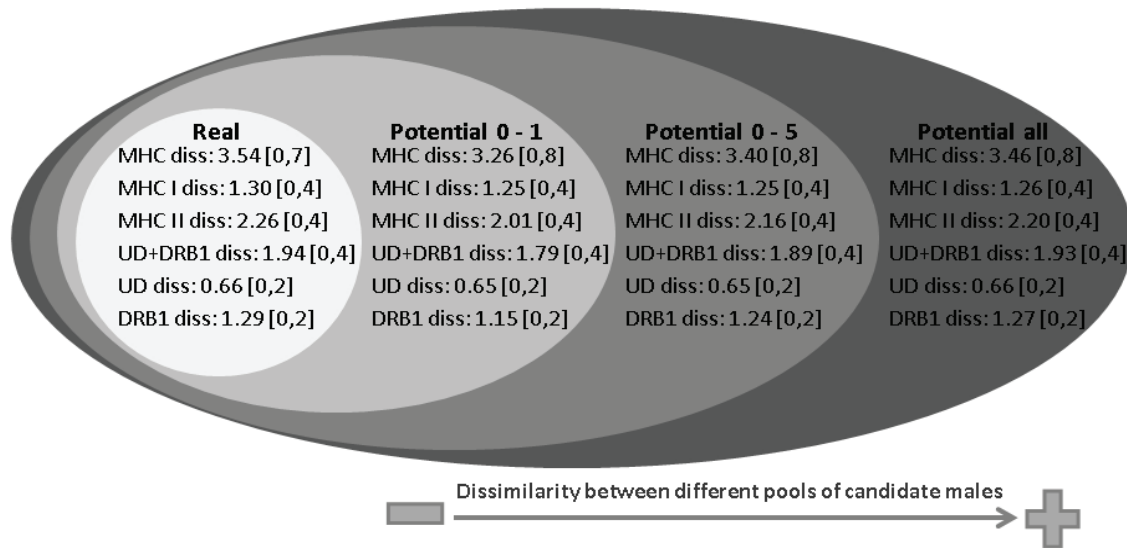
Social males were no more MHC diverse than expected under random mate choice, and this whatever the pool of candidate males considered (Table 29).

Social pairs showed a higher allelic dissimilarity at the four MHC loci, at the MHC class II loci, at the two MHC loci presenting protein polymorphism and at the *Mama-DRB1* locus than expected under random choice among the males residing in the same territory or in the immediate vicinity of a given female. However, given that MHC allelic dissimilarity increased among individuals residing farther apart (Figure 13), social pairs showed an allelic dissimilarity consistent with random choice among the males residing farther away (Table 29). Moreover, when estimated by the MHC pair genetic distance, social pairs' MHC dissimilarity did not differ from the one expected under random mate choice and this whatever the pool of candidate males considered (Table 29).

(a) Genetic characteristics	Candidate males											
	Observed mean	Residing in the same territory or in the immediate vicinity			Residing at a maximal distance of five territories away			Residing in the studied population				
		Observed difference	Simulated difference [CI 95%]	p-value	Observed difference	Simulated difference [CI 95%]	p-value	Observed difference	Simulated difference [CI 95%]	p-value		
MHC allelic diversity	6.43	-0.05	-0.01 [-0.13, 0.10]	0.50	-0.04	-0.01 [-0.13, 0.10]	0.59	-0.03	-0.01 [-0.12, 0.09]	0.78		
MHC class I allelic diversity	2.90	-0.07	-0.01 [0.09, 0.08]	0.15	-0.06	-0.01 [-0.09, 0.08]	0.21	-0.04	-0.01 [0.09, 0.07]	0.45		
MHC class II allelic diversity	3.54	0.03	0.001 [-0.09, 0.08]	0.51	0.03	0.002 [-0.08, 0.08]	0.47	0.02	0.001 [-0.08, 0.08]	0.62		
MHC genetic distance	0.04	0.0004	-0.0002 [-0.002, 0.002]	0.77	0.001	-0.0001 [-0.003, 0.003]	0.70	0.0004	<0.0003 [-0.003, 0.003]	0.78		
MHC class I genetic distance	0.01	-0.0004	<0.0001 [-0.001, 0.001]	0.15	-0.0004	<0.0001 [-0.001, 0.001]	0.22	-0.0003	<0.0001 [-0.0006, 0.0004]	0.47		
MHC class II genetic distance	0.04	0.001	<0.0001 [-0.002, 0.002]	0.46	0.001	<0.0001 [-0.003, 0.003]	0.40	0.001	<0.0001 [-0.003, 0.003]	0.55		
MHC allelic dissimilarity	3.54	<b>0.29</b>	<b>-0.0004</b> [-0.17, 0.17]	< <b>0.0001</b>	0.15	0.001 [-0.16, 0.16]	0.06	0.09	0.001 [-0.16, 0.14]	0.23		
MHC class I allelic dissimilarity	1.30	0.05	-0.02 [-0.11, 0.07]	0.36	0.05	-0.004 [-0.10, 0.09]	0.25	0.04	-0.01 [-0.10, 0.07]	0.43		
MHC class II allelic dissimilarity	2.26	<b>0.25</b>	<b>0.01</b> [-0.10, 0.13]	< <b>0.0001</b>	<b>0.11</b>	<b>0.0003</b> [-0.11, 0.10]	<b>0.05</b>	0.06	0.001 [-0.10, 0.10]	0.22		
MHC pair genetic distance	0.01	< 0.0001	<0.0001 [-0.0004, 0.0003]	0.10	< 0.0001	-0.0001 [-0.0004, 0.0002]	0.91	< 0.0001	<0.0001 [-0.0004, 0.0002]	0.99		
MHC class I pair genetic distance	0.002	< 0.0001	<0.0001 [-0.0002, 0.0001]	0.85	< 0.0001	<0.0001 [-0.0001, 0.0001]	0.72	< 0.0001	<0.0001 [-0.0001, 0.0001]	0.83		
MHC class II pair genetic distance	0.02	0.001	<0.0001 [-0.001, 0.001]	0.15	< 0.0001	-0.0002 [-0.001, 0.001]	0.99	-0.0001	-0.0001 [-0.001, 0.001]	0.99		

(b) Genetic characteristics	Observed mean	Candidate males									
		Residing in the same territory or in the immediate vicinity					Residing at a maximal distance of five territories away				
		Observed difference	Simulated difference [CI 95%]	p-value	Observed difference	Simulated difference [CI 95%]	p-value	Observed difference	Simulated difference [CI 95%]	p-value	Observed difference
UD and DRB1 allelic diversity	3.27	-0.01	0.001 [-0.06, 0.06]	0.84	-0.02	-0.02 [-0.08, 0.05]	0.93	-0.03	-0.02	-0.02 [-0.07, 0.05]	0.75
UD allelic diversity	1.46	-0.02	-0.01 [0.06, 0.04]	0.44	-0.03	-0.02 [-0.07, 0.03]	0.54	-0.03	-0.02	-0.02 [0.07, 0.03]	0.64
DRB1 allelic diversity	1.81	0.02	0.01 [-0.03, 0.06]	0.40	0.02	0.001 [-0.04, 0.04]	0.43	0.003	0.003	0.003 [-0.04, 0.04]	0.84
UD and DRB1 genetic distance	0.04	0.001	-0.0001 [-0.003, 0.003]	0.60	-0.001	-0.0003 [-0.003, 0.002]	0.66	0.0004	-0.0001	-0.0001 [-0.003, 0.003]	0.75
UD genetic distance	-0.0002	<0.0001	<0.0001 [-0.0004, 0.0002]	0.45	-0.0002	-0.0001 [-0.0004, 0.0002]	0.57	-0.0002	-0.0001	-0.0001 [-0.0004, 0.0002]	0.67
DRB1 genetic distance	0.004	<0.0001	<0.0001 [-0.0003, 0.0003]	0.90	0.001	<0.0001 [-0.003, 0.003]	0.40	0.001	<0.0001	<0.0001 [-0.003, 0.003]	0.39
UD and DRB1 allelic dissimilarity	1.94	<b>0.15</b>	<b>0.01</b> [-0.07, 0.09]	<b>0.001</b>	0.05	-0.02 [-0.10, 0.06]	0.24	-0.01	-0.02	-0.02 [-0.11, 0.06]	0.77
UD allelic dissimilarity	0.66	0.01	-0.003 [-0.06, 0.05]	0.78	0.01	-0.01 [-0.06, 0.04]	0.73	-0.001	-0.02	-0.02 [-0.07, 0.04]	0.99
DRB1 allelic dissimilarity	1.29	<b>0.15</b>	<b>0.01</b> [-0.05, 0.07]	< <b>0.0001</b>	0.05	-0.01 [-0.07, 0.05]	0.09	0.02	-0.01	-0.01 [-0.07, 0.05]	0.47
UD and DRB1 pair genetic distance	0.02	0.0004	<0.0001 [-0.001, 0.001]	0.17	< 0.0001	-0.0002 [-0.001, 0.0004]	0.99	-0.0001	-0.0002	-0.0002 [-0.001, 0.0004]	0.99
UD pair genetic distance	0.003	< 0.0001	<0.0001 [-0.0002, 0.0001]	0.76	< 0.0001	<0.0001 [-0.0002, 0.0001]	0.86	-0.0001	<0.0001	<0.0001 [-0.0002, 0.0000]	0.77
DRB1 pair genetic distance	0.03	0.001	<0.0001 [-0.001, 0.001]	0.17	0.0002	-0.0002 [-0.001, 0.001]	0.99	-0.0003	-0.0002	-0.0002 [-0.001, 0.001]	0.99

**Table 29.** Genetic characteristics of real social males and pairs compared to the potential candidate males and pairs residing in the same territory or in the immediate vicinity of a given female, residing at a maximal distance of five territories away or residing in the overall studied population. MHC diversity and dissimilarity estimators were calculated over the four MHC, the two MHC class I and the two MHC class II loci (a) or over the two MHC loci presenting protein polymorphism, the *Mama-UD* and the *Mama-DRB1* locus (b). P-values <0.05 are indicated in bold.



**Figure 13.** Spatial distribution of the MHC allelic dissimilarity estimated over all loci (MHC diss), over MHC class I loci (MHC I diss), over MHC class II loci (MHC II diss) over *Mama-UD* and *Mama-DRB1* loci (UD+DRB1 diss), over *Mama-UD* locus (UD diss) and over *Mama-DRB1* locus (DRB1 diss). Real: between females and males of real social pairs; Potential 0-1: between real social females and candidate males residing in the same territory or in the immediate vicinity; Potential 0-5: between real social females and candidate males residing at a maximal distance of five territories away of any real social females; Potential all: between real social females and candidate males residing in the studied population.

### Is EPP pattern dependent of social males' and social pairs' genetic characteristics?

Neither the presence nor the number of EPY within litters depended on the social male's MHC genetic diversity (Table 30).

Neither the presence nor the number of EPY within litters depended on the social pair MHC allelic dissimilarity when estimated over the four MHC loci, the MHC class I or the MHC class II loci (Table 30a) or on the MHC pair genetic distance (Table 30a, b). However, when the MHC allelic dissimilarity was further estimated only on the two MHC loci leading to a polymorphic proteins (*Mama-UD* and *Mama-DRB1*) or on *Mama-DRB1* locus only, the number of EPY within litters was found to decrease when the social pair MHC allelic dissimilarity increased (Table 30b).

Chapter 2: Mate choice for genetic characteristics

(a) Independent variable	Presence of EPY		Number of EPY		N
	estimate ± SE	P	estimate ± SE	P	
MHC allelic diversity	-0.13 ± 0.25	0.59	-0.06 ± 0.19	0.73	134
MHC class I allelic diversity	-0.18 ± 0.40	0.66	-0.23 ± 0.35	0.50	135
MHC class II allelic diversity	-0.17 ± 0.34	0.61	-0.13 ± 0.29	0.64	139
MHC genetic distance	-0.73 ± 11.27	0.95	1.39 ± 8.23	0.87	134
MHC class I genetic distance	-30.07 ± 67.45	0.66	-38.80 ± 57.67	0.50	135
MHC class II genetic distance	-4.18 ± 11.42	0.71	-6.59 ± 10.54	0.53	139
MHC allelic dissimilarity	-0.03 ± 0.17	0.85	-0.05 ± 0.13	0.71	134
MHC class I allelic dissimilarity	0.23 ± 0.28	0.42	0.13 ± 0.18	0.48	135
MHC class II allelic dissimilarity	-0.40 ± 0.28	0.14	-0.55 ± 0.32	0.09	138
MHC pair genetic distance	-94.26 ± 59.18	0.11	81.03 ± 59.54	0.17	134
MHC class I pair genetic distance	-125.71 ± 168.49	0.46	-174.28 ± 105.74	0.10	135
MHC class II pair genetic distance	-60.29 ± 31.67	0.06	-71.75 ± 39.65	0.07	138
(b) Independent variable	Presence of EPY		Number of EPY		N
	estimate ± SE	P	estimate ± SE	P	
UD and DRB1 allelic diversity	-0.07 ± 0.49	0.89	0.09 ± 0.41	0.83	134
UD allelic diversity	-0.05 ± 0.66	0.94	0.003 ± 0.54	0.99	135
DRB1 allelic diversity	-0.07 ± 0.72	0.92	0.32 ± 0.63	0.61	139
UD and DRB1 genetic distance	0.31 ± 12.43	0.98	2.87 ± 9.40	0.76	134
UD genetic distance	-7.87 ± 109.21	0.95	0.49 ± 89.66	0.99	135
DRB1 genetic distance	-66.47 ± 124.76	0.59	-101.23 ± 126.70	0.42	139
UD and DRB1 allelic dissimilarity	-0.21 ± 0.34	0.54	<b>-0.45 ± 0.21</b>	<b>0.03</b>	<b>134</b>
UD allelic dissimilarity	0.06 ± 0.54	0.91	-0.37 ± 0.31	0.23	135
DRB1 allelic dissimilarity	-0.78 ± 0.50	0.12	<b>-1.04 ± 0.49</b>	<b>0.03</b>	<b>138</b>
UD and DRB1 pair genetic distance	-51.94 ± 31.53	0.10	48.88 ± 30.38	0.11	134
UD pair genetic distance	-163.63 ± 175.34	0.35	-260.76 ± 148.13	0.08	135
DRB1 pair genetic distance	-31.95 ± 16.87	0.06	-38.67 ± 20.30	0.06	138

**Table 30.** Generalized estimating equation models showing the effects of the MHC allelic characteristics of the real social males and pairs on both the presence and the number of EPY. The MHC diversity and dissimilarity estimators were calculated over the four MHC, the two MHC class I and the two MHC class II loci (a) or over the two MHC loci presenting protein polymorphism, the *Mama-UD* and the *Mama-DRB1* locus (b). All models include the number of sexually mature male subordinates present in a given family, the litter size, the social pair relatedness and its associated quadratic term. P-values <0.05 are indicated in bold.

Are social and extra-pair males or pairs genetically different?

Neither social males were more MHC diverse than extra-pair males, nor social pairs were more MHC dissimilar than extra-pair pairs (Table 31).

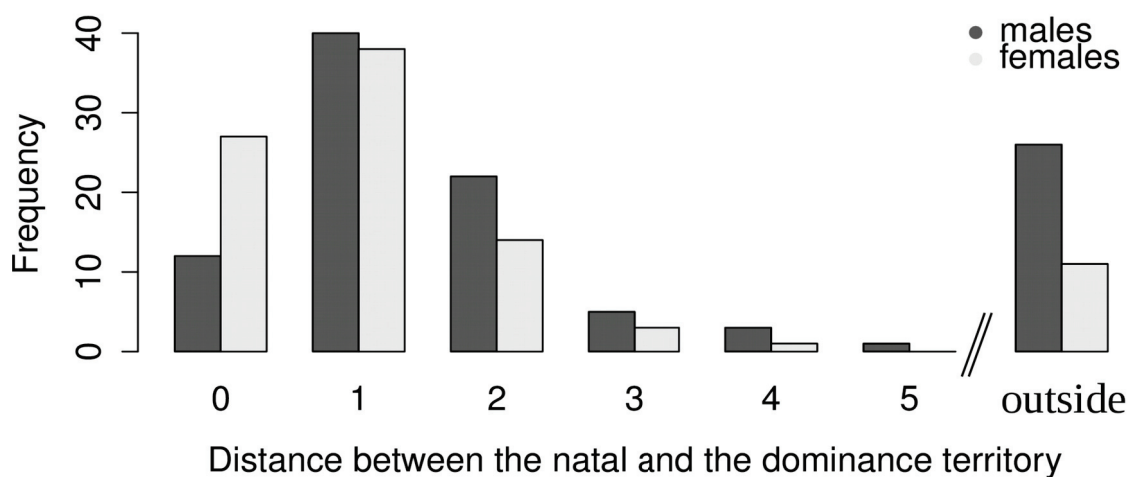
(a) Genetic characteristics	Nbr.	P	N
	social < extra-pair		
MHC allelic diversity	2	0.99	14
MHC class I allelic diversity	5	0.91	14
MHC class II allelic diversity	2	0.99	15
MHC allelic individual genetic distance	1	1.00	14
MHC class I allelic individual genetic distance	5	0.91	14
MHC class II allelic individual genetic distance	3	0.99	15
MHC allelic dissimilarity	1	1.00	14
MHC class I allelic dissimilarity	2	0.99	14
MHC class II allelic dissimilarity	2	0.99	14
MHC allelic pair genetic distance	1	1.00	14
MHC class I allelic pair genetic distance	1	1.00	14
MHC class II allelic pair genetic distance	2	0.99	14
(b) Genetic characteristics	Nbr.	P	N
	social < extra-pair		
UD and DRB1 allelic diversity	1	1.00	14
UD allelic diversity	0	1.00	14
DRB1 allelic diversity	1	1.00	15
UD and DRB1 individual genetic distance	1	1.00	14
UD individual genetic distance	0	1.00	14
DRB1 individual genetic distance	2	0.99	15
UD and DRB1 allelic dissimilarity	1	1.00	14
UD allelic dissimilarity	2	0.99	14
DRB1 allelic dissimilarity	0	1.00	14
UD and DRB1 pair genetic distance	1	1.00	14
UD pair genetic distance	1	1.00	14
DRB1 pair genetic distance	2	0.99	14

**Table 31.** Genetic characteristics of social males and pairs compared to the corresponding extra-pair males and pairs. MHC diversity and dissimilarity estimators were calculated over the four MHC, the two MHC class I and the two MHC class II loci (a) or over the two MHC loci presenting protein polymorphism, the *Mama-UD* and the *Mama-DRB1* locus (b).

#### *Appendix 4. Effect of the spatial distance on males and pairs genetic characteristics*

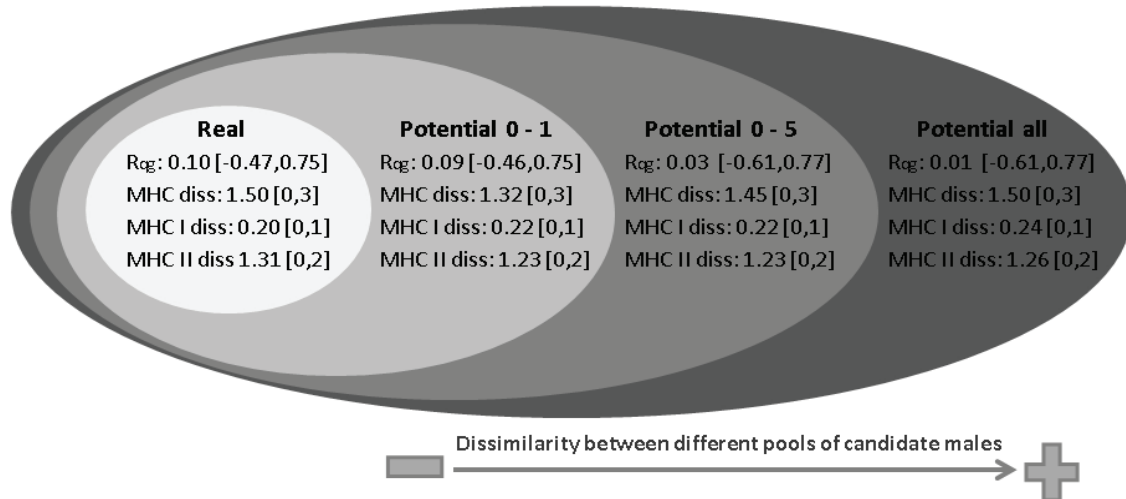
##### *Male dispersal distance*

Since 1990, the dispersal distance of a total of 109 males and 94 females that reached dominance in the study site was known. A total of 83 (73%) out of these 109 males and 83 (88%) out of these 94 females came from the study site and have either inherited dominance in their natal territory, moved to an adjacent territory or to another territory inside the study site to reach dominance (Figure 14), whereas the remaining 26 males (24%) and 11 females (12%) came from outside the study site (Figure 14). The dispersal distance between the natal and the dominance territory was short, with a mean distance  $\pm$  SD of  $1.40 \pm 1.01$  for the 83 males and a mean distance of  $0.95 \pm 0.87$  for the 83 females that were born and reached dominance inside our study site.



**Figure 14.** Dispersal distance (measured in the number of territories between the natal and the dominance territory) of the males and the females that reached dominance between 1990 and 2014 in the Nature Reserve of La Grande Sassièrè.





**Figure 15.** Spatial distribution of the relatedness estimated with the relatedness ( $R_{gg}$ ) and the MHC protein dissimilarity estimated over all loci (MHC diss), over MHC class I loci (MHC I diss) and over MHC class II loci (MHC II diss). Real: between females and males of real social pairs; Potential 0-1: between real social females and males residing in the same territory or in the immediate vicinity; Potential 0-5: between real social females and males residing at a maximal distance of five territories away; Potential all: between real social females and all males residing in the studied population.

Genetic characteristics	Candidate males									
	Observed mean	Residing at a maximal distance of five territories away				Residing in the studied population				P
		Observed difference	Simulated difference [CI 95%]	P	Observed difference	Simulated difference [CI 95%]	P	Observed difference	Simulated difference [CI 95%]	
SH	0.95	-0.01	0.01 [-0.02, 0.03]	0.08	-0.02	0.004 [-0.02, 0.03]	0.07			
MHC protein diversity	2.91	0.01	-0.01 [-0.06, 0.05]	0.62	-0.01	-0.01 [-0.06, 0.05]	0.80			
MHC class I protein diversity	1.15	0.001	-0.01 [-0.04, 0.03]	0.95	-0.01	-0.01 [-0.04, 0.02]	0.99			
MHC class II protein diversity	1.75	0.01	-0.001 [-0.04, 0.04]	0.69	-0.01	0.003 [-0.04, 0.04]	0.56			
MHC protein distance	0.05	0.003	-0.0002 [-0.004, 0.004]	0.88	-0.001	<0.0001 [-0.004, 0.004]	0.80			
MHC class I protein distance	0.003	<0.0001	-0.0002 [-0.001, 0.001]	0.96	-0.0002	-0.0002 [-0.0008, 0.0004]	0.99			
MHC class II protein distance	0.05	0.0002	-0.0002 [-0.004, 0.004]	0.93	-0.001	<0.0001 [-0.004, 0.004]	0.77			
$R_{\text{ig}}$	0.10	<b>0.08</b>	<b>0.01 [-0.02, 0.03]</b>	< <b>0.001</b>	<b>0.09</b>	<b>0.004 [-0.02, 0.02]</b>	< <b>0.0001</b>			
$R_{\text{ig}}-R_{\text{ig\_intermediate}}$	0.20	<b>-0.02</b>	<b>0.01 [-0.02, 0.01]</b>	<b>0.02</b>	<b>-0.02</b>	<b>0.01 [-0.02, 0.01]</b>	<b>0.001</b>			
MHC protein dissimilarity	1.50	0.05	-0.02 [-0.09, 0.06]	0.22	-0.003	-0.02 [-0.10, 0.05]	0.99			
MHC class I protein dissimilarity	0.20	-0.02	-0.01 [-0.04, 0.03]	0.54	-0.04	-0.02 [-0.05, 0.02]	0.15			
MHC class II protein dissimilarity	1.31	<b>0.08</b>	<b>-0.01 [-0.08, 0.05]</b>	<b>0.03</b>	0.05	-0.01 [-0.07, 0.05]	0.11			
MHC pair protein distance	0.02	0.001	-0.0001 [-0.001, 0.001]	0.05	0.001	-0.0002 [-0.001, 0.001]	0.16			
MHC class I pair protein distance	0.002	<0.0001	<0.0001 [-0.0003, 0.0002]	0.99	-0.0002	-0.0002 [-0.0004, 0.0001]	0.43			
MHC class II pair protein distance	0.04	0.002	-0.0001 [-0.002, 0.002]	0.03	0.002	-0.0002 [-0.002, 0.002]	0.99			

**Table 32.** Genetic characteristics of real social males and pairs compared to the ones of potential candidate males and pairs depending on the spatial distances between the candidate males and a given female. SH: standardized heterozygosity;  $R_{\text{ig}}$ : relatedness.  $R_{\text{ig}}-R_{\text{ig\_intermediate}}$ : absolute difference between the observed value and the intermediate value of relatedness (0.17) obtained from the generalized estimating equation model of the number of EPY within litters. P-values <0.05 are indicated in bold.

## Chapter 2: Mate choice for genetic characteristics

# **CHAPTER 3: MHC CHARACTERISTICS AND PARASITES**

---

**The role of MHC on intestinal parasite variations in Alpine marmots**

M Ferrandiz-Rovira, A Cohas, D Allainé and M-P Callait-Cardinal

*Draft*



## Abstract

The major histocompatibility complex (MHC) transcript genes, coding for antigen presenting molecules, represent an important functional complex of genes of vertebrates due to its key role in host-parasite interactions. Individuals with a high MHC diversity and/or individuals carrying specific MHC alleles are expected to be more resistant to parasite infection. Here, we investigated the influence of MHC characteristics on the abundance or the presence of three intestinal parasites (*Eimeria monacis*, *Eimeria sp.* and *Ctenotaenia marmotae*) in a wild population of Alpine marmots (*Marmota marmota*), an hibernating mammal. Our results reveal that (1) MHC diversity did not affect the abundance or the presence of the three intestinal parasites, (2) Alpine marmots carrying the specific allele *Mama-DRB1\*02* showed lower oocyst excretion of the coccidia *E. monacis* and (3) the presence specific alleles did not affect the presence of the coccidia *Eimeria sp.* nor the presence of the cestode *C. marmotae*. This results suggest that MHC may play a weak role on the associations between the Alpine marmots and their intestinal parasites and that other immune mechanisms may play a greatest role. We emphasize the need for empirical studies of host-parasite interactions in the wild to understand how hosts may resist and/or tolerate its parasites.

**Keywords:** host–parasite co-evolution, immunogenetics, *Marmota marmota*, MHC, parasite load, parasite prevalence, resistance

## 1. INTRODUCTION

Pathogens responsible for infectious diseases are potential agents of natural selection for living beings (Haldane 1949). Under natural conditions, individuals are constantly but unequally exposed to a range of parasites. Parasite loads observed at a given time and a given individual reflect both its own level of exposition and the performance of its immune system. Finally, the set of mechanisms linked with parasite's pathogenesis and host's immunity is able to cause injuries or diseases resulting in a decrease of host's health and fitness, or even lead to host's death. Quantifying and interpreting patterns in host-parasite interactions is essential for understanding the ecological and evolutionary implications of parasites on their hosts (Sol et al. 2003).

Although interactions between hosts and parasites are permanent, their relationships are far from being stable since co-evolution between hosts and parasites under the form of arms race exists. Specifically, the co-evolution between hosts and parasites, also referred as Red Queen Hypothesis (Van Valen 1973), is based on the necessity for the host to avoid or combat parasites, which leads to the evolution of alleles conferring resistance to parasites (resistance is defined here following Thomas et al. (2007) and Råberg et al. (2009) by the ability of hosts to prevent parasite infection or limit parasite load). Also, the co-evolution between hosts and parasites is based on the necessity for the parasites to encounter and infect its hosts, which leads to the evolution of parasite genes conferring more virulent capacities. But once infected, resistance is not the unique way in which hosts may protect themselves: some hosts are good to reducing parasite load but they are not necessary the healthiest (Råberg et al. 2009). Tolerance, defined as the ability to withstand the infection by limiting the harm caused by a given parasite burden (Restif and Koella 2004, Råberg et al.

2009) while paying a low fitness cost (Medzhitov et al. 2012, Sorci 2013), may be another way for hosts to protect themselves from parasitism (Restif and Koella 2004, Råberg et al. 2009). Although little is known regarding how tolerance affects the co-evolution between hosts and parasites, there are reasons to suspect that it exists (see Råberg et al. 2009 for a review analysing the current knowledge of resistance and tolerance in animals).

Among the wide range of identified immune genes (e.g. immunoglobulin receptors, interleukin receptors, Mannose binding lectin or Toll-like receptors), the major histocompatibility complex (MHC), has been largely studied during the past decades due to its importance in vertebrate immunity defence and mate choice (Bernatchez and Landry 2003, Piertney and Oliver 2006, Kamiya et al. 2014, Sin et al. 2014). Moreover, its structure and functions are well known. The MHC is a multi-gene family present in all jawed vertebrates and divided into three families named class I, II and III (Kelley et al. 2005, Acevedo-Whitehouse and Cunningham 2006). MHC contains the most polymorphic loci known in the vertebrate genome (reviewed in Kelley et al. 2005) and its primary function is the immune response. Indeed, MHC transcript molecules are receptors that bind foreign peptides and present them to T-cells (Doherty and Zinkernagel 1975). Then, if a T cell identifies the antigen peptide as a non-self, it triggers a cascade of immune reaction processes (Klein 1982) to control the infection. The majority of MHC polymorphism is located at the codons supposedly involved in antigen binding sites (ABS) (Garrigan and Hedrick 2003), which is the basis for a differential peptide binding (Reche and Reinherz 2003).

Although each functional transcript MHC allele is available to bind a range of structurally diverse parasites (Frank 2002), different ABS are supposed to be required to bind different antigen peptides from the wide range of parasite strains and species (Hughes and Yeager 1998). Given that MHC genes are co-dominantly expressed, genetically diverse individuals at MHC loci are expected to have a higher probability to recognize a wider array of parasites since individuals with high MHC diversity are expected to increase the diversity of antigens presented to T cells than less diverse individuals (Doherty and



Zingernagel 1975, Hughes and Nei 1989, Penn et al. 2002). Evidences of this hypothesis, also known as heterozygote advantage or overdominance, have been found in fifteen studies conducted on non-model species (Sin et al. 2014). For example, in striped mouse (*Rhabdomys pumilio*) MHC diversity was negatively correlated with helminth infection and helminth faecal egg count (Froeschke and Sommer 2005). Also, in European water vole (*Arvicola terrestris*), MHC diverse individuals were simultaneously co-infected by fewer parasite species than less MHC diverse water voles (Oliver et al. 2009). Alternatively, having too many MHC genes could also be disadvantageous since it could promote negative T-cell selection (i.e. a process removing T-cells that bind too strongly to self peptides) (Starr et al. 2003). If this process may prevent autoimmunity from occurring, negative T-cell selection could reduce the diversity of T-cell receptors (Lawlor et al. 1990, Nowak et al. 1992) and thus could reduce host's resistance to parasites. To counteract these selective forces, Nowak et al. (1992) suggested that an optimal number of MHC genes rather than a maximal number of MHC genes would allow conferring a higher parasite resistance. According to this hypothesis, in a study on bank voles (*Myodes glareolus*), the lower prevalence (i.e. the number of infected hosts divided by the number of examined hosts) of different nematode species and the lower prevalence of the most common nematode were found in individuals of intermediate MHC diversity (Kloch et al. 2010). Also, in three-spined sticklebacks (*Gasterosteus aculeatus*), the lower parasite load found after two experimentally rounds of infection by three core three-spined sticklebacks parasite species was found in individuals carrying an intermediate level of MHC diversity displayed (Wegner et al. 2003a).

Also, specific MHC alleles are expected to be associated to resistance or susceptibility to infectious diseases, as predicted by the negative frequency-dependent selection hypothesis (Bodmer 1972, Potts and Wakeland 1990, Slade and McCallum 1992). This hypothesis proposes that specific alleles (often rare alleles) are likely to offer a greater resistance against parasites than more common alleles because parasites are more likely to be adapted to the most common alleles (Bodmer 1972, Potts and Wakeland 1990, Slade and McCallum

1992), thus rare alleles may confer a selective advantage. Under these circumstances, co-evolution between hosts and parasites would drive a cyclic change in allelic frequencies. Compared to the heterozygote advantage hypothesis, the support for the negative frequency-dependent selection hypothesis accounts for more than thirty studies carried out on non-model species from a wide range of taxa (reviewed in Sin et al. 2014).

MHC class I and II translated alleles are trans-membrane proteins situated in cell surface, but have different functions. Whereas MHC class I molecules are present on the surface of all nucleated cells and bind small peptides derived from intra-cellular antigens, MHC class II molecules are present only on macrophages, B cells or dendritic cells, and bind greater peptides derived from extra-cellular phagocytosed antigens (Bjorkman and Parham 1990, Hughes and Yeager 1998, Neefjes et al. 2011). However, the dichotomy between MHC class I/intra-cellular pathogens and class II/extra-cellular pathogens is not as clear cut as it seems. Indeed, MHC class I transcript genes may also bind antigens derived from extra-cellular parasites in a process known as cross-presentation, a process already been reported (Heath and Carbone 2001, Ackerman and Cresswell 2004). Moreover, MHC class II may also bind pathogen proteins from intra-cellular parasites after its phagocytosis (Gregg 2011, Roche and Furuta 2015). In the specific case of intestinal parasites, both intra- and extra-cellular parasites have a contact with the gut associated lymphoid tissue. It is in in this tissue where there is an important presence of cells that are able to transcript MHC class II molecules (Yun et al. 2000, Mowat 2003), suggesting an important role of MHC class II genes to combat both intestinal helminths and protozoans. Alternative to the immune mechanisms associated with MHC genes, other mechanisms also play a key role to combat intestinal parasites. For example, inflammatory responses (i.e. a tissue reaction to injury or an antigen that may include pain, swelling, itching, redness, heat, and loss of function) have been found to be initiated once pathogens are colonized the intestinal surface (Mowat 2003). Also, phagocytosis mechanisms may also participate in the parasite elimination (Williams 2012). Moreover, MHC class I and II genes are supposed to be under

different selective pressures and thus may be differently selected according to those pressures (reviewed in Yeager and Hughes 1999, Kumánovics et al. 2003). Consequently, the study of the associations between MHC characteristics at both MHC class I and class II loci and parasites offers an excellent opportunity for a comprehensive study of the mechanisms underlying MHC-parasite associations.

Here, assuming that a part of the great variability of parasite load is explained by genetic characteristics of the immune system, we examined parasite abundance and prevalence (proportion of individuals infected by a given parasite species in the sample) of three common intestinal parasites detected by faecal egg counts (FEC) (two coccidia: *Eimeria monacis* and *Eimeria sp.*; and one cestode: *Ctenotaenia marmotae*) (Callait and Gauthier 2000), and their association with individual MHC functional characteristics at both MHC class I and class II in a wild population of Alpine marmots (*Marmota marmota*). The Alpine marmot is a medium-sized, long-lived, social and hibernating mammal living in family groups from 2 to 20 individuals (Allainé 2000). We tested whether (1) the abundance of *E. monacis*, (2) the prevalence of *Eimeria sp.* and (3) the prevalence of *C. marmotae* were associated with MHC diversity and/or the presence of specific alleles. We predicted that all intestinal parasites (*E. monacis*, *Eimeria sp.*, and *C. marmotae*) may be associated primarily with MHC class II loci, and that intra-cellular coccidian species (*E. monacis*, *Eimeria sp.*) may be also associated with MHC class I loci.

## 2. MATERIAL AND METHODS

### *Field methods and sample collection*

Data were collected between 2008 and 2014 in 24 family groups from a wild population of Alpine marmots located in the Nature Reserve of La Grande Sassièrè (at 2,340 m a.s.l., French Alps, 45°29'N, 65°90'E). Since 1990,

individuals from the population have been captured every year from mid-April to mid-July using two-door live traps baited with dandelions (*Taraxacum densleonis*) placed near the entrances of the main burrows in order to assign each captured individual to its family group. Once captured, faecal samples were collected into the traps and all baits and faeces were removed after each capture to reduce faecal contamination in the traps. Then, individuals were tranquillized with Zolétil 100 (combination of tiletamine and zolazepam, posology = 10 mg Kg<sup>-1</sup> by intramuscular route), sexed, aged from their size (up to 3 years), their social status was determined according to scrotal development for males and teat development for females. Captured marmots were individually marked with a transponder chip (model ID100, 0.9 cm long, <0.1 cm in diameter, Trovan Ltd, www.Trovan.com, Identifikationssysteme, Metternicher Straße 4, 53919 Weilerswist, Germany) injected under the skin of the neck for permanent individual recognition, and a numbered ear tag (1 cm x 3 mm).

The configuration of family groups, composed of a dominant couple; sexually mature (2 years and older) and immature (yearlings) subordinates of both sexes; and pups born that year (Allainé et al. 2000), was assessed from both capture–recapture data and intensive daily observations (see Cogas et al. 2006 for details on observations protocol). Moreover, scent marking behaviour was used to confirm the identity of the dominant pair (Bel et al. 1999).

### *Parasite identification and quantification*

340 faecal samples belonging to 194 individuals (mean faecal sample per individual  $\pm$  SD = 1.76  $\pm$  1.08; range: 1-7; faecal samples from the same individual were collected in different years) were used for parasite identification and quantification using faecal egg count (FEC). FEC is a non-destructive measure usually used for wildlife that allows realising several samples of the same host during several years. A modified Mac-Master technique (Cringoli et al. 2004, Pereckiené et al. 2007, Vadlejch et al. 2011) using saturated zinc sulphate solution (SG = specific gravity 1.36) was performed to count parasite's eggs,

oocysts or other parasite stages on a known quantity (3 or 5 g) of faeces from each sample (see Appendix 1 in Chapter 3 for details on Mac-Master technique). Parasite identifications were based on morphology and morphometrics of eggs and oocysts. The previously described species in Alpine marmots (two species of the genus *Eimeria* (coccidian protozoa), *Ctenotaenia marmotae* (cestode anoplocephalidae), *Ascaris laevis* (nematode ascaridae), *Citellina alpina* (nematode oxyuridae) and *Capillaria sp.* (nematode spiruridae)) were found in the present study (Bassano et al. 1992, Preleuthner et al. 1999, Callait and Gauthier 2000).

### *Studied host and parasite species*

#### *Alpine marmots*

Individuals of the same family group hibernate together between early October to mid-April and during their active season they accumulate enough body reserves to survive hibernation and initiate reproduction the following year (Körtner and Heldmaier 1995).

Before hibernation, the fresh tissue mass of the stomach, the small intestine, the caecum and the colon of Alpine marmots undergoes a strong decrease in mass leading to gastrointestinal tract probably operating at low level during hibernation (Carey and Cooke 1991, Hume et al. 2002) along with a great reduction of intestinal lumen. After emergence, the gastrointestinal tract increases until mid-summer (Hume et al. 2002). Using two different strategies, intestinal parasites of Alpine marmots need to be adapted when their environment becomes stressful (coinciding with hibernation of their host), either by passive avoidance or by active adaptation (Thomas et al. 1995). In the first case, parasites may be able to leave the gastro-intestinal tract by a “self curing” effect and to survive into the environment *via* extra-intestinal stages such as eggs or oocysts (Simitch and Petrovitch 1953, Davis 1969, Thomas et al. 1995). Then, a large decrease of parasite load is observed at the beginning of hibernation. In the second case,

parasites may be able to cope with low temperature and anatomical changes harbouring prolonged or arrested stages within host during hibernation (for example, Coggins et al. (1982) and Chute (1960) showed that helminths are able to survive in gut of bats and woodchucks (*Marmota monax*) respectively; although none of them clarify what are the survival forms). For Alpine marmot parasites, strategies during hibernation are probably different among parasite species. The biggest parasites, *C. marmotae* and *A. laevis*, leave the intestine at the beginning of hibernation; the small *C. alpina* may be able to survive in the caecum; nothing is proved about the coccidian species (Callait and Gauthier 2000). As a consequence, once Alpine marmots awake from hibernation and as soon as their digestive tracts increase in size and activity, parasite activity is expected to increase.

Despite the overdispersed and often high intestinal parasite numbers found in some Alpine marmots (e.g. in *C. marmotae* the mean  $\pm$  SD eggs per individual was  $102.8 \pm 131.9$  (N = 60) (MP Callait-Cardinal personal communication)), any visible pathological impact has been observed (Callait 1999, Bassano et al. 1992). Up to six intestinal parasites were found in the present study, but we only focus here on three species (two coccidia: *E. monacis* and *Eimeria sp.*; and the cestode *C. marmotae*), because the lack of data on *A. laevis*, *C. alpina* and *Capillaria sp.* due to the intrinsic parasite characteristics prevents us to integrate them in this study.

### *Coccidia*

The only oocysts of intestinal coccidian present in the collected faecal samples belong to the genus *Eimeria* (Bassano et al. 1992, Wilber et al. 1998), intracellular protozoans. All *Eimeria spp.* have a direct life cycle with an internal (parasitic) and external (environmental) phase (Daugochies and Najdrowski 2005). Individuals became infected by the direct ingestion of *Eimeria spp.* oocysts while they feed. They have both sexual and asexual multiplication phases within cells of the small or large intestines (Daugochies and Najdrowski 2005). Some *Eimeria spp.* cause severe intestinal lesions (e.g. severe diarrhoea with

sometimes lethal outcome) particularly in domestic animals (Daugochies and Najdrowski 2005 for cattle and De Gussem 2007 for poultry). However, clinical disease is not consistently associated with coccidial infection in wild mammals (Duszynski and Upton 2001) because the infection pressure is usually low.

Two species of the genus *Eimeria* were identified. The first species, corresponds to *E. monacis*, has small oocysts (20.3 x 17.8 µm) and is present in all faecal samples in our study (prevalence of 100%). The oocysts of the second species of the genus *Eimeria* are bigger (32.4 x 24.2 µm) and may correspond to the species *E. marmotae* (Galli-Valerio 1923). Unfortunately, this morphotype is not officially recognized as species because the original description does not meet the minimal criteria for complete species descriptions as set down in the International Code of Zoological Nomenclature *Eimeria spp.* from rodents of the Tribe Marmotini (Wilber et al. 1998). This second eimerian species is less common (prevalence of 23%) and it will be named here and throughout this thesis as *Eimeria sp.*

Only oocyst abundance (i.e. the number of faecal oocysts) of *E. monacis* was used to evaluate an eventual association with individual MHC characteristics since its prevalence was 100%. On the contrary, only the presence or absence of *Eimeria sp.* (coded as 1 and 0 respectively), was used to evaluate an eventual association with individual MHC characteristics. Indeed, the limited number of samples with the presence of *Eimeria sp.* (only 76 out of 332 faecal samples distributed between the 7 years of the study) makes difficult an accurate study of the factors that may be associated with the abundance of this parasite. We thus only use the presence of *Eimeria sp.* in this analysis.

### *Cestodes*

The only eggs of cestode present in the collected faecal samples belong to *Ctenotaenie marmotae*. This anoplocephalid tapeworm is an extra-cellular parasite with an indirect life cycle (Oribate mites are their intermediate hosts) (Ebermann 1976). Individuals became indirectly infected through the ingestion of the cysticercoïd stages present in the body cavity of the mites living in soils,



humus or grasses (Denegri 1993). Its complete development occurs in the small intestine of marmots, with a delay of almost 40 days between the ingestion of infested mites and the emission of eggs in the faeces. For this reason, and because all cestodes are eliminated during hibernation, marmot contamination restarts at the beginning of each active season and important seasonal variations of emission of eggs from *C. marmotae* are observed (Callait and Gauthier 2000). Indeed, no eggs of *C. marmotae* could be identified during the first month of the active season, followed by a gradual increase of prevalence to reach almost 55% at the beginning of hibernation. Since the entire mature proglottid (i.e. the segment of a tapeworm containing eggs) of this anoplocephalid cestode is expelled in the host-faeces (without real egg-laying), no correlation is usually observed between the number of eggs quantified in FEC and the number of tapeworm really present in host (Dever et al. 2015). As a consequence, the egg abundance could not be taken into account in our analyses and only its presence or absence (coded as 1 and 0 respectively), was used in the analysis.

### *MHC genotyping*

For 175 individuals, four MHC polymorphic loci - two from MHC class I exon 2 (*Mama-UB* and *Mama-UD*) and two from MHC class II DRB loci (*Mama-DRB1* and *Mama-DRB2*) - were genotyped using next generation sequencing, Sanger sequencing or were deduced based on mother-father-offspring triads (see Appendix 1 in Chapter 2 for details on genotyping methods and MHC characteristics). A total of 695 genotypes from these 175 individuals were obtained among the four MHC loci (502 using next generation sequencing, 178 using Sanger sequencing and 15 deduced on mother-father-offspring triads). Finally, a total of 5 expected genotype calls could not be obtained despite our efforts to do so.

### *Genetic characteristics estimators*

To test for an association between MHC characteristics and the abundance of *E.*



*monacis*, the prevalence of *Eimeria sp.* and the prevalence of *C. marmotae*, the MHC diversity as well as the presence/absence of specific MHC proteins was estimated for each individual. MHC estimators were calculated over the only two MHC loci, *Mama-UD* and *Mama-DRB1*, that led to polymorphic proteins (three and eight alleles resulting in two and seven proteins respectively), as well as over *Mama-UD* and *Mama-DRB1* considered separately because in Alpine marmots MHC class I and II genes are under different selective pressures (Kuduk et al. 2012). And, indeed, MHC class I and II estimators are not correlated in this species, suggesting that both classes could convey different information (Chapter 2). Only individuals successfully genotyped at both MHC loci, solely at the MHC class I locus or solely at the MHC class II locus were used to calculate the MHC, MHC class I and MHC class II estimators respectively.

Two indexes were used to estimate the MHC diversity: the MHC protein diversity and the MHC protein distance. The MHC protein diversity was estimated as the number of different MHC proteins an individual produces once the protein translated by the allele *Mama-DRB1\*08* was discarded since it has a stop codon and therefore the translated protein is non-functional (Kuduk et al. 2012). The MHC protein distance was calculated using the software MEGA version 6.06 (Tamura et al. 2013) as the sum of the p-distances (i.e. the proportion of amino acid sites that differs between two alleles) among the considered MHC loci except for the p-distance between a given protein and the one translated from *Mama-DRB1\*08* that was set to 0. All MHC diversity estimators were further calculated considering alleles instead of proteins (see Appendix 3 in Chapter 3 for details on the calculation of the MHC allelic diversity estimators). The different indexes used to estimate the MHC diversity were found to be highly correlated, except for MHC class I allelic and protein diversity, suggesting that in Alpine marmots these different estimators partly convey similar information (Chapter 2).

### *Statistical analysis*

Generalized Linear Mixed Models (GLMM) were used to investigate whether the abundance of *E. monacis*, the presence of *Eimeria sp.* or the presence of *C. marmotae* was influenced by individual MHC diversity and the presence of specific MHC functional alleles. Three series of GLMM models were conducted using the function “glmmPQL” of the R package “MASS” (Venables and Ripley 2002), with the abundance of *E. monacis*, the presence of *Eimeria sp.* or the presence of *C. marmotae* as dependent variables respectively. In the models with the abundance of *E. monacis*, GLMM were used with an identity link and a variance given by a Gaussian distribution whereas, in the models with the presence of *Eimeria sp.* and *C. marmotae*, GLMM were used with a logit link and a variance given by a binomial distribution. To control for individuals with multiple samples and for individuals from the same family group, individual identity and family group identity were included as random effects, respectively. To control for the changes in parasite load occurring during the active period of Alpine marmots (Callait 1999, Callait and Gauthier 2000) the sampling date (coded as the Julian date) was systematically included as a continuous variable.

Since the abundance and/or the presence of parasites in host faeces may depend on many factors other than MHC characteristics, the analysis was conducted in three steps. In the first step, we tested whether potential confounding variables other than MHC characteristics affected the parasite variables. Specifically, we tested whether (1) the year of sampling, (2) the local climatic variables (mean air winter temperatures, mean air spring temperatures, an index of vegetation drought), (3) the global climatic variables (the global Normalized Differential Vegetation Index (NDVI) and the North Atlantic Oscillation (NAO)), (4) the family group characteristics (orientation of a marmot's territory, family group size and family group density), and (5) the individual characteristics (age class, sex, social status and genome-wide diversity) influence parasite dynamics of each species (see Appendix 2 in Chapter 3 for details). Then, a baseline model for each parasite species was

selected by retaining only the confounding variables with statistically significant (or nearly statistically significant) effects to obtain a baseline model for subsequent analysis of each parasite species: *E. monacis* (Table 35 in Appendix 2, Chapter 3), *Eimeria sp.* (Table 36 in Appendix 2, Chapter 3) and *C. Marmotae* (Table 37 in Appendix 2, Chapter 3).

In the step two, we then investigated whether the abundance of *E. monacis*, the prevalence of *Eimeria sp.* or the prevalence of *C. marmotae* were related to MHC diversity. Using the baseline model for each parasite species, each genetic estimator was then added separately. Both a linear and a quadratic effects were considered for all genetic diversity estimators to test for an optimal number of MHC diversity.

In the step three, we investigated whether the abundance of *E. monacis*, the prevalence of *Eimeria sp.* or the prevalence of *C. marmotae* were related to the presence of a given protein. Using the baseline model for each parasite species, the presence or the absence of each MHC protein (coded as 1 and 0 respectively) was then added separately. To avoid statistical problems derived from both the low and the high number of individuals carrying a specific protein, we focused on the 6 proteins with a frequency higher than 5% and lower than 95% (see Table 39 in Appendix 3, Chapter 3 - for details on the frequency of each protein).

The abundance of *E. monacis* was log-transformed, to correct for heterogeneity of variance. All statistical analyses were conducted using the R software version 3.0.3 (R Development Core Team, 2013). To control for multiple tests, we applied a modified false discovery rate procedure by Benjamini and Yekutieli (2001), the alternative to the Bonferroni correction advocated for genetic and ecological studies (Verhoeven et al. 2005, Narum 2006). Therefore, the results were considered statistically significant if the p-value was lower than 0.01068 (p-value associated to Benjamini and Yekutieli (2001) correction). Given that the first *C. marmotae* eggs are observed at the middle of the capture period (*C. marmotae* egg emission is delayed because of parasite maturation in host before emitting eggs), the analysis of *C. marmotae* only included those faeces collected from the first day of *C. marmotae* egg emission. Moreover, samples

Chapter 3: MHC characteristics and parasites collected in 2013 could not be included in the statistical analysis of *C. marmotae* due to the small sample size (<10 faecal samples). Thus, a total of 225 out of 340 faecal samples were used in the statistical analysis of *C. marmotae*. Finally, since MHC genetic estimators were not available for all individuals, sample sizes vary in the statistical analysis (see Table 33 and Table 38 for sample sizes). Analysis conducted with the MHC diversity calculated considering alleles instead of proteins gave qualitatively similar results, therefore only the results obtained with proteins are presented (see Appendix 3 in Chapter 3, Table 38 for results obtained considering alleles).

### 3. RESULTS

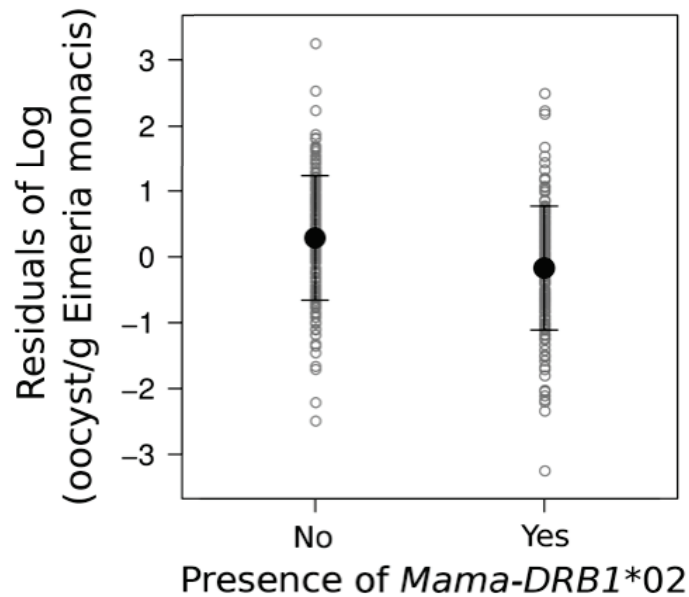
#### *E. monacis*

The mean number of *E. monacis* per gram of faeces was 24350 (SD  $\pm$  34185; range: 200 – 274200; N = 340). The abundance of *E. monacis* was found to depend on age class. Specifically, yearlings exhibited a higher abundance of *E. monacis* than did two-years old and adults, with no difference between these two last classes (Table 35 in Appendix 2, Chapter 3). Moreover, the abundance of *E. monacis* was found to vary between years (Table 35 in Appendix 2, Chapter 3).

The abundance of *E. monacis* did not depend on individual MHC genetic estimators considered (protein diversity or protein distance) whatever the loci considered (the two MHC loci, the MHC class I locus and the MHC class II locus) after Benjamini and Yekutieli (2001) correction (Table 33). Surprisingly, the abundance of *E. monacis* tended to depend on a quadratic relationship with the MHC protein diversity at the two MHC loci after Benjamini and Yekutieli (2001) (Table 33), being higher for intermediate MHC protein diversity levels.

Individuals carrying the protein *Mama-DRB1\*02* showed significantly lower abundance of *E. monacis* than individuals lacking this protein (Table 34, Figure

16). However, the abundance of *E. monacis* did not depend on the presence of any other specific MHC protein (Table 34).



**Figure 16.** Residual of the abundance of *Eimeria monacis* as a function of the presence of the *Mama-DRB1\*02* protein (N = 316). Residuals are corrected for confounding factors (the Julian date of sample collection, the year of sample collection and the age class). Open circles represent the observed residual abundance of *E. monacis* and the black dots represent observed data averaged over either the presence or the absence of the *Mama-DRB1\*02* protein.

### *Eimeria sp.*

From a total of 332 analysed faecal samples, the presence of *Eimeria sp.* was found in 76 faecal samples (mean number of *Eimeria sp.* per gram of faeces: 1772; SD  $\pm$  4706; range: 50 – 34550; N = 77). Thus, the apparent prevalence of *Eimeria sp.* in the studied population is 23%.

The presence of *Eimeria sp.* was found to depend on age class. Specifically, yearlings exhibited a higher probability to present *Eimeria sp.* than did two-years old and adults, with no difference between these two last classes (Table 36 in Appendix 2, Chapter 3). Moreover, presence of *Eimeria sp.* was found to increase lightly as the capture period progressed and to vary between years (Table 36 in Appendix 2, Chapter 3).

The presence of *Eimeria sp.* did not depend on individual MHC genetic estimators considered (protein diversity or protein distance) whatever the loci considered (the two MHC loci, the MHC class I locus and the MHC class II locus) (Table 33).

The presence of *Eimeria sp.* neither depended on the presence of any specific MHC protein (Table 34).

### *C. marmotae*

From a total of 225 analysed faecal samples, the presence of *C. marmotae* was found in 90 faecal samples. Thus, the prevalence of *C. marmotae* in the studied population is 40%. As expected, the number of individuals infected by *C. marmotae* was found to increase as the capture period progressed (Table 37 in Appendix 2, Chapter 3).

The presence of *C. marmotae* did not depend on individual MHC genetic estimators considered (protein diversity or protein distance) whatever the loci considered (the two MHC loci, the MHC class I locus and the MHC class II locus) (Table 33).

The presence of *C. marmotae* neither depended on the presence of any specific MHC protein (Table 34), although individuals carrying the protein *Mama-DRB1\*02* tended to show a higher presence of *C. marmotae* than individuals lacking this protein (Table 34).

Dependent variable	Independent variable	estimate $\pm$ SE	P	N
Abundance of <i>E. monacis</i>	MHC protein diversity	2.17 $\pm$ 0.97	0.03	316
	(MHC protein diversity) <sup>2</sup>	-0.38 $\pm$ 0.16	0.02	
	MHC class I protein diversity	-0.22 $\pm$ 0.21	0.29	321
	MHC class II protein diversity	-0.01 $\pm$ 0.18	0.94	316
	MHC protein distance	-1.59 $\pm$ 2.07	0.44	316
	MHC class I protein distance	-12.92 $\pm$ 12.10	0.29	321
	MHC class II protein distance	-1.30 $\pm$ 2.12	0.54	316
Presence of <i>Eimeria sp.</i>	MHC protein diversity	-0.23 $\pm$ 0.22	0.30	307
	MHC class I protein diversity	-0.13 $\pm$ 0.35	0.71	312
	MHC class II protein diversity	-0.37 $\pm$ 0.32	0.25	307
	MHC protein distance	-0.93 $\pm$ 3.68	0.80	307
	MHC class I protein distance	-7.86 $\pm$ 20.76	0.71	312
	MHC class II protein distance	-0.67 $\pm$ 3.80	0.86	307
	Presence of <i>C. marmotae</i>	MHC protein diversity	0.08 $\pm$ 0.25	0.74
MHC class I protein diversity		-0.48 $\pm$ 0.41	0.24	219
MHC class II protein diversity		0.60 $\pm$ 0.39	0.12	214
MHC protein distance		6.87 $\pm$ 4.30	0.11	214
MHC class I protein distance		-28.14 $\pm$ 24.03	0.24	219
MHC class II protein distance		8.31 $\pm$ 4.47	0.07	214

**Table 33.** Generalized linear mixed models showing the effects of the MHC protein characteristics of individuals on the abundance of *E. monacis*, the presence of *Eimeria sp.* and the presence of *C. marmotae*. The MHC diversity estimators were calculated over the two MHC loci, over the MHC class I locus and over the MHC class II locus. All the models with the abundance of *E. monacis* and the presence of *Eimeria sp.* include the year, the Julian date and age class; and all the models with the presence of *C. marmotae* include the year and the Julian date. The critical p-value after Benjamini and Yekutieli (2001) correction correspond to 0.01068. N refers to the samples examined.

Dependent variable	Independent variable	estimate $\pm$ SE	P	N
Abundance of <i>E. monacis</i>	Presence of <i>Mama-UD*02</i>	-0.23 $\pm$ 0.20	0.26	321
	Presence of <i>Mama-DRB1*01</i>	0.06 $\pm$ 0.17	0.73	316
	<b>Presence of <i>Mama-DRB1*02</i></b>	<b>-0.46 <math>\pm</math> 0.16</b>	<b>0.005</b>	316
	Presence of <i>Mama-DRB1*03</i>	0.30 $\pm$ 0.18	0.10	316
	Presence of <i>Mama-DRB1*06</i>	0.31 $\pm$ 0.21	0.14	316
	Presence of <i>Mama-DRB1*07</i>	0.08 $\pm$ 0.23	0.72	316
	Presence of <i>Mama-DRB1*08</i>	0.16 $\pm$ 0.28	0.58	316
	Presence of <i>Eimeria sp.</i>	Presence of <i>Mama-UD*02</i>	-0.03 $\pm$ 0.33	0.93
Presence of <i>Mama-DRB1*01</i>		0.06 $\pm$ 0.29	0.84	307
Presence of <i>Mama-DRB1*02</i>		-0.03 $\pm$ 0.29	0.92	307
Presence of <i>Mama-DRB1*03</i>		0.15 $\pm$ 0.31	0.62	307
Presence of <i>Mama-DRB1*06</i>		-0.20 $\pm$ 0.36	0.57	307
Presence of <i>Mama-DRB1*07</i>		-0.45 $\pm$ 0.42	0.29	307
Presence of <i>Mama-DRB1*08</i>		0.32 $\pm$ 0.46	0.48	307
Presence of <i>C. marmotae</i>		Presence of <i>Mama-UD*02</i>	-0.26 $\pm$ 0.41	0.53
	Presence of <i>Mama-DRB1*01</i>	-0.13 $\pm$ 0.35	0.71	216
	Presence of <i>Mama-DRB1*02</i>	0.57 $\pm$ 0.34	0.09	216
	Presence of <i>Mama-DRB1*03</i>	0.56 $\pm$ 0.38	0.15	216
	Presence of <i>Mama-DRB1*06</i>	0.07 $\pm$ 0.41	0.87	216
	Presence of <i>Mama-DRB1*07</i>	-0.73 $\pm$ 0.49	0.14	216
	Presence of <i>Mama-DRB1*08</i>	-0.53 $\pm$ 0.60	0.38	216

**Table 34.** Generalized linear mixed models showing the effects of the presence or absence of each allele leading to polymorphic proteins of individuals on the abundance of *E. monacis*, the presence of *Eimeria sp.* and the presence of *C. marmotae*. All the models with the abundance of *E. monacis* and the presence of *Eimeria sp.* include the year, the Julian date and age class; and all the models with the presence of *C. marmotae* include the year and the Julian date. P-values after Benjamini and Yekutieli (2001) correction (p-value = 0.01068) are indicated in bold. N refers to the samples examined.



## 4. DISCUSSION

In this study, we analysed the influence of individual MHC functional characteristics – MHC diversity and the presence of specific MHC alleles – on the abundance or the presence of three intestinal parasites (*E. monacis*, *Eimeria sp.* and *C. marmotae*) in a wild population of Alpine marmots. Once potential confounding factors were corrected (sampling date, year and age class, the later factor only for *E. monacis* and *Eimeria sp.* analysis), MHC characteristics were not associated with the examined parasites, except for an association between the abundance of *E. monacis* and the presence of the functional allele *Mama-DRB1\*02*. Specifically, individuals with *Mama-DRB1\*02* had a lower abundance of *E. monacis*. Neither the local climatic variables, the global climatic variables, the family group characteristics nor the individual characteristics (except age class for *E. monacis* and *Eimeria sp.* analysis) were found to influence parasite dynamics.

Our results indicate that MHC diversity was not associated with the abundance nor with the presence of any intestinal parasites. In the commonly used hypothesis that parasite load reflects the immune performance of their hosts, these results suggest a weak or an absent heterozygote advantage in Alpine marmots. The low diversity usually observed for the core parasite species of Alpine marmots (Bassano 1996) along with an apparently absence of pathogenicity of their intestinal parasites (Callait and Gauthier 2000), may explain the absence of heterozygote advantage, since this mechanism has been proposed as able to maintain high allelic diversity in populations exposed to multiple pathogens (Penn et al. 2002). This observation found in Alpine marmots contrasts with both higher parasite diversity (Cohn et al. 1986, Bassano 1996) and higher MHC diversity (Zhou et al 2003, Moreno-Cugnon et al. 2015) found in woodchucks (*Marmota monax*), another marmot species widely distributed in North America (Figure 1). Woodchuck parasites include intestinal parasites,

Chapter 3: MHC characteristics and parasites viruses and bacteria (Cohn et al. 1986, Bassano 1996), whereas in Alpine marmots viruses and bacteria have not been identified to date (Bassano 1996). In addition, woodchucks have 20 MHC class I and 15 MHC class II described alleles (Zhou et al 2003, Moreno-Cugnon et al. 2015), compared to the 5 MHC class I and 11 MHC class II described alleles in Alpine marmots (Kuduk et al. 2012, Chapter 1). Unfortunately, to date, whether or not MHC diversity participates in the defence against woodchuck parasites remains to be studied.

In agreement with the fact that resistance enable the host to limit parasite burden and to reduce parasite reproductive rates (Råberg et al. 2007), Alpine marmots carrying the allele *MamaDRB1\*02* have a low abundance of *E. monacis*, which suggests an increase of resistance conferred by this specific allele. Similarly, the presence of specific MHC class II alleles has also been associated with a decrease in the abundance of other coccidian species: *E. stiedai* in European rabbits (*Oryctolagus cuniculus*) (Oppelt et al. 2010) or *E. melis* in European badgers (*Meles meles*) (Sin et al. 2014). However, it is surprising to observe an association between the abundance of *E. monacis* and the specific allele *MamaDRB1\*02*, whereas no association is evidenced between the presence of the closer species *Eimeria sp.* and any specific MHC allele. This results accords with the absence of cross-protection usually observed between different *Eimeria spp.* in the same host (Smith et al. 2002, Chapman et al. 2013). As a consequence, immunity towards a given eimerian species would not confer protection against other eimerian species. Similarly, a study on the European rabbit (Oppelt et al. 2010) found an association between a specific allele and the abundance of *E. stiedai*, whereas no association was found with *Eimeria sp.*. In Alpine marmots, the very contrasted infestation patterns between the two eimerian species may be explained by alternative mechanisms. First, whereas all Alpine marmots are infested by *E. monacis*, a lower number of individuals are infested by *Eimeria sp.* (prevalence of 23%). However, given that *Eimeria sp.* is equally distributed among family groups (*Eimeria sp.* was present across years in 22 out of 24 examined family groups), differences among territories should not be responsible for differences among prevalence of these two parasites. Second,

*E. monacis* excretion is more abundant than *Eimeria sp.* (mean oocyst excretion of *E. monacis* is thirteen times more abundant than the mean oocyst excretion of *Eimeria sp.*). The low prevalence associated with the low abundance of *Eimeria sp.* suggests that Alpine marmots may have a better resistance towards *Eimeria sp.* than towards *E. monacis*. If true, we would expect an association between a specific MHC allele and *Eimeria sp.* rather than with *E. monacis*. This unexpected result may be also due to the difference in immune response generated by host if *Eimeria spp.* are not equally located in the gut (Yun et al. 2000). A difference in immune response depending on *Eimeria spp.* localization may be generated because of (1) the difference of lymphoid tissue distribution according to the intestine areas (Mowat 2003) and (2) the difference of MHC class II molecules expression, which seem to be strongly expressed in the small intestinal epithelium than in the large intestine (Bland 1988, Steiniger et al. 1989, Pitman and Blumberg 2000 in rats; German et al. 1998 in dogs). Exploration of the precisely location of the multiplication phases of the two *Eimeria spp.* of Alpine marmots by the use of histological methods seems thus essential to understand the existing differences between these two closer species. Finally, resistance towards *Eimeria spp.* should depend on immune responses other than MHC characteristics. For example, oocyst output has been found to be rapidly reduced with more rapid inflammatory response, without involvement of MHC antigen presentation in experimental studies in mice during a primary infection with *Eimeria sp.* (Ovington et al. 1995, Schito et al. 1998, Chapman et al. 2013). Also, in an experimentally infection in chickens, Toll-like receptors (a family of proteins involved in innate immune response Owen et al. 2013) and playing important roles in the reduction and clearance of pathogens) were found to be associated with *E. tenella* (Zhang et al. 2012). We would like to highlight the idiosyncratic nature of the response to parasites and to emphasize that, even for related species, it may be hazardous to extrapolate host-parasite interactions from one species to another.

The presence of *C. marmotae* was found not to be associated with any MHC allele, indicating that resistance towards this intestinal parasite seem not to be

due to MHC characteristics. Compared to *Eimeria spp.*, the cestode *C. marmotae* is a strict extra-cellular intestinal parasite. As a consequence, the immune pathways to combat adult cestodes should largely differ from those used to combat intra-cellular or intra-tissular parasites. First, strict intestinal parasites are separated from the mucosa by the enteric epithelial barrier, which difficult the host capacity to detect their presence (McKay and Webb 2006). Second, local immunity has been hypothesized to be a main immune pathway to combat cestodes closer to *C. marmotae* such as *Anoplocephala perfoliata*, a cestode that parasites horses (Pittaway et al. 2014). Third, in order to combat large adult cestodes, hosts may use alternative defensive mechanisms such as neurotransmitters (McKay 2010), immune modulation (Wang and McKay 2005, McKay 2010), modifications of the stomachal pH or production of mucus (McKay and Webb 2006). Four, for cestodes such as *A. perfoliata* in horses or *Hymenolepis diminuta* in rats, the role of acquired immunity (which includes the role of MHC) in controlling cestode infestation is still unproven (Pittaway et al. 2014). Overall, the complexity of immune mechanisms used to combat adult cestodes suggests that it might be difficult to show an association between any of the involved mechanisms, including MHC characteristics, in studies carried out in the wild. Given that immune pathways to combat adult cestodes have largely been understudied (McKay and Webb 2006, Terrazas et al. 2012, Pittaway et al. 2014), further studies are thus required.

According to a common observed pattern in host–parasite assemblages (reviewed in Hudson and Dobson 1997), age class was associated with *E. monacis* and *Eimeria sp.*. Specifically, yearlings exhibited a higher abundance of *E. monacis* and a higher presence of *Eimeria sp.* than did two-years old and adults, although no excretion differences were observed between two-years old and adults. Similarly, *Eimeria spp.* FEC was found to decrease with age in European badgers (Sin et al. 2014), European rabbits (Oppelt et al. 2010) and cattles (Bangoura et al. 2012). A high parasite load in young individuals has been hypothesized to be due the development of an acquired immunity in young individuals ('immunity' hypothesis, Hudson and Dobson 1997). As a

consequence, they are supposed to be more prone to parasite infections than are adults. Two experimental studies investigating *Eimeria spp.* oocyst emission in mice (Schito et al. 1998) and rats (Shi et al. 2000) found that, during the primary infection there is a quickly peak of oocyst excretion followed by a decline of oocyst excretion (individuals nearly reach zero excreted oocysts), whereas during the secondary infection there is an oocyst peak excretion significantly lower and shorter. If we extrapolate from known information in model species (Schito et al. 1998, Shi et al. 2000) and given that Alpine marmots are systematically re-infected, one may parsimoniously think that the observed decrease of the abundance of *E. monacis* and *Eimeria sp.* with age is related to the establishment of acquired immunity. However, caution is required since the demonstration of a real effect of the acquired immunity is extremely complex in the wildlife (Bradley and Jackson 2008) because co-infections, which are common in the wild (Petney and Andrews 1998), may alter or modify the immune responses of their hosts. The absence of a decrease in oocyst emission between two years-old and adults may be due to a possible impact of hibernation, which may drastically reduce all types of circulating leukocytes during torpor and thus reduce host's innate and adaptive immune system capacities (Bouma et al. 2010). The same processes may explain the absence of decrease in *C. marmotae* prevalence with age class. Indeed, infection with *C. marmotae* may occurs from weaning (5 out of 6 examined juveniles were infected with *C. marmotae* before hibernation – MP Callait-Cardinal personal communication) and may systematically re-start throughout their life (Callait and Gauthier 2000). Disentangling a possible impact of hibernation on the maintenance of immunity remains an important issue that requires further investigation.

We have focused here to search for an association between MHC characteristics and parasite infection level, a measure of host's resistance capacity as a function of the acquired immune system aiming to detect and eliminate invading pathogens. Our results suggest that MHC characteristics play a weak role on the associations between the Alpine marmots and their intestinal parasites. However, host's defensive capacity against parasites may not only

depend on resistance, but also on parasite tolerance (Restif and Koella 2004, Råberg et al. 2009). This defensive strategy acts reducing the detrimental effects of parasites without altering its development and without reducing its reproductive rates (Restif and Koella 2004). In this view, tolerance should have a neutral or positive effect on parasite prevalence (Råberg et al. 2009) and, once established, would be able to be not so costly than resistance (Miller et al. 2006). As a consequence, tolerance may be useful when hosts need to cope with hard environmental pressures and favour the less costly strategy (Read et al. 2008). Finally, tolerance has been hypothesized to be evolutionary advantageous (Boots et al. 2009, Schulenburg et al. 2009). Usually, the contribution of tolerance in a host-parasite system can be demonstrated by plotting parasite burden against health outcome (or fitness): the flatter the slope, the better the tolerance (Råberg, 2007, Read et al. 2008). So, unlike resistance, tolerance cannot be measured on a single animal, but across individuals of a given host group. It may be very interesting to study tolerance in host-parasite systems with high prevalence associated with apparent low negative impact from parasites (seemingly commensalism) (Miller et al. 2006) as the Alpine marmot and its intestinal parasites. In our wild host-parasite system, the difficulty for further investigations on tolerance is to find a relevant and easily measure of health or fitness.

If to date Alpine marmots seem not to suffer from parasites, climate change may alter this apparent equilibrium. Climate change has been associated to an increased risk of infectious disease outbreaks due to the extension of the seasonal parasite activity and to the increase of transmission rate (Hernandez et al. 2013). Moreover, hibernating species have been hypothesized to be directly threatened by climate change and the introduction of foreign parasites (Geiser 2013). Indeed, hibernating species greatly decrease, although do not complete suppress, their immunological functions during hibernation (Bouma et al. 2010). As a consequence, they may be more vulnerable during hibernation periods. For example, populations of hibernating bats in North America are being devastated due to a fungal infection, known as white nose syndrome, that may easily invade body tissues due to the low immunological functions during hibernation (Cryan

Chapter 3: MHC characteristics and parasites et al. 2010). Quantifying and interpreting how climate change may alter host-parasite interactions in hibernating species is essential for understanding the ecological and evolutionary implications of parasites on their hosts (Hueffer et al. 2011) that awaits further investigation.

## 5. APPENDIX

### *Appendix 1. Mac-Master technique used to identify and quantify the presence/absence and the abundance of Alpine marmot parasite species*

Each aliquot was placed in a labelled shell vial and thoroughly homogenized with 15 mL/g faeces of zinc sulphate solution; the faecal suspension was then sieved into a new vial to remove large debris; during agitation of the filtrate, an aliquot was taken from the centre of the vial and rapidly pipetted into a single chamber of a Mac-Master slide (2 chambers McMaster Slides (acrylic microscope slides) ; size: 7.5 cm /3 cm; Thomas E Krecek, 908 Ladove Drive, College Station, texas 77845, USA ) to avoid formation of bubbles; the suspension was agitated a second time to fill the second chamber. Each filled Mac-Master slide was left to stand for 3-5 min. before microscopic examination, in order to allow the eggs to rise to the surface of the Mac-Master slide. The slide was examined with the 10x lens, focusing on the top layer and each type of eggs, oocysts or other parasite stages were counted separately. The number of eggs or oocysts was always counted in the two grids of the Mac-Master slide (0.3 mL = 2 x 0.15 mL). The counts, expressed as eggs/g faeces (EPG) or oocysts/g faeces (OPG), were averaged using the Mac-Master multiplication factor depending on the faecal sample dilution and the Mac-Master slide area (volume) examined as follows: EPG or OPG =  $n (d/v)$  where  $n$  is the number of eggs or oocysts counted under



the volume examined,  $d$  the dilution factor (here, 15) and  $v$  the volume examined (here, 0.3 mL).

In order to detect eggs or oocysts when they are in small quantity (near the detection limit of 50 EPG or OPG), a flotation coverslip was performed using the rest of the suspension poured into a 15 mL centrifuge tube until a slight positive meniscus was formed; the coverslip was placed, delicately, on the tube before centrifuge at 2000 rpm for 2 minutes; after let it stand for 3-5 minutes, the coverslip was then removed to a microscope slide and was entirely examined using the 10x or 40x objective lens to confirm the diagnosis. If one or several eggs or oocysts from one parasite species were detected in the flotation coverslip but were absent in the Mac-Master slide, the count was <50 EPG or OPG.

## *Appendix 2. Detailed statistical analyses to select a baseline model for each parasite species*

### *Description of confounding variables*

#### Temporal and climatic variables

Parasite dynamics depend on environmental factors varying in time (Schmid-Hempel 2011). In mountains, environmental conditions vary sharply from one year to the next. Thus, parasite prevalence and/or abundance are likely also to vary from one year to the next. Consequently, the year of sampling was included to account for the inter-annual variability. Moreover, a total of three local (mean winter air temperatures, mean spring air temperatures and the Bagnouls Gaussens Index (BGI, Gaillard et al. 1997); all measures were daily recorded from the weather stations of Tignes and Val d'Isère, located 5 km away from the study site) and two global indices (the global Normalized Differential Vegetation Index (NDVI) and the North Atlantic Oscillation (NAO, Hurrell 1995)) were used to assess whether climate characteristics affect either the abundance of *E. monacis*, the presence of *Eimeria* sp. or the presence of *C. marmotae*. Winter variables included mean air temperatures and the NAO from December to March as a



measure of winter severity, which determine the onset of vegetation and in turn may impact parasite dynamics. Spring variables included mean air temperatures from April to mid-June and the NDVI, a proxy of vegetation productivity (Pettorelli et al. 2005). NDVI was measured in the second half of April, to detect any change in the timing of spring vegetation onset and thus in the forage availability and parasite transmission for Alpine marmots. Finally, the summer variable included the BGI, an index of vegetation drought. Indeed, oocysts and eggs survival is likely to be affected by drought. For example, a study on *E. zurnii* of cattle, oocysts were found to be susceptible to desiccation (Marquardt 1960). All climatic characteristics of the year preceding the collection of each faecal sample were used to analyse the influence of past climate variations on parasite dynamics.

#### Family group characteristics

On one hand, the orientation of a marmot's territory (i.e., in the valley or oriented South) may contribute in parasite infection because in the valley snow lasts longer, so that parasite dynamics from these territories are expected to differ than parasites in the South. Thus, the orientation of family groups was included in the analysis. On the other hand, for many parasites, transmission depends on host population density. Accordingly, in two phylogenetically controlled frameworks, population density has been shown to correlate with parasite abundance across mammals (Arneberg et al. 1998, Nunn et al. 2003). In a study on yellow-bellied marmots (*Marmota flaviventris*), the relationships between parasite prevalence and host population size were diverse and complex (Lopez et al. 2013). Whereas the prevalence of *Ascaris* sp. and parasite diversity was associated to yellow-bellied marmot colony size, the prevalence of other parasites (*Entamoeba* sp., *Trypanosoma Lewisi*, *Eimeria* sp. and *Thrassis stanfordi*) was not (Lopez et al. 2013). In the specific case of Alpine marmots we expect a positive relationship between family group size and the presence and/or abundance of intestinal parasites. Thus, family group size of the year preceding the collection of each faecal sample the year was included in the analysis.

### Individual characteristics

Three age classes (yearlings, two-years old and adults; parasite identification and quantification could not be evaluated for juveniles because their faecal collection is often null or too small to be reliable with the same FEC method used for all samples) were included in the models because higher parasite load in young individuals is a common pattern observed in host–parasite interactions (reviewed in Hudson and Dobson 1997). The sex was also included in the analysis since hormonally mediated sex differences in parasite susceptibility are expected to in mammals (Schalk and Forbes 1997). The social status (subordinate or dominant) was included since subordinates and dominants are likely to be under different stress pressures, a factor that may make hosts more susceptible to parasitism. Finally, the genetic diversity at neutral loci (measured using 16 microsatellites) was included since genetically more diverse individuals are expected to (1) have a higher probability of carrying adaptive alleles implicated in immune response (Coltman et al. 1999, Keller and Waller 2002, Puurtinen et al. 2004) and (2) have a lower probability of diseases susceptibility driven by overdominance at genes not directly involved in immune response, but involved in the capacity to clear and survive an infection (Coltman et al. 1999). To this purpose, all 194 individuals were genotyped at 16 microsatellites: SS-Bib1, SS-Bib18, SS-Bib20, SS-Bib31, SS-Bib4 (Klinkicht 1993); MS41, MS45, MS47, MS53, MS56, MS6, ST10 (Hanslik and Kruckenhauser 2000); Ma002, Ma018, Ma066, Ma091 (Da Silva et al. 2003) (see see Appendix 2 in Chapter 1 for details on genotyping methods and microsatellite characteristics). Genetic diversity at neutral loci was estimated for each individual as the standardized heterozygosity (SH, Coltman et al. 1999) calculated with the R function “GENHET v3.1” (Coulon 2010).

### *Results*

We took into account all these potentially confounding effects and retained the

statistically or nearly statistically significant ones to obtain a baseline model for subsequent analysis of each parasite species: *E. monacis* (Table 35), *Eimeria sp.* (Table 36) and *C. marmotae* (Table 37).

(a) Baseline model		Log <i>Eimeria monacis</i> = Julian date + F(year) + F(age class) + (1 family group)+ (1 id)						
(b)		Parameters of the baseline model						
Dependent variable	Distribution	Independent variable	estimate ± SE	p-value	Wald test			
					$\chi^2$	df	p-value	
Log <i>Eimeria monacis</i>	Binomial	Intercept	9.03 ± 0.99	<0.01	85.30	1	<0.01	
		Julian date	0.01 ± 0.01	0.49	0.49	1	0.48	
		Year	2009	0.34 ± 0.30	0.25	17.10	6	0.01
			2010	0.28 ± 0.30	0.36			
			2011	0.02 ± 0.36	0.95			
			2012	0.01 ± 0.30	0.99			
			2013	-1.01 ± 0.44	0.02			
			2014	0.23 ± 0.34	0.51			
			Age class	2 years	-0.68 ± 0.19	<0.01	14.90	2
		Adults	-0.43 ± 0.16	0.01				

**Table 35.** Baseline model (a) obtained after selection of the confounding variables for the analysis of the abundance of *Eimeria monacis* and its associated parameters. *F(variable)* indicates that the variable is entered as a factor in the model. *(1|variable)* indicates that the variable is entered as a random effect on the intercept (b) (N = 340). *F(variable)* indicates that the variable is entered as a factor in the model. *(1|variable)* indicates that the variable is entered as a random effect on the intercept. We used Wald's test to measure the contribution of each categorical factor to the overall model, and include the statistic ( $\chi^2$ ) and the degrees of freedom (df).

(a) Baseline model		Presence of <i>Eimeria sp.</i> = Julian date + F(year) + F(age class) + (1 family group)+ (1 id)						
(b)		Parameters of the baseline model						
Dependent variable	Distribution	Independent variable	estimate ± SE	p-value	Wald test			
					χ <sup>2</sup>	df	p-value	
Presence of <i>Eimeria sp.</i>	Binomial	Intercept	-5.49 ± 2.05	<0.01	7.4	1	0.01	
		Julian date	0.02 ± 0.01	0.04	4.6	1	0.03	
		Year	2009	1.36 ± 0.68	0.05	11.20	6	0.08
			2010	1.06 ± 0.70	0.13			
			2011	2.15 ± 0.78	0.01			
			2012	1.03 ± 0.69	0.14			
			2013	2.28 ± 0.94	0.02			
			2014	1.38 ± 0.76	0.07			
			Age class	2 years	-0.89 ± 0.41	0.03	12.0	2
		Adults	-1.04 ± 0.32	<0.01				

**Table 36.** Baseline model (a) obtained after selection of the confounding variables for the analysis of the presence of *Eimeria sp.* and its associated parameters. *F(variable)* indicates that the variable is entered as a factor in the model. *(1|variable)* indicates that the variable is entered as a random effect on the intercept. (b) (N = 331). *F(variable)* indicates that the variable is entered as a factor in the model. *(1|variable)* indicates that the variable is entered as a random effect on the intercept. We used Wald's test to measure the contribution of each categorical factor to the overall model, and include the statistic (χ<sup>2</sup>) and the degrees of freedom (df).

(a) Baseline model		Presence of <i>C. marmotae</i> = Julian date + F(year) + (1 family group)+ (1 id)						
(b)		Parameters of the baseline model						
Dependent variable	Distribution	Independent variable	estimate ± SE	p-value	Wald test			
					χ <sup>2</sup>	df	p-value	
Presence of <i>C. marmotae</i>	Binomial	Intercept	-13.52 ± 3.20	<0.01	18.5	1	<0.01	
		Julian date	0.08 ± 0.02	<0.01	22.0	1	<0.01	
		Year	2009	-0.75 ± 0.61	0.22	2.70	5	0.74
			2010	-0.26 ± 0.61	0.68			
			2011	-0.06 ± 0.71	0.93			
			2012	-0.28 ± 0.63	0.66			
			2014	-0.38 ± 0.81	0.64			

**Table 37.** Baseline model (a) obtained after selection of the confounding variables for the analysis of the presence of *C. marmotae* and its associated parameters. *F(variable)* indicates that the variable is entered as a factor in the model. *(1|variable)* indicates that the variable is entered as a random effect on the intercept. (b) (N = 225). *F(variable)* indicates that the variable is entered as a factor in the model. *(1|variable)* indicates that the variable is entered as a random effect on the intercept. We used Wald's test to measure the contribution of each categorical factor to the overall model, and include the statistic (χ<sup>2</sup>) and the degrees of freedom (df).

### *Appendix 3. MHC genetic diversity estimators and results obtained with these estimators*

#### *MHC genetic diversity estimators*

All MHC estimators were calculated over the four MHC allelic polymorphic loci. MHC estimators were calculated over MHC class I and class II loci separately because MHC class I and II genes are under different selective pressures leading to MHC class I and II estimators not to be correlated and suggesting that in Alpine marmots both classes could convey different information (Chapter 2). Only individuals successfully genotyped at the four MHC loci, at the two MHC class I loci or at the two MHC class II loci were used to calculate the MHC, MHC class I and MHC class II estimators respectively.

Two indexes were used to estimate the MHC genetic diversity: the MHC allelic diversity and the MHC genetic distance. The MHC allelic diversity was calculated as the number of different MHC alleles an individual possesses. The genetic distance was calculated using the software MEGA version 6.06 (Tamura et al. 2013) as the sum of the p-distances (i.e. the proportion of nucleotide sites that differs between two alleles) among the considered MHC loci.

The different index used to calculate MHC genetic diversity – MHC allelic diversity and MHC genetic distance – estimators are highly correlated, suggesting that in Alpine marmots different index convey the same information (Chapter 2).

#### *Results*

Neither the abundance of *E. monacis*, the presence of *Eimeria sp.* nor the presence of *C. marmotae* depend on individual MHC allelic characteristics whatever the loci considered (the four MHC loci, the two MHC class I loci and the two MHC class II loci) or the genetic estimators considered (allelic diversity or genetic distance) (Table 38).

Dependent variable	Independent variable	estimate $\pm$ SE	p-value	N
Abundance of <i>E. monacis</i>	MHC allelic diversity	0.05 $\pm$ 0.07	0.41	312
	MHC class I allelic diversity	0.06 $\pm$ 0.10	0.54	321
	MHC class II allelic diversity	0.10 $\pm$ 0.11	0.33	312
	MHC genetic distance	1.76 $\pm$ 2.82	0.53	312
	MHC class I genetic distance	6.92 $\pm$ 14.83	0.64	321
	MHC class II genetic distance	1.75 $\pm$ 2.89	0.55	312
Presence of <i>Eimeria sp.</i>	MHC allelic diversity	-0.11 $\pm$ 0.12	0.34	303
	MHC class I allelic diversity	0.01 $\pm$ 0.17	0.95	312
	MHC class II allelic diversity	-0.25 $\pm$ 0.19	0.19	303
	MHC genetic distance	-0.78 $\pm$ 5.00	0.88	303
	MHC class I genetic distance	-2.55 $\pm$ 25.97	0.92	312
	MHC class II genetic distance	-0.45 $\pm$ 5.19	0.93	303
Presence of <i>C.</i> <i>marmotae</i>	MHC allelic diversity	0.13 $\pm$ 0.14	0.32	214
	MHC class I allelic diversity	0.02 $\pm$ 0.20	0.94	219
	MHC class II allelic diversity	0.35 $\pm$ 0.22	0.12	214
	MHC genetic distance	8.89 $\pm$ 5.82	0.13	214
	MHC class I genetic distance	3.52 $\pm$ 29.82	0.90	219
	MHC class II genetic distance	9.50 $\pm$ 6.02	0.12	214

**Table 38.** Generalized linear mixed models showing the effects of the MHC allelic characteristics of individuals on the abundance of *E. monacis*, the presence of *Eimeria sp.* and the presence of *C. marmotae*. The MHC diversity estimators were calculated over the two MHC loci, over the MHC class I locus and over the MHC class II locus. All the models with the abundance of *E. monacis* and the presence of *Eimeria sp.* include the year, the Julian date and age class; and all the models with the presence of *C. marmotae* include the year and the Julian date. The critical p-value after Benjamini and Yekutieli (2001) correction correspond to 0.01068. N refers to the samples examined.

Loci	Allele	N	Percentage
<i>Mama-UD</i>	<i>Mama-UD*01</i> and/or <i>Mama-UD*03</i> <sup>1</sup>	170	97%
	<b><i>Mama-UD*02</i></b>	45	<b>26%</b>
<i>Mama-DRB1</i>	<b><i>Mama-DRB1*01</i></b>	<b>94</b>	<b>55%</b>
	<b><i>Mama-DRB1*02</i></b>	<b>83</b>	<b>48%</b>
	<b><i>Mama-DRB1*03</i></b>	<b>48</b>	<b>28%</b>
	<i>Mama-DRB1*04</i>	3	2%
	<i>Mama-DRB1*05</i>	3	2%
	<b><i>Mama-DRB1*06</i></b>	<b>37</b>	<b>22%</b>
	<b><i>Mama-DRB1*07</i></b>	<b>29</b>	<b>17%</b>
	<b><i>Mama-DRB1*08</i></b> <sup>2</sup>	<b>18</b>	<b>10%</b>

**Table 39.** Number (N) and percentage of individuals carrying the 10 polymorphic proteins. Retained proteins for the statistical analysis (percentage of individuals carrying a given protein > 5% and < 95%) are indicated in bold: 1: Given that *Mama-UD\*01* and *Mama-UD\*03* produce the same protein are counted together 2: non-functional protein.

# **CHAPTER 4: GENETIC CHARACTERISTICS AND SURVIVAL**

---

**Evidences for a switch of the effect of genetic characteristics on juvenile  
Alpine marmot survival**

M Ferrandiz-Rovira, D Allainé, M-P Callait-Cardinal, P Dupont and A Cohas

*Draft*



## Chapter 4: Genetic characteristics and survival

## Abstract

There is compelling evidence about the effects of genetic diversity on individual fitness. Individuals with high genetic diversity at neutral and adaptive loci are supposed to have higher survival probabilities. However, climate change is expected to change selective pressures and, accordingly, it might change the effects of genetic characteristics on individual fitness. In this study, we used a 23-year monitoring study of a wild population of Alpine marmots (*Marmota marmota*) challenged by climate change to test whether juvenile survival depends on individual multilocus heterozygosity at neutral loci (based on 16 microsatellites) and/or the genetic diversity of specific adaptive loci (based on two functional polymorphic major histocompatibility complex (MHC) loci involved in immunity response and pathogen resistance) and, if these potential influences vary between two periods (from 1990 to 2001 and from 2002 to 2012). Our results indicated that both neutral and MHC diversity were associated to juvenile survival, but genetic diversity effects varied between the two periods. Whereas during the first period juvenile survival was positively but not significantly associated with increasing multilocus microsatellite heterozygosity and negatively associated with MHC diversity, the reverse patterns were found during the second period for both neutral and MHC genetic diversity. Our study illustrates the importance of investigating both neutral and adaptive genetic characteristics in populations challenged by climate change when studying heterozygosity-fitness correlations.

**Keywords:** adaptation, climate change, fitness, genetic diversity, heterozygosity, major histocompatibility complex, *Marmota marmota*, microsatellites, survival

## 1. INTRODUCTION

The study of the impact of individual genetic variability on fitness has long been a major interest for conservation and evolutionary biologists. Indeed, compelling evidence has demonstrated that genetic diversity is positively associated with fitness (Keller and Waller 2002). Individuals with high levels of genetic diversity at neutral loci often present higher fitness (Hansson and Westerberg 2002, Chapman et al. 2009, Szulkin et al. 2010). Accordingly, heterozygosity at neutral loci has been found to affect survival (e.g. in great reed warblers (*Acrocephalus arundinaceus*) (Hansson et al. 2001), in American crows (*Corvus brachyrhynchos*) (Townsend et al. 2009)). Although often debated, multiple non-exclusive hypotheses have been proposed to explain the relationship between fitness-related traits and individual multilocus heterozygosity at microsatellite markers - the so-called heterozygosity-fitness correlation (David 1998, Hansson and Westerberg 2002, Szulkin et al. 2010): (1) the direct effect hypothesis, which results from functional overdominance at the genotyped loci *per se* and thus have a direct effect on fitness; (2) the general effect hypothesis, which results from effects of homozygosity at genome-wide fitness loci (a potential consequence of inbreeding depression); and (3) the local effect hypothesis, which is expected to occur when the genotyped markers are in linkage disequilibrium with closely linked fitness loci. Also, heterozygous individuals have a higher adaptive potential, which may increase their chances to face changes in either biotic or abiotic conditions (Frankham 2005).

In addition to neutral genetic diversity, compelling evidence has demonstrated that immune genetic diversity is positively associated with fitness because of its role in variation in immune function and pathogen prevalence. There is now an increasing body of evidence for associations between immune

gene diversity at genes such as the major histocompatibility complex (MHC, a gene family that encodes for genes able to bind to peptide fragments derived from pathogens and display them on the cell surface for recognition by the appropriate T-cells) and pathogen resistance/susceptibility in wild mammalian populations. For instance, a negative association between MHC diversity and parasite load has been reported in different species (Penn et al. 2002, Froeschke and Sommer 2005, Lenz et al. 2009; see Sin et al. 2014 for a recent review). While investigations under natural conditions remain rare, there is now an increasing body of evidence for positive associations between MHC diversity and survival in wild mammalian populations (e.g. birds: Seychelles warblers (*Acrocephalus sechellensis*) (Brouwer et al. 2010); common yellowthroats (*Geothlypis trichas*) (Dunn et al. 2013); and artiodactyls: Soay sheep (*Ovis aries*) (Paterson et al. 1998). Given that MHC genes are co-dominantly expressed, individuals with high levels of genetic diversity at MHC are expected to present higher fitness because they should increase their probability to bind any pathogen-derived antigen (Doherty and Zinkernagel 1975, Penn et al. 2002). However, too much MHC diversity may be costly since it may promote negative selection on self-reacting lymphocytes, which is likely to appear with elevated MHC diversity levels (Nowak et al. 1992, Woelfing et al. 2009). Accordingly, individuals with intermediate MHC diversity rather than those with maximal MHC diversity are supposed to display a higher pathogen resistance/susceptibility, a prediction that has been confirmed in bank voles, *Myodes glareolus*, (Kloch et al. 2010) or in three-spined sticklebacks (*Gasterosteus aculeatus*) (Wegner et al. 2003a,b). However, to our knowledge, no quadratic relationship between MHC diversity and survival has yet been evidenced. Thus, research in this field awaits further investigation.

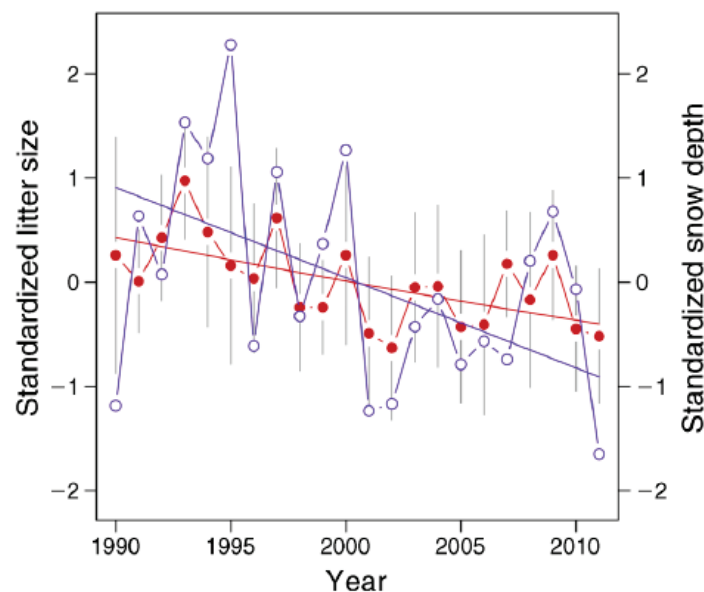
It is now clear that the current climate change weakens numerous populations due to changes in both biotic (e.g. new diseases, changes in parasites, changes in predators or changes in competitors) and abiotic (e.g. environmental changes derived from human activity) conditions. These abrupt changes are a potential source of new and/or altered selection pressures on traits important for fitness of

the major part of wildlife populations (Jump et al. 2008, Gienapp et al. 2008, Pauls et al. 2013). As a consequence of these changes, population genetic diversity has been hypothesised to either increase or decrease (Pauls et al. 2013). One of the few studies that have evaluated the changes in genetic diversity due to climate change found that genetic diversity in prairie grass species declined significantly following 10 years of experimentally increased intra-annual variability in growing season precipitation regimes to simulate the new conditions driven by climate change (Avolio et al. 2013). Despite the reduction of overall population genetic diversity, Avolio et al. (2013) found greater genome dissimilarity among individuals, characteristic that was positively correlated with aboveground productivity. However, whether climate change alter the associations between genetic characteristics (either at neutral and/or on MHC genetic diversity) and survival remains, to date, unexplored.

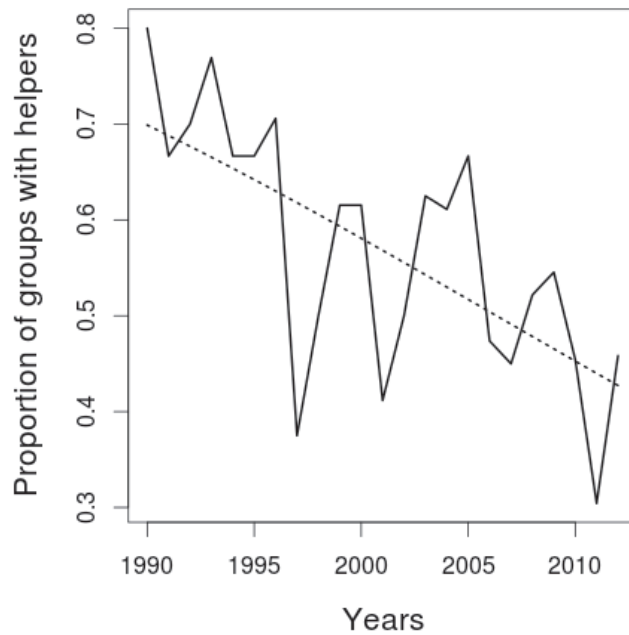
Alpine marmots (*Marmota marmota*) are socially monogamous mammals living in family groups from two to 20 individuals (Allainé 2000) composed of a dominant couple, sexually mature (two years of age and older) and immature (yearlings) subordinates of both sexes, and pups born the given year (Allainé et al. 2000). They are cooperative social breeders hibernating in family groups from mid-October until early April. Subordinate males are called helpers because they increase pup survival during hibernation through social thermoregulation (Allainé and Theuriau 2004). In Alpine marmots, as in other hibernating species, body fat accumulated during the active period is the unique source of energy during hibernation and they lose up to 30% of their body mass during this period (Arnold 1990). Because juveniles have less time to accumulate fat reserved before hibernation (Arnold 1990), this age class may be considered as the most critical for Alpine marmot survival. Accordingly, Alpine marmot survival has been found to increase with age from around 60% for juveniles to over 80% for adult dominant marmots (Cohas et al. 2009, Rézouki et al. in prep).

A large body of evidence indicates that climate change has induced changes in Alpine marmots followed since 1990 at the Grande Sassièrè Natural Reserve (at 2,340 m a.s.l., 45°29'N, 65°90'E) situated in the French Alps. Indeed, a

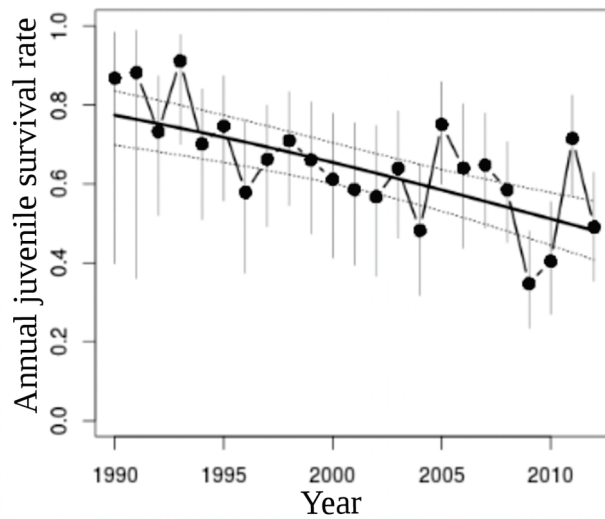
continuous decrease in litter size has been observed between 1990 and 2011 (Tafani et al. 2013b, Figure 17). This decrease is associated to the continuously thinning of winter snow cover (average snow depth from December to March decreased by  $1.87 \pm 0.68$  (mean  $\pm$  SE) cm per year from 1991 to 2011) (Tafani et al. 2013b, Figure 17). On the other hand, a continuous decrease in the proportion of families with helpers during hibernation was found, from a proportion of 80% of groups with helpers in 1990 to 46% in 2013 (Rézouki et al. in prep, Figure 18). This finding evidences a modification of the social composition of the families during hibernation (Rézouki et al. in prep). Because the number of helpers affects juvenile survival, we observe consequently a continuous decrease in juvenile survival between 1990 and 2013 (from a juvenile survival probability of  $0.85 \pm 0.16$  in 1990 to  $0.42 \pm 0.07$  in 2013), a decrease that is directly related with an increase in winter harshness over the studied period (Rézouki et al. in prep, Figure 19).



**Figure 17.** Yearly variation ( $\pm$  95% CI) in standardized litter size of Alpine marmots (*Marmota marmota*) and standardized snow cover in winter (from December to March) at the Grande Sassi re Natural Reserve (French Alps). Litter size is represented in red (with confidence interval around the average in gray), and snow depth in blue (from Tafani et al. 2013b).



**Figure 18.** Yearly changes in the proportion of groups having helpers during winter between 1990 and 2013 at the Grande Sassièrè Natural Reserve (French Alps). Dotted lines represent the best-fitted linear prediction over time (from Rézouki et al. in prep).



**Figure 19.** Annual estimates ( $\pm$  95% CI, vertical segments) of juvenile (from 0 to 1 year of age) survival rate of Alpine marmots at the Grande Sassièrè Natural Reserve from 1990 to 2012, derived from a full-time dependent capture-mark-recapture model, and the corresponding temporal trends ( $\pm$  95% CI, dashed lines) (from Rézouki et al. in prep).

Taking advantage of this long-term monitoring study of Alpine marmots at the Grande Sassièrè Natural Reserve, we investigated whether juvenile survival depended on neutral and/or on MHC genetic diversity, using 16 microsatellites and two functional polymorphic MHC loci, respectively. Earlier work on the studied population revealed that the higher neutral genetic diversity (measured by these 16 microsatellites), the higher the survival for juveniles (Cohas et al. 2009). In contrast, survival of all other age classes (yearlings, two-years old individuals and adults) did not depend on neutral genetic diversity (Cohas et al. 2009). Here, we investigated whether the potential impact of genetic polymorphism at neutral and/or on MHC loci on juvenile survival changed along the studied period (from 1990 to 2012). We hypothesized that the altered environmental conditions may affect the contribution of genetic diversity at neutral and/or at MHC loci to juvenile survival.

## 2. MATERIAL AND METHODS

### *Field methods and sample collection*

Here we considered 837 individuals captured as pups between 1990 to 2012 in 25 family groups from a wild population of Alpine marmots located in the Nature Reserve of La Grande Sassièrè in the French Alps (at 2,340 m a.s.l., 45°29'N, 65°90'E). Every year, individuals from the population were monitored, from mid-April to mid-July, using both capture-mark-recapture and daily observations. Juveniles were counted and trapped either by smaller two-door live traps or by hand within the three days following their first emergence from the burrows (i.e. approximately 40 days after their birth). Around 80% of older individuals were captured at least once every year, between mid-April to mid-July, using two-door live traps baited with dandelions (*Taraxacum densleonis*). Once captured, individuals were tranquilized with Zolétil 100 (0.1 mL Kg<sup>-1</sup>), sexed, aged from



their size (up to 3 years), and their social status was determined according to scrotal development for males and teat development for females. Captured marmots were individually marked with a transponder chip (model ID100, 0.9 cm long, <0.1 cm in diameter, Trovan Ltd, [www.Trovan.com](http://www.Trovan.com), Identifikationssysteme, Metternicher Straße 4, 53919 Weilerswist, Germany) injected under the skin of the neck for permanent individual recognition, and a numbered ear tag (1 cm x 3 mm). For genetic analysis, hair samples and skin biopsies from the flank of individuals were collected.

The composition of family groups (juveniles, yearlings and adults from both sexes) was assessed from both capture–recapture data and intensive daily observations (see Cohas et al. 2006 for details on observation protocol). Moreover, scent marking behaviour was used to confirm the identity of the dominant pair (Bel et al. 1999).

Given that recapture probability is really high in the first stages of Alpine marmot life (Cohas et al. 2009) and that yearlings and two-years old are easily identified during intensive daily observations, we considered individuals to be dead when they were never recaptured or observed.

### *Microsatellite genotyping*

The 837 individuals were genotyped at 16 microsatellites: SS-Bibl1, SS-Bibl18, SS-Bibl20, SS-Bibl31, SS-Bibl4 (Klinkicht 1993); MS41, MS45, MS47, MS53, MS56, MS6, ST10 (Hanslik and Kruckenhauser 2000); Ma002, Ma018, Ma066, Ma091 (Da Silva et al. 2003) (see Appendix 1 in Chapter 2 for details on genotyping methods and microsatellite characteristics).

### *MHC genotyping*

823 out of 837 individuals were genotyped at four MHC polymorphic loci - two from MHC class I exon 2 (*Mama-UB* and *Mama-UD*) and two from MHC class II DRB loci (*Mama-DRB1* and *Mama-DRB2*) - using next generation sequencing (error rate = 0.21%), Sanger sequencing or were deduced based on mother-father-

Chapter 4: Genetic characteristics and survival offspring triads (see Appendix 1 in Chapter 2 for details on genotyping methods and MHC characteristics). A total of 3147 genotypes from these 823 individuals were obtained among the four MHC loci (2521 using next generation sequencing: 476 *Mama-UB*, 649 *Mama-UD*, 696 *Mama-DRB1* and 700 *Mama-DRB2*; 530 using Sanger sequencing: 239 *Mama-UB*, 124 *Mama-UD*, 76 *Mama-DRB1* and 91 *Mama-DRB2*; and 96 were deduced based on mother-father-offspring triads: 76 at *Mama-UB*, 11 at *Mama-UD*, 1 at *Mama-DRB1* and 8 at *Mama-DRB2*).

### *Genetic diversity estimators*

Neutral genetic diversity estimators were calculated over the 16 microsatellite loci, while MHC diversity estimators were calculated separately over the two MHC loci that led to polymorphic proteins, *Mama-UD* and *Mama-DRB1*, (three and eight alleles resulting in two and seven proteins respectively) because in Alpine marmots MHC class I and II genes are under different selective pressures (Kuduk et al. 2012). Indeed, MHC class I and II diversity estimators are not correlated in this species, suggesting that both classes could convey different information (Chapter 2).

The neutral genetic heterozygosity was estimated by three commonly used measures of individual multilocus heterozygosity using the 16 genotyped microsatellites (but see Miller et al. 2014 for the number of microsatellites that should be used to assess better estimates of individual multilocus heterozygosity) calculated with the R function “GENHET v3.1” (Coulon 2010): the standardized heterozygosity (SH, Coltman et al. 1999), the internal relatedness (IR, Amos et al. 2001) and the homozygosity by locus (HL, Aparicio et al. 2006). However, because of the high intercorrelation among these estimates (Spearman correlation between all pairs of heterozygosity measures:  $\rho > 0.94$ ,  $P < 0.001$ ), we only report results here for the standardized heterozygosity.

Two indexes were used to estimate the MHC diversity: the MHC protein diversity and the MHC protein distance. The MHC protein diversity was

estimated as the number of different MHC proteins an individual produces once the protein translated by the allele *Mama-DRB1\*08* was discarded since it has a stop codon and therefore the translated protein is non-functional (Kuduk et al. 2012). The MHC protein distance was calculated using the software MEGA version 6.06 (Tamura et al. 2013) as the sum of the p-distances (i.e. the proportion of amino acid sites that differs between two alleles) among the considered MHC loci. Given that *Mama-DRB1\*08* translated protein is non-functional, the p-distance between a given protein and the one translated from *Mama-DRB1\*08* was set to 0 (i.e. equivalent to consider an heterozygous individual carrying *Mama-DRB1\*08* as homozygous). Moreover, all MHC diversity estimators were also calculated considering alleles instead of proteins and were calculated over the four MHC loci and over MHC class I and class II loci separately (see Appendix 1 for details on the calculation of the MHC allelic diversity estimators).

### *Statistical analysis*

To investigate if the influence of genetic diversity changed between 1990 to 2012, we divided the population-specific time series into two periods of nearly the same duration (twelve and eleven years respectively): the first period lasts from 1990 to 2001 and the second period, from 2001 to 2012 (see Table 40 for details on individuals included in both periods). Both periods differ in the juvenile survival, who decreased between the two periods (estimate  $\pm$  SE =  $0.34 \pm 0.14$ ; p-value = 0.02; N = 837, Table 40). Given that the number of sampled families increased between 1990 and 2012, the number of juveniles included in the analysis was higher during the second period (Table 40).

	First period (1990 - 2001)	Second period (2002 - 2012)
Number of juveniles	337	500
Survival	55%	47%
Mean $\pm$ SD standardised heterozygosity	0.94 $\pm$ 0.24	0.91 $\pm$ 0.25
Median [1 <sup>st</sup> - 3th quartile] MHC class I protein diversity	1 [1, 2]	1 [1, 2]
Median [1 <sup>st</sup> - 3th quartile] MHC class II protein diversity	2 [1, 2]	2 [1, 2]
Mean $\pm$ SD MHC class I protein distance	0.001 $\pm$ 0.004	0.003 $\pm$ 0.01
Mean $\pm$ SD MHC class II protein distance	0.04 $\pm$ 0.04	0.04 $\pm$ 0.04
Mean $\pm$ SD snow depth (cm)	107.60 $\pm$ 19.59	97.56 $\pm$ 16.18
Mean $\pm$ SD winter temperatures ( $^{\circ}$ C)	-3.21 $\pm$ 0.67	-4.19 $\pm$ 1.06
Mean $\pm$ SD North Atlantic Oscillation	0.65 $\pm$ 0.73	-0.12 $\pm$ 0.95

**Table 40.** Characteristics of juvenile individuals and whether characteristics between the two studied periods. Snow depth and winter temperatures were recorded daily, from from December to March, from the weather stations of Tignes (73296400) and Val d'Isère (73304404), located 5 km away from the study site. The North Atlantic Oscillation index (Hurrell 1995) was calculated each year from December to March, which is a large scale seasonal index which appears as a good predictor of ecological variation (negative values colder temperatures that favored frost rather than snow).

Generalized Linear Mixed Models (GLMM) were used to investigate whether the juvenile survival was influenced by genetic diversity in each time period. Two series of GLMM models were constructed using the package “MASS” (Venables and Ripley 2002) of R version 3.1.1 (R Development Core Team 2014) with the juvenile survival (coded as 0 for dead and 1 for alive individuals) during the first period and the juvenile survival (coded as 0 for dead and 1 for alive individuals) during the second period as the dependent variables, respectively. A logit link and a binomial error were used for GLMM. To control for the non-independence of the data resulting from all pups of a same litter sharing a common environment and juvenile survival varying sharply from one year to the next (Rézouki et al. in prep), the litter and the year were entered as random terms in each model. Moreover, male helpers (coded as 0, 1 and 2 for absent, present and unknown, respectively) were added as a categorical variable since they can increase the juvenile survival (Arnold 1990, Allainé and Theuriau 2004). Then, genetic diversity estimators were added separately. A linear effect was considered for standardized heterozygosity while both a linear and a

Chapter 4: Genetic characteristics and survival

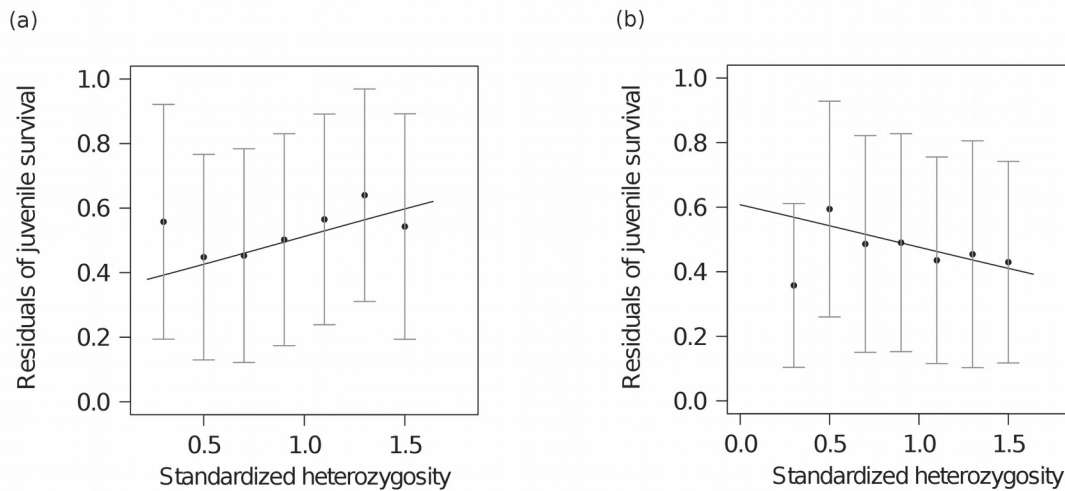
quadratic effects were considered for all genetic diversity estimators to test for an increase on survival according to intermediate MHC diversity levels. Finally, to test if the influence of genetic diversity changed between 1990 to 2012, the genetic diversity values of each studied period (1990-2001 vs. 2002-2012) were compared *a posteriori* using Wald tests (Agresti 2007). Since genetic diversity estimators were not available for all individuals, sample sizes varied in the statistical analyses from 316 to 337 individuals in the first period and from 456 to 500 individuals in the second period.

Unless otherwise stated, analyses conducted with the MHC diversity estimators calculated considering alleles instead of proteins gave qualitatively similar results, therefore only the results obtained with proteins are presented below (see Appendix 1, Chapter 4 for results obtained considering alleles).

### 3. RESULTS

A difference in the effect of the presence of male helpers on juvenile survival was evidenced between the two periods ( $\chi^2 = 7.50$ ,  $df = 2$ ,  $p\text{-value} = 0.02$ ). As expected, during the first period, the juvenile survival was found to depend on male helpers during hibernation ( $\chi^2 = 11.4$ ,  $df = 2$ ,  $p\text{-value} = 0.003$ ;  $N = 337$ ). In this first period, the presence of helpers increased juvenile survival (estimate  $\pm$  SE =  $1.14 \pm 0.40$ ;  $N = 337$ ). However, during the second period, the juvenile survival did not depend on male helpers during hibernation ( $\chi^2 = 0.93$ ,  $df = 2$ ,  $p\text{-value} = 0.63$ ;  $N = 500$ ).

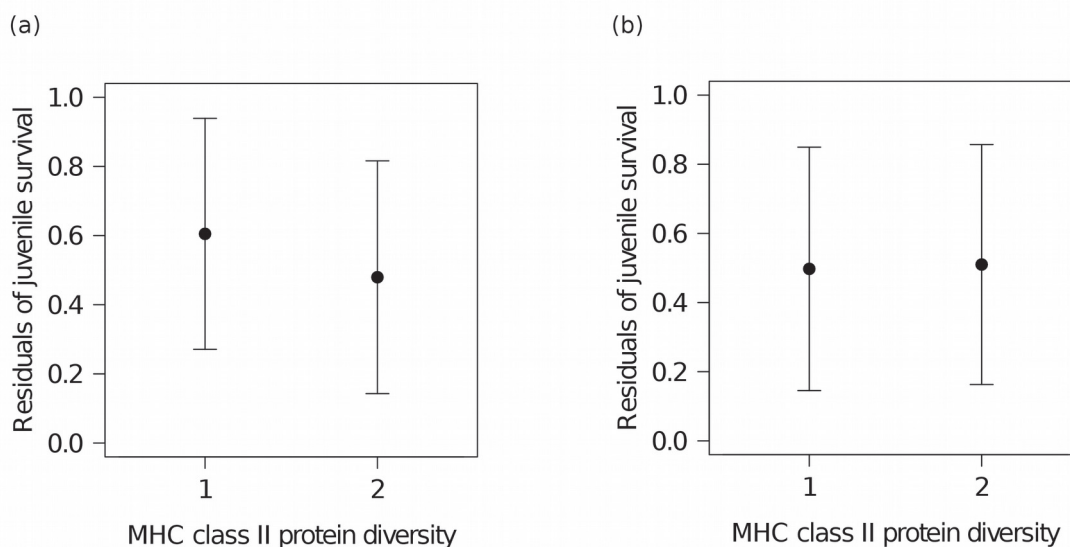
The effect of standardized heterozygosity on juvenile survival differed between the two periods ( $\chi^2 = 4.10$ ,  $df = 1$ ,  $p\text{-value} = 0.04$ , Figure 20). Whereas during the first period, juvenile survival tended to be positively correlated with standardized heterozygosity, during the second period, juvenile survival tended to be negatively correlated (Table 41, Figure 20).



**Figure 20.** Residual of the juvenile survival as a function of standardized heterozygosity during the first period (1990-2001) (a) and during the second period (2002-2012) (b). Residuals are corrected for the presence of male helpers. Thick lines represent predictions of the model and solid circles represent mean residual juvenile survival per standardized heterozygosity (class with 0.20) with their associated standard error bars.

No difference in the effect of MHC class I locus on juvenile survival was evidenced between the two periods whatever the MHC diversity estimators considered (MHC class I protein diversity:  $\chi^2 = 0.14$ ,  $df = 1$ ,  $p$ -value = 0.70; MHC class I protein distance:  $\chi^2 = 0.14$ ,  $df = 1$ ,  $p$ -value = 0.70). During both the first and the second period, juvenile survival did not depend on MHC class I locus (Table 41).

A clear difference in the effect of the diversity of MHC class II locus on juvenile survival was evidenced between the two periods whatever the MHC diversity estimators considered (MHC class II protein diversity:  $\chi^2 = 6.00$ ,  $df = 1$ ,  $p$ -value = 0.01, Figure 21; MHC class II protein distance:  $\chi^2 = 16.40$ ,  $df = 1$ ,  $p$ -value = <0.01). Whereas, juvenile survival was negatively and linearly correlated with MHC class II locus during the first period, it tended to be positively and linearly correlated with MHC class II locus during the second period, and this whatever the MHC diversity estimators considered (Table 41, Figure 21).



**Figure 21.** Residual of the juvenile survival as a function of MHC class II protein diversity during the first period (1990-2001) (a) and during the second period (2002-2012) (b). Residuals are corrected for the presence of male helpers. Solid circles represent mean residual juvenile survival per MHC class II protein diversity with their associated standard error bars.

Independent variable	First period (1990 - 2001)			Second period (2002 - 2012)		
	estimate $\pm$ SE	p-value	N	estimate $\pm$ SE	p-value	N
Standardized heterozygosity	0.69 $\pm$ 0.61	0.26	337	-0.53 $\pm$ 0.45	0.24	500
MHC class I protein diversity	0.10 $\pm$ 0.57	0.86	316	0.32 $\pm$ 0.30	0.30	468
MHC class II protein diversity	<b>-0.60 <math>\pm</math> 0.29</b>	<b>0.04</b>	<b>317</b>	0.11 $\pm$ 0.23	0.64	456
MHC class I protein distance	6.03 $\pm$ 33.57	0.86	316	18.69 $\pm$ 17.87	0.30	468
MHC class II protein distance	<b>-10.75 <math>\pm</math> 3.84</b>	<b>0.01</b>	<b>317</b>	4.70 $\pm$ 3.06	0.13	456

**Table 41.** Generalized Linear Mixed Models showing the effects of the genetic characteristics on the juvenile survival on both the first and the second period. The MHC diversity estimators were calculated over the MHC class I locus and over the MHC class II locus. All models include helpers as a categorical variable (absence, presence and unknown, coded as 0, 1 and 2 respectively). P-values  $<$  0.05 are indicated in bold. N: number of individuals.

## 4. DISCUSSION

While the study of how genetic diversity affects survival has now been studied in wild populations, how the influence of these genetic characteristics may change



over time and along with climate change remains unknown. We investigated whether individual genetic diversity at neutral and at MHC loci impacted juvenile survival in a wild population of Alpine marmots challenged by climate change. Taking advantage of a long-term monitoring study, we found that juvenile Alpine marmots benefited from the presence of male helpers during the first studied period (1990-2001), whereas they did not benefited from their presence during the second period (2002-2012). The influence of genetic diversity at both neutral and MHC loci also changed between the two periods. During the first period, genetic diversity at neutral loci was positively but not significantly associated with juvenile survival, while genetic diversity at MHC class II locus was negatively associated. On the contrary, during the second period, genetic diversity at neutral loci was negatively but not significantly associated with juvenile survival, while genetic diversity at MHC class II locus was positively but not significantly associated. Our results thus provide the first evidence that the impact of both neutral and MHC diversity on juvenile Alpine marmot survival has changed between 1990 and 2012.

The presence of male helpers during the first hibernation of Alpine marmots was found to increase juvenile survival during the first examined period but their presence no longer impacted juvenile survival during the second examined period. Until now, cooperative breeding, which has evolved in the Alpine marmot (Barash 1976, Armitage 1999), allowed Alpine marmot juveniles to decrease its winter mortality through the energetic benefit of social thermoregulation mainly carried out by subordinate males (Allainé et al. 2000, Allainé and Theuriau 2004). A likely explanation for the lack of effect of the presence of male helpers during the second period is that the hibernation conditions may have become too harsh. As a consequence of the increase of harsher hibernation conditions (Tafari et al. 2013b, Rézouki et al. in prep), the effect of the presence of male helpers, whose presence has dramatically decreased (Rézouki et al. in prep), might be cancelled.

Previous work on this population suggested that juvenile Alpine marmots with higher genetic diversity at neutral loci (measured using the same 16



microsatellites) had a higher survival than juveniles with lower genetic diversity in a period encompassing cohorts born from 1990 to 2006 (Cohas et al. 2009). Accordingly, our results indicate that juveniles with higher genetic diversity at neutral loci (microsatellites) were positively associated with higher survival during the first period. However, this positive association of genetic diversity at neutral loci appeared weaker in our study (the effect was only significant in the study of Cohas et al. (2009)). This difference may be due to stochastic differences in sample composition or to the fact that we used GLMM rather than multistate CMR models (Lebreton and Pradel 2002). In contrast, genetic diversity at neutral loci was negatively but not significantly associated with juvenile survival during the second period. At least two reasons may explain the change on the effect of genetic diversity at neutral loci on juvenile survival. First, given that winter conditions are so harsh, juveniles with higher genetic diversity at neutral loci may have no selective advantage with respect to less genetic diverse juveniles and this assuming that genetic diversity at neutral loci does not affect juvenile probability of overcoming extreme winter conditions. However, this explanation seems more reasonable to explain a lack of association between genetic diversity at neutral loci and survival than to explain the negative (although not significantly) association found between these terms. Additionally (or alternatively), the change regarding the effect of neutral genetic diversity on juvenile survival may be due to micro-evolutionary processes that took place during the second period. Indeed, a way to deal successfully with the increasingly winter harshness that Alpine marmots should face (Tafari et al. 2013b, Rézouki et al. in prep) may involve micro-evolution *via* the selection of local genotypes better adapted to changing environmental conditions (e.g. in Alpine marmots those individuals that may reduce their energy consumption during hibernation). As a consequence, the selection of local genotypes may tend to reduce genetic variation at the selected loci (Pauls et al. 2013). However, in this case, genetic diversity at neutral loci should be associated with these specific loci that confer a selective advantage to changing environmental conditions. According to Pauls et al. (2013), which hypothesised that populations genetic diversity may either

increase or decrease as a consequence of climate change, the opposite pattern was found in the only study, to our knowledge, that has evaluated the changes in genetic diversity due to climate change. Specifically, in a 30-years monitoring study in a decreasing population of fur seals (*Arctocephalus gazella*) in which climate change has reduced prey availability and caused a significant decline in seal birth weight, Forcada and Hoffman (2014) found that mortality rate of homozygous individuals (measured by nine microsatellites) has increased along the studied period leading to the more heterozygous individuals to better survive to breeding age.

In agreement with the fluctuating selection hypothesis, which proposes that temporal parasite diversity and virulence could lead to an heterogeneous selection of different MHC alleles through time (i.e. individuals that are advantaged by a given MHC characteristic at time  $t$  may not be equally advantaged by these MHC characteristics at time  $t+1$ ) (Hill 1991, Hedrick 2002), the effect of MHC diversity on Alpine marmot juvenile survival was found to differ between the first and the second period. MHC diversity was found to be negatively associated with juvenile survival during the first period, whereas it tended to be positively associated with juvenile survival during the second period. Juvenile Alpine marmots with high MHC diversity had lower survival probabilities during the first period, suggesting that an excess of MHC alleles might have been maladaptive in the Nature Reserve of La Grande Sassi re. Indeed, the cost of maintaining high MHC diversity levels (e.g. the promotion of negative T-cell selection), may have been higher than benefits associated with the MHC heterozygote advantage during the first period. Surprisingly, selection may have favoured less MHC diverse individuals rather than individuals with intermediate levels of MHC diversity as it is predicted by the MHC optimality hypothesis (Nowak et al. 1992). The low diversity usually observed for the core parasite species of Alpine marmots (Bassano 1996) along with an apparent absence of pathogenic activity of their intestinal parasites (Callait and Gauthier 2000) and a low role of MHC genes in intestinal parasite resistance (Chapter 3), may explain the negative effect of MHC diversity in juvenile Alpine marmots.

This negative association between MHC diversity and juvenile survival in Alpine marmots is similar to the negative association found between encoding Toll-like receptors diversity and survival in pale-headed brushfinch (*Atlapetes pallidiceps*) (Hartmann et al. 2014). If selective pressures may have favoured juvenile Alpine marmots with low MHC diversity during the first period, these selective pressures may have changed in recent years because of the potential effects of climate change, which has been associated to an increased exposure to new and/or more virulent parasites (Hernandez et al. 2013). Accordingly, these changes may have favoured individuals with higher MHC diversity during the second period given that these individuals are expected to have higher chances to bind a higher number of pathogen-derived antigens (Doherty and Zinkernagel 1975, Penn et al. 2002). This change resulting in the effect of MHC diversity on Alpine marmot juvenile survival should be associated with changes in selective pressures (i.e. the detection of new and/or more virulent parasites of Alpine marmots), a fact that apparently does not have occurred in the studied population (Chapter 3).

While the evidence that climate change is driving changes in wild populations is indisputable, it is less evident how genetic diversity may influence individual fitness in a changing environment. The aim of this study was to investigate the role of genetic diversity at neutral and MHC loci in the juvenile survival of Alpine marmots exposed to a changing climate. We found that juvenile survival depended or tended to depend on both neutral and MHC genetic diversity. However, the effect of genetic characteristics seem to be context-dependent as indicated by the fact that the genes involved in juvenile survival between the first and the second periods differ. Exploration of the association between genetic diversity at either neutral or MHC loci in long-term monitored wild populations is a promising avenue of research that awaits further investigation.

## 5. APPENDIX

### *Appendix 1. MHC genetic diversity estimators and results obtained with these estimators*

#### *MHC genetic diversity estimators*

All MHC diversity estimators were calculated over the four MHC allelic polymorphic loci (*Mama-UB*, *Mama-UD*, *Mama-DRB1* and *Mama-DRB2*). MHC diversity estimators were calculated over MHC class I and class II loci separately because MHC class I and II genes are under different selective pressures leading to MHC class I and II estimators not to be correlated (Table 24) and suggesting that both classes could convey different information. Only individuals successfully genotyped at the two MHC class I loci or at the two MHC class II loci were used to calculate the MHC class I and MHC class II estimators respectively.

Two indexes were used to estimate the MHC genetic diversity: the MHC allelic diversity and the MHC genetic distance. The MHC allelic diversity was calculated as the number of different MHC alleles an individual possesses. The genetic distance was calculated using the software MEGA version 6.06 (Tamura et al. 2013) as the sum of the p-distances (i.e. the proportion of nucleotide sites that differs between two alleles) among the considered MHC loci. Both indexes used to calculate MHC genetic diversity – MHC allelic diversity and MHC genetic distance – estimators were found to be highly correlated (Table 26), suggesting that different index convey the same information.

#### *Results*

No difference in the effect of MHC class I locus on juvenile survival was evidenced between the two periods whatever the MHC diversity estimators considered (MHC class I allelic diversity:  $\chi^2 = 0.01$ , degrees of freedom = 1, p-

value = 0.91; MHC class I genetic distance:  $\chi^2 = 0.01$ , degrees of freedom = 1, p-value = 0.91). In both the first and the second period, juvenile survival did not depend on MHC class I locus whatever the MHC diversity estimators considered, except for a positive effect of MHC class I genetic distance on juvenile survival during the second period (Table 42).

A clear difference in the effect of MHC class II locus on juvenile survival was evidenced between the two periods whatever the MHC diversity estimators considered (MHC class II allelic diversity:  $\chi^2 = 6.00$ , degrees of freedom = 1, p-value = 0.02; MHC class II genetic distance:  $\chi^2 = 12.40$ , degrees of freedom = 1, p-value = 0.001). Whereas during the first period, juvenile survival was negatively correlated with MHC class II locus, during the second period juvenile survival was positively correlated with MHC class II genetic distance (Table 42). However, no effect was evidenced between juvenile survival and MHC allelic diversity during the second period (Table 42).

Independent variable	First period (1990 - 2001)			Second period (2002 - 2012)		
	estimate $\pm$ SE	p-value	N	estimate $\pm$ SE	p-value	N
MHC class I allelic diversity	0.26 $\pm$ 0.18	0.14	306	0.24 $\pm$ 0.13	0.06	460
MHC class II allelic diversity	<b>-0.32 <math>\pm</math> 0.16</b>	<b>0.05</b>	<b>313</b>	0.06 $\pm$ 0.13	0.65	444
MHC class I genetic distance	44.27 $\pm$ 29.74	0.14	306	<b>40.83 <math>\pm</math> 20.28</b>	<b>0.04</b>	<b>460</b>
MHC class II genetic distance	<b>-9.59 <math>\pm</math> 4.81</b>	<b>0.05</b>	<b>313</b>	7.22 $\pm$ 4.08	0.08	444

**Table 42.** Generalized Linear Mixed Models showing the effects of the MHC allelic characteristics on the juvenile survival on both the first and the second period. The MHC diversity estimators were calculated over the two MHC class I loci and over the two MHC class II loci. All models include helpers as a categorical variable (absence, presence and unknown, coded as 0, 1 and 2 respectively). P-values  $< 0.05$  are indicated in bold. N: number of individuals.

## **DISCUSSION**

---



The aim of this thesis was to investigate the role of the MHC characteristics in a natural population of Alpine marmots. By taking advantage of a long-term individually based monitoring of a wild population situated at the Nature Reserve of the Grande Sassi re, in the first chapter I reliably genotyped (error rate = 0.21%) more than a thousand of individuals at four MHC loci belonging to two MHC class I (*Mama-UB* and *Mama-UD*) and two MHC class II (*Mama-DRB1* and *Mama-DRB2*) loci (Chapter 1). To process amplicon-based NGS data, an easy to use methodology (freely available at <https://github.com/tbigot/alFinder>) was developed, a procedure that can be applied to any genetic system and any organism, except when extensive copy-number variation occurs (Chapter 1). In the second chapter, I investigated whether female mate choice is based on genetic characteristics at either neutral and MHC loci. Using sixteen microsatellites as a proxy of neutral genetic characteristics and four MHC loci, the results of this study revealed that whereas social pairs are chosen based on MHC characteristics solely, the occurrence of EPP depended on both neutral and MHC characteristics. Overall, these results suggest that the observed pattern of female mate choice may be driven by population genetic structure, dispersal patterns and parasite pressures (Chapter 2). In the third chapter, I explored whether the presence and/or the abundance of three intestinal parasites (*E. monacis*, *Eimeria sp.* and *C. marmotae*) depended on individual MHC characteristics. The results of this study revealed that neither the MHC diversity nor the presence of MHC specific alleles were associated with intestinal parasites, except for *E. monacis*. Specifically, Alpine marmots carrying the allele *Mama-DRB1\*02* showed lower abundance of *E. monacis*. Overall, these results suggest a weak role of MHC on intestinal parasites. Rather, Alpine marmots may combat the analyzed intestinal parasites by immune pathways other than MHC genes. In the last chapter, I investigated whether juvenile survival depended on genetic characteristics at either neutral and MHC loci and, when it did, whether their influence varied between 1990 and 2012. This study revealed that juvenile survival tended to depend on both neutral and MHC diversity but both effects differed between two periods (from 1990 to 2001 and from 2002 to 2012). Whereas neutral genetic diversity tended to



increase juvenile survival and MHC diversity decreased juvenile survival during the 1<sup>st</sup> period, neutral genetic diversity tended to decrease juvenile survival and MHC diversity tended to increase juvenile survival during the 2<sup>nd</sup> period. Overall, these results suggest that selective pressures have probably changed in a 23-years period of time, as indicated by the fact that genes involved in juvenile survival between the 1<sup>st</sup> and the 2<sup>nd</sup> period differed.

In this thesis I explored and presented a complex picture of MHC associations with mate choice, intestinal parasites and juvenile survival in the wild population of Alpine marmots followed since 1990 at the Nature Reserve of the Grande Sassi re. Given the complexity of the role of MHC found in this thesis and the new opportunities offered by NGS techniques allowing to obtain large sample sizes often lacking in ecological studies, this work is also a call for further studies exploring the role of MHC characteristics in long-term monitored wild populations. Finally, I would like to discuss below some additional points on the issues addressed in this thesis and propose new challenges that would allow us to better understand the roles exerted by the MHC genes and other functional genes in the wild.

## **1. ADDITIONAL THOUGHTS ABOUT MATE CHOICE**

In Alpine marmots, dominant males only reproduce with the dominant female of its family group given that no pup was fathered by a dominant male outside its family group. As a consequence, male reproductive success depends on its capacity to acquire dominance (subordinates rarely reproduce). However, the access to a dominant female does not ensure fertilization of all offspring since EPP exist (Goossens et al. 1998, Cohan et al. 2006, 2008, Chapter 2). When EPP occurs, males suffer from a decrease of their reproductive success. At least three different processes are involved in the decrease of male reproductive success and

are described below.

First, an increase in the number of sexually mature subordinate males present in the family group has been found to increase the male's probability to be cuckolded (Cohas et al. 2006, 2008, Lardy et al. 2012, Chapter 2). As suggested by Lardy et al. (2012), the increase of the occurrence of EPP with an increase of the number of sexually mature subordinate males present in Alpine marmot family groups could fit with the “limited control” hypothesis (Clutton-Brock 1998, Reeve et al. 1998). This hypothesis postulates that because the size and the composition of social groups may determine the intensity of male-male competition, dominant males cannot control all the reproduction of the group despite their efforts to do so when the number of sexually mature individuals increases (Clutton-Brock 1998, Reeve et al. 1998). Accordingly, this hypothesis has also been found in other species (e.g. in banded mongooses, *Mungos mungo* (Cant 2000) or in meerkats, *Suricata suricatta* (Clutton-Brock et al. 2001)). In Alpine marmots, the presence of subordinate males in the family group is a double-edged sword. Indeed, although it enhances the male's probability to be cuckolded and a decrease of dominant male reproductive success, it contributes to the juvenile survival during the first winter (Arnold 1990, Allainé and Theuriau 2004). Thus, the presence of subordinate males indirectly contributes to increase the lifetime reproductive success (LRS, defined as the number of young produced by a given individual during its reproductive life that survive to sexual maturity (Arnold and Wade 1984a,b)) of dominant males by increasing the juvenile survival of its offspring, the most critical period for Alpine marmots survival.

Second, dominant male reproductive success may decrease as a result of sperm competition (i.e. competition between spermatozoa of two or more different males to fertilize the same ova (Parker 1970)) when females engage in extra-pair copulations. Under the sperm competition hypothesis, male's probability to fertilize ova is supposed to be proportional to the relative quantity of sperm delivered by the male (Parker 1998, Pizzari and Parker 2009). A given male is thus supposed to gain in sperm competition if it copulates more times

with the female than its rival. However, females may have different copulation behaviours towards the social and the extra-pair male. For example, females may copulate more times with the extra-pair male rather than with the social male, which may result in a decrease of dominant male reproductive success *via* sperm competition. Unfortunately, little is known regarding an hypothetical difference of copulation behaviour in Alpine marmots because, as many ground-dwelling sciurids, they normally mate underground (King and Allainé 1998). Indeed, despite the 24 years of fieldwork in the Nature Reserve of the Grande Sassièrè, a single copulatory behaviour has been reported (King and Allainé 1998). Besides that male's probability to fertilize ova is supposed to be proportional to the relative quantity of sperm delivered by the male, sperm quality may also play a role in the outcome of sperm competition (Snook 2005, Pizzari and Parker 2009, see Appendix E for detailed information regarding sperm competition). Thus, the sperm quality, may also impact the male's reproductive success. One way to study sperm competition in Alpine marmots would be the study of sperm ejaculates (both sperm quantity and quality, see Appendix E for more details) thanks to electro-ejaculation techniques. Electro-ejaculation has become a standard collection technique for breeding programs and a large number of protocols exists for different species. Although, to my knowledge, no protocol for electro-ejaculation of Alpine marmots is available, it might be developed and applied to several individuals in experimental conditions to control for all candidate males a female may mate with and to monitor the offspring outcome during pregnancy and just after birth.

Third, females may display a post-copulatory cryptic mate choice, which may also affect male's reproductive success. In a post-copulatory cryptic mate choice, females may be able to bias fertilization towards the chosen males during or after copulation when they mate with more than one male in the same reproductive event (Jennions and Petrie 2000, Birkhead and Pizzari 2002). For example, evidences for a post-copulatory cryptic mate choice for MHC characteristics has been found in red junglefowl (*Gallus gallus*). In this study, the number of sperm counted on eggs was found to be biased in favour to MHC-

dissimilar males (Løvlie et al. 2013). The post-copulatory cryptic mate choice may also explain why, in Alpine marmots, the number of EPY but not the presence of EPY increased when social pairs share a high number of MHC alleles (Chapter 2). To determine if the evidence found for a biased mate choice in Alpine marmots is the result of a cryptic female choice, controlled mating experiments are required. This project might be conducted with the same individuals used to study the sperm competition in Alpine marmots (see the previous paragraph) since both sperm competition and cryptic female choice are two processes operating at the same time.

## **2. THE DIFFICULTIES OF THE STUDY OF PARASITES**

Host-parasites interactions are a central part of ecological communities and a key to understand ecosystems. For this reason, the study of host-parasites interactions is a major interest for a large number of researchers from numerous fields (e.g. veterinary, ecology, evolutionary biology or molecular ecology). Despite the large interest that host-parasites interactions could generate, multiple difficulties are associated to the study of host-parasites interactions in the wildlife. I would like to highlight here some of these difficulties as well as to propose some clues for further investigation that could be applicable to our study.

Hosts may be infected by multiple parasite species throughout their lives. Moreover, hosts simultaneously infected with several parasite species, are common in the wild (Petney and Andrews 1998). Thus, the study of the overall parasite community throughout the host's life seem to offer an excellent opportunity to understand the exact nature of host-parasites interactions. However, most of the studies examining associations between MHC characteristics and parasites have used few number of parasite species (Sin et al. (2014)). The low number of studies using multiple parasite species is probably

due to the difficulty in determining and quantifying all the parasite species (see below for possible methods to overcome this challenge depending on parasite species). An additional difficulty is that it is often a challenge to obtain large number of samples from multiple individuals throughout their entire life, particularly in non-model species.

In the present thesis, three intestinal parasites (*E. monacis*, *Eimeria* sp. and *C. marmotae*) of Alpine marmots were evaluated (Chapter 3), but this study failed to include blood parasites (if any) or other intestinal parasites such as nematodes. Blood parasites were not integrated in this work despite that blood samples are collected when Alpine marmots are captured. The main problem for the study of blood parasites in Alpine marmots is that, tot date, little is known about its blood parasites (Bassano 1996). Thus, the first step before the study of MHC characteristics and blood parasites is to screen for blood parasite determination. One way to determine blood parasites could be the use of NGS techniques for taxonomic identification *via* the design of a DNA barcoding project (see Pompanon and Samadi 2013 for more detailed information on the use of NGS for taxonomic identification). The advancements of NGS techniques (Chapter 1) and the development of DNA barcoding projects (Pompanon and Samadi 2013) should enable the application of this technique in Alpine marmots.

Also, the nematode *Ascaris laevis* (the most common nematode egg found in the analysed faeces) was not integrated in this work because the egg emission period does not match with the main capture period (most of the captures are realized between mid-May to mid-June whereas egg emission start mid-July). Indeed, all adult *A. laevis* are probably eliminated from the small intestine during hibernation and then there is a delay between contamination (starts at the beginning of the active period) and egg-laying because there is a visceral migratory phase of larvae in lung and liver before the arrival of the adults in the small intestine (Callait and Gauthier 2000).

A way to solve the lack of data regarding the presence and the abundance of *A. laevis* in faecal samples but also to solve the lack of data regarding the presence and the abundance of the overall parasite community throughout the

year would be to repeatedly capture individuals during the overall active period. Although around 80% of Alpine marmots followed in the Nature Reserve of the Grande Sassi re are captured at least once per year, individuals are more difficult to capture as the active period progresses. Moreover, multiple captures of the same individuals during the same year are rare (mean number of captures per year and individual ( $\pm$ SD):  $1.31 \pm 0.85$ ). Interestingly, the yellow-bellied marmots (*M. flaviventris*) followed in the Rocky Mountain Biological Laboratory are regularly trapped throughout the active period (around twice per month from mid-April to early September), which has allowed to collect faecal and blood samples during their overall active period (Lopez et al. 2013). Given that such procedure is not viable in Alpine marmots, alternative sampling methods may be explored. For example, molecular genotyping of faecal samples could resolve this lack of data if faecal samples can be reliably identified. To do so, fresh faecal samples could be collected in the external latrines of different family groups for DNA fingerprinting. DNA fingerprinting is a powerful tool in conservation genetics because it enables to monitor the dynamics of endangered populations, obtain estimates of temporal trends of demographic processes and obtain genetic parameters difficult to obtain by traditional capture-recapture programs (Caniglia et al. 2014). Because numerous research groups have already been working on DNA fingerprint, the progress of this technique allows nowadays to reliably associate a faecal samples to a unique individual at an affordable cost (Kraus et al. 2015). Although DNA fingerprint could be realized using microsatellites, the use of microsatellites may suffer from several drawbacks such as the lack of standardisation between laboratories and high error rates (see Kraus et al. 2015 for developed arguments). For this reason, Kraus et al. (2015) recently developed a genotyping protocol that uses single-nucleotide polymorphism (SNP)-based marker and they successfully applied it to reliably assign faecal samples to grey wolves (*Canis lupus*) using a 96-SNP panel. Although the use of such protocol in Alpine marmots necessitates the development of a large panel of SNP's, the discovery of around one hundred of SNP's should not represent a challenge due to the advancements of NGS techniques (Chapter 1).

### 3. MHC-CHARACTERISTICS AND FITNESS

MHC is considered as a set of functional genes that have a key role on individual fitness (Bernatchez and Landry 2003). Fitness could be defined as either the lifetime reproductive success (LRS), which includes both survival and reproductive success (Clutton-Brock 1988) or the individual contribution to population growth (Coulson et al. 2006)). The same way that survival may depend on MHC characteristics (see Chapter 4), other components of individual fitness may also depend on MHC-characteristics.

First, MHC diverse individuals are expected to have a higher resistance to parasitic infections because they are supposed to recognize a wide range of antigens (Doherty and Zinkernagel 1975, Hughes and Nei 1989, Penn et al. 2002). Also, the possession of specific MHC alleles is expected to confer a high resistance to specific parasites (Bodmer 1972, Slade and McCallum 1992). As a consequence, individual MHC characteristics are expected to play a role on the LRS (1) by enabling MHC diverse individuals and/or individuals carrying specific MHC alleles to increase their life expectancy, which in turn is expected to increase their number of reproductive events, and/or (2) by enabling individuals possessing advantageous MHC genes to increase their offspring survival *via* the transmission of these advantageous MHC characteristics from the parents to their offspring.

Second, in species where female mate choice is based on MHC characteristics (Yamazaki et al. 1976, Wedekind et al. 1995, Olsson et al. 2003, Radwan et al. 2008, Schwensow et al. 2008a,b, Eizaguirre et al. 2009, Juola and Dearborn 2012, see Chapter 2 for more detailed examples), the reproductive success is expected to depend on the MHC characteristics. On one hand, if individuals are chosen according to the good genes hypotheses (Brown 1997,

Neff and Pitcher 2005), only MHC diverse males and/or males carrying specific MHC genotypes are expected to be chosen as sexual mates. As a consequence, MHC diverse males and/or males with specific MHC genotypes are expected to have a higher reproductive success. On the contrary, if individuals are chosen according to the genetic compatibility hypothesis, all individuals are expected to reproduce if they find their corresponding sexual partner because mate choice is based on a relative criterion specific to each female (Trivers 1972, Zeh and Zeh 1996, 1997, Neff and Pitcher 2005). In this particular case, as suggested in the Chapter 2 for Alpine marmots, all individuals should be able to reproduce.

Third, there is some evidence that MHC characteristics of a given couple may affect its reproductive success *via* the maternal-foetal interactions. Indeed, the gametic union of too similar MHC genotypes has been suggested to produce maternal-foetal interactions that trigger spontaneous abortions. For example, female abortion in humans (Ober et al. 1998) or in pigtailed macaque (*Macaca nemestrina*, Knapp et al. 1996) has been associated with a high number of shared MHC alleles between the couple. In these cases, sexual partners with high MHC similarity are expected to have a low reproductive success.

While the associations between MHC characteristics and survival are still being studied in wild populations (e.g. Paterson et al. 1998, Brouwer et al. 2010, Dunn et al. 2013, Sepil et al. 2013), the existence of associations between MHC characteristics and other components of fitness such as reproductive success or LRS have still been poorly studied under natural conditions. For example, a study on rhesus macaques (*Macaca mulatta*) found that males with higher MHC diversity had higher reproductive success (measured as the number of sired offspring) than males with less MHC diversity (Sauermann et al. 2001). Also, a study on Egyptian vultures (*Neophron percnopterus*) found that birds with higher MHC diversity had a higher breeding success (measured as the number of successful reproduction attempts divided by the number of reproductive attempts) (Agudo et al. 2012). Additionally, a study on great tits (*Parus major*) found that the presence of three different MHC supertypes (i.e. alleles with similar functional effects) was associated to survival, annual recruitment and



LRS (Sepil et al. 2013). On the contrary, a study on collared flycatchers (*Ficedula albicollis*) found no evidence for an association between LRS and MHC (Radwan et al. 2012). With this weak number of studies (the above cited examples represent nearly all published studies) it is difficult to draw general conclusions about the significance of eventual associations between MHC characteristics and individual fitness. Further exploration in an enlarged number of species is thus required.

There are probably few studies that have analyzed the associations between MHC characteristics and fitness, not because this topic lack of interest, but because obtaining reliable measures of fitness in the wild is still a challenge for ecology and evolutionary biologists (Clutton-Brock and Sheldon 2010). Indeed, research projects in the wild that attempt to measure fitness necessitates long-term and individual-based study systems where individuals are marked, breeding attempts are recorded and the offspring outcome is known (Clutton-Brock and Sheldon 2010). If collecting reliable data is in itself a challenge (see Clutton-Brock and Sheldon 2010 for detailed description of challenges), the major problem that researchers should face is the difficulty associated to the maintenance of funding without interruptions. Although scarce, some long-term studies with such kind of data exist (see Clutton-Brock and Sheldon 2010, Kappeler and Watts 2012 for a list of nearly all individually-based long-term field studies). Fortunately, the high quality, long-term, individual-based study carried out on Alpine marmots living in the Nature Reserve of the Grande Sassièrè is one of these. Moreover, this project offers an excellent opportunity to test whether MHC characteristics affect LRS and should be the subject of future research now that all MHC characteristics of the followed individuals are already available (Chapter 1).

#### 4. THE STUDY OF INTRA-SPECIFIC GENETIC DIVERSITY

The study of a high quality and detailed long-term individually based monitoring of Alpine marmots of the Nature Reserve of the Grande Sassi re has allowed a comprehension of the role of MHC characteristics in the population (Chapters 2 to 4). However, environmental heterogeneity, parasite pressure heterogeneity or demographic heterogeneity that exist between populations suggest that the role of MHC diversity may also differ (Bernatchez and Landry 2003). For example, in a study carried out on eight natural populations of three-spined sticklebacks (*Gasterosteus aculeatus*), populations being exposed to a higher parasite diversity had higher MHC diversity than populations with a lower parasite diversity (Wegner et al. 2003b).

Interestingly, the Alpine marmots of the Pyrenees became extinct probably during the Pleistocene / Holocene transition (Preleuthner and Pinsker 1993, Rassmann et al. 1994) and the Alpine marmots currently occupying the Pyrenees were introduced into the northern Pyrenees during the second half of the twentieth century from individuals captured in the French Alps (Herrero et al. 2000, L pez et al. 2010, see material and methods for more details). Since their introduction in the Pyrenees, this new meta-population has probably faced parasite species or parasite strains different from the meta-population of the Alps. Thus, a fast co-evolution under the form of arms-races (Van Valen 1973) between the marmots living in the Pyrenees and their parasites may have arisen. As a consequence, we expect MHC characteristics to be different between the two mountain ranges (Alps and Pyrenees) and between the different populations within a given mountain range.

To test this hypothesis, we started a project aiming to compare the individual and population MHC characteristics in marmots living in the Alps and in the Pyrenees (Appendix C). In a collaboration with a research group of the Universitat Aut noma de Barcelona (Catalonia), we are using the collected

information from the studied population of the present thesis (located in the Nature Reserve of the Grande Sassi re, Alps) and another high quality and highly detailed dataset from a population located at the Pyrenees. Indeed, a population of Alpine marmots situated in the Llosa valley (at 1,900 m a.s.l., Catalan Pyrenees, 42° 26'N, 1°42'E) is followed since 2008 by the research group of the Universitat Aut noma de Barcelona with the same methodologies used in the Nature Reserve of the Grande Sassi re. Also, we have collected genetic, faeces and blood samples from 7 Alpine and 3 Pyrenean contemporaneous populations in order to compare genetic characteristics and data on parasite variability and distribution among different marmot populations (Appendix C). A post-doc is currently genotyping the overall collected samples to perform statistical analysis which will allow us to better understand how different selective pressures impact the MHC characteristics of the different populations in the two mountain ranges.

### *A look at the past*

The low intra-specific genetic diversity found at several genetic markers such as allozymes (Preleuthner and Pinsker 1993), minisatellites ( Kruckenhauser et al. 1997, Rassmann et al. 1994), microsatellites (Cohas et al. 2009, but see Goossens et al. 2001) and MHC loci (Kuduk et al. 2012, Chapter 1) in Alpine marmots living in the whole Alpine arc is an important concern of the wildlife managers of the La Vanoise National Park (France). Indeed, populations with low intra-specific genetic diversity may be at risk of having a reduced capacity to adapt to changing environments (Keller et al. 1994). The low genetic diversity found in Alpine marmots has been hypothesized to be the result of bottlenecks that may have occurred during the Holocene period (Preleuthner and Pinsker 1993, Rassmann et al. 1994). Moreover, the drastic reduction of effectives during the Holocene period has even been advanced to have caused the extinction of this species in the Pyrenees. Low levels of intra-specific genetic diversity have been also observed in species that have undergone recent or historical bottlenecks such as cheetahs (*Acinonyx jubatus*, O'Brien et al. 1983), brown bears (*Ursus arctos*,

Valdiosera et al. 2008), or Iberian lynxs (*Lynx pardinus*, Rodríguez et al. 2011).

During my thesis, we started a collaboration with the Institute of Evolutionary Biology (CSIC – Universitat Pompeu Fabra, Catalonia) to investigate when the decline of intra-specific genetic diversity in Alpine marmots occurred (i.e. if the genetic diversity decreased during a bottleneck coinciding with the Pleistocene/Holocene transition, if the genetic diversity decreased much later or if the genetic diversity has been low throughout the last 100 000 years) and to evaluate if Alpine and Pyrenean populations experienced different demographic histories (Appendix C). To do this, the genotyping of a series of sub-fossil samples from different geographic locations within the Alpine and the Pyrenean arc is needed (Appendix C). The study of the mtDNA control region appears as a likely genetic system to carry out such study. Indeed, mtDNA control region is the most variable region of the mitochondrial genome (Brown et al. 1993), which enables a good assessment of the intra-specific genetic diversity. Additionally, mtDNA has more chances to be successfully genotyped than nuclear DNA (mtDNA is generally better conserved than nuclear DNA in sub-fossil samples). Given that this is a current project, some of the steps are already done while others still remain to be done. First, we found around 100 sub-fossil samples (bones and teeth) spanning more than 100 000 years from Alpine and Pyrenean populations that are located in different museums (Table 43). The collection of geology located at the Université Claude Bernard Lyon I, which contains some of the available sub-fossil samples, ceded to this project three bones belonging to different individuals from the Châtillon-Saint-Jean population (Table 43) (Appendix C). Their DNA was extracted from the three bones in the Institute of Evolutionary Biology to avoid DNA contamination with the contemporaneous samples, processed at the Laboratoire de Biometrie et Biologie Evolutive (Université Claude Bernard Lyon 1, France). Then, we attempted to sequence these sub-fossil samples at the four MHC polymorphic loci described for contemporaneous populations (Kuduk et al. 2012, Chapter 1). Unfortunately, we failed to genotype these samples, probably because of the advanced state of degradation of the genetic material. The next steps of this project are to attempt

to genotype these sub-fossil samples at mtDNA (work in progress) as well as to genotype all sub-fossil samples available (Table 43).

Geographic region	Number of samples	Age	Site	Country
Alps	1	Middle pleistocene	Val d'Arno	Italy
Alps	13	Late pleistocene	Châtillon-Saint-Jean	France
Alps	1	Mesolithic	Culoz	France
Alps	2	NA	Poncin	France
Alps	2	Pleistocene	Villereversure	France
Alps	2	Holocene	Les Déserts	France
Alps	3	Pleistocene	Solutre-Pouilly	France
Alps	1	Quaternary	Aillon-le-Jeune	France
Alps	1	Quaternary	Evosges	France
Alps	1	Quaternary	Flaine	France
Alps	1	NA	Etrigny	France
Alps	1	NA	Valais	Switzerland
Alps	2	NA	Massif de la Vanoise	France
Alps	2	12-9 Ky	Rencurel	France
Alps	14	12-9 Ky	Presles	France
Pyrenees	2	Pleistocene	Isòbol	Catalonia
Pyrenees	10	130-74 Ky	Huermece	Spain
Pyrenees	23	125-10 Ky	Mondragón	Spain
Pyrenees	10	Quaternary	Cantábrilo	Spain

**Table 43.** Number of subfossil samples from universities or museum collections and their corresponding location. NA: non-available; Ky: Kilo Years.

## 5. FROM MHC TO THE STUDY OF OVERALL IMMUNE GENES

Although the almost universal acceptance of MHC as a significant factor in vertebrate parasite resistance (see Sin et al. 2014 for a recent review) and sexual

selection (see reviews in Bernatchez and Landry 2003, Brennan and Kendrick 2006, Milinski 2006), other immune-related proteins (e.g. immunoglobulin receptors, interleukin receptors, Mannose binding lectin or Toll-like receptors) also participate in immune response against parasites (Acevedo-Whitehouse and Cunningham 2006, Schmid-Hempel 2011) and may also be sexually selected (but see Penn and Potts 1999).

All living organisms (from prokaryotes, plants and invertebrates to early and jawed vertebrates) have innate immune systems (jawed vertebrates have also acquired immunity systems) which can be used for self-defence. These immune systems depend on a great number of genes that encode for different protein families: immunoglobulin-superfamily, leucine-rich repeats, lectins, tumour necrosis factor family, cytokine receptor families, chemokine receptor families, peptidoglycan-recognition proteins, gram-negative-binding proteins, scavenger receptors, down syndrome cell adhesion, fibrinogen-related protein, variable domain chitin-binding proteins or anti-microbial peptides (see Schmid-Hempel 2011 for detailed information on characteristics and functions of these main immune defence protein families). Across the immense diversity of living organisms, all species possess at least some of the above cited protein families. Interestingly, most of these immune defence proteins have been maintained in phylogenetically distant species (Schmid-Hempel 2011). As in MHC, high levels of allelic polymorphism are expected in genes encoding for these immune-related proteins; and mutation, balancing selection or gene conversion are likely to favour their genetic polymorphism (Schmid-Hempel 2011, Pespeni et al. 2012).

To date, few studies of immunogenetics in non-model vertebrate species have examined immune genes other than MHC. For example, in a study on field voles (*Microtus agrestis*) Turner et al. (2012) found that, in multiple loci encoding cytokines and three non-cytokines immune-genes, the diversity at these loci is consistent with balancing selection maintaining genetic polymorphism. Interestingly, in a study on the threatened Pale-headed brushfinch (*Atlapetes pallidiceps*) and two non-threatened brushfinch species (*Atlapetes latinuchus*, *Buarremon torquatus*) Hartmann et al. (2014) found that Toll-like receptor (TLR)

diversity in *A. pallidiceps* is significantly lower than TLR diversity in both *A. latinuchus* and *B. torquatus*. However, the low levels of TLR found in *A. pallidiceps* do not seem to compromise the long-term fitness of the population given that in this species a high TLR diversity was found to be linked to a decreased survival probability (Hartmann et al. 2014). However, the limited number of studies investigating the role of non-MHC genes does not allow for generalizations about the importance of these genes in the wild. The incorporation of genes other than MHC to the study of the different roles of immunogenetic diversity in wild vertebrates is thus required (Acevedo-Whitehouse and Cunningham 2006, Vinkler and Albrecht 2009, Turner et al. 2012, Pespeni et al. 2012).

The extension of knowledge to a wider range of immune defence proteins in vertebrates, but also to the wide diversity of organisms should allow us to better understand the role of the overall immune system in the wild, and this in a great variety of host-parasite contexts. Thus, I would like to highlight the importance of considering additional genetic regions other than MHC and move from the study of a single genetic family (here MHC) towards the study of a fuller picture of the immune defence related genes. Moreover, the study of other genes such as those encoding for TLR have recently become a promising genetic system. Contrary to the high number of pseudogenes and the recent duplications that complicate MHC loci sequencing (especially in birds), TLR can be easily amplified in diverse bird species (Alcaide and Edwards 2011, Grueber et al. 2012). If such kind of studies seemed unimaginable few years ago due to the technical difficulties associated with polymorphic genes genotyping (see Chapter 1 for more details), the advances of NGS techniques coupled with bioinformatic progress that allow to reliably genotype a high number of individuals (Galan et al. 2010, Babik et al. 2009, Radwan et al. 2012, Huchard et al. 2012, Lighten et al. 2014b, Stutz and Bolnick 2014, Chapter 1) should enable the use of a broad number of genes in future immunogenetic studies.

## 6. CONCLUSIONS

This thesis presents a body of work that answers many questions related to the genotyping process of highly polymorphic loci and the role of intra-specific genetic diversity at MHC loci in a wild population of Alpine marmots. Yet, this thesis also results in a series of questions that identify opportunities for further investigation, both in this species and in a wide range of organisms. Throughout both personal and team work projects I hope I made an original and valuable contribution to increase knowledge in the fields of molecular ecology and evolutionary ecology. To conclude, I would like to highlight that this thesis is just a start - I hope I will continue working in this field - and as the catalan poet Miquel Martí i Pol (Pol 1981) said, *tot està per fer i tot és possible* (everything is yet to be done and everything is possible).





## **APPENDIX A:**

# **THE DEAR ENEMY PHENOMENON**

---

**The dear enemy phenomenon in Alpine marmots: responses from orbital gland secretions from dominant males**

M Ferrandiz-Rovira, V Berger, and A Cohas

*Draft of the introduction and material and methods sections*



*This appendix contains the theoretical framework and the material section of a project investigating the existence of the dear enemy phenomenon in Alpine marmots. MFR and AC conceived and designed this project as an adjacent work of this thesis.*

## **INTRODUCTION**

A territory is an area occupied by individuals or a group of individuals and it is defended from incursions by other members of its species (Maher and Lott 1995). Assuming that intra-specific competition occurs for limited resources, owning a territory provides an advantage regarding the exploitation of resources. Indeed, owning a territory may confer benefits such as the access to nest sites, to food resources or to mate opportunities. In multiple-purpose foraging territories (e.g. feeding and breeding sites), the possession of a territory is a key factor for the owner fitness (Temeles 1994). Individuals use acoustic, olfactory or visual signalization as territorial advertisement to mark territorial boundaries to their conspecifics. These advertisements are supposed to discourage the entry of intruders in the territory and therefore entail benefits such as the avoidance of fights. However, the defence of a territory implies costs of time and can potentially increase the risk of predation (Bradbury and Vehrenkamp 1998). If such advertisements do not discourage intruders, chases and fighting follow, with an associated risk of injury or death (Bradbury and Vehrenkamp 1998). Territorial behaviour can therefore only be developed when obtained benefits are higher than the cost of defence (Brown 1964).

To minimize the risk of injury or death in a fight for a territory, individuals may evaluate the opponents' characteristics and modulate their reactivity according to the threat that opponents represent. For example, assuming that owners receive information about advertisements from their neighbours, the owner of a territory would not assess their neighbours as intruders because they already own a territory. Then, individuals can consider neighbours as less

threatening than stranger individuals. On the contrary, strangers may be considered as a greater threat with respect to neighbours because they may be potential territory usurpers and try to evict the owner of a territory by fighting (Temeles 1994). Indeed, the cost of landless stranger's fight for gaining a territory is reduced compared to the cost of the actual owner for losing a territory. Therefore, behavioural differences towards strangers and neighbours, termed as the dear enemy phenomenon (Fisher 1954, Temeles 1994), are expected, with more aggressively behaviour towards strangers than towards neighbours.

The dear enemy phenomenon has been found in a great variety of taxa. For example, in invertebrates (Heinze et al. 1996, Booksmythe et al. 2010), in birds (Stoddard et al. 1990) or in mammals (Palphramand and White 2007, Monclús et al. 2014; see Temeles 1994 for a review of other studies and other taxa). Species validating the dear phenomenon are mostly organised in multiple-purpose foraging territories (Temeles 1994), with relatively stable territories where the discrimination between neighbours and strangers may be facilitated (Stoddard et al. 1990, Rosell and Bjørkøyli 2002). However, some experiments failed to observe the dear enemy phenomenon (e.g. in strawberry dart-poison frog (*Oophaga pumilio*) (Bee 2003), in buff-banded rail (*Gallirallus philippensis*) (Lachish and Goldizen 2004), see Temeles 1994 for a review of studied species that did not exhibit the dear enemy phenomenon).

In mammals, advertisements regarding the ownership of a territory are supposed to be mainly realized by the use of chemical signals. Indeed, chemical compounds may contain honest information about the sender, they can inform about the age (Buesching et al. 2002, MacDonald et al. 2008, Kean et al. 2011), the sex (Zhang et al. 2005, Setchell et al. 2010), the social status (Hayes et al. 2001, Burgener et al. 2009) or the body condition (Buesching et al. 2002). Moreover, scent marking may alert about the occupation of a territory, even in the absence of the owner (Ralls 1971). Evidences of dear enemy phenomenon have been found for different types of gland secretions. For example, in Eurasian beavers (*Castor fiber*) (Rosell and Bjørkøyli 2002) and in Columbian ground

squirrels (*Uroditellus columbianus*) (Raynaud and Dobson 2011), the dear enemy phenomenon was validated with anal and orbital gland secretions respectively. However, in yellow-bellied marmots (*Marmota flaviventris*) no evidence of the existence of the dear enemy phenomenon was found with anal gland secretions (Cross et al. 2013).

Alpine marmots (*Marmota marmota*) is a socially monogamous and a cooperative breeding species who live in family groups from 2 to 20 individuals (Allainé 2000). Family groups are composed of a dominant couple, subordinates of both sexes, and pups of the year (Allainé et al. 2000). The dear enemy phenomenon is likely to exist in this species for at least three reasons. First, they live in a multi-purpose territories (territories are their feeding and breeding sites). Territories have adjacent boundaries between families that facilitate the contact and the recognition of neighbours. The low turnover of dominant individuals could lead to the stable relations between owners and neighbours (Temeles 1994). Second, strangers represent a important threat. Indeed, mature subordinates (2 years and older) of both sexes start dispersing to reach dominance by replacing a dominant and the aggressive encounters between dispersal and dominant individuals from the same sex are common and can cause severe infections as well as the eventual loss of a territory for the dominants. Moreover, landless individuals are highly subject to fitness costs. Indeed dispersing individuals failing to reach dominance (Stephens et al. 2002) or dominant individuals losing their territory have high mortality rates (Lardy et al. 2011). Consequently, stranger individuals may represent a higher threat for dominant marmots than their neighbours already owning a territory. Third, territory advertisement through scent-marking behaviour could be a key factor to prevent aggressive encounters between individuals. In Alpine marmots, cheek rubbing by orbital glands is the main territory marking mode (Bel et al. 1995, 1999).

Here, we investigated the existence of the dear enemy phenomenon in dominant individuals of both sexes living in a wild population of Alpine marmots

using orbital gland secretions collected on dominant males. Although dominants are only replaced by individuals of the same sex, dominant females may also display the dear enemy phenomenon through males to avoid the higher costs in term of fitness associated to mate change. Indeed, in Alpine marmots a change of dominant male lead to the decrease of annual reproductive success (no young at emergence and infanticide of male yearlings) (Lardy et al. 2011). We thus conducted a series of two-way choice experiments to determine if Alpine marmots respond differently towards neighbour and stranger scents and, respond differently depending on the sex of the individual interacting with the scent. We predicted that Alpine marmots would spend more time smelling (as a measure of identification of the smell as well as an measure of interest) and/or marking (as a measure of aggressiveness) the scents from strangers than those from neighbours. We also predict distinct responses towards scents of dominants males depending on their sex.

## **MATERIAL**

### *Studied species*

Alpine marmots territory size varies between 0.9 ha and 2.8 ha (Perrin et al. 1993b). They have a main burrow used to hibernate from mid-October to mid-April, to sleep during the night and to give birth (Perrin et al. 1993b) and have secondary burrows mainly used to escape from eventual predators. Moreover, all their food source is drawn from their territories.

Sexually mature subordinates of both sexes can reach dominance by replacing its father or mother. Otherwise, mature subordinates start dispersing to reach dominance mainly by replacing a dominant from another territory or rarely by establishing a new territory, (only five cases observed between 1990 and 2014 in the studied population). The main dispersal period is just after hibernation and early summer, where aggressive encounters between dispersal and dominant individuals from the same sex are common (MFR, VB and AC personal

observations).

Orbital gland is an exocrine gland located in the dermis and its secretive activity is age-dependent, with no secretion observed in juvenile Alpine marmots (Bel et al. 1999). Previous studies on Alpine marmots showed that (1) scent-marking mainly occurs between the breeding period and emergence of offspring (Bel et al. 1995, Lenti Boero 1995); (2) scent-marking is spatially distributed and mainly involves boundaries of the territory and main burrows (Bel et al. 1995, Lenti Boero 1995); (3) although all mature individuals of a family group scent-mark, its mainly done by dominant individuals (Bel et al. 1995); and (4) foreign scent marks are removed or replaced (Bel et al. 1995). Consequently, in Alpine marmots orbital glands may be used to assess information about intruders and a signalling mode that would enable the presence of the dear enemy phenomenon this species.

### *Field methods*

Data were collected from 2012 to 2014 in a wild population of Alpine marmots located in the Nature Reserve of La Grande Sassi re (at 2,340 m a.s.l., French Alps, 45°29'N, 65°90'E). Over 80% of the marmots from this population have been captured since 1990 at least once every year, between mid-April to mid-July, using two-door live traps baited with dandelions (*Taraxacum densleonis*). Once captured, individuals were tranquillized with Zol til 100 (0.1 mL Kg<sup>-1</sup>), sexed, aged from their size (up to 3 years) and their social status was determined according to scrotal development for males and teat development for females. Individuals were tagged with one metal tags at their ear for visual identification of sex (right side for females and left side for males). Moreover, only dominants were also tagged with a coloured plastic tag (right side for males and left side for females, see Figure 22 for an illustration). The composition of family groups was assessed from both capture–recapture data and intensive observations (see Cohas et al. 2006 for details on observations protocol) and scent marking behaviour was used to confirm the identity of the dominant pair (Bel et al. 1999). Moreover,



borders of each family group as well as main burrows were determined from intensive observations.



**Figure 22.** Dominant male (with a plastic tag on the right ear indicating its social status and sex) smelling one of two glass tubes placed during an experiment.

### *Sample collection*

Scent samples were collected from 34 dominant males (13 at 2012, 22 at 2013 and 12 at 2014). First, we shaved orbital gland area with sterile cotton swab saturated with ethanol and we waited full evaporation to avoid polluting samples. Second, we pressed orbital gland area and sampled outgoing secretions by rubbing the distal half of a clean glass tube (20 mm long x 25 mm outside Ø, vwr® reference number 212-1126, Figure 22). We manipulated glass tubes with aluminium foil to avoid human scent contamination, we wrapped the tubes in aluminium foil and we maintained them at ambient temperature (approximately 5°C) until the experiment. We kept manipulated tubes no more than three days after the sample collection. Because of fieldwork constraints, we collected between one and six glass tubes per dominant.

### *Experimental design*

The test device consisted of two wooden sticks covered by glass tubes (20 mm

long x 25 mm outside Ø, vwr® reference number 212-1126, Figure 22). We placed glass tubes 50 cm apart to allow individuals to differentiate between scents of each tube and at approximately 50 cm from the main burrow entrance to maximize the chances that a given marmot will be involved in the experiment (Figure 22). We placed permanently two clean glass tubes in each marmot territory to avoid exploration behaviours due to the introduction of a new object in their territory. We replaced these tubes by experimental glass tubes when the experiment started. We cleaned glass tubes with water and alcohol and, once naturally dried, we wrapped them in aluminium foil. We always manipulated glass tubes with aluminium foil to avoid human scent contamination.

We distinguished three types of tubes for the experiments:

- 1) Stranger tube: a glass tube with orbital gland secretion from a dominant stranger male. We considered individuals as strangers when any boundary between the two territories was adjacent. Intensive observations of Alpine marmots allowed us to determine that marmots do not have a discovering behaviour outside their own territory once they become dominants and we thus considered stranger individuals as unfamiliar individuals.
- 2) Neighbour tube: a glass tube with orbital gland secretion from a neighbour dominant male. We considered individuals as neighbours when at least one boundary between the two territories was adjacent to the boundary of the family.
- 3) Empty tube: a cleaned glass tube to control for a purely interest on this object.

We carried out three types of experiments: (1) stranger vs. neighbour; (2) stranger vs. empty; and (3) neighbour vs. empty. We randomly assigned glass tubes to each wooden stick to control for a potential bias of individual behaviour through one side. We carried out all experiments between 18th Mai and 27th June from 2012 to 2014. This period corresponds to the end of gestation and lactation, and until the emergence of the pups of the year. We performed experiments during the main activity period of the active season (between 8:00 a.m. and 6:00 p.m). Given that since 1990 researchers penetrate into the marmot territories at least once per day to bait the traps during the annually capture season (around 45

days per year per family group), we assumed that human presence during the experiment installation did not disturb marmots. Although individuals entered into their burrows when humans approach their territory, they typically emerged from their burrows within minutes once researchers moved away from their territory. From the moment of experiment installation until the end of the experiment, between one and three people were continuously observing around the experiment with 10 x 42 binoculars and/or 20 x 60 telescopes. We identified individuals thanks to the ear tags. Moreover, we recorded experiments with a digital video camera (Sony® Handycam model DCR-DVD650 or JVC® digital video model GZ-E 209).

Each experiment was considered as a success when a member of the dominant couple interacted with at least one glass tube (smelled or marked). An experiment was finished when the given individual was either out from the surrounding area where the experiment was placed or once it was inside the main burrow. After a successful experiment, we replaced the glass tubes by fresh ones in order to try a successful experiment with the other member of the dominant couple. We did not take into account those experiments in which subordinates interacted with glass tubes and when it occurred, we replaced these tubes by fresh glass tubes to avoid contamination. In the case of dominants were too far from the tubes within three hours after experiment installation, we stopped the experiment and we repeated it another day with new scent samples.

### *Response measurements*

We displayed video recordings with Microsoft Windows Media Player (Microsoft®). First, we viewed all videos in real time and noted all behaviours. The ethogram of behaviours exhibited by dominant marmots during the experiments was established (see Table 44 for the behaviours included in the ethogram). Second, we reviewed all videos in slow motion (x 0.5) to ensure an accurate identification of behaviours as well as their duration (with an accuracy of 0.5 s). Finally, we recorded the total duration in seconds of the behaviours

implicated on the dear enemy phenomenon: smelling (considered as a measure of identification of the smell as well as an measure of interest) and marking (considered as a measure of aggressiveness). Moreover, we also recorded the number of marks and considered as a complementary measure of aggressiveness.

Behaviours and actions	
Smell right tube*	Scratches or make a hole
Smell left tube*	Walk or run
Mark right tube	Sitting or lying
Mark left tube	Standing
Smell near right tube§	Grooming
Smell near right tube§	Eat
Smell another place	Run away
Smell while walking	Return to the burrow
Mark another place	Go out from the burrow
Move the tail	Watch over
Interaction with another marmot	

**Table 44.** Ethogram of behaviours and actions exhibited by Alpine marmots during the dear enemy experiments. \*: The maximal distance between Alpine marmot nose and the tube is around 5 cm. §: Distance between Alpine marmot nose and the tube from 5 cm.

We analysed a total of 108 videos (23 for a neighbour tube vs. a stranger tube experiment; 36 for a neighbour tube vs. an empty tube experiment; and 49 for a stranger tube vs. an empty tube experiment. A total of 23 dominant males and 29 dominant females interacted with the glass tubes.



**APPENDIX B:  
GENETIC VARIATION AND  
ALARM CALLS**

---

*Poster presented at "Journée FR41" Lyon, France , February 27th 2014*





## LES CARACTÉRISTIQUES GÉNÉTIQUES AFFECTENT-ELLES LES CRIS D'ALARME CHEZ LA MARMOTTE ALPINE?

T. Lengagne<sup>1</sup>, D. Allainé<sup>2</sup>, D. Lepetit<sup>2</sup>, M. Ferrandiz-Rovira<sup>2</sup>, A. Cohas<sup>2</sup>

<sup>1</sup>Université de Lyon, F-69000, Lyon; Université Lyon 1; CNRS, UMR5023, Laboratoire d'Ecologie des Hydrosystèmes Naturels et Anthropisés, F-69622, Villeurbanne, France

<sup>2</sup>Université de Lyon, F-69000, Lyon; Université Lyon 1; CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, F-69622, Villeurbanne, France.

E-mail: [mariona.ferrandiz-rovira@univ-lyon1.fr](mailto:mariona.ferrandiz-rovira@univ-lyon1.fr)

### INTRODUCTION

L'émission des cris d'alarme fait prendre un risque à l'émetteur, le gain en valeur sélective de ce comportement est à rechercher dans le cadre de la coopération. Les variations individuelles liées à la propension à alarmer sont non négligeables suggérant l'existence de personnalités plus ou moins coopératives. Ces différences peuvent trouver leur origine dans les caractéristiques génétiques des individus.

Notre projet vise à déterminer comment les cris d'alarme sont affectés par les caractéristiques génétiques des individus. Nous cherchons donc:

- 1) à déterminer si les individus diffèrent génétiquement au niveau du gène Avpr1 impliqué dans les comportements sociaux;
- 2) à déterminer les caractéristiques des signaux acoustiques émis par ces individus.

### MODÈLE BIOLOGIQUE

La marmotte alpine (*Marmota marmota*):

- Espèce sociale et territoriale vivant en familles élargies (2-20 individus).
- Cris d'alarme en situations de danger (aigle royal, renard, etc.).
- 26 familles capturées et suivies dans la Réserve Naturelle de la Grande Sassièrre (Savoie).



### ENREGISTREMENT ET ANALYSE DU SIGNAL ACOUSTIQUE

- Utilisation d'un microphone omnidirectionnel Sennheiser ME62 connecté à un enregistreur numérique Tascam FR2LE (Fe 44.1kHz, dynamique 16 bits).
- 13 individus (78 cris) enregistrés (Fig. 1).
- Analyse des cris avec SASLAB. À partir d'un sonagramme, chaque cri est décrit par une série de variables fréquentielles et temporelles (étoiles rouges sur la Fig. 2).



Fig.1

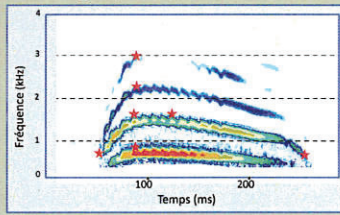


Fig.2

### RÉSULTATS DES ANALYSES DU SIGNAL ACOUSTIQUE

Il existe des différences individuelles importantes dans la structure acoustique des cris:

- 84.2% de la variance totale expliquée par les trois premières fonctions discriminantes (Fig. 3).
- Un chant tiré au hasard dans le jeu de données est attribué à la marmotte qui l'a effectivement émis dans 82% des cas.

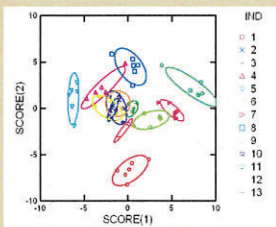
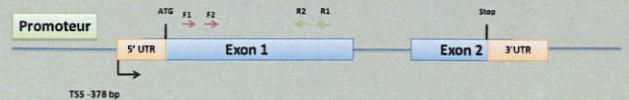


Fig.3

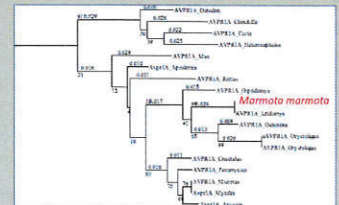
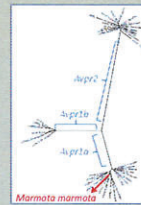
### CARACTÉRISATION DU PROMOTEUR DU GÈNE AVPR1a

Le gène AVPR-1a est responsable de l'expression de récepteurs du système nerveux central qui influencent les fonctions cérébrales chez les mammifères.

- Caractérisation de l'exon 1 du gène Avpr-1a: 27 génomes/transcriptomes de rongeurs disponibles
- Alignement des CDS, protéines, promoteur, mRNA
- Obtention de la structure intron/exon du gène.



- Séquençage de l'exon 1 supposé.
- Alignement et vérification qu'il s'agit de l'exon 1 du gène Avpr-1a chez la marmotte alpine.



- Technique PCR de marche sur le chromosome (FPNI-PCR)

### RÉSULTATS SÉQUENÇAGE DU PROMOTEUR

- 3.2 kB du promoteur caractérisés.
- 2 régions polymorphiques dans le promoteur et 1 région polymorphique dans la région 5' UTR.

### CONCLUSIONS

Notre projet a permis de mettre au point les analyses génétiques ainsi que les méthodes d'enregistrement et d'analyse des signaux d'alarme. Il est maintenant nécessaire de génotyper tous les individus qui ont été enregistrés afin de préciser si les différences individuelles dans les caractéristiques acoustiques correspondent à des niveaux d'expression différents du gène Avpr1.

### REMERCIEMENTS

Nous remercions toutes les personnes impliquées dans la capture des marmottes ainsi que les autorités du Parc National de la Vanoise.





**APPENDIX C:**  
**FOLLOW THE ALPINE MARMOT**  
**TRACKS**

---

*Poster presented at "Journée FR41" Lyon, France , February 27th 2014*



# PISTER LA MARMOTTE ALPINE DANS LE TEMPS ET L'ESPACE - COMPRENDRE L'ÉVOLUTION DE SA DIVERSITÉ INTRA-SPÉCIFIQUE -

M. Ferrandiz-Rovira<sup>1,2</sup>, S. Sauzet<sup>1</sup>, C. Bichet<sup>1</sup>, D. Allainé<sup>1</sup>, O. Ramírez<sup>3</sup>, E. Robert<sup>4</sup> and A. Cohas<sup>1</sup>



<sup>1</sup>Laboratoire de Biométrie et Biologie Evolutive, UMR-CNRS 5558, Université Lyon 1, France

<sup>2</sup>VetAgro Sup, Université Lyon 1, France.

<sup>3</sup>Institute of Evolutionary Biology (CSIC-UPF), Barcelone, Espagne.

<sup>4</sup>Laboratoire de Géologie de Lyon Terre, Planètes, Environnement, UMR-CNRS 5276, Université Lyon 1, France

E-mail: mariona.ferrandiz-rovira@univ-lyon1.fr



## INTRODUCTION

Diversité génétique intra-spécifique → Potentiel adaptatif  
**A PRESERVER**

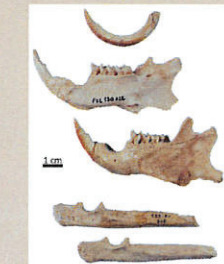
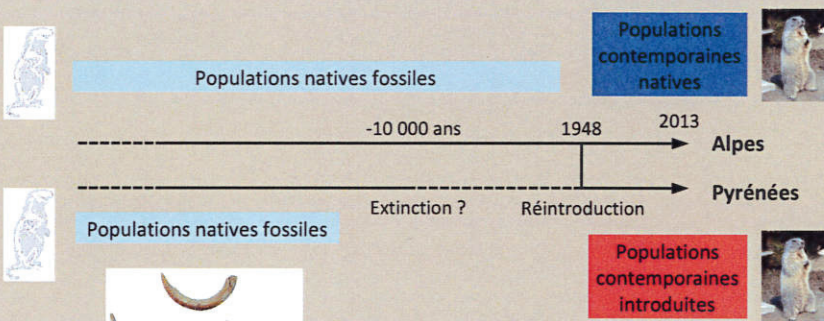
Variation, polymorphisme  
Distribution  
Maintien, transmission  
**A COMPRENDRE**

## OBJECTIFS

- Quantifier l'impact d'événements démographiques sur la diversité génétique
- Identifier et dater les goulets d'étranglement dans les Alpes et les Pyrénées
- Comprendre l'extinction de l'espèce dans les Pyrénées vs. son maintien dans les Alpes

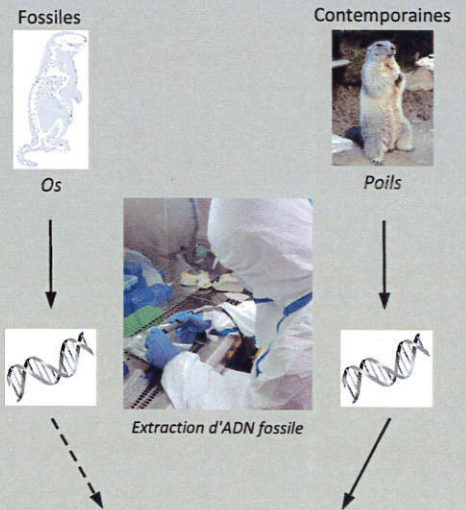
## MODELE BIOLOGIQUE

La marmotte alpine (*Marmota marmota*)



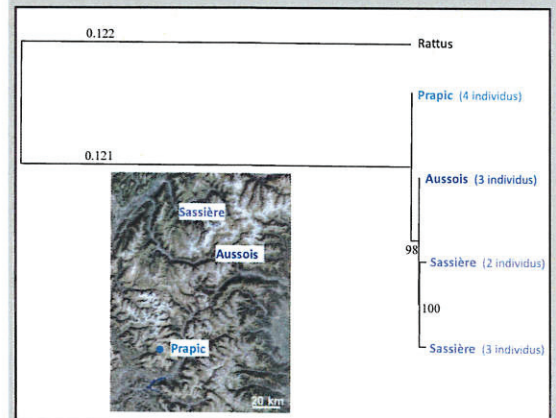
Fossiles (os) de marmottes alpines

## METHODES & RESULTATS



1. Gènes du **CMH** (4 locus) = résistance parasitaire → **sélection**  
Séquençés pour 866 individus issus de 3 populations

2. **ADNmt**, gène du **cytochrome b** = horloge moléculaire → **événements démographiques**



Arbre (NJ, méthode J-C) après séquençage et alignements de 12 individus issus de 3 populations alpines contemporaines. Longueur des branches et poids des nœuds (Bootstrap 1000 répliquations)

## A SUIVRE...

Contemporaines



→ Poursuite du séquençage CMH et cytochrome b

Fossiles



→ Poursuite de l'extraction d'ADN  
→ Amplification du cytochrome b et séquençage  
→ Design des primers et amplification du CMH

## ANALYSES

- L'histoire démographique (logiciel BEAST)
- Diversité génétique (logiciel DNAsp)
- Temps de divergence, patrons de migration (logiciel IMA2)





**APPENDIX D:  
MHC GENOTYPING IN A  
NON-MODEL ANURAN**

---

*Poster presented at "Journée FR41" Lyon, France , February 27th 2014*



# Complexe majeur d'histocompatibilité (CMH) et choix de partenaire chez un amphibien

Hugo CAYUELA<sup>1</sup>, Mariona FERRANDIZ-ROVIRA<sup>2</sup>, Jérôme BRIOLAY<sup>3</sup>, Pierre JOLY<sup>1</sup>, Jean-Paul LENA<sup>1</sup>,

Thierry LENGAGNE<sup>1</sup>, Aurélie COHAS<sup>2</sup>

<sup>1</sup>Laboratoire d'Ecologie des Hydrosystèmes Naturels et Anthropisés

<sup>2</sup>Laboratoire de Biométrie et de Biologie évolutive

<sup>3</sup>Plateau technique du DTAMB, Université Lyon 1



## CONTEXTE

Les gènes du complexe majeur d'histocompatibilité jouent un rôle crucial dans l'initiation et la régulation des réponses immunitaires et sont supposés contribuer à la sélection de partenaires sexuels sur la base de la compatibilité génétique. Si un choix de partenaire CMH-basé a été mis en évidence chez de nombreux vertébrés, son utilisation par certains groupes tels que les amphibiens reste énigmatique. Nous abordons cette question chez un Anoures, le Sonneur à ventre jaune (*Bombina variegata*).

## HYPOTHESES

**H1** : Le choix des femelle est-il CMH-basé ?

**H2** : Le choix de la femelle est-il répétable ?

**H3** : Les femelles sélectionnent-elles les mâles les plus génétiquement dissimilaires sur la base du CMH ?

**H4** : Les femelles sélectionnent-elles les mâles porteurs d'allèles rares du CMH ?



## DESIGN EXPERIMENTAL

- Taille de l'échantillon N = 24 femelles testées.
- Chaque femelle est confrontée à deux épreuves de reproduction pour tester la répétabilité du choix.
- Au cours de chaque épreuve, la femelle effectue un choix entre deux mâles de taille et de condition corporelle comparable. Ce même choix est répliqué 24 h plus tard.
- Soit 48 épreuves de reproduction réalisées dans 10 enclos expérimentaux (4 m x 2 m) composé de deux compartiments à une mare.

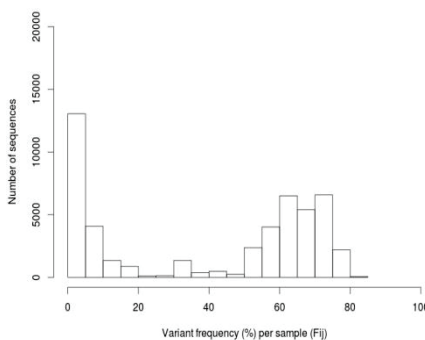


## CARACTERISATION MOLECULAIRE DU CMH

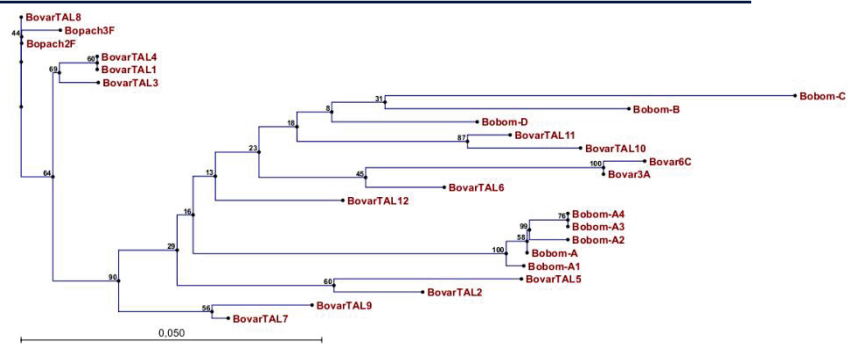
Amplification d'un fragment de 157 bp CMH Class II DRB exon 2 à l'aide des primers BobomMHCIIEXON2F2 et BobomMHCIIEXON2R1. Séquençage haut-débit (454 GS-Junior, Roche Diagnostic). 87 échantillons.

*Etape du processus de minimisation des erreurs de séquençage*

Processing description	Read number (amplicon number)
Number of read (all mixed loci)	64.821 (87)
Elimination of reads without both forward and reverse individual tags and loci markers	60.101 (86)
Elimination of singletons and reads with <95 % and > 105 % of known fragment size	49.133 (86)
Elimination of amplicons < 12 reads for a given locus	49.131 (85)
Elimination of <3 reads/variant	44.089 (85)



Distribution des fréquences de chaque variant *j* dans chaque amplicon *i* après élimination des amplicons avec moins de 12 reads.



Arbre phylogénétique incluant les allèles identifiés (BovarTAL) et les séquences de références chez *Bombina variegata* et deux espèces-sœur, *B. bombina* et *B. pachypus*.

## RESULTATS ACTUELS

- 12 nouveaux allèles identifiés appartenant à au moins 3 loci dont un monomorphe
- Chaque individu possède 2 à 4 allèles, avec une moyenne de 3 allèles/individu dans la population

## FINALISATION DES ANALYSES

Assignement des 12 nouveaux allèles à chaque loci et test des hypothèses relatives au choix de partenaire à venir...





## **APPENDIX E:**

### **PRE- AND POST-COPULATORY TRAITS**

---

**Do pre- and post-copulatory traits covary in large herbivores?**

M Ferrandiz-Rovira, J-F. Lemaître, S. Lardy, B. Claramunt-López and A. Cohas *BMC Evolutionary Biology* (2014), 14: 79.



RESEARCH ARTICLE

Open Access

# Do pre- and post-copulatory sexually selected traits covary in large herbivores?

Mariona Ferrandiz-Rovira<sup>1</sup>, Jean-François Lemaître<sup>1</sup>, Sophie Lardy<sup>1,2</sup>, Bernat C López<sup>3,4</sup> and Aurélie Cohas<sup>1\*</sup>

## Abstract

**Background:** In most species, males compete to gain both matings (via pre-copulatory competition) and fertilizations (via post-copulatory competition) to maximize their reproductive success. However, the quantity of resources devoted to sexual traits is finite, and so males are predicted to balance their investment between pre- and post-copulatory expenditure depending on the expected pay-offs that should vary according to mating tactics. In *Artiodactyla* species, males can invest in weapons such as horns or antlers to increase their mating gains or in testes mass/sperm dimensions to increase their fertilization efficiency. Moreover, it has been suggested that in these species, males with territory defence mating tactic might preferentially increase their investment in post-copulatory traits to increase their fertilization efficiency whereas males with female defence mating tactic might increase their investment in pre-copulatory sexually selected traits to prevent other males from copulating with females. In this study, we thus test the prediction that male's weapon length (pre-copulatory trait) covaries negatively with relative testes size and/or sperm dimensions (post-copulatory traits) across *Artiodactyla* using a phylogenetically controlled framework.

**Results:** Surprisingly no association between weapon length and testes mass is found but a negative association between weapon length and sperm length is evidenced. In addition, neither pre- nor post-copulatory traits were found to be affected by male mating tactics.

**Conclusions:** We propose several hypotheses that could explain why male ungulates may not balance their reproductive investment between pre- and post-copulatory traits.

**Keywords:** Pre-copulatory competition, Post-copulatory competition, Secondary sexual traits, Sexual selection, Sperm competition, Weapon length

## Background

Understanding the diversity and evolution of male sexual traits is a major interest in evolutionary ecology [1,2], and there is now substantial evidence that sexual competition has shaped these traits [1,3]. Firstly, males compete for mating opportunities (i.e. pre-copulatory competition) and individuals who invest in secondary sexual traits such as ornaments or weapons are often better competitors and gain more mating events (e.g. [4,5]). Secondly, in species where females mate with more than one male in the same reproductive bout, males compete to fertilize a set of ova through sperm competition (i.e. post-copulatory competition) [3]. In many cases, sperm competition can be compared to a raffle where the male probability to fertilize

eggs is proportional to the relative number of sperm delivered by the male [6,7]. Sperm traits such as sperm dimensions and characteristics can also increase the probability of fertilization through their impact on sperm mobility and viability [8]. Such sperm characteristics can be considered as sperm quality characteristics and can also play an important role in the outcome of sperm competition [7,8]. Consequently, sperm competition has selected for adaptations that increase both sperm quantity and/or quality. For example, in mammalian taxa, testes mass (relative to body mass), a proxy of sperm quantity [9], is strongly associated with different proxies of sperm competition such as social group size in bats [10] or the percentage of within-litter multiple paternity in rodents [11]. Moreover, although sperm quantity has been traditionally emphasized, recent studies testing simultaneously the contribution of both sperm quantity and quality on fertility have shown that, under some circumstances, sperm quality

\* Correspondence: aurelie.cohas@univ-lyon1.fr

<sup>1</sup>Laboratoire Biométrie et Biologie Evolutive, Université de Lyon, CNRS, UMR5558, Université Lyon 1, F-69622, Villeurbanne, F-69000 Lyon, France  
Full list of author information is available at the end of the article

could be more important than quantity. For example, one study found that in the white-footed mouse (*Peromyscus leucopus noveboracensis*), the probability of siring litters was more affected by sperm quality (measured as sperm viability and percentage of sperm presenting abnormalities) than sperm quantity [12].

The expression and maintenance of sexual traits involved in both pre- and post-copulatory competition are energy and time consuming (e.g. [13,14]) and are generally associated with various costs. For example, conspicuous secondary sexual traits often cause high predation risks [15] and male investment in both pre- and post-copulatory traits decreases immune efficiency [16,17]. Since males can allocate only a limited amount of resources to sexual competition [6], theoretical models of sperm competition predict that, when the relative intensity of pre- or post-copulatory pressures change, males should modulate their reproductive investment between traits involved in competition for mating and traits involved in the production of high quality ejaculates depending on the expected pay-offs [6,18]. Basically, when females are likely to mate with different males (i.e. high level of sperm competition), males should preferentially increase their investment in ejaculate expenditure by decreasing the investment to pre-copulatory male sexually selected traits; whereas when female propensity to mate multiply is weak (i.e. low level of sperm competition), males should predominantly increase their investment to pre-copulatory male sexually selected traits by decreasing the investment in ejaculate expenditure [6,18]. So far, evidence for such relationships at the inter-specific level comes mainly from two studies: (1) Simmons and Emlen [19], who found in beetles from the genus *Onthophagus*, that species with the steepest allometric slopes of horn size on body size also display the shallowest allometric slopes of testes mass on body size (and vice versa), and (2) Fitzpatrick et al. [20], who recently revealed a negative relationship between sexual size dimorphism and both baculum length and testes mass across pinniped species. Although male development and maintenance of pre-copulatory male sexually selected traits other than body size or mass are costly (i.e. [21]), and can potentially covary with the expression of ejaculate characteristics (i.e. [22]), inter-specific studies involving male armaments or ornamental traits remain scarce. Moreover, in this context, investment in sperm quality has never been considered, maybe because models of allocation to sperm competition according to the level of pre-copulatory competition are principally focused on sperm quantity [6,18].

Ungulates are well-suited to examine covariation between investment in sexual traits since males in this group face intense competition both to secure matings and fertilize ova and developed traits to increase their success under both pre- and post-copulatory competition

[23,24]. The predominant mating system of ungulate species is polygynous [25] and conspicuous weapons such as horns for *Bovidae* and antlers for *Cervidae* have been sexually selected in this group because they provide an advantage for males in gaining matings and ultimately increase fitness [23]. For example, in Soay sheep (*Ovis aries*) larger horns enhance the probability of being observed in consort, which in turn is related to mating success [4]. In addition, ungulate species where males face a high level of sperm competition have developed larger testes [26,27], and sperm velocity is also likely to be under strong selective pressure since this parameter is directly related to variation in male fertility (e.g. red deer in Malo et al. [28]). Moreover, it has been hypothesized that the level of intramale competition and ultimately the relative investment in both pre- and post-copulatory traits could potentially differ between mating tactics exhibited by males in polygynous ungulates [23]. In the female defence mating tactic, males follow females and attempt to guard them during oestrus. Therefore, male mating opportunities might depend predominantly on their dominance hierarchy and abilities to prevent other males from copulating with females mainly from the use of their weapons [24]. On the contrary, in the territory (or resource) defence tactic, males defend areas to attract females but do not monopolize females that can freely range over several male territories. As a consequence, the risk of sperm competition might be higher in species with territory defence tactic and males should increase their investment in ejaculate expenditure as predicted by a recent model of sperm competition [18]. Thus, the variation found in ungulate sexual traits (weapon length in Bro-Jørgensen [23] and Plard et al. [29] and both testes mass and sperm length in Gomendio et al. [30]) could ultimately be the result of different levels of sperm competition between mating tactics exhibited by males.

In this study, we first use a comparative analysis on ungulates to test for covariation between male investment in weapon (horns or antlers) length and male investment in sperm quantity and/or quality (relative testes mass and/or sperm dimensions). We expect that the relative size of these pre- and post-copulatory traits should be negatively correlated [18]. Then we tested the hypothesis that male mating tactic (female or territory defence) can mediate the relative investment in these pre- and post-copulatory traits. Following Bro-Jørgensen [23], we expect species with a female defence mating tactic to invest in pre-copulatory male sexually selected traits at the expense of their investment in post-copulatory traits, while we expect the reverse pattern for species with a territory defence mating tactic.

## Results

Regarding post-copulatory traits, no significant relationships between relative testes mass and sperm dimensions

were observed (Additional file 1: Table S1a). However, positive relationships between sperm dimensions were evidenced (see Additional file 1: Table S1b and S1c).

No significant relationship between relative weapon length and relative testes mass was found (Table 1). However, there was evidence for a strong significant negative relationship ( $r = -0.27$ ) between weapon and total sperm length (Figure 1a, Table 1). Although lengths of all sperm components (head, midpiece and tail) are significantly longer as total sperm length increases (see Additional file 1: Table S1b and S1c), only tail length was found to be strongly negatively associated with weapon length ( $r = -0.30$ ) although this relation was found to be significant only with the phylogenetic tree derived from Bininda-Emonds et al. [31] ( $p = 0.03$  with Bininda-Emonds et al. [31];  $r = -0.26$ ,  $p = 0.07$  with Agnarsson and May-Collado [32] tree) (Figure 1b, Table 1, Additional file 2: Table S2a). No significant associations were found between weapon length and head or midpiece length, nor between weapon length and midpiece volume (Table 1). The taxonomic family had no effect on the evidenced relationships (see Additional file 3: Table S3).

No relationships between mating tactics and any of the pre- and post-copulatory traits were evidenced (Table 2). Furthermore, the mating tactics did not modulate the relationships between weapons length and the different post-copulatory traits (Table 3).

## Discussion

Surprisingly, no evidence of covariation between relative weapon length and relative testes mass was found across ungulate species. However our results show evidence of a negative association between relative weapon length

and sperm length. Moreover, we found that allocation to pre- and post-reproductive traits is independent of the mating tactic.

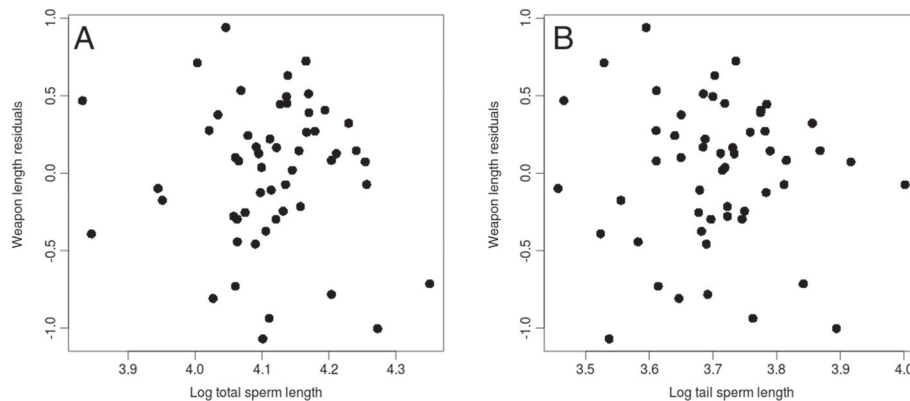
The absence of a negative covariation between relative weapon length and relative testes mass contrasts with both theoretical predictions [18] and recent empirical findings from Fitzpatrick et al. [20], who found a negative relationship between sexual size dimorphism and relative testes mass across thirteen species of pinnipeds. One possible explanation for the apparent discrepancy between these two studies could come from a much lower intensity of post-copulatory competition in ungulates compared to pinnipeds as suggested by their small relative testes mass [30]. However, scrutiny on Fitzpatrick et al. [20] data reveals that the gonadosomatic index (testes mass/body mass  $\times 100$ ) of pinnipeds is very close to the gonadosomatic index of ungulates from our dataset (mean  $\pm$  SD gonadosomatic index [range]:  $0.06 \pm 0.03$  [0.02-0.13] in pinnipeds ( $N = 14$ ) and  $0.10 \pm 0.19$  [0.02-1.23] in ungulates ( $N = 45$ ); Mann–Whitney–Wilcoxon test:  $W = 330$ ;  $p = 0.80$ ). Therefore, the absence of a negative association between weapon length and relative testes mass in ungulates is unlikely to come from a lower post-copulatory sexual selection in this group, which is not surprising since evidence of female multiple mating in ungulates are now compelling (e.g. [4,33]).

Our results further show evidence of a negative association between relative weapon length and sperm length across species of ungulates. When sperm components were analysed separately, this negative association was found only with the length of the sperm tail. In mammals, sperm tail length is correlated with sperm velocity [34], which can potentially increase male fertilization

**Table 1 Phylogenetically corrected models testing the relationships between male pre- and post-copulatory sexually selected traits**

Dependent variable	Independent variables	Beta $\pm$ SE	t	P	N	df	r	$\lambda$
Weapon length	Body mass	0.48 $\pm$ 0.09	5.58	< 0.001	45	42	0.65	0.69 <sup>0.36/0.21</sup>
	Testes mass	0.04 $\pm$ 0.08	0.50	0.62			0.08	
Weapon length	Body mass	0.56 $\pm$ 0.06	10.01	< 0.001	54	51	0.81	1.00 <sup>&lt; 0.001/1</sup>
	Sperm length	-0.95 $\pm$ 0.47	-2.04	0.05			-0.27	
Weapon length	Body mass	0.58 $\pm$ 0.06	10.43	< 0.001	54	51	0.83	0.98 <sup>0.01/&lt; 0.001</sup>
	Sperm head length	-0.04 $\pm$ 0.29	-0.12	0.90			-0.02	
Weapon length	Body mass	0.56 $\pm$ 0.06	9.10	< 0.001	53	50	0.79	0.98 <sup>&lt; 0.001/0.26</sup>
	Sperm midpiece length	-0.11 $\pm$ 0.31	-0.35	0.73			-0.05	
Weapon length	Body mass	0.51 $\pm$ 0.06	8.37	< 0.001	47	44	0.78	0.98 <sup>&lt; 0.001/0.21</sup>
	Sperm midpiece volume	0.02 $\pm$ 0.07	0.35	0.73			0.05	
Weapon length	Body mass	0.57 $\pm$ 0.05	10.58	< 0.001	54	51	0.83	1.00 <sup>&lt; 0.001/1</sup>
	Sperm tail length	-0.80 $\pm$ 0.36	-2.24	0.03			-0.30	

Models are tested with Bininda-Emonds et al. [31] phylogeny across ungulates species (both *Bovidae* and *Cervidae* pooled).  $r$  represents effect size values.  $\lambda$  represents the index of phylogenetic covariance (see Methods section). The superscripts following the  $\lambda$  value indicate p-value of likelihood ratio tests against models with  $\lambda = 0$  (first position) and  $\lambda = 1$  (second position). All variables were log transformed.



**Figure 1** Weapon length (after correction for body mass allometry) in relation to total sperm length (a) and tail length (b) without considering phylogeny for 54 species (both *Bovidae* and *Cervidae* pooled).

success under competitive conditions [7,8]. Therefore, although an increase in sperm length when the intensity of post-copulatory sexual selection increases is potentially adaptive, such negative covariation between sexual traits such as weapons and microscopic structures like sperm across species is striking and deserves to be studied in other groups before drawing any definitive conclusion. Indeed, this is the first time that a relationship between pre-copulatory male sexually selected traits and sperm characteristics is documented at the inter-specific level although such covariations have been repeatedly investigated at the species level (see [35] for a recent compilation of these studies). For example, in coho salmon (*Onchorhynchus kisutch*) males with more intense spawning colouration, in other words that allocate heavily in secondary sexual characters, had lower sperm velocities than males with less pronounced spawning coloration [36]. Conversely, intra-specific studies have also documented positive relationships between secondary sexual

traits and measures of ejaculate quality (e.g. [37]). However, a meta-analysis has recently failed to find a general pattern of covariation between pre- and post-copulatory traits [35], suggesting that species ecological and biological characteristics might play a significant role on the direction and the strength of the covariation.

It is particularly surprising to observe a negative relationship between weapon length and sperm length while no association was found between weapon length and testes mass since testes mass is typically considered by far as the most robust indicator of investment in post-copulatory competition [38]. However, as emphasized by Simmons and Fitzpatrick [20] in their recent review on the evolution of male fertility, caution is required before considering relative testes mass as an absolute proxy of sperm competition level. Indeed, testes can perform functions other than sperm production and allocation in testes mass could be favoured even in the absence of variation in the level of sperm competition [39]. In ungulates, testes

**Table 2** Phylogenetically corrected models testing for differences in male pre- and post-copulatory sexually selected traits with different mating tactics

Dependent variables	Independent variables	Beta ± SE	t	P	N	df	r	λ
Weapon length*	Body mass*	0.53 ± 0.07	7.67	< 0.001	58	55	0.72	0.98 <sup>0.001/0.30</sup>
	Tactic	-0.22 ± 0.17	-1.29	0.20			-0.17	
Testes mass*	Body mass*	0.67 ± 0.12	5.51	< 0.001	45	42	0.65	0.00 <sup>1/&lt; 0.001</sup>
	Tactic	0.16 ± 0.27	0.60	0.55			0.09	
Sperm length	Tactic	-1.02 ± 1.94	-0.53	0.60	54	52	-0.07	0.67 <sup>0.06/&lt; 0.001</sup>
Sperm head length	Tactic	-0.16 ± 0.41	-0.38	0.70	54	52	-0.05	0.58 <sup>0.01/&lt; 0.001</sup>
Sperm midpiece length	Tactic	0.04 ± 0.78	0.05	0.96	53	51	0.01	0.92 <sup>&lt; 0.001/0.01</sup>
Sperm midpiece volume	Tactic	-0.13 ± 0.39	-0.34	0.74	47	45	-0.05	0.00 <sup>1/0</sup>
Sperm tail length	Tactic	-1.23 ± 1.50	-0.82	0.42	54	52	-0.11	0.52 <sup>0.24/&lt; 0.001</sup>

Female defence species is coded as 0 and territory defence is coded as 1. Models are tested with Bininda-Emonds et al. [31] phylogeny across ungulates species (both *Bovidae* and *Cervidae* pooled). r represents effect size values and λ represents the index of phylogenetic covariance (see Methods section). The superscripts following the λ value indicate p-value of likelihood ratio tests against models with λ = 0 (first position) and λ = 1 (second position). \*log transformed variables.

**Table 3 Phylogenetically corrected models testing the relationship between weapon length and testes mass, sperm dimensions and mating tactics**

Dependent variable	Independent variables	beta ± SE	t	P	N	df	r	λ
Weapon length	Body mass	0.44 ± 0.09	4.70	< 0.001	45	41	0.59	0.27 <sup>0.64/0.11</sup>
	Testes mass	-0.04 ± 0.12	-0.31	0.76			-0.05	
	Tactic	-0.74 ± 0.60	-1.22	0.23			-0.19	
	Testes mass × Tactic	0.11 ± 0.14	0.74	0.46			0.12	
Weapon length	Body mass	0.56 ± 0.06	8.40	< 0.001	54	50	0.77	1.00 <sup>&lt; 0.001/0.81</sup>
	Sperm length	-1.92 ± 1.05	-1.83	0.07			0.25	
	Tactic	-5.12 ± 4.74	-1.08	0.28			0.15	
	Sperm length × Tactic	1.20 ± 1.16	1.04	0.30			0.15	
Weapon length	Body mass	0.53 ± 0.06	8.53	< 0.001	54	50	0.77	0.99 <sup>&lt; 0.001/0.38</sup>
	Sperm head length	0.90 ± 0.76	1.18	0.24			0.16	
	Tactic	2.14 ± 1.69	1.26	0.21			0.18	
	Sperm head length × Tactic	-1.11 ± 0.81	-1.37	0.18			-0.19	
Weapon length	Body mass	0.52 ± 0.07	7.50	< 0.001	53	49	0.73	0.98 <sup>&lt; 0.001/0.36</sup>
	Sperm midpiece length	-0.26 ± 0.46	-0.56	0.58			0.08	
	Tactic	-0.59 ± 1.46	-0.41	0.69			-0.06	
	Sperm midpiece length × Tactic	0.16 ± 0.58	0.28	0.78			0.04	
Weapon length	Body mass	0.49 ± 0.07	7.26	< 0.001	47	43	0.74	0.98 <sup>&lt; 0.001/0.20</sup>
	Sperm midpiece volume	0.08 ± 0.17	0.45	0.65			0.07	
	Tactic	-0.07 ± 0.23	-0.28	0.78			-0.04	
	Sperm midpiece volume × Tactic	-0.06 ± 0.19	-0.34	0.74			-0.05	
Weapon length	Body mass	0.55 ± 0.06	8.91	< 0.001	54	50	0.78	1.00 <sup>&lt; 0.001/1</sup>
	Sperm tail length	-1.44 ± 0.75	-1.92	0.06			-0.26	
	Tactic	-3.06 ± 3.14	-0.98	0.33			-0.14	
	Sperm tail length × tactic	0.78 ± 0.85	0.91	0.37			0.13	

Female defence species is coded as 0 and territory defence is coded as 1. Models are tested with Bininda-Emonds et al. [31] phylogeny across ungulates species (both *Bovidae* and *Cervidae* pooled). *r* represents effect size values and  $\lambda$  represents the index of phylogenetic covariance (see Methods section). The superscripts following the  $\lambda$  value indicate p-value of likelihood ratio tests against models with  $\lambda = 0$  (first position) and  $\lambda = 1$  (second position). All variables were log transformed.

mass varies between monogamous and polygynous species [26]. However, following the male mating rate hypotheses [39], polygynous species could have heavier testes than monogamous species as a result of a higher number of females to fertilize even in the absence of variation in the level of sperm competition. On the contrary, whether sperm characteristics respond to post-copulatory competition is still a debated topic, especially when, like in our study, no relationship between sperm dimensions and relative testes mass was found. Even if studies reporting either negative relationships or no influence of sperm competition on sperm/tail length have prevented the emergence of a clear picture [40,41], it appears that the majority of comparative studies performed over the past 20 years have reported positive relationships between the level of sperm competition and total sperm length (see Simmons and Fitzpatrick [20] for a review). We thus cannot rule out that other post-

copulatory traits such as sperm dimensions could respond to post-copulatory sexual selection [8].

Finally, preferential investment in pre- or post-copulatory traits could be obscured by variation in the pre- and post-copulatory selective pressures resulting from their mating tactics. Indeed in the moth (*Plodia interpunctella*) mating tactic is associated with differences in ejaculate expenditure. Moth males with female defence mating tactic have larger heads and thoraxes, have smaller testes and produce fewer sperm than males without female defence mating tactic [42]. However, our results do not show distinct pattern of allocation to pre- and post-reproductive traits according to the mating tactics. Indeed, none of the pre- and post-copulatory traits investigated differ between mating tactics nor is the covariation between sperm dimensions (total sperm and tail length) and weapon length affected by mating tactics. Discrepancies between our predictions and our results might have at least two



explanations. First, the classification of mating tactics used (female defence versus territory defence) might not reflect adequately the level of pre- and post-copulatory competition in ungulate species. For instance, in the great kudu (*Tragelaphus strepsiceros*), a non-territorial species, the level of pre-copulatory competition might be, contrary to our hypothesis, not so elevated since in this species males fight only occasionally [43]. Conversely, the level of post-copulatory competition can also be high in species commonly classified as female defence mating tactic such as the white-tailed deer (*Odocoileus virginianus*). Indeed, in this species, DNA microsatellite markers reveal that multiple paternities are widespread within litters [33]. Second, in ungulates, males can show strong plasticity in their reproductive behaviour shifting from one mating tactic to another (see [44] for a review on this topic). In this study, we have used the main mating tactic generally associated with the species, but we cannot exclude that the degree of the occurrence of alternative mating tactics might exert selective pressures. For example, within a given species, individuals adopting a territory defence tactic are predicted to have large weapons and low testes mass while individuals adopting a female defence mating strategy are predicted to have small weapons and high testes mass. These individual variations in mating tactic might obscure the relationships between pre- and post-copulatory traits at the species level. Unfortunately high quality data on reproductive tactics in ungulates are currently not sufficient to obtain a classification of mating tactics that would better reflect the selective pressures exerted on pre- and post-copulatory traits, limiting the conclusions that can be drawn from the absence of distinct pattern of allocation to pre- and post-reproductive traits reported.

## Conclusions

Our results reveal the absence of a negative association between pre-copulatory male sexually selected traits and relative testes mass. Nevertheless, a negative association between pre-copulatory male sexually selected traits and sperm tail length was found. Covariations at the inter-specific level between sexual traits or others are sometimes interpreted as an evolutionary trade-off (e.g. [20]) although such interpretations require extreme caution [18]. Indeed, these conclusions are often based on the assumption that the quantity of resources devoted to the production and maintenance of sexual traits should be the same across species [6,45]. In ungulates, the variance in mass-specific metabolic rate appears to be small compared to other mammalian taxa [30]. Therefore, the amount of resources that males can allocate to sexual traits is likely to be roughly similar between these species although fine scale measurements would be needed before drawing any definitive conclusion on the absence or presence of an evolutionary trade-off. Further studies

on the relative costs and benefits associated with investment in different sexual traits in ungulates (see also [20]), through for instance the use of experimental manipulations (e.g. [19]) and mating trials, are now required before drawing any conclusion on the presence of an evolutionary trade-off.

## Methods

### Dataset

Data were collected on adult males of *Bovidae* and *Cervidae* species, the two main families of weaponed *Artiodactyla*. We first conducted two separate literature surveys: one on weapon length (horn length for *Bovidae* and antler length for *Cervidae*). Once the maximum number of data on weapon length was collected, we focused our literature survey on testes mass and sperm length, as well as sperm head length, sperm midpiece length and volume, and sperm tail length, since sperm is a complex cell divided in three main structures. Additionally, following the classification proposed by Bro-Jørgensen [23] and later used by Plard et al. [29], each species was characterized as having either a female defence mating tactic when they guard a female during their receptive period (level 0 in statistical analysis) or a territory defence mating tactic (level 1 in statistical analysis) depending on the males main mating tactic. For territory defence category, both studies [23,29] merge territorial species (where individuals defend large territories with food resources) and lekking species (where individuals defend small territories with no food resources but females) since lekking could be an alternative mating tactic of territorial species. Moreover, no intra-specific differences in weapon size have been found between territorial and lekking populations and thus combining both groups is justified [23,29]. The dataset was then supplemented using information from more specific sources. All data used were checked from the original source and are provided in the Additional file 4: Table S4.

The final dataset contains information on weapon length, body mass, paired testes mass, sperm dimensions and mating tactics for 58 *Artiodactyla* species including 40 species of *Bovidae* and 18 species of *Cervidae*. Since testes mass and all sperm dimensions were not available for all species, sample sizes vary from 31 to 58 species in the statistical analysis.

### Comparative method

Species may share characteristics as a result of a common ancestry. This could create dependency among the data, which potentially compromises statistical tests [46]. On *Artiodactyla*, no consensus phylogeny has been reached yet and we thus used two different phylogenies with topology and branch length (Additional file 5: Figure S1).

The first phylogenetic tree was derived from the phylogenetic supertree of mammals of Bininda-Emonds et al. [31] and is commonly used in inter-specific studies focused on *Artiodactyla*. However, this tree displays an important number of polytomies and we thus repeated our analyses with another phylogenetic tree derived from Agnarsson and May-Collado [32]. This second phylogenetic tree presents no polytomies. Unfortunately, five species from our dataset were absent from this tree: sunda sambar (*Cervus timorensis*), blue wildebeest (*Connochaetes taurinus*), lechwe (*Kobus leche*), red brocket (*Mazama americana*), and common duiker (*Sylvicapra grimmia*). Consequently, both trees were used to compare consistency between models using a tree containing the whole set of species but containing polytomies [31] and using a tree containing no polytomies but without five species from our dataset [32].

We then used phylogenetic generalized linear models (PGLS). This statistical method estimates an index called  $\lambda$  that indicates whether the phylogeny correctly predicts the patterns of covariance among species on a given trait.  $\lambda$  varies between 0 (complete absence of phylogenetic structure, i.e. the phylogenetic structure can be represented by a star phylogeny) and 1 (phylogenetic structure in agreement with a Brownian model of the evolution of the considered traits, i.e. the phylogenetic structure can be represented by the previously constructed tree with unmodified topology and branch length) [47].  $\lambda$  is then introduced in the model to control for the phylogenetic effect by multiplying all the off-diagonal values of the variance-covariance matrix extracted, with the R-package 'ape' [48], from the constructed phylogenetic tree. Then, the generalized linear model is fitted with this modified variance-covariance matrix [47].

### Statistical analysis

All statistical analyses were conducted with the two phylogenetic trees described above. However, unless otherwise stated, the results were qualitatively identical with these two phylogenetic trees. Therefore, only results from Bininda-Emonds et al. [31] are presented in the manuscript, although those from Agnarsson and May-Collado [32] can be found in Additional file 2: Table S2.

Since weapons differ in structure and growth rate between *Bovidae* and *Cervidae*, we first tested for an effect of the taxonomic family on each studied trait. We thus built PGLS models with each male sexual trait (pre- and post-copulatory) considered as the dependent variable and family (coded as 0 for *Bovidae* and 1 for *Cervidae*) as the independent variable. Since no difference in any of the traits studied could be found between *Bovidae* and *Cervidae* families, both families were considered together in all subsequent analyses (Additional file 3: Table S3).

Due to the role of sperm quantity and quality in fertilization success [7], both traits may have co-evolved leading to dependencies between the associated traits. In order to test the relationship between these traits, we first analysed, using the PGLS regressions, the relationship between relative testes mass and sperm dimensions as well as between sperm dimensions.

Once these analyses were conducted, we proceeded to the test of our two hypotheses. First, to test for a negative association between traits involved in pre- and post-copulatory sexual competition, we constructed a series of PGLS models including weapon length as a dependent variable and either testes mass, total sperm length or any sperm component dimensions as independent variables. To further control for a potential confounding effect of the taxonomic family on the association between pre- and post-copulatory traits, we included an additive or an interactive effect of species family with each post-copulatory traits in the model described above. The absence of an additive or an interactive effects of the family on the relationship between traits (see Additional file 3: Table S3), further confirm that both families can be considered together in the presented analyses. Furthermore, analyses conducted on the species belonging to the *Bovidae* species only (the number of species of *Cervidae* being too small ( $N = 18$ ) for separate analyses) gave qualitatively similar results and are thus not presented.

Second, to test for an effect of the mating tactic on investment in pre- and post-copulatory traits, we also constructed PGLS models with each trait (weapon length, testes mass, total sperm length or each sperm components length) considered as the dependent variables and mating tactics (coded as 0 for female defence mating tactic and coded as 1 for territory defence mating tactic) as the independent variable. We further investigated whether the association between traits involved in pre- and post-copulatory sexual competition could be modulated by the mating tactics. For this purpose, we tested for an interactive effect of the mating tactics with either relative testes mass, total sperm length or any sperm component dimensions on weapon length. Here, again, analyses conducted on the species belonging to the *Bovidae* species only gave qualitatively similar results and are not presented.

Whenever relationships between two traits or more were investigated, all variables were log transformed to linearize the relationship. Log transformed male body mass was included as an independent variable to control for allometric relationship in any of the above models involving weapon length or testes mass, but not in models involving sperm dimensions since preliminary analyses failed to evidence any allometric relationship between sperm dimensions and body mass. Finally, for each model, normality of the residuals was checked using

standard diagnostic plots (density plots of the distribution of the residuals and normal Q-Q plots) and the fit to the data was assessed graphically by plotting the fitted values against both the residuals and the observed data.

All statistical analyses were conducted using the packages 'ape', 'mvtnorm', 'adephylo', 'caper' and 'phylobase' of R version 2.15.0 R [49]. Effect sizes ( $r$ ) were calculated from the  $t$ -values and degrees of freedom from PGLS models [50]. Unless otherwise stated, all tests were two-tailed, the level of significance was set to 0.05, and parameter estimates are given  $\pm$  SE.

## Additional files

**Additional file 1: Table S1.** Phylogenetically corrected models testing (a) the relations between testes mass as dependent variable and the dimensions of the different sperm components (total sperm length, head length, midpiece length, tail length and midpiece volume) across ungulate species considering phylogenetic tree based on Bininda-Emonds et al. [31] and Agnarsson and May-Collado [32]; and (b) between the dimensions of the different sperm components (total sperm length, head length, midpiece length, tail length and midpiece volume) across ungulate species obtained when considering the phylogenetic tree based on Bininda-Emonds et al. [31]; and (c) Agnarsson and May-Collado [32]. The superscripts following the  $\lambda$  value indicate p-value of likelihood ratio tests against models with  $\lambda = 0$  (first position) and  $\lambda = 1$  (second position).

**Additional file 2: Table S2.** Phylogenetically corrected models with Agnarsson and May-Collado [32] phylogeny across ungulates species showing (a) the relationships between pre- (weapon length) and post-copulatory traits (testes mass, total sperm length, head length, midpiece length, tail length and midpiece volume); (b) differences in the size of pre- and post-copulatory traits between species with different mating tactics; and (c) relationships between weapon length and both post-copulatory traits (testes mass, total sperm length, head length, midpiece length, tail length and midpiece volume) and mating tactics. For mating tactics, 'female defence' is coded as 0 and 'territory defence' is coded as 1.  $\lambda$  represents the index of phylogenetic covariance (see Methods section). The superscripts following the  $\lambda$  value indicate p-value of likelihood ratio tests against models with  $\lambda = 0$  (first position) and  $\lambda = 1$  (second position). \*variables without log transformation.

**Additional file 3: Table S3.** Phylogenetically corrected models showing no effect of the taxonomic family (*Cervidae* or *Bovidae*) to which a species belongs to on the pre-copulatory traits (weapon length) and the post-copulatory traits (testes mass, total sperm length, head length, midpiece length, tail length and midpiece volume) (a) or on the relations between pre-copulatory and post-copulatory traits (b). Only the additive models were presented since no significant interactions were found. The superscripts following the  $\lambda$  value indicate p-value of likelihood ratio tests against models with  $\lambda = 0$  (first position) and  $\lambda = 1$  (second position). \*means that variables were log transformed.

**Additional file 4: Table S4.** Dataset used in the analyses.

**Additional file 5: Figure S1.** Phylogenetic reconstruction for the 58 ungulate species used in the phylogenetically corrected models. These reconstructions were based on Bininda-Emonds et al. [31] (a); and Agnarsson and May-Collado [32] (b). The five species in bold are missing from the phylogenetic tree derived from Agnarsson and May-Collado [32], which contains 53 species.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

Conceived and designed the experiments: AC, JFL and MFR. Analysed the data: MFR and AC. Wrote the paper: MFR, JFL, AC and BC. Collected the data: MFR, SL and AC. All authors read and approved the final manuscript.

## Acknowledgments

We thank Prof. Dominique Allainé, Dr. Sébastien Devillard, Dr. Camille Ferdenzi and Dr. Steven Ramm for helpful discussions and comments on the manuscript. We are grateful to Dr. Igñi Agnarsson for kindly providing the phylogeny. MFR is supported by a scholarship for postgraduate studies (Obra Social Fundació "La Caixa" and VetAgro Sup). JFL is supported by a post-doctoral grant from the Fyssen Foundation, Paris.

## Author details

<sup>1</sup>Laboratoire Biométrie et Biologie Evolutive, Université de Lyon, CNRS, UMR5558, Université Lyon 1, F-69622, Villeurbanne, F-69000 Lyon, France. <sup>2</sup>Groupe Ecologie et conservation des vertébrés, Université d'Angers, Faculté des Sciences, 49045 Angers, France. <sup>3</sup>CREAF, Cerdanyola del Vallès, 08193 Catalunya, Spain. <sup>4</sup>Universitat Autònoma de Barcelona, Cerdanyola del Vallès, 08193 Catalunya, Spain.

Received: 8 January 2014 Accepted: 13 March 2014

Published: 10 April 2014

## References

- Clutton-Brock TH: **Sexual selection in males and females.** *Science* 2007, **318**:1882–1885.
- Leonard JL, Córdoba-Aguilar A: *The Evolution of Primary Sexual Characters in Animals.* Oxford: Oxford Univ. Press; 2010.
- Birkhead TR, Møller A: *Sperm Competition and Sexual Selection.* London: London Academic Press; 1998.
- Preston BT, Stevenson IR, Pemberton JM, Coltman DW, Wilson K: **Overt and covert competition in a promiscuous mammal: the importance of weaponry and testes size to male reproductive success.** *Proc R Soc Lond B Biol Sci* 2003, **270**:633–640.
- Loyau A, Gomez D, Moureau B, Thery M, Hart NS, Saint Jalme M, Bennett ATD, Sorci G: **Iridescent structurally based coloration of eyespots correlates with mating success in the peacock.** *Behav Ecol* 2007, **18**:1123–1131.
- Parker GA: **Sperm competition and the evolution of ejaculates: towards a theory base.** In *Sperm Competition and Sexual Selection.* Edited by Birkhead TR, Møller AP. London: London Academic Press; 1998.
- Pizzari T, Parker GA: **Sperm competition and sperm phenotype.** In *Sperm Biology: An Evolutionary Perspective.* Edited by Birkhead TR, Hosken DJ, Pitnick S. London: London Academic Press; 2009.
- Snook RR: **Sperm in competition: not playing by the numbers.** *Trends Ecol Evol* 2005, **20**:46–53.
- Møller AP: **Ejaculate quality, testes size and sperm production in mammals.** *Funct Ecol* 1989, **3**:91–96.
- Hosken DJ: **Sperm competition in bats.** *Proc R Soc Lond B Biol Sci* 1997, **264**:385–392.
- Ramm SA, Parker GA, Stockley P: **Sperm competition and the evolution of male reproductive anatomy in rodents.** *Proc R Soc Lond B Biol Sci* 2005, **272**:949–955.
- Malo AF, Martinez-Pastor F, Alaks G, Dubach J, Lacy RC: **Effects of genetic captive- breeding protocols on sperm quality and fertility in the white-footed mouse.** *Biol Reprod* 2010, **83**:540–548.
- Pitnick S: **Investment in testes and the cost of making long sperm in Drosophila.** *Am Nat* 1996, **148**:57–80.
- Ramm SA, Stockley P: **Sperm competition and sperm length influence the rate of mammalian spermatogenesis.** *Biol Lett* 2010, **6**:219–221.
- Zuk M, Kolluru GR: **Exploitation of sexual signals by predators and parasitoids.** *Q Rev Biol* 1998, **73**:415–438.
- Simmons LW, Roberts B: **Bacterial immunity traded for sperm viability in male crickets.** *Science* 2005, **309**:2031–2041.
- Garvin JC, Dunn PO, Whittingham LA, Steeber DA, Hasselquist D: **Do male ornaments signal immunity in the common yellowthroat?** *Behav Ecol* 2008, **19**:54–60.
- Parker GA, Lessells CM, Simmons LW: **Sperm competition games: a general model for precopulatory male-male competition.** *Evolution* 2013, **67**:95–109.

19. Simmons LW, Emlen DJ: **Evolutionary trade-off between weapons and testes.** *Proc Natl Acad Sci U S A* 2006, **103**:16346–16351.
20. Fitzpatrick JL, Almbro M, Gonzalez-Voyer A, Kolm N, Simmons LW: **Male contest competition and the coevolution of weaponry and testes in Pinnipeds.** *Evolution* 2012, **66**:3595–3604.
21. Allen BJ, Levinton JS: **Costs of bearing a sexually selected ornamental weapon in a fiddler crab.** *Funct Ecol* 2007, **21**:154–161.
22. Rowe M, Swaddle JP, Pruett-Jones S, Webster MS: **Plumage coloration, ejaculate quality and reproductive phenotype in the red-backed fairy-wren.** *Anim Behav* 2010, **79**:1239–1246.
23. Bro-Jørgensen J: **The intensity of sexual selection predicts weapon size in male bovids.** *Evolution* 2007, **61**:1316–1326.
24. Bro-Jørgensen J: **Intra- and intersexual conflicts and cooperation in the evolution of mating strategies: lessons learnt from ungulates.** *Evol Biol* 2011, **38**:28–41.
25. Clutton-Brock TH: **Mammalian mating systems.** *Proc R Soc Lond B Biol Sci* 1989, **236**:339–372.
26. Ginsberg JR, Rubenstein DI: **Sperm competition and variation in zebra mating-behavior.** *Behav Ecol Sociobiol* 1990, **26**:427–434.
27. Lemaître J-F, Ramm SA, Barton RA, Stockley P: **Sperm competition and brain size evolution in mammals.** *J Evol Biol* 2009, **22**:2215–2221.
28. Malo AF, Garde JJ, Soler AJ, García AJ, Gomendio M, Roldan ERS: **Male fertility in natural populations of red deer is determined by sperm velocity and the proportion of normal spermatozoa.** *Biol Reprod* 2005, **72**:822–829.
29. Plard F, Bonenfant C, Gaillard J-M: **Revisiting the allometry of antlers among deer species: male-male sexual competition as a driver.** *Oikos* 2011, **120**:601–606.
30. Gomendio M, Tourmente M, Roldan ERS: **Why mammalian lineages respond differently to sexual selection: metabolic rate constrains the evolution of sperm size.** *Proc R Soc Lond B Biol Sci* 2011, **278**:3135–3141.
31. Bininda-Emonds ORP, Cardillo M, Jones KE, MacPhee RDE, Beck RMD, Grenyer R, Price SA, Vos RA, Gittleman JL, Purvis A: **The delayed rise of present-day mammals.** *Nature* 2007, **446**:507–512.
32. Agnarsson I, May-Collado LJ: **The phylogeny of *Cetartiodactyla*: the importance of dense taxon sampling, missing data, and the remarkable promise of cytochrome b to provide reliable species-level phylogenies.** *Mol Phylogenet Evol* 2008, **48**:964–998.
33. DeYoung RW, Demarais S, Gonzales RA, Honeycutt RL, Gee KL: **Multiple paternity in white-tailed deer *Odocoileus virginianus* revealed by DNA microsatellites.** *J Mammal* 2002, **83**:884–892.
34. Tourmente M, Gomendio M, Roldan ERS: **Sperm competition and the evolution of sperm design in mammals.** *BMC Evol Biol* 2011, **11**:12.
35. Mautz BS, Møller AP, Jennions MD: **Do male secondary sexual characters signal ejaculate quality? A meta-analysis.** *Biol Rev* 2013, **88**:669–682.
36. Pitcher TE, Doucet SM, Beausoleil JMJ, Hanley D: **Secondary sexual characters and sperm traits in coho salmon *Oncorhynchus kisutch*.** *J Fish Biol* 2009, **74**:1450–1461.
37. Chargé R, Jalme MS, Lacroix F, Cadet A, Sorci G: **Male health status, signalled courtship display, reveals ejaculate quality and hatching success in a lekking species.** *J Anim Ecol* 2010, **79**:843–850.
38. Simmons LW, Fitzpatrick JL: **Sperm wars and the evolution of male fertility.** *Reproduction* 2012, **144**:519–534.
39. Vahed K, Parker DJ: **The evolution of large testes: sperm competition or male mating rate?** *Ethology* 2012, **118**:107–117.
40. Gomendio M, Roldan ERS: **Implications of diversity in sperm size and function for sperm competition and fertility.** *Int J Dev Biol* 2008, **52**:439–447.
41. Pitnick S, Hosken DJ, Birkhead TR: **Sperm morphological diversity.** In *Sperm Biology: An Evolutionary Perspective*. Edited by Birkhead TR, Hosken DJ, Pitnick S. London: London Academic Press; 2009.
42. Gage MJG: **Continuous variation in reproductive strategy as an adaptive response to population density in the moth *Plodia interpunctella*.** *Proc R Soc Lond B Biol Sci* 1995, **261**:25–30.
43. Owen-Smith N: **Mortality rates of male and female kudus: the costs of sexual size dimorphism.** *J Anim Ecol* 1993, **62**:428–440.
44. Lott DF: *Intraspecific Variation in the Social Systems of Wild Vertebrates*. Cambridge: Cambridge Univ Press; 1991.
45. Shutler D: **Sexual selection: when to expect trade-offs.** *Biol Lett* 2011, **7**:101–104.
46. Harvey PH, Pagel MD: *The Comparative Method in Evolutionary Biology*. Oxford: Oxford Univ Press; 1991.
47. Freckleton RP, Harvey PH, Pagel M: **Phylogenetic analysis and comparative data: a test and review of evidence.** *Am Nat* 2002, **160**:712–726.
48. Paradis E, Claude J, Strimmer K: **APE: Analyses of phylogenetics and evolution in R language.** *Bioinformatics* 2004, **20**:289–290.
49. R Development Core Team: *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2012. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
50. Nakagawa S, Cuthill IC: **Effect size, confidence interval and statistical significance: a practical guide for biologists.** *Biol Rev* 2007, **82**:591–605.

doi:10.1186/1471-2148-14-79

**Cite this article as:** Ferrandiz-Rovira et al.: Do pre- and post-copulatory sexually selected traits covary in large herbivores? *BMC Evolutionary Biology* 2014 **14**:79.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)







## REFERENCES

- Acevedo-Whitehouse K, Cunningham AA (2006) Is MHC enough for understanding wildlife immunogenetics? *Trends in Ecology & Evolution*, 21, 433-438.
- Ackerman AL, Cresswell P (2004) Cellular mechanisms governing cross-presentation of exogenous antigens. *Nature Immunology*, 5, 678-684.
- Agresti A (2007) Analysis of ordinal categorical data. John Wiley & Sons, New Jersey.
- Agudo R, Carrete M, Alcaide M, Rico C, Hiraldo F, Donazar JA (2012) Genetic diversity at neutral and adaptive loci determines individual fitness in a long-lived territorial bird. *Proceedings of the Royal Society B-Biological Sciences*, 279, 3241-3249.
- Aguilar A, Roemer G, Debenham S, Binns M, Garcelon D, Wayne RK (2004) High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 3490-3494.
- Akira S, Takeda K, Kaisho T (2001) Toll-like receptors: critical proteins linking innate and acquired immunity. *Nature Immunology*, 2, 675-680.
- Alcaide M, Edwards SV (2011) Molecular evolution of the Toll-like receptor multigene family in birds. *Molecular Biology and Evolution*, 28, 1703-1715.
- Allainé D, Rodrigue I, Le Berre M, Ramousse R (1994) Habitat preferences of Alpine marmots, *Marmota marmota*. *Canadian Journal of Zoology - Revue Canadienne de Zoologie*, 72, 2193-2198.
- Allainé D (2000) Sociality, mating system and reproductive skew in marmots: evidence and hypotheses. *Behavioural Processes*, 51, 21-34.
- Allainé D, Brondex F, Graziani L, Coulon J, Till-Bottraud I (2000) Male-biased sex ratio in litters of Alpine marmots supports the helper repayment hypothesis. *Behavioral Ecology*, 11, 507-514.
- Allainé D (2004) Sex ratio variation in the cooperatively breeding Alpine marmot *Marmota marmota*. *Behavioral Ecology*, 15, 997-1002.
- Allainé D, Theuriau F (2004) Is there an optimal number of helpers in Alpine marmot family groups? *Behavioral Ecology*, 15, 916-924.
- Altizer S, Harvell D, Friedle E (2003) Rapid evolutionary dynamics and disease threats to biodiversity. *Trends in Ecology & Evolution*, 18, 589-596.
- Amos W, Wilmer JW, Fullard K, Burg TM, Croxall JP, Bloch D, Coulson T (2001). The influence of parental relatedness on reproductive success. *Proceedings of the Royal Society of London B-Biological Sciences*, 268, 2021-2027.
- Apanius V, Penn D, Slev PR, Ruff LR, Potts WK (1997) The nature of selection on the major histocompatibility complex. *Critical Reviews in Immunology*, 17, 179-224.
- Aparicio JM, Ortego J, Cordero PJ (2006). What should we weigh to estimate heterozygosity, alleles or loci? *Molecular Ecology*, 15, 4659-4665.
- Arcaro KF, Eklund A (1998) A review of MHC-based mate preferences and fostering experiments in two congenic strains of mice. *Genetica*, 104, 241-244.
- Armitage KB (1999) Evolution of sociality in marmots. *Journal of Mammalogy*, 80, 1-10.
- Armitage KB (2014) Marmot biology: sociality, individual fitness, and population dynamics. Cambridge University Press, Cambridge.
- Arneberg P, Skorping A, Grenfell B, Read AF (1998) Host densities as determinants of abundance in parasite communities. *Proceedings of the Royal Society of London. Series B-Biological Sciences*, 265, 1283-1289.

- Arnold SJ, Wade MJ (1984a) On the measurement of natural and sexual selection: theory. *Evolution*, 38, 709-719.
- Arnold SJ, Wade MJ (1984b) On the measurement of natural and sexual selection: applications. *Evolution*, 38, 720-734.
- Arnold W (1988) Social thermoregulation during hibernation in Alpine marmots (*Marmota marmota*). *Journal of Comparative Physiology B*, 158, 151-156.
- Arnold W (1990) The evolution of marmot sociality, II: costs and benefits of joint hibernation. *Behavioral Ecology and Sociobiology*, 27, 239-246.
- Arnold W, Lichtenstein AV (1993) Ectoparasite loads decrease the fitness of Alpine marmots (*Marmota marmota*) but are not a cost of sociality. *Behavioral Ecology*, 4, 36-39.
- Arnold W, Dittami J (1997) Reproductive suppression in male Alpine marmots. *Animal Behaviour*, 53, 53-66.
- Avolio ML, Beaulieu JM, Smith MD (2013) Genetic diversity of a dominant C4 grass is altered with increased precipitation variability. *Oecologia*, 171, 571-581.
- Babik W, Taberlet P, Ejsmond MJ, Radwan J (2009) New generation sequencers as a tool for genotyping of highly polymorphic multilocus MHC system. *Molecular Ecology Resources*, 9, 713-719.
- Babik W (2010) Methods for MHC genotyping in non-model vertebrates. *Molecular Ecology Resources*, 10, 237-251.
- Babik W, Kawalko A, Wojcik JM, Radwan J (2012) Low major histocompatibility complex Class I (MHC I) variation in the European Bison (*Bison bonasus*). *Journal of Heredity*, 103, 349-359.
- Bahr A, Wilson AB (2012) The evolution of MHC diversity: evidence of intralocus gene conversion and recombination in a single-locus system. *Gene*, 497, 52-57.
- Bangoura B, Mundt H-C, Schmäscke R, Westphal B, Dauschies A (2012) Prevalence of *Eimeria bovis* and *Eimeria zuernii* in German cattle herds and factors influencing oocyst excretion. *Parasitology Research*, 110, 875-881.
- Barash DP (1974) The evolution of marmot societies: a general theory. *Science*, 185, 415-420.
- Bassano B, Sabatier B, Rossi L, Macchi E (1992) Parasitic fauna of the digestive tract of *Marmota marmota* in the western Alps. In: *First International Symposium on Alpine marmot (Marmota marmota) and Genus Marmota* (ed. Bassano B, Durio P, Gallo Orsi U, Macchi E), pp. 13-24. Saint Vincent (Aosta - Italy).
- Bassano B (1996) Sanitary problems related to marmot-other animals cohabitation in mountain areas. In: *Biodiversity in Marmots / Biodiversité chez les Marmottes* (ed. Le Berre M, Ramousse R, Le Guelte L), pp. 135-140. International marmot network, Moscow, Lyon.
- Bateson P (1983) Optimal outbreeding. In: *Mate choice* (ed. Bateson P), Cambridge University Press, Cambridge.
- Bee MA (2003) A test of the 'dear enemy effect' in the strawberry dart-poison frog (*Dendrobates pumilio*). *Behavioral Ecology and Sociobiology*, 54, 601-610.
- Bel M-C, Porteret C, Coulon J (1995) Scent deposition by cheek rubbing in the Alpine marmot (*Marmota marmota*) in the French Alps. *Canadian Journal of Zoology*, 73, 2065-2071.
- Bel M, Coulon J, Sreng L, Allainé D, Bagnères AG, Clement JL (1999) Social signals involved in scent-marking behavior by cheek-rubbing in Alpine marmots (*Marmota marmota*). *Journal of Chemical Ecology*, 25, 2267-2283.
- Belkhir K, Castric V, Bonhomme F (2002). IDENTIX, a software to test for relatedness

- in a population using permutation methods. *Molecular Ecology Notes*, 2, 611-614.
- Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, 29, 1165-1188.
- Berger V, Lemaître J-F, Gaillard J-M, Cohas A. (2015) How do animals optimize the size-number trade-off when aging? Insights from reproductive senescence patterns in marmots. *Ecology*, 96, 46-53.
- Bernatchez L, Landry C (2003) MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *Journal of Evolutionary Biology*, 16, 363-377.
- Besson J (1971) Introduction de la marmotte dans les Pyrénées Occidentales. *Proceedings of the 96ème Congrès National des Sociétés Savantes, Toulouse, France*, 3, 397-399.
- Bichet C, Penn DJ, Moodley Y, Dunoyer L, Cellier-Holzem E, Belvalette M, Grégoire A, Garnier S, Sorci G (2014) Females tend to prefer genetically similar mates in an island population of house sparrows. *BMC Evolutionary Biology*, 14, 47.
- Birkhead TR, Pizzari T. (2002) Postcopulatory sexual selection. *Nature Reviews Genetics*, 3, 262-273.
- Bjorkman PJ, Parham P (1990) Structure, function, and diversity of class I major histocompatibility complex molecules. *Annual Review of Biochemistry*, 59, 253-288.
- Bland P (1988) MHC class II expression by the gut epithelium. *Immunology Today*, 9, 174-178.
- Blumstein DT, Henderson SJ (1996) Cheek-rubbing in golden marmots (*Marmota caudata aurea*). *Journal of Zoology*, 238, 113-123.
- Bodmer W (1972) Evolutionary significance of HL-A system. *Nature*, 237, 139-183.
- Bollmer JL, Vargas FH, Parker PG (2007) Low MHC variation in the endangered Galapagos penguin (*Spheniscus mendiculus*). *Immunogenetics*, 59, 593-602.
- Bollmer JL, Dunn PO, Freeman-Gallant CR, Whittingham LA (2012) Social and extra-pair mating in relation to major histocompatibility complex variation in common yellowthroats. *Proceedings of the Royal Society B-Biological Sciences*, 279, 4778-4785.
- Booksmythe I, Jennions MD, Backwell PRY (2010) Investigating the 'dear enemy' phenomenon in the territory defence of the fiddler crab, *Uca mjoebergi*. *Animal Behaviour*, 79, 419-423.
- Boots M, Best A, Miller MR, White A (2009) The role of ecological feedbacks in the evolution of host defence: what does theory tell us? *Philosophical Transactions of the Royal Society B-Biological Sciences*, 364, 27-36.
- Bouma HR, Carey HV, Kroese FG (2010) Hibernation: the immune system at rest? *Journal of Leukocyte Biology*, 88, 619-624.
- Bradbury JW, Vehrencamp SL (1998) *Principles of animal communication*. Sunderland, MA: Sinauer Associates.
- Bradley J, Jackson J (2008) Measuring immune system variation to help understand host-pathogen community dynamics. *Parasitology*, 135, 807-823.
- Brennan PA, Kendrick KM (2006) Mammalian social odours: attraction and individual recognition. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 361, 2061-2078.
- Brouwer L, Barr I, Van De Pol M, Burke T, Komdeur J, Richardson DS (2010) MHC-dependent survival in a wild population: evidence for hidden genetic benefits gained through extra-pair fertilizations. *Molecular Ecology*, 19, 3444-3455.



- Brown JL (1964) The evolution of diversity in avian territorial systems. *Wilson Bulletin*, 76, 160-169.
- Brown JL (1997) A theory of mate choice based on heterozygosity. *Behavioral Ecology*, 8, 60-65.
- Brown JR, Beckenbach AT, Smith MJ (1993) Intraspecific DNA sequence variation of the mitochondrial control region of white sturgeon (*Acipenser transmontanus*). *Molecular Biology and Evolution*, 10, 326-341.
- Buesching CD, Waterhouse JS, Macdonald DW (2002) Gas-chromatographic analyses of the subcaudal gland secretion of the European badger (*Meles meles*) Part I: Chemical differences related to individual parameters. *Journal of Chemical Ecology*, 28, 41-56.
- Burgener N, Dehnhard M, Hofer H, East ML (2009) Does anal gland scent signal identity in the spotted hyaena? *Animal Behaviour*, 77, 707-715.
- Callait M (1999) Relations hôte-parasite dans un écosystème alpin: dynamique annuelle et structures spatiales chez la marmotte Alpine (*Marmota marmota* Linne 1758). PhD thesis, University of Savoie, Chambéry, France.
- Callait M, Gauthier D (2000) Parasite adaptations to hibernation in Alpine marmots (*Marmota marmota*). In: *Life in the cold. 11th international hibernation symposium* (ed. Heldmaier G, Klingenspor M), pp 139-146, Springer, Heidelberg.
- Cammen K, Hoffman JI, Knapp LA, Harwood J, Amos W (2011) Geographic variation of the major histocompatibility complex in Eastern Atlantic grey seals (*Halichoerus grypus*). *Molecular Ecology*, 20, 740-752.
- Caniglia R, Fabbri E, Galaverni M, Milanese P, Randi E (2014) Noninvasive sampling and genetic variability, pack structure, and dynamics in an expanding wolf population. *Journal of Mammalogy*, 95, 41-59.
- Cant MA (2000) Social control of reproduction in banded mongooses. *Animal Behaviour*, 59, 147-158.
- Canty A, Ripley BD (2014) boot: Bootstrap R (S-Plus) Functions.
- Carey HV, Cooke HJ (1991) Effect of hibernation and jejunal bypass on mucosal structure and function. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, 261, G37-G44.
- Carlin JB, Wolfe R, Brown CH, Gelman A (2001) A case study on the choice, interpretation and checking of multilevel models for longitudinal binary outcomes. *Biostatistics*, 2, 397-416.
- Carrington M, Nelson GW, Martin MP, Kissner T, Vlahov D, Goedert JJ, Kaslow R, Buchbinder S, Hoots K, O'Brien SJ (1999) HLA and HIV-1: heterozygote advantage and B\* 35-Cw\* 04 disadvantage. *Science*, 283, 1748-1752.
- Castillo S, Srithayakumar V, Meunier V, Kyle CJ (2010) Characterization of major histocompatibility complex (MHC) DRB exon 2 and DRA exon 3 fragments in a primary terrestrial rabies vector (*Procyon lotor*). *PLoS ONE*, 5, e12066.
- Castro-Prieto A, Wachter B, Sommer S (2011) Cheetah paradigm revisited: MHC diversity in the world's largest free-ranging population. *Molecular Biology and Evolution*, 28, 1455-1468.
- Chapman HD, Barta JR, Blake D, Gruber A, Jenkins M, Smith NC, Suo X, Tomley FM (2013) A selective review of advances in coccidiosis research. *Advances in Parasitology*, 83, 93-171.
- Chapman JR, Nakagawa S, Coltman DW, Slate J, Sheldon BC (2009) A quantitative review of heterozygosity–fitness correlations in animal populations. *Molecular Ecology*, 18, 2746-2765.

- Charlesworth D, Awadalla P (1998) Flowering plant self-incompatibility: the molecular population genetics of Brassica S-loci. *Heredity*, 81, 1-9.
- Charpentier MJE, Boulet M, Drea CM (2008) Smelling right: the scent of male lemurs advertises genetic quality and relatedness. *Molecular Ecology*, 17, 3225-3233.
- Chute RM (1960) Overwintering of helminths in hibernating animals. *Journal of Parasitology*, 46, 855-858.
- Clutton-Brock TH (1988) Reproductive success: studies of individual variation in contrasting breeding systems. University of Chicago Press, Chicago.
- Clutton-Brock TH (1998) Reproductive skew, concessions and limited control. *Trends in Ecology & Evolution*, 13, 288-292.
- Clutton-Brock TH, Brotherton P, Russell A, O'riain M, Gaynor D, Kansky R, Griffin A, Manser M, Sharpe L, McIlrath G, et al. (2001) Cooperation, control, and concession in meerkat groups. *Science*, 291, 478-481.
- Clutton-Brock TH, Sheldon BC (2010) Individuals and populations: the role of long-term, individual-based studies of animals in ecology and evolutionary biology. *Trends in Ecology & Evolution*, 25, 562-573.
- Coggins JR, Tedesco JL, Rupprecht CE (1982) Seasonal changes and overwintering of parasites in the bat, *Myotis lucifugus* (Le Conte), in a Wisconsin hibernaculum. *American Midland Naturalist*, 107, 305-315.
- Cohas A, Yoccoz NG, Da Silva A, Goossens B, Allainé D (2006) Extra-pair paternity in the monogamous Alpine marmot (*Marmota marmota*): the roles of social setting and female mate choice. *Behavioral Ecology and Sociobiology*, 59, 597-605.
- Cohas A, Bonenfant C, Gaillard JM, Allainé D (2007) Are extra-pair young better than within-pair young? A comparison of survival and dominance in Alpine marmot. *Journal of Animal Ecology*, 76, 771-781.
- Cohas A, Yoccoz NG, Bonenfant C, Goossens B, Genton C, Galan M, Kempnaers B, Allainé D (2008) The genetic similarity between pair members influences the frequency of extrapair paternity in Alpine marmots. *Animal Behaviour*, 76, 87-95.
- Cohas A, Bonenfant C, Kempnaers B, Allainé D (2009) Age-specific effect of heterozygosity on survival in Alpine marmots, *Marmota marmota*. *Molecular Ecology*, 18, 1491-1503.
- Cohn DL, Erb HN, Georgi JR, Tennant BC (1986) Parasites of the laboratory woodchuck (*Marmota monax*). *Laboratory Animal Science*, 36, 298-302.
- Coltman D, Pilkington J, Smith J, Pemberton J (1999) Parasite-mediated selection against inbred Soay sheep in a free-living, island population. *Evolution*, 53, 1259-1267.
- Coltman D, Slate J (2003) Microsatellite measures of inbreeding: a meta-analysis. *Evolution*, 57, 971-983.
- Coulon A (2010) GENHET: an easy-to-use R function to estimate individual heterozygosity. *Molecular Ecology Resources*, 10, 167-169.
- Coulson T, Benton T, Lundberg P, Dall S, Kendall B, Gaillard J-M (2006) Estimating individual contributions to population growth: evolutionary fitness in ecological time. *Proceedings of the Royal Society B-Biological Sciences*, 273, 547-555.
- Couturier M (1955) Acclimatation et acclimatement de la Marmotte des Alpes, *Marmota m. marmota* (Linné 1758) dans les Pyrénées françaises. *Säugetierkd. Mitteilungen*, 3, 105-108.
- Cringoli G, Rinaldi L, Veneziano V, Capelli G, Scala A (2004) The influence of flotation solution, sample dilution and the choice of McMaster slide area (volume) on the reliability of the McMaster technique in estimating the faecal egg counts of

- gastrointestinal strongyles and *Dicrocoelium dendriticum* in sheep. *Veterinary Parasitology*, 123, 121-131.
- Crnokrak P, Roff DA (1999) Inbreeding depression in the wild. *Heredity*, 83, 260-270.
- Cross HB, Blumstein DT, Rosell F (2013) Do marmots display a 'dear enemy phenomenon' in response to anal gland secretions? *Journal of Zoology*, 289, 189-195.
- Cryan PM, Meteyer CU, Boyles JG, Blehert DS (2010) Wing pathology of white-nose syndrome in bats suggests life-threatening disruption of physiology. *BMC Biology*, 8, 135.
- Csilléry K, Johnson T, Beraldi D, Clutton-Brock TH, Coltman D, Hansson B, Spong G, Pemberton JM (2006) Performance of marker-based relatedness estimators in natural populations of outbred vertebrates. *Genetics*, 173, 2091-2101.
- Cutrera AP, Zenuto RR, Lacey EA (2011) MHC variation, multiple simultaneous infections and physiological condition in the subterranean rodent *Ctenomys talarum*. *Infection, Genetics and Evolution*, 11, 1023-1036.
- Da Silva A, Luikart G, Allainé D, Gautier P, Taberlet P, Pompanon F (2003) Isolation and characterization of microsatellites in European Alpine marmots (*Marmota marmota*). *Molecular Ecology Notes*, 3, 189-190.
- Dauguschies A, Najdrowski M (2005) Eimeriosis in cattle: current understanding. *Journal of Veterinary Medicine, Series B*, 52, 417-427.
- David P (1998) Heterozygosity–fitness correlations: new perspectives on old problems. *Heredity*, 80, 531-537.
- Davis SD (1969) Hibernation: intestinal protozoa populations in ground squirrels. *Experimental Parasitology*, 26, 156-165.
- De Gussem M (2007) Coccidiosis in poultry: review on diagnosis, control, prevention and interaction with overall gut health. In: *Proceedings of the 16th European Symposium on Poultry Nutrition*, pp. 253-261, Beekbergen.
- Denegri GM (1993) Review of oribatid mites as intermediate hosts of tapeworms of the Anoplocephalidae. *Experimental and Applied Acarology*, 17, 567-580.
- Dever M, Kahn L, Doyle E (2015) Removal of tapeworm (*Moniezia spp.*) did not increase growth rates of meat-breed lambs in the Northern Tablelands of NSW. *Veterinary Parasitology*, 208, 190-194.
- Doherty PC, Zinkernagel RM (1975) Enhanced immunological surveillance in mice heterozygous at H-2 gene complex. *Nature*, 256, 50-52.
- Dunn PO, Bollmer JL, Freeman-Gallant CR, Whittingham LA (2013) MHC variation is related to a sexually selected ornament, survival, and parasite resistance in common yellowthroats. *Evolution*, 67, 679-687.
- Duszynski DW, Upton SJ (2001) Enteric Protozoans: *Cyclospora*, *Eimeria*, *Isospora*, and *Cryptosporidium spp.* In: *Parasitic diseases of wild mammals* (ed. Samuel WM, Pybus MJ, Kocan AA), pp. 416-459. Iowa State University, Iowa.
- Ebermann E (1976) Oribatids (Oribatei, Acari) as intermediate hosts of *Ctenotaenia marmotae* (Frolich, 1802), a tapeworm parasitic in marmots (author's transl). *Zeitschrift für Parasitenkunde*, 50, 303-312.
- Edwards SV, Hedrick PW (1998) Evolution and ecology of MHC molecules: from genomics to sexual selection. *Trends in Ecology & Evolution*, 13, 305-311.
- Eizaguirre C, Yeates SE, Lenz TL, Kalbe M, Milinski M (2009) MHC-based mate choice combines good genes and maintenance of MHC polymorphism. *Molecular Ecology*, 18, 3316-3329.
- Ejmsmond MJ, Radwan J (2011) MHC diversity in bottlenecked populations: a simulation

- model. *Conservation Genetics*, 12, 129-137.
- Ejmsmond MJ, Radwan J, Wilson AB (2014) Sexual selection and the evolutionary dynamics of the major histocompatibility complex. *Proceedings of the Royal Society B-Biological Sciences*, 281, 20141662.
- Ekblom R, Sæther SA, Grahn M, Fiske P, Kålås JA, Höglund J (2004) Major histocompatibility complex variation and mate choice in a lekking bird, the great snipe (*Gallinago media*). *Molecular Ecology*, 13, 3821-3828.
- Ellis JS, Turner LM, Knight ME (2012) Patterns of selection and polymorphism of innate immunity genes in bumblebees (Hymenoptera: Apidae). *Genetica*, 140, 205-217.
- Ellstrand NC, Elam DR (1993) Population genetic consequences of small population size: implications for plant conservation. *Annual Review of Ecology and Systematics*, 24, 217-242.
- Etter PD, Bassham S, Hohenlohe PA, Johnson EA, Cresko WA (2011) SNP discovery and genotyping for evolutionary genetics using RAD sequencing. *Molecular Methods for Evolutionary Genetics*, 772, 157-178.
- Evans ML, Neff BD (2009) Major histocompatibility complex heterozygote advantage and widespread bacterial infections in populations of Chinook salmon (*Oncorhynchus tshawytscha*). *Molecular Ecology*, 18, 4716-4729.
- Feng D-F, Doolittle RF (1987) Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *Journal of Molecular Evolution*, 25, 351-360.
- Fisher J (1954) Evolution and bird sociality. In: *Evolution as a process* (ed. Huxley A, Ford E), Allen and Unwin, London.
- Forcada J, Hoffman JI (2014) Climate change selects for heterozygosity in a declining fur seal population. *Nature*, 511, 462-465.
- Frank SA (2002) Immunology and evolution of infectious disease. Princeton University Press, Princeton.
- Frankel O, Soulé M (1981) Conservation and evolution. Cambridge University Press, Cambridge.
- Frankham R (1996) Relationship of genetic variation to population size in wildlife. *Conservation Biology*, 10, 1500-1508.
- Frankham R (2005) Stress and adaptation in conservation genetics. *Journal of Evolutionary Biology*, 18, 750-755.
- Froeschke G, Sommer S (2005) MHC class II DRB variability and parasite load in the striped mouse (*Rhabdomys pumilio*) in the southern Kalahari. *Molecular Biology and Evolution*, 22, 1254-1259.
- Gaillard J-M, Boutin J-M, Delorme D, Van Laere G, Duncan P, Lebreton J-D (1997) Early survival in roe deer: causes and consequences of cohort variation in two contrasted populations. *Oecologia*, 112, 502-513.
- Galan M, Guivier E, Caraux G, Charbonnel N, Cosson J-F (2010) A 454 multiplex sequencing method for rapid and reliable genotyping of highly polymorphic genes in large-scale studies. *BMC Genomics*, 11, 296.
- Galli-Valerio B (1923) Parasitologische untersuchungen und beiträge zur parasitologischen technik. *Zentralblatt fuer Bakteriologie Parasitenkunde*, 91, 120-125.
- Galtier N, Piganeau G, Mouchiroud D, Duret L (2001) GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. *Genetics*, 159, 907-911.
- Garrigan D, Hedrick PW (2003) Perspective: detecting adaptive molecular

- polymorphism: lessons from the MHC. *Evolution*, 57, 1707-1722.
- Geiser F (2013) Hibernation. *Current Biology*, 23, R188-R193.
- Gensac P, Rothé B (1974) Carte de la végétation de la Réserve de la Grande Sassièrè. *Travaux Scientifiques du Parc National de la Vanoise*, 5, 77-103.
- German AJ, Bland PW, Hall EJ, Day MJ (1998) Expression of major histocompatibility complex class II antigens in the canine intestine. *Veterinary Immunology and Immunopathology*, 61, 171-180.
- Gienapp P, Teplitsky C, Alho JS, Mills JA, Merilä J (2008) Climate change and evolution: disentangling environmental and genetic responses. *Molecular Ecology*, 17, 167-178.
- Glenn TC (2011) Field guide to next-generation DNA sequencers. *Molecular Ecology Resources*, 11, 759-769.
- Goossens B, Graziani L, Waits LP, Farand E, Magnolon S, Coulon J, Bel M-C, Taberlet P, Allainé D (1998) Extra-pair paternity in the monogamous Alpine marmot revealed by nuclear DNA microsatellite analysis. *Behavioral Ecology and Sociobiology*, 43, 281-288.
- Goossens B, Chikhi L, Taberlet P, Waits L, Allainé D (2001) Microsatellite analysis of genetic variation among and within Alpine marmot populations in the French Alps. *Molecular Ecology*, 10, 41-52.
- Goüy de Bellocq J, Charbonnel N, Morand S (2008) Coevolutionary relationship between helminth diversity and MHC class II polymorphism in rodents. *Journal of Evolutionary Biology*, 21, 1144-1150.
- Gregg BS (2011) *Toxoplasma gondii*--Antigen presentation and interactions in the intestinal mucosa. PhD thesis, Faculties of the University of Pennsylvania, Pennsylvania.
- Grueber CE, Wallis GP, King TM, Jamieson IG (2012) Variation at innate immunity Toll-like receptor genes in a bottlenecked population of a New Zealand robin. *PLoS ONE*, 7, e45011.
- Hackländer K, Mostl E, Arnold W (2003) Reproductive suppression in female Alpine marmots, *Marmota marmota*. *Animal Behaviour*, 65, 1133-1140.
- Haldane J (1949) Disease and evolution. *Ricerca Scientifica*, 19, 325-334.
- Halekoh U, Højsgaard S, Yan J (2006) The R package geepack for generalized estimating equations. *Journal of Statistical Software*, 15, 1-11.
- Hamilton W, Zuk M (1982) Heritable true fitness and bright birds - a role for parasites. *Science*, 218, 384-387.
- Hanslik S, Kruckenhauser L (2000) Microsatellite loci for two European sciurid species (*Marmota marmota*, *Spermophilus citellus*). *Molecular Ecology*, 9, 2163-2165.
- Hansson B, Bensch S, Hasselquist D, Åkesson M (2001). Microsatellite diversity predicts recruitment of sibling great reed warblers. *Proceedings of the Royal Society of London B-biological Sciences*, 268, 1287-1291.
- Hansson B, Westerberg L (2002) On the correlation between heterozygosity and fitness in natural populations. *Molecular Ecology*, 11, 2467-2474.
- Harf R, Sommer S (2005) Association between major histocompatibility complex class II DRB alleles and parasite load in the hairy-footed gerbil, *Gerbillurus paeba*, in the southern Kalahari. *Molecular Ecology*, 14, 85-91.
- Hartmann SA, Schaefer HM, Segelbacher G (2014) Genetic depletion at adaptive but not neutral loci in an endangered bird species. *Molecular Ecology*, 23, 5712-5725.
- Hayes R, Richardson B, Wyllie S (2001) Increased social dominance in rabbits, *Oryctolagus cuniculus*, is associated with increased secretion of 2-phenoxyethanol



- from the chin gland. In: *Chemical signals in vertebrates 9* (ed. Marchlewska-Koi A, Lepri J, Müller-Schwarze D), Plenum Press, New York.
- Heath WR, Carbone FR (2001) Cross-presentation in viral immunity and self-tolerance. *Nature Reviews Immunology*, 1, 126-134.
- Hedrick PW (1992) Female choice and variation in the major histocompatibility complex. *Genetics*, 132, 575-581.
- Hedrick PW (1994) Evolutionary genetics of the major histocompatibility complex. *American Naturalist*, 143, 945-964.
- Hedrick PW, Parker KM, Gutiérrez-Espeleta GA, Rattink A, Lievers K (2000) Major histocompatibility complex variation in the Arabian oryx. *Evolution*, 54, 2145-2151.
- Hedrick PW (2002) Pathogen resistance and genetic variation at MHC loci. *Evolution*, 56, 1902-1908.
- Hedrick PW, Hurt CR (2012) Conservation genetics and evolution in an endangered species: research in Sonoran topminnows. *Evolutionary Applications*, 5, 806-819.
- Heinze J, Foitzik S, Hippert A, Holldobler B (1996) Apparent dear-enemy phenomenon and environment-based recognition cues in the ant *Leptothorax nylanderi*. *Ethology*, 102, 510-522.
- Herdegen M, Babik W, Radwan J (2014) Selective pressures on MHC class II genes in the guppy (*Poecilia reticulata*) as inferred by hierarchical analysis of population structure. *Journal of Evolutionary Biology*, 27, 2347-2359.
- Hernandez AD, Poole A, Cattadori IM (2013) Climate changes influence free-living stages of soil-transmitted parasites of European rabbits. *Global Change Biology*, 19, 1028-1042.
- Herrero J (2000) La invasión pirenaica de la marmota Alpina. *Biológica*, 47, 22-29.
- Hill AVS (1991). HLA associations with malaria in Africa: some implications for MHC evolution. In: *Molecular evolution of the major histocompatibility complex* (ed. Klein J, Klein D), pp. 403-419. Springer, Berlin.
- Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD Tags. *PLoS Genetics*, 6, e1000862.
- Horton NJ, Lipsitz SR (1999) Review of software to fit generalized estimating equation regression models. *The American Statistician*, 53, 160-169.
- Huchard E, Albrecht C, Schliehe-Diecks S, Baniel A, Roos C, Peter PMK, Brameier M (2012) Large-scale MHC class II genotyping of a wild lemur population by next generation sequencing. *Immunogenetics*, 64, 895-913.
- Huchard E, Baniel A, Schliehe-Diecks S, Kappeler PM (2013) MHC-disassortative mate choice and inbreeding avoidance in a solitary primate. *Molecular Ecology*, 22, 4071-4086.
- Huchard E, Pechouskova E (2014) The major histocompatibility complex and primate behavioral ecology: New tools and future questions. *International Journal of Primatology*, 35, 11-31.
- Hudson P, Dobson A (1997) Host-parasite processes and demographic consequences. In: *Host-parasite evolution: general principles and avian models* (ed. Clayton DH, Moore J), pp. 128-154. Oxford University Press, Oxford.
- Hueffer K, O'Hara TM, Follmann EH (2011) Adaptation of mammalian host-pathogen interactions in a changing arctic environment. *Acta Veterinaria Scandinavica*, 53, 17.
- Hughes AL, Nei M (1989) Evolution of the major histocompatibility complex:

- independent origin of nonclassical class I genes in different groups of mammals. *Molecular Biology and Evolution*, 6, 559-579.
- Hughes AL, Nei M (1992) Models of host-parasite interaction and MHC polymorphism. *Genetics*, 132, 863-864.
- Hughes AL, Yeager M (1998) Natural selection at major histocompatibility complex loci of vertebrates. *Annual Review of Genetics*, 32, 415-435.
- Hume I, Beiglböck C, Ruf T, Frey-Roos F, Bruns U, Arnold W (2002) Seasonal changes in morphology and function of the gastrointestinal tract of free-living Alpine marmots (*Marmota marmota*). *Journal of Comparative Physiology B*, 172, 197-207.
- Hurrell JW (1995) Decadal trends in the North Atlantic Oscillation: regional temperatures and precipitation. *Science*, 269, 676-679.
- Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM (2007) Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biology*, 8, R143.
- Jarne P, Lagoda P (1996) Microsatellites, from molecules to populations and back. *Trends in Ecology & Evolution*, 11, 424-429.
- Jennions MD (1997) Female promiscuity and genetic incompatibility. *Trends in Ecology & Evolution*, 12, 251-253.
- Jennions MD, Petrie M (2000) Why do females mate multiply? A review of the genetic benefits. *Biological Reviews of the Cambridge Philosophical Society*, 75, 21-64.
- Jepson A, Banya W, SisayJoof F, HassanKing M, Nunes C, Bennett S, Whittle H (1997) Quantification of the relative contribution of major histocompatibility complex (MHC) and non-MHC genes to human immune responses to foreign antigens. *Infection and Immunity*, 65, 872-876.
- Jones J (2001) Putting knowledge of plant disease resistance genes to work. *Current Opinion in Plant Biology*, 4, 281-287.
- Jordan WC, Bruford MW (1998) New perspectives on mate choice and the MHC. *Heredity*, 81, 127-133.
- Jørgensen MH, Lagesen K, Mable BK, Brysting AK (2012) Using high-throughput sequencing to investigate the evolution of self-incompatibility genes in the Brassicaceae: strategies and challenges. *Plant Ecology & Diversity*, 5, 473-484.
- Jump AS, Marchant R, Peñuelas J (2009) Environmental change and the option value of genetic diversity. *Trends in Plant Science*, 14, 51-58.
- Juola FA, Dearborn DC (2012) Sequence-based evidence for major histocompatibility complex-disassortative mating in a colonial seabird. *Proceedings of the Royal Society B-Biological Sciences*, 279, 153-162.
- Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology*, 16, 1099-1106.
- Kamiya T, O'Dwyer K, Westerdahl H, Senior A, Nakagawa S (2014) A quantitative review of MHC-based mating preference: the role of diversity and dissimilarity. *Molecular Ecology*, 23, 5151-5163.
- Kappeler PM, Watts DP (2012) Long-term field studies of primates. Springer, New York.
- Kean EF, Muller CT, Chadwick EA (2011) Otter scent signals age, sex, and reproductive status. *Chemical Senses*, 36, 555-564.
- Keller LF, Arcese P, Smith JN, Hochachka WM, Stearns SC (1994) Selection against inbred song sparrows during a natural population bottleneck. *Nature*, 372, 356-357.

- Keller LF, Waller DM (2002) Inbreeding effects in wild populations. *Trends in Ecology & Evolution*, 17, 230-241.
- Kelley J, Walter L, Trowsdale J (2005) Comparative genomics of major histocompatibility complexes. *Immunogenetics*, 56, 683-695.
- Kempnaers B (2007) Mate choice and genetic quality: a review of the heterozygosity theory. *Advances in the Study of Behavior*, 37, 189-278.
- King W, Allainé D (1998) Copulatory behaviour of Alpine marmots (*Marmota marmota*). *Mammalia*, 62, 439-441.
- Klein J (1982) Evolution and function of the major histocompatibility complex. In: *Histocompatibility Antigens* (ed. Parham P, Strominger J), pp. 223-239. Chapman and Hall, London.
- Klein J, Sato A, Nagl S, O'hUigin C (1998) Molecular trans-species polymorphism. *Annual Review of Ecology and Systematics*, 29, 1-21.
- Klinkicht M (1993) Untersuchungen zum Paarungssystem des Alpenmurmeltiers, *Marmota M. marmota* mittels DAN Fingerprinting. PhD Thesis, University of Munich, Munich, Germany.
- Kloch A, Babik W, Bajer A, Sinski E, Radwan J (2010) Effects of an MHC-DRB genotype and allele number on the load of gut parasites in the bank vole *Myodes glareolus*. *Molecular Ecology*, 19, 255-265.
- Knapp LA, Ha JC, Sackett GP (1996) Parental MHC antigen sharing and pregnancy wastage in captive pigtailed macaques. *Journal of Reproductive Immunology*, 32, 73-88.
- Knapp LA (2005) The abcs of mhc. *Evolutionary Anthropology: Issues, News, and Reviews*, 14, 28-37.
- Kokko H, Brooks R, Jennions MD, Morley J (2003) The evolution of mate choice and mating biases. *Proceedings of the Royal Society of London B-Biological Sciences*, 270, 653-664.
- Kokko H, Ots I (2006) When not to avoid inbreeding. *Evolution*, 60, 467-475.
- Körtner G, Heldmaier G (1995) Body-weight cycles and energy-balance in the Alpine marmot (*Marmota marmota*). *Physiological Zoology*, 68, 149-163.
- Kraus RH, Von Holdt B, Cocchiararo B, Harms V, Bayerl H, Kühn R, Förster DW, Fickel J, Roos C, Nowak C (2014) A single-nucleotide polymorphism-based approach for rapid and cost-effective genetic wolf monitoring in Europe based on noninvasively collected samples. *Molecular Ecology Resources*, 15, 295-305.
- Kruckenhauser L, Miller W, Preleuthner M, Pinsker W (1997) Differentiation of Alpine marmot populations traced by DNA fingerprinting. *Journal of Zoological Systematics and Evolutionary Research*, 35, 143-149.
- Kuduk K, Johanet A, Allainé D, Cohas A, Radwan J (2012) Contrasting patterns of selection acting on MHC class I and class II DRB genes in the Alpine marmot (*Marmota marmota*). *Journal of Evolutionary Biology*, 25, 1686-1693.
- Kumánovics A, Takada T, Lindahl KF (2003) Genomic organization of the mammalian MHC. *Annual Review of Immunology*, 21, 629-657.
- Lachish S, Goldizen AW (2004) Responses to neighbours and non-neighbours in the buff-banded rail (*Gallirallus philippensis*): no dear-enemy relationships. *Australian Journal of Zoology*, 52, 369-378.
- Lardy S, Cohas A, Figueroa I, Allainé D (2011) Mate change in a socially monogamous mammal: evidences support the forced divorce hypothesis. *Behavioral Ecology*, 22, 120-125.
- Lardy S (2012) Evolutionary consequences of intra-sexual competition in social species.



- the example of the Alpine marmot (*Marmota marmota*). PhD thesis, Department of Evolutionary Ecology, University of Lyon, Villeurbanne, France.
- Lardy S, Cohas A, Desouhant E, Tafani M, Allainé D (2012) Paternity and dominance loss in male breeders: the cost of helpers in a cooperatively breeding mammal. *PLoS ONE*, 7, e29508.
- Lawlor DA, Zemmour J, Ennis PD, Parham P (1990) Evolution of class-I Mhc genes and proteins - From natural-selection to thymic selection. *Annual Review of Immunology*, 8, 23-63.
- Le Berre M (1994). Bases de la gestion des marmottes. In: *2ème Journée d'Etude sur la Marmotte Alpine* (ed. Ramousse R, Le Berre M), pp. 31-38.
- Lebreton JD, Pradel R (2002) Multistate recapture models: modelling incomplete individual histories. *Journal of Applied Statistics*, 29, 353-369.
- Leclaire S, Merckling T, Raynaud C, Mulard H, Bessière J-M, Lhuillier É, Hatch SA, Danchin É (2012) Semiochemical compounds of preen secretion reflect genetic make-up in a seabird species. *Proceedings of the Royal Society B-Biological Sciences*, 279, 1185-1193.
- Lenti Boero D (1995) Scent-deposition behaviour in Alpine marmots (*Marmota marmota* L.): its role in territorial defence and social communication. *Ethology*, 100, 26-38.
- Lenz TL, Becker S (2008) Simple approach to reduce PCR artefact formation leads to reliable genotyping of MHC and other highly polymorphic loci - Implications for evolutionary analysis. *Gene*, 427, 117-123.
- Lenz TL, Wells K, Pfeiffer M, Sommer S (2009) Diverse MHC IIB allele repertoire increases parasite resistance and body condition in the long-tailed giant rat (*Leopoldamys sabanus*). *BMC Evolutionary Biology*, 9, 269.
- Liang KY, Zeger SL (1986) Longitudinal data-analysis using generalized linear-models. *Biometrika*, 73, 13-22.
- Lighten J, Oosterhout C, Paterson IG, McMullan M, Bentzen P (2014a) Ultra-deep Illumina sequencing accurately identifies MHC class IIb alleles and provides evidence for copy number variation in the guppy (*Poecilia reticulata*). *Molecular Ecology Resources*, 14, 753-767.
- Lighten J, Van Oosterhout C, Bentzen P (2014b) Critical review of NGS analyses for de novo genotyping multigene families. *Molecular Ecology*, 23, 3957-3972.
- Ljungqvist M, Åkesson M, Hansson B (2010) Do microsatellites reflect genome-wide genetic diversity in natural populations? A comment on Väli et al. (2008). *Molecular Ecology*, 19, 851-855.
- López BC, Pino J, López A (2010) Explaining the successful introduction of the Alpine marmot in the Pyrenees. *Biological Invasions*, 12, 3205-3217.
- Lopez J, Wey T, Blumstein D (2013) Patterns of parasite prevalence and individual infection in yellow-bellied marmots. *Journal of Zoology*, 291, 296-303.
- Løvlie H, Gillingham MA, Worley K, Pizzari T, Richardson DS (2013) Cryptic female choice favours sperm from major histocompatibility complex-dissimilar males. *Proceedings of the Royal Society B: Biological Sciences*, 280, 20131296.
- Lynch M, Ritland K (1999). Estimation of pairwise relatedness with molecular markers. *Genetics*, 152, 1753-1766.
- Macdonald EA, Fernandez-Duque E, Evans S, Hagey LR (2008) Sex, age, and family differences in the chemical composition of owl monkey (*Aotus nancymaae*) subcaudal scent secretions. *American Journal of Primatology*, 70, 12-18.
- Maher CR, Lott DF (1995) Definitions of territoriality used in the study of variation in

- vertebrate spacing systems. *Animal Behaviour*, 49, 1581-1597.
- Magnolon S (1999). Dispersion natale chez la marmotte Alpine (*Marmota marmota*). modalites et effet de quelques facteurs proximaux. PhD thesis, University of Tours, Tours, France.
- Marquardt WC (1960) Effect of high temperature on sporulation of *Eimeria zurnii*. *Experimental Parasitology*, 10, 58-65.
- Martinsohn JT, Sousa AB, Guethlein LA, Howard JC (1999) The gene conversion hypothesis of MHC evolution: a review. *Immunogenetics*, 50, 168-200.
- Mays HL, Hill GE (2004) Choosing mates: good genes *versus* genes that are a good fit. *Trends in Ecology & Evolution*, 19, 554-559.
- Mays HL, Albrecht T, Liu M, Hill GE (2008) Female choice for genetic complementarity in birds: a review. *Genetica*, 134, 147-158.
- McClelland EE, Penn DJ, Potts WK (2003) Major histocompatibility complex heterozygote superiority during coinfection. *Infection and Immunity*, 71, 2079-2086.
- McKay DM, Webb R (2006) Cestode infection: immunological considerations from host and tapeworm perspectives. In: *Parasitic flatworms: molecular biology, biochemistry, immunology and physiology* (ed. Maule A, Marks N), pp. 193-209. CAB International, Wallingford.
- McKay DM (2010) The immune response to and immunomodulation by *Hymenolepis diminuta*. *Parasitology*, 137, 385-394.
- McNeely JA, Miller KR, Reid WV, Mittermeier RA, Werner TB (1990) Conserving the world's biological diversity. IUCN, World Resources Institute, Conservation International, WWF-US and the World Bank, Washington, DC.
- Mead LS, Arnold SJ (2004) Quantitative genetic models of sexual selection. *Trends in Ecology & Evolution*, 19, 264-271.
- Medzhitov R, Schneider DS, Soares MP (2012) Disease tolerance as a defense strategy. *Science*, 335, 936-941.
- Milinski M (2006) The major histocompatibility complex, sexual selection, and mate choice. *Annual Review of Ecology Evolution and Systematics*, 37, 159-186.
- Miller HC, Moore JA, Nelson NJ, Daugherty CH (2009) Influence of major histocompatibility complex genotype on mating success in a free-ranging reptile population. *Proceedings of the Royal Society B: Biological Sciences*, 276, 1695-1704.
- Miller JM, Malenfant RM, David P, Davis CS, Poissant J, Hogg JT, Festa-Bianchet M, Coltman DW (2014) Estimating genome-wide heterozygosity: effects of demographic history and marker type. *Heredity*, 112, 240-247.
- Miller MR, White A, Boots M (2006) The evolution of parasites in response to tolerance in their hosts: the good, the bad, and apparent commensalism. *Evolution*, 60, 945-956.
- Monclús R, Saavedra I, de Miguel J (2014) Context-dependent responses to neighbours and strangers in wild European rabbits (*Oryctolagus cuniculus*). *Behavioural Processes*, 106, 17-21.
- Moreno-Cugnon L, Esparza-Baquer A, Larruskain A, García-Etxebarria K, Menne S, González-Aseguinolaza G, Jugo BM (2015) Characterization and genotyping of the DRB1 gene of the major histocompatibility complex (MHC) in the *Marmota monax*, animal model of hepatitis B. *Molecular Immunology*, 63, 505-512.
- Morgan M, Anders S, Lawrence M, Aboyoun P, Pages H, Gentleman R (2009) ShortRead: a bioconductor package for input, quality assessment and exploration

- of high-throughput sequence data. *Bioinformatics*, 25, 2607-2608.
- Mowat AM (2003) Anatomical basis of tolerance and immunity to intestinal antigens. *Nature Reviews Immunology*, 3, 331-341.
- Narum SR (2006) Beyond Bonferroni: less conservative analyses for conservation genetics. *Conservation Genetics*, 7, 783-787.
- Neefjes J, Jongstra ML, Paul P, Bakke O (2011) Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nature Reviews Immunology*, 11, 823-836.
- Neff BD, Pitcher TE (2005) Genetic quality and sexual selection: an integrated framework for good genes and compatible genes. *Molecular Ecology*, 14, 19-38.
- Nei M, Gu X, Sitnikova T (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proceedings of the National Academy of Sciences*, 94, 7799-7806.
- Nikolich-Zugich J, Fremont DH, Miley MJ, Messaoudi I (2004) The role of mhc polymorphism in anti-microbial resistance. *Microbes and Infection*, 6, 501-512.
- Nowak MA, Tarczyhornoch K, Austyn JM (1992) The optimal number of major histocompatibility complex-molecules in an individual. *Proceedings of the National Academy of Sciences of the United States of America*, 89, 10896-10899.
- Nunn CL, Altizer S, Jones KE, Sechrest W (2003) Comparative tests of parasite species richness in primates. *The American Naturalist*, 162, 597-614.
- O'Brien SJ, Wildt DE, Goldman D, Merrill CR, Bush M (1983) The cheetah is depauperate in genetic variation. *Science*, 221, 459-462.
- Ober C, Hyslop T, Elias S, Weitkamp LR, Hauck WW (1998) Human leukocyte antigen matching and fetal loss: results of a 10 year prospective study. *Human Reproduction*, 13, 33-38.
- Oh KP, Badyaev AV (2006) Adaptive genetic complementarity in mate choice coexists with selection for elaborate sexual traits. *Proceedings of the Royal Society B-biological Sciences*, 273, 1913-1919.
- Oliver MK, Telfer S, Piortney SB (2009) Major histocompatibility complex (MHC) heterozygote superiority to natural multi-parasite infections in the water vole (*Arvicola terrestris*). *Proceedings of the Royal Society B-biological Sciences*, 276, 1119-1128.
- Olsson M, Madsen T, Nordby J, Wapstra E, Ujvari B, Wittsell H (2003) Major histocompatibility complex and mate choice in sand lizards. *Proceedings of the Royal Society of London Series B-Biological Sciences*, 270, S254-S256.
- Oomen RA, Gillett RM, Kyle CJ (2013) Comparison of 454 pyrosequencing methods for characterizing the major histocompatibility complex of nonmodel species and the advantages of ultra deep coverage. *Molecular Ecology Resources*, 13, 103-116.
- Oppelt C, Starkloff A, Rausch P, Von Holst D, Rödel HG (2010) Major histocompatibility complex variation and age-specific endoparasite load in subadult European rabbits. *Molecular Ecology*, 19, 4155-4167.
- Ovington KS, Alleva LM, Kerr EA (1995) Cytokines and immunological control of *Eimeria spp.* *International journal for parasitology*, 25, 1331-1351.
- Owen JA, Punt J, Stranford SA (2013) *Kuby immunology*. WH Freeman, New York.
- Palphramand KL, White PCL (2007) Badgers, *Meles meles*, discriminate between neighbour, alien and self scent. *Animal Behaviour*, 74, 429-436.
- Parker GA (1970) Sperm competition and its evolutionary consequences in insects. *Biological Reviews of the Cambridge Philosophical Society*, 45, 525-567.

- Parker GA (1998) Sperm competition and the evolution of ejaculates: towards a theory base. In: *Sperm competition and sexual selection* (ed. Birkhead TR, Møller AP), pp. 3–54. London Academic Press, London.
- Paterson S, Wilson K, Pemberton JM (1998) Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population (*Ovis aries* L.). *Proceedings of the National Academy of Sciences of the United States of America*, 95, 3714-3719.
- Pauls SU, Nowak C, Bálint M, Pfenninger M (2013) The impact of global climate change on genetic diversity within populations and species. *Molecular Ecology*, 22, 925-946.
- Penn DJ, Potts WK (1998) How do major histocompatibility complex genes influence odor and mating preferences? *Advances in Immunology*, 69, 411-436.
- Penn DJ, Potts WK (1999) The evolution of mating preferences and major histocompatibility complex genes. *American Naturalist*, 153, 145-164.
- Penn DJ (2002) The scent of genetic compatibility: sexual selection and the major histocompatibility complex. *Ethology*, 108, 1-21.
- Penn DJ, Damjanovich K, Potts WK (2002) MHC heterozygosity confers a selective advantage against multiple-strain infections. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 11260-11264.
- Pereckienė A, Kaziūnaite V, Vyšniauskas A, Petkevičius S, Malakauskas A, Šarkūnas M, Taylor M (2007) A comparison of modifications of the McMaster method for the enumeration of *Ascaris suum* eggs in pig faecal samples. *Veterinary Parasitology*, 149, 111-116.
- Perrin C, Coulon J, Le Berre M (1993a) Social behavior of Alpine marmots (*Marmota marmota*): seasonal, group, and individual variability. *Canadian Journal of Zoology*, 71, 1945-1953.
- Perrin C, Allainé D, Le Berre M (1993b) Socio-spatial organization and activity distribution of the Alpine marmot *Marmota marmota*: preliminary Results. *Ethology*, 93, 21-30.
- Perrin N, Mazalov V (2000) Local competition, inbreeding, and the evolution of sex-biased dispersal. *The American Naturalist*, 155, 116-127.
- Pespeni MH, Garfield DA, Manier MK, Palumbi SR (2012) Genome-wide polymorphisms show unexpected targets of natural selection. *Proceedings of the Royal Society B: Biological Sciences*, 279, 1412-1420.
- Petney TN, Andrews RH (1998) Multiparasite communities in animals and humans: frequency, structure and pathogenic significance. *International Journal for Parasitology*, 28, 377-393.
- Pettorelli N, Vik JO, Myrsterud A, Gaillard J-M, Tucker CJ, Stenseth NC (2005) Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends in Ecology & Evolution*, 20, 503-510.
- Piertney S, Oliver M (2006) The evolutionary ecology of the major histocompatibility complex. *Heredity*, 96, 7-21.
- Pitman RS, Blumberg RS (2000) First line of defense: the role of the intestinal epithelium as an active component of the mucosal immune system. *Journal of Gastroenterology*, 35, 805-814.
- Pittaway CE, Lawson AL, Coles GC, Wilson AD (2014) Systemic and mucosal IgE antibody responses of horses to infection with *Anoplocephala perfoliata*. *Veterinary Parasitology*, 199, 32-41.
- Pizzari T, Parker GA (2009) Sperm competition and sperm phenotype. In: *Sperm*

- biology: an evolutionary perspective* (ed. Birkhead TR, Hosken DJ, Pitnick S). London Academic Press, London.
- Pol MM (1981) Ara mateix. In: L'àmbit de tots els àmbits. Llibres del Mall, Barcelona.
- Pompanon F, Samadi S (2013) Chapitre VIII: Caractériser la diversité du vivant. In: *Les cahiers prospectives Génomique environnementale (INEE-CNRS)*. pp. 43-52.
- Potts W, Wakeland E (1990) Evolution of diversity at the major histocompatibility complex. *Trends in Ecology & Evolution*, 5, 181-187.
- Preleuthner M, Pinsker W (1993) Depauperated gene pools in *Marmota m. marmota* are caused by an ancient bottleneck: electrophoretic analysis of wild populations from Austria and Switzerland. *Acta Theriologica*, 38, 121-139.
- Preleuthner M, Calderola S, Lanfranchi P, Prosl H (1999) Parasiten des Alpenmurmeltieres (*Marmota m. marmota*): Systematik, Entwicklung, Verbreitung. *Stapfia*, 146, 77-92.
- Promerová M, Babik W, Bryja J, Albrecht T, Stuglik M, Radwan J (2012) Evaluation of two approaches to genotyping major histocompatibility complex class I in a passerine - CE-SSCP and 454 pyrosequencing. *Molecular Ecology Resources*, 12, 285-292.
- Puurtinen M, Hytönen M, Knott KE, Taskinen J, Nissinen K, Kaitala V (2004) The effects of mating system and genetic variability on susceptibility to trematode parasites in a freshwater snail, *Lymnaea stagnalis*. *Evolution*, 58, 2747-2753.
- Queller DC, Goodnight KF (1989) Estimating relatedness using genetic-markers. *Evolution*, 43, 258-275.
- R Development Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing Vienne, Austria.
- Råberg L, Sim D, Read AF (2007) Disentangling genetic variation for resistance and tolerance to infectious diseases in animals. *Science*, 318, 812-814.
- Råberg L, Graham AL, Read AF (2009) Decomposing health: tolerance and resistance to parasites in animals. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 364, 37-49.
- Radwan J, Kawalko A, Wojcik J, Babik W (2007) MHC-DRB3 variation in a free-living population of the European bison, *Bison bonasus*. *Molecular Ecology*, 16, 531-540.
- Radwan J, Tkacz A, Kloch A (2008) MHC and preferences for male odour in the bank vole. *Ethology*, 114, 827-833.
- Radwan J, Biedrzycka A, Babik W (2010) Does reduced MHC diversity decrease viability of vertebrate populations? *Biological Conservation*, 143, 537-544.
- Radwan J, Zagalska-Neubauer M, Cichon M, Sendek J, Kulma K, Gustafsson L, Babik W (2012) MHC diversity, malaria and lifetime reproductive success in collared flycatchers. *Molecular Ecology*, 21, 2469-2479.
- Ralls K (1971) Mammalian scent marking. *Science*, 171, 443-449.
- Rassmann K, Arnold W, Tautz D (1994) Low genetic-variability in a natural Alpine marmot population (*Marmota marmota*, Sciuridae) revealed by dna-fingerprinting. *Molecular Ecology*, 3, 347-353.
- Rausch R, Bridgens J (1989) Structure and function of sudoriferous facial glands in Nearctic marmots, *Marmota spp.* (Rodentia: Sciuridae). *Zoologischer Anzeiger*, 223, 265-282.
- Raynaud J, Dobson SF (2011) Scent communication by female Columbian ground squirrels, *Urocitellus columbianus*. *Behavioral Ecology and Sociobiology*, 65, 351-358.



- Read AF, Graham AL, Råberg L (2008) Animal defenses against infectious agents: is damage control more important than pathogen control. *PLoS Biology*, 6, e1000004.
- Reche P, Reinherz E (2003) Sequence variability analysis of human class I and class II MHC molecules: Functional and structural correlates of amino acid polymorphisms. *Journal of Molecular Biology*, 331, 623-641.
- Reeve HK, Emlen ST, Keller L (1998) Reproductive sharing in animal societies: reproductive incentives or incomplete control by dominant breeders? *Behavioral Ecology*, 9, 267-278.
- Restif O, Koella JC (2004) Concurrent evolution of resistance and tolerance to pathogens. *The American Naturalist*, 164, E90-E102.
- Rézouki C, Tafani M, Cohas A, Gaillard J-M, Allainé D, Bonenfant C (*in prep*). Temporal trends and ecological correlates of survival in the Alpine marmot with climate change.
- Richardson DS, Komdeur J, Burke T, Von Schantz T (2005) MHC-based patterns of social and extra-pair mate choice in the Seychelles warbler. *Proceedings of the Royal Society B: Biological Sciences*, 272, 759-767.
- Roche PA, Furuta K (2015) The ins and outs of MHC class II-mediated antigen processing and presentation. *Nature Reviews Immunology*, 15, 203-216.
- Rodríguez R, Ramírez O, Valdiosera CE, García N, Alda F, Madurell-Malapeira J, Marmi J, Doadrio I, Willerslev E, Götherström A et al. (2011) 50,000 years of genetic uniformity in the critically endangered Iberian lynx. *Molecular Ecology*, 20, 3785-3795.
- Rosell F, Bjørkøyli T (2002) A test of the dear enemy phenomenon in the Eurasian beaver. *Animal Behaviour*, 63, 1073-1078.
- Rudatis A, De Battisti R (2005) Alimentazione di *Marmota marmota* in praterie altimontane delle dolomiti bellunesi. *Hystrix, the Italian Journal of Mammalogy*, 16, 135-142.
- Sabatier B (1989) Les parasites de la marmotte Alpine: étude dans les Alpes françaises et synthèse bibliographique. PhD Thesis, Ecole Nationale Vétérinaire de Lyon, Marcy-l'Étoile, France.
- Sauermann U, Nurnberg P, Bercovitch FB, Berard JD, Trefilov A, Widdig A, Kessler M, Schmidtke J, Krawczak M (2001) Increased reproductive success of MHC class II heterozygous males among free-ranging rhesus macaques. *Human Genetics*, 108, 249-254.
- Savage AE, Zamudio KR (2011) MHC genotypes associate with resistance to a frog-killing fungus. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 16705-16710.
- Schalk G, Forbes MR (1997) Male biases in parasitism of mammals: effects of study type, host age, and parasite taxon. *Oikos*, 78, 67-74.
- Schito ML, Chobotar B, Barta JR (1998) Cellular dynamics and cytokine responses in BALB/c mice infected with *Eimeria papillata* during primary and secondary infections. *The Journal of Parasitology*, 84, 328-337.
- Schlupf NA, Schule R, Klimpe S, Karle KN, Synofzik M, Schicks J, Riess O, Schols L, Bauer P (2011) Amplicon-based high-throughput pooled sequencing identifies mutations in CYP7B1 and SPG7 in sporadic spastic paraplegia patients. *Clinical Genetics*, 80, 148-160.
- Schmid-Hempel P (2011) *Evolutionary parasitology: the integrated study of infections, immunology, ecology, and genetics*. Oxford University Press, Oxford.

- Schulenburg H, Kurtz J, Moret Y, Siva-Jothy MT (2009) Introduction. *Ecological immunology*. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 364, 3-14.
- Schwensow N, Fietz J, Dausmann K, Sommer S (2008a) MHC-associated mating strategies and the importance of overall genetic diversity in an obligate pair-living primate. *Evolutionary Ecology*, 22, 617-636.
- Schwensow N, Eberle M, Sommer S (2008b) Compatibility counts: MHC-associated mate choice in a wild promiscuous primate. *Proceedings of the Royal Society B-biological Sciences*, 275, 555-564.
- Schwensow N, Eberle M, Sommer S (2010) Are there ubiquitous parasite-driven major histocompatibility complex selection mechanisms in gray mouse lemurs? *International Journal of Primatology*, 31, 519-537.
- Sepil I, Moghadam HK, Huchard E, Sheldon BC (2012) Characterization and 454 pyrosequencing of major histocompatibility complex class I genes in the great tit reveal complexity in a passerine system. *BMC Evolutionary Biology*, 12, 68.
- Sepil I, Lachish S, Sheldon BC (2013) Mhc-linked survival and lifetime reproductive success in a wild population of great tits. *Molecular Ecology*, 22, 384-396.
- Setchell JM, Charpentier MJE, Abbott KM, Wickings EJ, Knapp LA (2010) Opposites attract: MHC-associated mate choice in a polygynous primate. *Journal of Evolutionary Biology*, 23, 136-148.
- Setchell JM, Huchard E (2010) The hidden benefits of sex: evidence for MHC-associated mate choice in primate societies. *Bioessays*, 32, 940-948.
- Setchell JM, Vaglio S, Abbott KM, Moggi-Cecchi J, Boscaro F, Pieraccini G, Knapp LA (2011) Odour signals major histocompatibility complex genotype in an Old World monkey. *Proceedings of the Royal Society B-biological Sciences*, 278, 274-280.
- Shelley EL, Blumstein DT (2005) The evolution of vocal alarm communication in rodents. *Behavioral Ecology*, 16, 169-177.
- Shi M, Huther S, Burkhardt E, Zahner H (2000) Immunity in rats against *Eimeria separata*: oocyst excretion, effects on endogenous stages and local tissue response after primary and challenge infections. *Parasitology Research*, 86, 891-898.
- Shokralla S, Spall JL, Gibson JF, Hajibabaei M (2012) Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology*, 21, 1794-1805.
- Siddle HV, Kreiss A, Eldridge MD, Noonan E, Clarke CJ, Pyecroft S, Woods GM, Belov K (2007) Transmission of a fatal clonal tumor by biting occurs due to depleted MHC diversity in a threatened carnivorous marsupial. *Proceedings of the National Academy of Sciences*, 104, 16221-16226.
- Siddle HV, Marzec J, Cheng Y, Jones M, Belov K (2010) MHC gene copy number variation in Tasmanian devils: implications for the spread of a contagious cancer. *Proceedings of the Royal Society B: Biological Sciences*, 277, 2001-2006.
- Simberloff D (1988) The contribution of population and community biology to conservation science. *Annual review of Ecology and Systematics*, 19, 473-511.
- Simitch T, Petrovitch Z (1953) Intestinal parasites of the *Citellus citellus* of Yugoslavia during hibernation and activity. *Annales de Parasitologie Humaine et Comparée*, 28, 29-32.
- Sin YW, Annavi G, Dugdale HL, Newman C, Burke T, MacDonald DW (2014) Pathogen burden, co-infection and major histocompatibility complex variability in the European badger (*Meles meles*). *Molecular Ecology*, 23, 5072-5088.
- Slade R, McCallum H (1992) Overdominant vs frequency-dependent selection at MHC

- loci. *Genetics*, 132, 861-862.
- Slate J, Pemberton J (2002) Comparing molecular measures for detecting inbreeding depression. *Journal of Evolutionary Biology*, 15, 20-31.
- Smith AL, Hesketh P, Archer A, Shirley MW (2002) Antigenic diversity in *Eimeria maxima* and the influence of host genetics and immunization schedule on cross-protective immunity. *Infection and Immunity*, 70, 2472-2479.
- Snook RR (2005) Sperm in competition: not playing by the numbers. *Trends in Ecology & Evolution*, 20, 46-53.
- Sol D, Jovani R, Torres J (2003) Parasite mediated mortality and host immune response explain age-related differences in blood parasitism in birds. *Oecologia*, 135, 542-547.
- Sommer S (2005) The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Frontiers in Zoology*, 2, 16.
- Sommer S, Courtiol A, Mazzoni CJ (2013) MHC genotyping of non-model organisms using next-generation sequencing: a new methodology to deal with artefacts and allelic dropout. *BMC Genomics*, 14, 542.
- Sorci G (2013) Immunity, resistance and tolerance in bird-parasite interactions. *Parasite Immunology*, 35, 350-361.
- Spurgin LG, Richardson DS (2010) How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings of the Royal Society B-Biological Sciences*, 277, 979-988.
- Starr TK, Jameson SC, Hogquist KA (2003) Positive and negative selection of T cells. *Annual Review of Immunology*, 21, 139-176.
- Steiniger B, Falk P, Lohmüller M, Van der Meide PH (1989) Class II MHC antigens in the rat digestive system. Normal distribution and induced expression after interferon-gamma treatment in vivo. *Immunology*, 68, 507.
- Stephens PA, Frey-Roos F, Arnold W, Sutherland WJ (2002) Model complexity and population predictions. The Alpine marmot as a case study. *Journal of Animal Ecology*, 71, 343-361.
- Steppan SJ, Kenagy GJ, Zawadzki C, Robles R, Lyapunova EA, Hoffmann RS (2011) Molecular data resolve placement of the Olympic marmot and estimate dates of trans-Beringian interchange. *Journal of Mammalogy*, 92, 1028-1037.
- Stoddard PK, Beecher MD, Horning CL, Willis MS (1990) Strong neighbor-stranger discrimination in song sparrows. *Condor*, 92, 1051-1056.
- Stutz WE, Bolnick DI (2014) Stepwise threshold clustering: a new method for genotyping MHC loci using next-generation sequencing technology. *PLoS ONE*, 9, e100587.
- Sutton JT, Nakagawa S, Robertson BC, Jamieson IG (2011) Disentangling the roles of natural selection and genetic drift in shaping variation at MHC immunity genes. *Molecular Ecology*, 20, 4408-4420.
- Sutton JT, Robertson B C, Grueber CE, Stanton JAL, Jamieson IG (2013) Characterization of MHC class II B polymorphism in bottlenecked New Zealand saddlebacks reveals low levels of genetic diversity. *Immunogenetics*, 65, 619-633.
- Szulkin M, Bierne N, David P (2010) Heterozygosity-fitness correlations: a time for reappraisal. *Evolution*, 64, 1202-1217.
- Tafari M, Cohas A, Bonenfant C, Gaillard J-M, Lardy S, Allainé D (2013a) Sex-specific senescence in body mass of a monogamous and monomorphic mammal: the case of Alpine marmots. *Oecologia*, 172, 427-436.
- Tafari M, Cohas A, Bonenfant C, Gaillard J-M, Allainé D (2013b) Decreasing litter size



- of marmots over time: a life history response to climate change? *Ecology*, 94, 580-586.
- Takahata N, Nei M (1990) Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics*, 124, 967-978.
- Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, 30, 2725-2729.
- Temeles EJ (1994) The role of neighbors in territorial systems - when are they dear enemies. *Animal Behaviour*, 47, 339-350.
- Terrazas CA, Rodríguez-Sosa M, Terrazas LI (2012) Cestoda: tapeworm infection. In: *Immunity to parasitic infection* (ed. Lamb TJ), John Wiley & Sons, Ltd.
- Thomas DM, Stanton NL, Seville RS (1995) A stable eimerian assemblage in Wyoming ground squirrels (*Spermophilus elegans elegans*): maintaining viability over winter. *Journal of the Helminthological Society of Washington*, 62, 1-5.
- Thomas F, Renaud F, Guégan J-F, (2007) *Écologie et évolution des systèmes parasitaires*. De Boeck Université, Bruxelles, Belgium.
- Thornhill NW (1993) *The natural history of inbreeding and outbreeding: theoretical and empirical perspectives*. University of Chicago Press, Chicago, USA.
- Townsend AK, Clark AB, McGowan KJ, Buckles EL, Miller AD, Lovette IJ (2009) Disease-mediated inbreeding depression in a large, open population of cooperative crows. *Proceedings of the Royal Society B-biological Sciences*, rspb-2008.
- Tregenza T, Wedell N (2000) Genetic compatibility, mate choice and patterns of parentage: Invited review. *Molecular Ecology*, 9, 1013-1027.
- Trivers RL (1972) Parental investment and sexual selection. In: *Sexual selection and the descent of man* (ed. Campbell B). Aldine, Chicago.
- Turner AK, Begon M, Jackson JA, Paterson S (2012) Evidence for selection at cytokine loci in a natural population of field voles (*Microtus agrestis*). *Molecular Ecology*, 21, 1632-1646.
- Väli Ü, Einarsson A, Waits L, Ellegren H (2008) To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? *Molecular Ecology*, 17, 3808-3817.
- Vadlejch J, Petrtl M, Zaichenko I, Čadková Z, Jankovská I, Langrová I, Moravec M (2011) Which McMaster egg counting technique is the most reliable? *Parasitology Research*, 109, 1387-1394.
- Valdiosera CE, Garcia-Garitagoitia JL, Garcia N, Doadrio I, Thomas MG, Hänni C, Arsuaga J-L, Barnes I, Hofreiter M, Orlando L et al. (2008) Surprising migration and population size dynamics in ancient Iberian brown bears (*Ursus arctos*). *Proceedings of the National Academy of Sciences*, 105, 5123-5128.
- Van Valen L (1973) A new evolutionary law. *Evolutionary Theory*, 1, 1-30.
- Van Vuren D, Armitage KB (1991) Duration of snow cover and its influence on life-history variation in Yellow-bellied marmots. *Canadian Journal of Zoology - Revue Canadienne de Zoologie*, 69, 1755-1758.
- Vasemägi A, Primmer CR (2005) Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Molecular Ecology*, 14, 3623-3642.
- Venables WN, Ripley BD (2002) *Modern Applied Statistics with S*. Springer, New York.

- Verhoeven KJ, Simonsen KL, McIntyre LM (2005) Implementing false discovery rate control: increasing your power. *Oikos*, 108, 643-647.
- Vinkler M, Albrecht T (2009) The question waiting to be asked: innate immunity receptors in the perspective of zoological research. *Folia Zoologica*, 58, 15-28.
- Wang A, McKay DM (2005) Immune modulation by a high molecular weight fraction from the rat tapeworm *Hymenolepis diminuta*. *Parasitology*, 130, 575-585.
- Warnes G (2012) genetics: Population genetics. R package version 1.3.8.1.
- Weatherall DJ (2003) Genomics and global health: time for a reappraisal. *Science*, 302, 597-599.
- Wedekind C, Seebeck T, Bettens F, Paepke AJ (1995) Mhc-dependent mate preferences in humans. *Proceedings of the Royal Society B-Biological Sciences*, 260, 245-249.
- Wegner KM, Kalbe M, Kurtz J, Reusch TBH, Milinski M (2003a) Parasite selection for immunogenetic optimality. *Science*, 301, 1343-1343.
- Wegner KM, Reusch TBH, Kalbe M (2003b) Multiple parasites are driving major histocompatibility complex polymorphism in the wild. *Journal of Evolutionary Biology*, 16, 224-232.
- Westerdahl H (2004) No evidence of an MHC-based female mating preference in great reed warblers. *Molecular Ecology*, 13, 2465-2470.
- Wilber PG, Duszynski D, Upton S, Seville R, Corliss J (1998) A revision of the taxonomy and nomenclature of the *Eimeria spp.* (Apicomplexa: Eimeriidae) from rodents in the Tribe Marmotini (Sciuridae). *Systematic Parasitology*, 39, 113-135.
- Williams AE (2011). *Immunology: mucosal and body surface defences*. John Wiley & Sons, Oxford.
- Windsor DA (1998) Controversies in parasitology, most of the species on Earth are parasites. *International Journal for Parasitology*, 28, 1939-1941.
- Winternitz JC, Minchey SG, Garamszegi LZ, Huang S, Stephens PR, Altizer S (2013) Sexual selection explains more functional variation in the mammalian major histocompatibility complex than parasitism. *Proceedings of the Royal Society B-biological Sciences*, 280, 20131605.
- Winternitz J, Promerova M, Polakova R, Vinker M, Schnitzer J, Munclinger P, Babik W, Radwan J, Bryja J, Albrecht T (2015) Effects of heterozygosity and MHC diversity on patterns of extra-pair paternity in the socially monogamous scarlet rosefinch. *Behavioral Ecology and Sociobiology*, 65, 459-469.
- Worley K, Collet J, Spurgin LG, Cornwallis C, Pizzari T, Richardson DS (2010) MHC heterozygosity and survival in red junglefowl. *Molecular Ecology*, 19, 3064-3075.
- Wu J, Saupe S, Glass N (1998) Evidence for balancing selection operating at the het-c heterokaryon incompatibility locus in a group of filamentous fungi. *Proceedings of the National Academy of Sciences of the United States of America*, 95, 12398-12403.
- Yamazaki K, Boyse EA, Mike V, Thaler HT, Mathieson BJ, Abbott J, Boyse J, Zayas ZA, Thomas L (1976) Control of mating preferences in mice by genes in major histocompatibility complex. *Journal of Experimental Medicine*, 144, 1324-1335.
- Yamazaki K, Beauchamp GK (2007). Genetic basis for MHC-dependent mate choice. In: *Genetics of sexual differentiation and sexually dimorphic behaviors* (ed. Yamamoto D). Academic press, Oxford.
- Yang D, Lu M, Hao L, Roggendorf M (2000) Molecular cloning and characterization of major histocompatibility complex class I cDNAs from woodchuck (*Marmota monax*). *Tissue Antigens*, 55, 548-557.

## References

- Yeager M, Hughes AL (1999) Evolution of the mammalian MHC: natural selection, recombination, and convergent evolution. *Immunological Reviews*, 167, 45-58.
- Yun C, Lillehoj H, Lillehoj E (2000) Intestinal immune responses to coccidiosis. *Developmental & Comparative Immunology*, 24, 303-324.
- Zagalska-Neubauer M, Babik W, Stuglik M, Gustafsson L, Cichon M, Radwan J (2010) 454 sequencing reveals extreme complexity of the class II Major Histocompatibility Complex in the collared flycatcher. *BMC Evolutionary Biology*, 10, 395.
- Zeh JA, Zeh DW (1996) The evolution of polyandry I: intragenomic conflict and genetic incompatibility. *Proceedings of the Royal Society B-Biological Sciences*, 263, 1711-1717.
- Zeh JA, Zeh DW (1997) The evolution of polyandry II: post-copulatory defences against genetic incompatibility. *Proceedings of the Royal Society B-Biological Sciences*, 264, 69-75.
- Zhang J, Soini H, Bruce K, Wiesler D, Woodley S, Baum M, Novotny M (2005) Putative chemosignals of the ferret (*Mustela furo*) associated with individual and gender recognition. *Chemical Senses*, 30, 727-737.
- Zhang L, Liu R, Ma L, Wang Y, Pan B, Cai J, Wang M (2012) *Eimeria tenella*: expression profiling of toll-like receptors and associated cytokines in the cecum of infected day-old and three-week old SPF chickens. *Experimental Parasitology*, 130, 442-448.
- Zhou J, Ferencik S, Rebmann V, Yang D, Lu M, Roggendorf M, Grosse-Wilde H (2003) Molecular genetic and biochemical analysis of Woodchuck (*Marmota monax*) MHC class I polymorphism. *Tissue Antigens*, 61, 240-248.
- Ziegler A, Kentenich H, Uchanska-Ziegler B (2005) Female choice and the MHC. *Trends in Immunology*, 26, 496-502.
- Zuur A, Ieno E, Walker N, Saveliev A, Smith G (2009) Mixed effects models and extensions in ecology with R. Springer, New York.

## ACKNOWLEDGEMENTS

To Aurélie Cohas, I would warmly thank you the chance you offered to me and the energy you spend to the Alpine marmot project, without your unending power my thesis would not be what it is. To Dominique Allainé, for accepting coming along with me during this thesis and to provide me with precious advice concerning my work. To Marie-Pierre Callait-Cardinal, you gave me the chance to work with parasites and to learn at your side. To Marie Charpentier, Gabriele Sorci, Elise Huchard and Pierre Joly for taking the time to judge my work and be a member of my thesis jury. To Thierry Lengagne, for all the work we have done together, the adventure is not yet finished. To Marie Charpentier, Frank Cézilly and Sophie Beltran, for your contribution on the *comité de pilotage* meetings. To the “Obra Social Fundació La Caixa” and VetAgro Sup, for their financial support, crucial for my PhD adventure. To the authorities of the Vanoise National Park, for granting us permission to work in La Grande Sassièrre Nature Reserve. To the Alpine marmotworld people, Simon Klein, Gwendal Rannou, Aline Mure, Julia Bentin, Martin Thoris, Camille Amelin, Camille Dewaele, Camille Labarrere, Quentin Febvy, Ben Rey, Cindy Canale, Aude Landré, Olivier Bastianelli, Carole et Denis Favre-Bonvin, the workers of the Golf of Tignes, Manu from the *Réfuge du Col du Lautaret* and all other people involved in the Alpine marmot project since 1990.

From the LBBE... To Vérane, for all special moments shared, at the fieldwork and at the office, but specially for the cold moments waiting for any marmot activity while we dreamed being in a warm beach. To Jeff, because your way of being helped me and provided me a good working atmosphere. To those I shared the office and very special moments: Sophie Lardy, Mathieu Douhard, Marlène Gamelon, Marion Germain, Floriane Plard, Morgane Tidière, Célia Rézouki, Marie-Léa Travert and William Gaudry. To all people I could share beautiful moments in the lab: Christophe Bonenfant, Mélissa Martin, Aitor, Lucille Palazy, Arzhela Hemery, Cécile Gotteland, Eléonor Hellard, Anne Nguyen, Cécile Vanpé, Favrice Hibert, Nicolas Lieury, Laurent Crespín, Marion Roussel, Pascale Chevret, Aurélie Siberchicot, Laurent Jacob, Emilie Delava, PAF, David Fouchet, Jodie Martin, Etienne Rajon, Charlotte Recapet, Christophe Plantamp, Jennifer Morinay, Nicolas Ferry, Thibault Gayet, Thomas Bigot, Sandrine Sauzet, Nathalie Arbasetti, Aline Maitrias, Odile Mulet-Marquis, Hélène Henri, Stéphane Delmotte, Bruno Spataro, Simmon Pennel and Lionel Humblot.

To the Lyon's family... To my neighbour, friend and *gairebé germà* Roger Fusté... for our evenings of conversation, for our bottles of wine, for our splendid meals... To Marion Tafani, for the smile and the joy with which you received me when I arrived in Lyon, smile and joy that have endured over these five years. To Yaelle Rosillo, for your crazy laughs, your Mexican meals and friendship. To Irene, for our chats on current political affairs and for our crazy dances when we listen ska-p. To Pierre Dupont, you rapidly changed your simple *étagère* condition to become a very important piece of furniture ;). To Flo(rent) Pedros and the great moments we spent together, specially our trips either by bike or by *Scarlet*, our visits to the Sant Jean market during these last months... To Fred, for the festivals we have made together and the patience - or not ;) - you've had when I switched to a daily-life style. To Aldo, because your struggle to be a doctor and has been and is a source of inspiration. To Sue, for your enthusiasm, because you make me discover Korean cuisine and because you come to encourage me when I took part in the triathlon. To my cousin Júlia, because in addition to share a part of the family, we share many other things. To Alba Lázaro, for the courage you had coming to

live in Lyon and the great number of *soirées* we spent together in the UAB, Lyon, Sabadell, Granada... To Clara Muñoz and our yoga sessions, very useful at the end of the thesis! I would like to thank many other important people I met during these five years in Lyon: Albulena, Balam, Ben, Gilles, Judith, Pablo, Anna Servat, Marie, Roman, Martin, Alex (Kiki), Damien, Marie-Claude Cohas, Alain Cohas, Nelly and Anan.

To the UAB people... To Bernat Claramunt, because you gave me the opportunity to be trapped into the marmotworld. To Mo, for all the moments of study and work, but especially for our encounters *a la festa major o a menjar croissants planxats*. To Anama, for sharing your knowledge and your good humour. I would like to thank many other people I met in the UAB: Anselm Rodrigo, Andrea, Lore, Joan, Neus, Xevi, Pepo, Mireia Vila, Eli, Alba Guiral, Puri, Pau Sunyer, Boi, Isabelle, Anna Torné, Irene Figueroa... To Josep Fortuny, for our several adventures in Nostra Llar, la UAB, Sabadell, la Roca, Lyon or Aix en Provance (*jus d'orange*). To Lujan and Salva, you made my blood run faster and drive correctly the *crashcauet*, but above all I warmly thank you for being there and make me smile every day whenever we are together.

To my second family from Sabadell... To Laia, for your unconditional support, specially when we started the *cala iaiona* adventure! To Clara, because although we live many kilometres away we have always found the time to do something together frequently. To Marina, for plenty of laughs that allowed me to arrive in France speaking French. To the members of *el palau*, for letting me live in your house: Deu, Luis, Suro, Palou, Ana and Èric. To the *lyonnaises* (Marta, Núria, Laia and Clara), for the unforgettable trips and moments we have passed together. To the rest of the Clotxes (Aina, Alba, Berni, Dedi, Geri, Gisela, Guim, Irene, Laura, Many, Mónica, Palou, Pau, Pere, Raimon i Sans), for all the moments that we have spent together (*caps d'any, calçotades, tionades, space 2cinus, olympigfest, castanyades, vermut, piscinas, Puigverts, futbols, riures, concerts, sopars perquè ens ve de gust, birres...*).

Last but not least, none of what I have done would have been possible without the support of my entire family. To my mother, because among many of the things you taught me, you taught me what a species is just looking the trees from the window of our house and because you pushed me to explore the research world. To my father, you taught me to love nature during our walks *al nostre riu*, our *caçadors de bolets* sessions, our hikes or our bike tours. To both of you I would like to warmly thank your unconditional support and the motivation you have always transmitted to me. To Èric, for accepting to road together despite knowing that we would be far away for a long time. I would warmly thank many thinks but specially your visits in the Alps (the surprise of the 4G!) and your ability to make me see the things in a more optimistic way. To *àvia* Eulàlia, for your big hugs and the *lacasitos* that give me the strength to go back to Lyon. To *avi* Josep, for the hugs that we learn to give us. To, *avi* Ramon, for your optimistic point of view of seeing and living the life. To *àvia* Joaquina, for your delicious meals and for the effort you made to be ready to come in Lyon this July. Finally, to the rest of my family: Miquel, Montse, Anna, Marc i Gemma; Pep, Antònia, Mireia, Ferran i Xavi; Eulàlia, Lluís, Adriana i Guillem; Àngels, Manolo, Marc, David i Anna; Julieta i Nevià; Miose, Marc, Noe, Laura, Pau, Emma, David, Neus i Khay.