The major histocompatibility complex, mate choice and pathogen resistance in the European badger *Meles meles*

A thesis submitted for the degree

Doctor of Philosophy

Yung Wa Sin

Wolfson College, University of Oxford

Trinity Term 2012

Yung Wa Sin **Doctor of Philosophy**

Wolfson College, Oxford **Trinity Term 2012**

The major histocompatibility complex, mate choice and pathogen resistance in the European badger *Meles meles*

Abstract

Studies of the evolution of the major histocompatibility complex (MHC) have been central to the understanding sexual selection and pathogen-mediated selection. The European badger *Meles meles* is well suited for exploring such questions because of its life history characteristics, reproductive biology and mating system. In this thesis, I examined both MHC class I and class II genes. Seven putatively functional sequences were found for class I genes and four for class II DRB genes. Evidence of past balancing selection of both genes was demonstrated by the d_N/d_S ratio, by positive selection at the antigen-binding site (ABS) and by trans-species polymorphism of alleles within other mustelids and carnivores. MHC class I genes also showed evidence of concerted evolution, but domains showed different evolutionary histories. MHC genes may influence microbiota and odour of an individual and influence mating preferences. I examined the bacterial community of the subcaudal gland secretion and demonstrated a high number of bacterial species (56 operational taxonomic units), which cubs exhibited a higher diversity than adults. The microbiota may lead to an individual-specific odour as a cue signaling the MHC genotype of potential mating partners. I report the first evidence for a MHC-based mating preference in carnivores. Female badgers showed a MHC-assortative mate choice towards breeding with males that had functionally similar MHC genes, for MHC class II DRB genes. This applied to neighbouring-group matings. I also found considerable annual fluctuation in the occurrence of MHC-based mate choice. Based on genome-wide background in the same mating randomizations I found no evidence of inbreeding, which indicated that MHC similarity was apparently the actual target of mate choice. In line with MHC-assortative mate choice, MHC heterozygosity had no influence on the co-infection status. Individual MHC alleles did, however, associate with resistance and susceptibility to specific pathogens, suggesting that MHC diversity may be driven and maintained by pathogen-mediated selection through rare-allele advantages and/or fluctuating selection. My study of genetic characteristics, mate choice and pathogen pressures in a wild population revealed past and contemporary evolutionary process of the MHC genes. This increases knowledge of how the MHC may affect mating behaviour and sexual selection, ultimately influencing population processes.

Acknowledgements

Funding for my university fees and maintenance was generously provided by the Croucher Foundation (Hong Kong). My laboratory work was funded by the Natural Environment Research Council (NERC) Biomolecular Analysis Facility (NBAF), Sheffield. The People's Trust for Endangered Species provides long-term funding for the Wytham Badger Project. I am also grateful to the Wolfson College for funding that enabled me to present my work at conferences.

I thank my supervisor David Macdonald for tremendous support, constructive advice and prompt response throughout my doctoral studies. I am extremely grateful to David for providing such a fantastic research group in which to work.

Great thanks to my co-supervisors Chris Newman and Hannah Dugdale for their enthusiasm and tremendous input into this project. They spent much time throughout my studies from giving me valuable advice at the beginning of my DPhil to constructive commenting on my thesis. I cannot imagine without their support.

Thanks also go to Terry Burke for his constructive support and advice throughout the project, from the days when we were planning our application to laboratory work in the NBAF-Sheffield. The NBAF-Sheffield is an extremely welcoming, helpful and friendly place and I am very grateful to Terry for providing me such an opportunity to work there.

I would like to thank everyone for their assistance with laboratory work at the NBAF-Sheffield and for the joy working with them, in particular Laura Brettell, Debs Dawson, Alain Frantz, Juan Galindo, Emma Gregson, Gav Horsburgh, Camille Jan, Claudius Kerth, Kang-Wook Kim, Andy Krupa, Maria-Elena Mannarelli, Alex Mestre, Celine Pagnier, Maria Jose Rivas Quelle and David Wright. Andy Krupa always provides prompt assistance. I thank Maria-Elena Mannarelli for helping me with the RSCA, Emma Gregson for helping me with the real-time qPCR, and Cindy Smith and Maggi Killion for teaching me T-RFLP.

I thank the Wytham Woods badger team, past and present members, for running this valuable study for many years (now in its 25th year). Special thanks to Geetha Annavi, Kirstin Bilham, Christina Buesching, Steven Ellwood, Yayoi Kaneko, Chris Newman, Pierre Nouvellet, Youbing Zhou and volunteers for assistance with badger trapping. Particular thanks to Christina for organizing the scent project. I thank Paul Johnson's advice on analysis, Inigo Montes's teaching of LCC measurement, and Dawn Burnham to ensure that things ran smoothly. I'd also like to thank Christos Astaras, Joanna Bagniewska, Magdalena Bennett, Femke Broekhuis, Merryl Gelling, Kerry Kilshaw, Lucy Tallents, and everyone at WildCRU.

Most importantly, sincere thanks to my family and friends, both in Britain and in Hong Kong. Thanks Emily for being with me and supporting me along the way.

Table of contents

Abstract II

Acknowledgements III

Table of contents V

List of tables VIII

List of figures IX

CHAPTER 1

Introduction 1

1.1 Sexual selection and immunity 2

1.2 The Major Histocompatibility Complex: An Overview 3

1.3 MHC and mate choice 6

1.4 MHC and pathogen resistance 10

1.5 Other MHC evolution hypotheses 13

1.6 European badgers: A review of mating systems and disease susceptibility 14

1.7 Thesis aims and organization 18

CHAPTER 2

MHC class II genes in the European badger (*Meles meles*): Characterization, patterns of variation and transcription analysis 20

2.1 Abstract 21

2.2 Introduction 22

2.3 Materials and methods 24

2.4 Results 31

2.5 Discussion 37

CHAPTER 3

Evolution of MHC class I genes in the European badger (*Meles meles*) 46

3.1 Abstract 47

3.2 Introduction 48

3.3 Materials and methods 51

3.4 Results 56

3.5 Discussion 67

CHAPTER 4

Molecular characterization of the microbial communities in the subcaudal gland secretion of the European badger 75

4.1 Abstract 76

4.2 Introduction 77

4.3 Materials and methods 79

4.4 Results 84

4.5 Discussion 90

CHAPTER 5

MHC-based mate choice: Female preference for MHC-similar male European badgers (*Meles meles*) 98

5.1 Abstract 99

5.2 Introduction 100

5.3 Materials and methods 102

5.4 Results 109

5.5 Discussion 125

CHAPTER 6

Pathogen burden, co-infection and major histocompatibility complex variability in the European badger (*Meles meles*) 129

6.1 Abstract 130

6.2 Introduction 131

6.3 Materials and methods 134

6.4 Results 140

6.5 Discussion 146

CHAPTER 7

Discussion 166

7.1 Abstract 167

7.2 Mate choice and immunity 168

7.3 MHC diversity and evolution 169

7.4 Subcaudal gland secretion: Bacterial community and scent profile 171

7.5 Mate choice based on the MHC 172

7.6 MHC and pathogen burden 175

7.7 Conclusions 177

7.8 Future research 178

References 180

Appendix 1

Evidence for a Role of the Host-Specific Flea (*Paraceras melis*) in the Transmission of *Trypanosoma* (*Megatrypanum*) *pestanai* to the European Badger 219

Appendix 2

The effects of pouch bacteria on the chemical composition of subcaudal gland secretions in the European badger *Meles meles*: Maintaining the balance between shared group-odour and individual-specific information in scent-marks 228

Appendix 3

Socio-ecological correlates of extra-group paternity in the European badger (*Meles meles*) 248

Appendix 4

Choosing mates for offspring genetic quality: Extra-group paternity does not enhance fitness in the European badger (*Meles meles*) 279

List of tables

Table 2.1 MHC class II-specific primers for European badgers 26

Table 2.2 Sequence polymorphism of MHC class II genes delineated by exon 32

Table 2.3 Polymorphisms of four MHC class II molecules according to their domains 33

Table 2.4 Rates (± standard error) of non-synonymous (d_N) and synonymous (d_S) substitutions for antigen-binding site (ABS), non-ABS, and combined (ABS + non-ABS) at the European badger *Meles meles* MHC loci 34

Table 2.5 CODEML detection of amino acid sites under positive selection 35

Table 2.6 Presence of MHC class II sequences in genomic DNA (gDNA) and complementary DNA (cDNA) from seven European badgers *Meles meles* 35

Table 3.1 MHC class I-specific primers for *M. meles* 53

Table 3.2 Presence of MHC class I sequences in genomic DNA (gDNA) and complementary DNA (cDNA) from 11 *M. meles* 59

Table 3.3 Sequence polymorphism of MHC class I genes/molecules delineated by exon/domain 60

Table 3.4 Positively selected sites (PSS), parameter estimates, log-likelihood values of different models of codon evolution and summary of test statistics for the likelihood-ratio test (LRT) of the α1 and α2-domain of MHC class I genes in *M. meles* 65

Table 4.1 Details of analyzed *M. meles* subcaudal gland secretion samples 80

Table 4.2 DFA and ANOSIM comparing T-RFLP profiles of different categories (age, sex, body condition, reproductive status and season of sampling) of *M. meles* subcaudal gland secretion samples 88

Table 4.3 The percentage contribution of the top 10 TRFs to the overall difference between cubs and adults 91

Table 5.1 Overall significance of MHC-based mating and inbreeding avoidance across six years (1993, 2004, 2005, 2008, 2009 and 2010), calculated using Fisher's method of combining probabilities 111

Table 5.2 The association between MHC class II DRB amino acid distance (all amino acid sites) and DRB allele sharing, Q&G's index of pairwise relatedness, within/neighbouring-group pairing and year 112

Table 5.3 The association between MHC class I gene amino acid distance (all amino acid sites) and MHC class I allele sharing, Q&G's index of pairwise relatedness, within/neighbouring-group pairing and year 112

List of figures

Figure 2.1 Amino acid sequence identity for the European badger *Meles meles* class II MHC clones, mustelid, phocine, and canine DRβ sequences 28

Figure 2.2 Amino acid sequence identity for the European badger class II MHC clones, mustelid, phocine, canine, and human DQβ sequences 29

Figure 2.3 Amino acid sequence identity for the European badger class II MHC clones, mustelid, phocine, canine, and human DQα sequences 29

Figure 2.4 Amino acid sequence identity for the European badger class II MHC clones, mustelid, phocine, canine, and human DRα sequences 30

Figure 2.5 Phylogenetic tree of MHC class II DRB and DQB exon 2 sequences from *Meles meles*, *Taxidea taxus*, and other mammal species including mustelid, phocine, canine, equine, and human 37

Figure 2.6 Phylogenetic tree of MHC class II DQA and DRA exon 2 sequences from *Meles meles*, *Taxidea taxus*, and other mammal species including mustelid, phocine, canine, equine, and human 38

Figure 3.1 Schematic representation of the positions of the primers used for amplification of MHC class I sequences from cDNA/gDNA 53

Figure 3.2 Amino acid sequence identity for class I sequences of *M. meles* and seven other mammals 57

Figure 3.3 Nucleotide sequence identity for the MHC class I intron 2 of *M. meles* clones and *C. lupus familiaris* 59

Figure 3.4 Neighbour-Net of MHC class I (a) exon 2/α1-domain, (b) exon 3/α2-domain, and (c) intron 2 sequences from *M. meles* and other mammalian species 61

Figure 3.5 Rates (± standard error) of non-synonymous (d_N; solid marker) and synonymous (d_S; open marker) substitutions and ratio of d_N to d_S (ω; grey bar) for antigen-binding site (ABS), non-ABS, and combined (ABS + non-ABS) at the *M. meles* MHC class I loci 63

Figure 3.6 Mean ω at each codon position across (a) exon 2 and (b) exon 3 of *M. meles* class I genes 66

Figure 3.7 Phylogenetic tree of MHC class I exon 4/α3-domain sequences from *M. meles* and other mammalian species 67

Figure 4.1 Phylogenetic tree showing different OTUs found in the subcaudal gland secretions of nine *M. meles* based on 16S rRNA gene sequences 85

Figure 4.2 Pie chart showing the relative abundance of different bacterial phyla/classes found in subcaudal gland secretion of *M. meles* adults and cubs 86

Figure 4.3 Discriminant function scores (mean ± SE) from DFA of 79 subcaudal gland secretion samples differentiating T-RFLP profiles of cubs (age 0; n=6), yearlings (age=1; n=12), young adults (age 2-5; n=33) and old adults (age>5; n=17) 89

Figure 4.4 Mean (± SE) difference of (a) Shannon index of diversity of cubs (n=6) and adults (n=73); (b) richness of cubs and adults; (c) Shannon index of diversity of non-breeding (n=14) and breeding females (n=30); and (d) richness of non-breeding and breeding females 90

Figure 5.1 Three MHC class II DRB haplotypes and five MHC class I haplotypes in the study population of European badger (*Meles meles*) 109

Figure 5.2 Mean MHC class I allele sharing, amino acid distance (all amino acid sites and ABS only) and Queller & Goodnight's index of pairwise relatedness (*R*) of assigned parent-pairs, compared with the frequency distribution of mean values generated from 1000 simulations of random pairings between breeding females and all of their within-group or neighbouring-group candidate fathers 113

Figure 5.3 Mean MHC class II DRB allele sharing, amino acid distance (all amino acid sites and ABS only) and Queller & Goodnight's index of pairwise relatedness (*R*) of assigned parent-pairs, compared with the frequency distribution of mean values generated from 1000 simulations of random pairings between breeding females and all of their within-group or neighbouring-group candidate fathers 119

Figure 6.1 Model averaged parameter estimates and their 95% confidence intervals for the eight MHC predictors (MHC class I and II genes) associated with the prevalence of seven pathogens 143

Figure 6.2 Model averaged parameter estimates and their 95% confidence intervals for the eight MHC predictors (MHC class I and II genes) associated with the infection intensity of eight pathogens 144

Figure 6.3 Model averaged parameter estimates and their 95% confidence intervals for the ten MHC predictors (MHC class I and II genes, MHC genes heterozygosity) associated with the co-infection by 13 pathogens 145

Chapter 1

Introduction

The aims of this doctorial research are to investigate the evolution of major histocompatibility complex (MHC) and how it is related to pathogen-mediated selection and sexual selection, in a wild population of European badgers *Meles meles*. This species is especially well suited for exploring such questions because of its life history characteristics, reproductive biology and mating system (Neal & Cheeseman, 1996; Buesching *et al.*, 2002c, b; Buesching *et al.*, 2003; Carpenter *et al.*, 2005; Yamaguchi *et al.*, 2006; Dugdale *et al.*, 2007; Macdonald *et al.*, 2009).

Studies on the relationships between MHC genotype, pathogen burden and mate choice have generally focused on laboratory animal models (Penn & Potts, 1999; Piertney & Oliver, 2006), where the subtleties of mate choice may be obscured in an artificial environment. By contrast, this study considers a population of wild badgers, with large sample sizes, where we observe the status of MHC and mate choice in an undisturbed system.

My thesis aims to:

1) examine the evolution and diversity of MHC genes in *M. meles*,
2) investigate whether *M. meles* select mates on the basis of the MHC, and
3) examine the relationship between MHC constitutions and pathogen burden.

1.1 Sexual selection and immunity

Female mate choice is widely recognized and different mechanisms have been proposed to explain its occurrence (Andersson, 1994; Andersson & Simmons, 2006). When there are no obvious direct benefits associated with mate choice, indirect genetic benefits are frequently inferred (Kokko *et al.*, 2003; Consuegra & Garcia-de-Leaniz, 2008). The overall importance of indirect genetic benefits in mate choice has sparked considerable debate (Jennions & Petrie, 2000; Andersson & Simmons, 2006; Charmantier & Sheldon, 2006). Indirect genetic benefits may be gained through the acquisition of 'good genes' or compatible genes by mate choice, in order to increase offspring fitness (Zeh & Zeh, 1996; Penn & Potts, 1999; Neff & Pitcher, 2005). Genetic compatibility can provide genetic benefits if males vary in their genetic compatibility with females and there is a mechanism by which more

compatible males are likely to sire more offspring, or if these offspring have higher fitness (Tregenza & Wedell, 2000).

Under the Red Queen hypothesis (vanValen, 1973), individuals are expected to choose mates to maximize the adaptive benefit of sexual reproduction, with respect to the evolutionary interaction between pathogens and their hosts. An individual is predicted to select a mate that has genes that, in combination with the primary individual's own genes, provide offspring with the best immune response against the pathogens. The immune system of potential mates can influence adaptive mating decisions, which has led to investigations of mating behavior and consequences for mate choice and fitness. Much of this research has focused on the role of the MHC genes on mate choice (e.g. Potts *et al.*, 1991; vonSchantz *et al.*, 1996; Milinski *et al.*, 2005; Drury, 2010; Dunn *et al.*, 2012).

Compared with other coding genes, the MHC genes generally show extremely high levels of polymorphism (Janeway *et al.*, 2005), maintained through a form of pathogen-driven selection and/or sexual selection (Penn & Potts, 1999; Jeffrey & Bangham, 2000; Birkhead & Pizzari, 2002; Ziegler *et al.*, 2005; Spurgin & Richardson, 2010). This high sequence variation at MHC loci confers individuals the ability to identify and process a large number of pathogenic antigens, and thus enables them to combat a wider range of immune challenges (Hughes & Nei, 1992; Nikolich-Žugich *et al.*, 2004). The highly diverse MHC has been linked to being discriminable odours (Yamazaki *et al.*, 1982; Ziegler *et al.*, 2005), which has led to the proposition that the MHC may provide a general mechanism of individual and kin recognition across vertebrates, in the context of mate choice (Egid & Brown, 1989; Brown & Eklund, 1994; Brennan & Kendrick, 2006). The role of MHC genes is thus key to understanding mate choice and sexual selection (Penn & Potts, 1999).

1.2 The Major Histocompatibility Complex: An Overview

1.2.1 Antigen presentation and immunological mechanisms

The major histocompatibility complex (MHC) is a multi-gene family that occurs in all vertebrates (Klein, 1986; Potts & Wakeland, 1990; Edwards & Hedrick, 1998). It is the most gene-dense region of

the mammalian genome and plays an important role in the immune system and auto-immunity. Cell-surface glycoproteins encoded by MHC genes are involved the recognition and binding of peptides and their presentation to T-cells, to initiate an immune response (Klein, 1986; Brown *et al.*, 1993). Thus, MHC plays a central role in both humoral and cell-mediated immune responses. As a consequence of this antigen presenting ability, the particular set of MHC molecules expressed by an individual influences the specific repertoire of antigens that an individual's MHC molecules can present (Penn & Potts, 1999). During T-cell development in the thymus, T-cell receptors (TCRs) are produced by rearranged genes, then MHC modulated thymic selection shapes this T-cell repertoire (Bevan *et al.*, 1994; Goldrath & Bevan, 1999; Huseby *et al.*, 2005), determining the antigen(s) that T-cells can respond to. The MHC thus partly establishes the response of an individual to the antigens of infectious pathogens, and hence determines susceptibility to disease.

1.2.2 MHC molecules

The MHC molecules activate different specific immunological effectors, such as cytotoxic T-cells, helper T-cells, natural killer cells, macrophage, and antibody-secreting B-cells (Klein, 1986; Janeway & Travers, 1994; Penn & Potts, 1999). The MHC genes are organized into regions encoding two classes, termed class I and class II, of antigen-presenting molecules.

Class I MHC molecules are heterodimers, which comprise a 45-kilodalton (kDa) α chain, associated non-covalently with a 12-kDa β_2-microglobulin molecule (Goldsby *et al.*, 2002). The α chain of class I MHC molecules is organized into three external domains (α1, α2, and α3), each containing approximately 90 amino acids, and is encoded by different exons from a single gene (Hughes & Nei, 1988; Reche & Reinherz, 2003). The α1 and α2 domains interact to form an antigen-binding site (ABS), located on the top surface of the class I MHC molecule, to bind a peptide of 8-10 amino acids. Class I MHC molecules are expressed on the surface of nearly all nucleated cells and their major function is in the presentation of endogenously derived peptides to CD8+ cytotoxic T-cells; thus, they play an essential role in defense against intra-cellular pathogens such as viruses (Hill, 1998).

Class II MHC molecules contain two different polypeptide chains, a 33-kDa α chain and a 28-kDa β chain that associate by non-covalent interaction and are encoded by two separate genes (Brown *et al.*, 1993; Reche & Reinherz, 2003). The membrane-distal portion of a class II molecule is composed of the α1 and β1 domains and forms the ABS for the binding of processed peptides. Class II MHC molecules are expressed primarily on antigen-presenting cells, e.g. macrophages, dendritic cells, and B cells. They present exclusively processed antigenic peptides from extracellular pathogens and parasites, which generally contain 13-18 amino acid residues, to CD4+ T-helper cells (Penn & Potts, 1999; Goldsby *et al.*, 2002).

1.2.3 Genetic diversity and selection

The central role of MHC genes in the immune system and their extremely high diversity (Garrigan & Hedrick, 2003) make them ideal candidates for studying how selection drives and maintains genetic diversity. Most research, to date, has focused on the ABS-encoding exon (e.g. Cheng *et al.*, 2009; deBellocq & Leirs, 2010), which is functionally important and shows high levels of variation, both in the number of alleles and the extent of sequence variation between alleles (Hughes & Yeager, 1998). Many studies have also demonstrated that the ABS displays higher rates of non-synonymous than synonymous substitutions ($d_n > d_s$) (reviewed in Bernatchez & Landry, 2003). Phylogenetic reconstructions also reveal trans-species polymorphism (Klein, 1987; Klein *et al.*, 1998; Klein *et al.*, 2007), where MHC alleles between species are more closely related than alleles within species. The persistence of ancestral allelic diversity could be maintained over long periods of time across speciation events, relative to neutral genetic variation (Richman, 2000). As a consequence, the general view is that balancing selection plays a major role in shaping nucleotide diversity in MHC genes (Bergström & Gyllensten, 1995). Due to the pivotal role of the MHC in the immune system, balancing selection is likely to act through pathogen-mediated selection (Spurgin & Richardson, 2010), which may be optimized by sexual selection (Penn & Potts, 1999).

1.3 MHC and mate choice

1.3.1 Advantages of MHC-dependent mate preference

MHC-dependent mate preference can enhance progeny fitness in a variety of ways, for example increasing MHC heterozygosity, which can potentially combat a wider range of immune challenges; and by avoiding inbreeding between close relatives (Penn & Potts, 1999). These processes maintain high MHC polymorphism and generate the heterozygote excess and linkage disequilibrium characteristic of the MHC in natural populations (Klein, 1986; Hedrick, 1992; Apanius *et al.*, 1997).

Two types of indirect genetic benefits, heterozygotes and rare alleles (Penn & Potts, 1999; Tregenza & Wedell, 2000; Neff & Pitcher, 2005), might be favoured during mate selection. MHC heterozygotes have an advantage over homozygotes because more peptides can be presented to the T-cells (Hughes & Nei, 1992). Rare MHC alleles might confer advantages if the animal is infected by a newly mutated pathogen that is able to evade the most common MHC molecules. These two mechanisms could contribute to the high diversity of MHC (Hughes & Nei, 1992; Penn & Potts, 1999; Neff & Pitcher, 2005), and both would also lead to the evolution of mate preference for MHC heterozygous and MHC dissimilarity (Potts, 2002). There is also empirical evidence shows that there exists an optimal number of MHC alleles at an intermediate level of MHC diversity (Wegner *et al.*, 2003; Woelfing *et al.*, 2009).

Choosing a MHC-dissimilar mating partner has also been suggested as a benefit of inbreeding avoidance. There is evidence that MHC-based cues can aid in the kin recognition (Potts, 2002); for example female house mice prefer sisters as communal nesting partners, which have similar MHC genes (Manning *et al.*, 1992), suggesting a role of MHC in kin recognition, which may help identify cooperative breeders and avoid inbreeding.

The extremely high levels of polymorphism inherent in the MHC are also able to provide the variability necessary for a genetically based recognition system (Grafen, 1990). Since the primary function of the MHC is to distinguish self- from non-self at the cellular level it may link to a

mechanism that also helps to distinguish relatives from non-relatives at the organismal level. Therefore, in addition to the crucial function of MHC molecules in immune response, they may also have a role in behavioural contexts (Keverne, 2004) and reproduction.

Alternative adaptive hypotheses have been proposed to explain how MHC influences reproductive success, which involve pre-copulatory sexual selection (Penn & Potts, 1999) and post-copulatory sexual selection (Wedekind *et al.*, 1996). Also MHC-dependent selective abortion could happen after fertilization (Alberts & Ober, 1993).

1.3.2 Pre-copulatory sexual selection

MHC genes and chemoreceptors (Mombaerts, 2004) are involved in various stages of female mate choice. Pre-copulatory sexual selection has been observed in a variety of different vertebrate species (Bernatchez & Landry, 2003). As the highly diverse MHC is associated with easily discriminable odours (Yamazaki *et al.*, 1982; Ziegler *et al.*, 2005), this has led to the suggestion that the MHC may provide a general mechanism of individual (Brennan & Kendrick, 2006) and kin (Brown & Eklund, 1994) recognition across vertebrates, in the context of mate choice (Egid & Brown, 1989). Another possibility is MHC genes may influence microflora (Singh *et al.*, 1990) that contribute to specific odour generation through product decomposition (Albone & Perry, 1975). This could be used as a direct cue to avoid inbreeding through the selection of particular MHC genotypically derived odour cues, or may provide an indication of infection status.

1.3.3 MHC-dependent mating preference

In mice, MHC-congenic strains prefer to mate with MHC-dissimilar individuals (Yamazaki *et al.*, 1976). Egid & Brown (1989) further established that oestrus female mice prefer the odour of MHC-dissimilar males, and female mice prefer to mate with MHC-dissimilar males. Several studies have in addition found that cross-fostering mice onto MHC dissimilar parents reversed the preference for MHC-dissimilar mates (Beauchamp *et al.*, 1988; Eklund, 1997; Penn & Potts, 1998a), which provides strong evidence for MHC-dependent mating preference and familial imprinting. Using human subjects, the "T-shirt study" (Wedekind *et al.*, 1995; Wedekind & Füri, 1997) exposed that humans

prefer the odour of MHC-dissimilar individuals. But this preference can be reversed by oral contraceptives (Wedekind *et al.*, 1995), which mimic the effect of pregnancy, when it would be better to associate with MHC-similar individuals who are likely to be kin and potential helpers. Gravid female stickleback were not found to prefer male MHC genotyped dissimilar to their own, but the odour of males with more MHC class-IIB alleles (Reusch *et al.*, 2001). The attractiveness of male odour to the female sticklebacks could be predictably modified by MHC peptide ligands (Milinski *et al.*, 2005). In ring-necked pheasants, females prefer to mate with males with long-spurs, which is associated with MHC genotype and viability (vonSchantz *et al.*, 1997).

Studies have also shown that human couples are more likely to be dissimilar in MHC haplotypes than by random mating (Ober, 1995; Ober *et al.*, 1997), MHC-similar couples had longer inter-birth intervals (Ober, 1995), and there was a deficit of MHC homozygotes at birth (Kostyu *et al.*, 1993). The underlying causes of these observations may result from selective abortion, or avoidance of mating with individuals carrying similar MHC haplotypes. In a study of wild-derived mice in a semi-natural enclosure, fewer MHC homozygotes were born than expected from random mating, implicating selective abortion and mating preference (Potts *et al.*, 1991).

Females, in a variety of taxa, will often engage in have extra-pair matings to enhance the heterozygosity and fitness of their offspring (Griffith *et al.*, 2002; Richardson *et al.*, 2005). Females may prefer compatible mates or mates that can provide 'good genes' during extra-pair matings but this must be reconciled against establishing themselves among good economic resources when choosing social partners (Buss, 1989). A study in the Seychelles warbler has shown that the MHC diversity of the extra-pair male was significantly higher than that of the females' cuckolded social mate, suggesting females use extra-pair mating to look for 'good genes' for their offspring while continuing to benefit from established resource quality in their breeding range (Richardson *et al.*, 2005). In badgers, plural mating behaviour and extra-group paternity (Carpenter *et al.*, 2005; Dugdale *et al.*, 2007) are common, the question of whether the high occurrence of this is related to MHC is investigated in this thesis.

1.3.4 Post-copulatory sexual selection

Post-copulatory sexual selection may occur by cryptic female selection, where the success of sperm in fertilizing the oocyte could be affected by the female's physiology (Eberhard, 1996; Birkhead & Pizzari, 2002). As the energy invested in embryo development is substantial, it makes evolutionary sense that females may have evolved a physiological mechanism that screens sperm for suitable genetic contribution before fertilization. The sperm receptor selection (SRS) hypothesis (Ziegler *et al.*, 2002) proposes a possible mechanism for the females to select between spermatozoa. The polygynandrous mating system (Carpenter *et al.*, 2005; Dugdale *et al.*, 2007) and reproductive biology (Cresswell *et al.*, 1992; Neal & Cheeseman, 1996; Yamaguchi *et al.*, 2006) of badgers should promote sperm competition and provide a good basis for females to select between sperm.

In addition to the selection of specific sperm during fertilization (Swanson & Vacquier, 2002), females may also select sperm MHC haplotypes after fertilization. This was indicated by an in vitro fertilization experiment in mice (Wedekind *et al.*, 1996) that showed the MHC type of the fertilizing spermatozoon would affect the second meiotic division of the oocyte, which in-turn leads to non-random MHC combination in the embryos. Furthermore Rülicke *et al.* (1998) showed that hepatitis virus-infected mice produced more MHC-heterozygous embryos, indicating parental infection could influence the outcome of fertilization and degree of MHC heterozygosity in the offspring. In badgers the oestrous cycle continue during embryonic diapause, suggesting females are capable of superfoetation (Canivenc & Bonnin, 1979; Yamaguchi *et al.*, 2006), together with the plural mating behaviour (Carpenter *et al.*, 2005; Dugdale *et al.*, 2007), all these provide an environment for the females to select between embryos.

MHC may also play a role in the successful development of the foetus and spontaneous abortion. Studies on humans and mice have reported that females tend to abort MHC-similar foetuses (Ober, 1992; Apanius *et al.*, 1997). Also, problems occurring during human infertility treatments may be related to MHC similarity within couples, whose in vitro fertilizations are more likely to fail (Ho *et al.*, 1994). Re-absorption and abortion also occur in badgers (Anderson & Trewhella, 1985), and whether it is related to MHC or not still requires investigation.

The adaptive hypotheses of pre- and post-copulatory sexual selection mentioned above explain how MHC influences on reproductive success are a paradigm for how MHC diversity can be maintained at high levels in natural populations through MHC-dependent sexual selection (Brown & Eklund, 1994; Jordan & Bruford, 1998; Penn & Potts, 1999). To understanding how MHC-dependent mate preferences may result in higher offspring fitness and drive the diversity of MHC genes, I will consider how pathogens impose selection on MHC genes.

1.4 MHC and pathogen resistance

MHC genes play an important role in determining the disease resistance and susceptibility of an animal (Nikolich-Žugich *et al.*, 2004). A reduction in MHC diversity within a species may thus predispose that species to infectious disease (Radwan *et al.*, 2010). For example, the increased susceptibility of cheetahs to various viral diseases is believed to result from a reduction in the number of different MHC molecules available to the species as a whole (due to evolutionary bottle-necking; O'Brien & Yuhki, 2006) and a corresponding limitation on the range of processed antigens with which these MHC molecules can interact. Thus high MHC diversity may be advantageous in providing a broad range of antigen-presenting MHC molecules that protect the species from a wide range of infectious diseases.

1.4.1 Pathogen-mediated selection on MHC

Pathogen-mediated selection results in the maintenance of large numbers of alleles in populations, as well as in greatly enhanced persistence of allelic diversity over long periods of time relative to neutral genetic variation (Richman, 2000). A number of studies (Bernatchez & Landry, 2003; Spurgin & Richardson, 2010) have highlighted the relationships between MHC genotype and pathogen burden. The heterozygote advantage hypothesis, the rare-allele advantage hypothesis, and the fluctuating selection hypothesis, have been proposed as important in maintaining high levels of MHC diversity (Hedrick, 1999; Bernatchez & Landry, 2003; Spurgin & Richardson, 2010).

1.4.2 Heterozygote Advantage Hypothesis

Over-dominance is proposed to confer a heterozygote advantage mechanism, promoting MHC diversity. In the case of over-dominance, heterozygotes are expected to have higher fitness than either parental homozygotes, especially if confronted with multiple species or strains of pathogens or parasites (Hughes & Nei, 1992). Heterozygous individuals possess a larger number of different MHC molecules and consequently can recognise and present a wider range of antigens to the immune system, hence increase the relative fitness of MHC heterozygotes compared with homozygotes (Takahata & Nei, 1990). In certain instances heterozygosity might only give an advantage if individuals are infected with multiple (i.e. a diversity of species or strains) pathogens (Apanius *et al.*, 1997), or rapidly evolving pathogen species or strains (Weidt *et al.*, 1995). While some studies support the heterozygote advantage (Thursz *et al.*, 1997; Penn *et al.*, 2002; Froeschke & Sommer, 2005), other studies have found no association between MHC heterozygosity and pathogen resistance (e.g. Langefors *et al.*, 2001; Harf & Sommer, 2005; Meyer-lucht & Sommer, 2005; Schad *et al.*, 2005; Tollenaere *et al.*, 2008). Also, the optimal diversity of MHC molecules expressed in an individual may not be completely heterozygous, but could represent an intermediate level (Wegner *et al.*, 2003; Kurtz *et al.*, 2004). Even though individuals with greater MHC diversity present more antigens to the immune system, they have a concomitant smaller T-cell repertoire because of thymic selection, in-so-far that there is a trade-off between maximizing the antigen number presented by MHC and in T-cell recognition capacity. A study on the polyploid *Xenopus* (African clawed frog) found that most naturally occurring species of this genus switch off all but one set of diploid loci, suggesting a cost of expressing too many MHC genes (DuPasquier *et al.*, 1989). This optimal MHC-heterozygosity hypothesis is consistent with evidence showing that in some instances MHC heterozygotes demonstrate no advantages, or even disadvantages (Hill *et al.*, 1991; Apanius *et al.*, 1997; Paterson *et al.*, 1998). If optimal intermediate MHC diversity is superior for combating infection and hence provides higher fitness (i.e. survival and subsequent reproductive success), then females should prefer to mate with males that result in offspring with optimal levels of MHC. Hedrick (2004), however, argued that selection for an intermediate optimum does not generally maintain polymorphisms and

should result in a monomorphic population with respect to allele numbers (but see Wegner *et al.*, 2004).

1.4.3 Rare Allele Advantage Hypothesis

The rare-allele advantage hypothesis (vanValen, 1973; Penn & Potts, 1999) proposes that rare alleles provide hosts with advantages when challenged by pathogens. This hypothesis proposes balancing selection, through a frequency-dependent co-evolutionary arms race, between the hosts and pathogens. Pathogens adapt to infect the most common host genotype, leaving rare genotypes least infected (Lively & Dybdahl, 2000). If alleles that are rare and more resistant to pathogens are favoured, but are selected against when they spread through the population and become common, a balanced polymorphism results. This leads to a hypothetical cycle of changing fitness of different alleles/haplotypes and results in the maintenance of high genetic diversity (Slade & McCallum, 1992). Penn and Potts (1999) suggested that MHC-disassortative mating preference (i.e. preference for MHC-dissimilar mates) would provide a moving target to those rapidly evolving parasites that otherwise evade immune recognition through shifting MHC presentation holes (different sets of antigens presented in the progeny) and shifting T-cell recognition holes (different T-cell repertoire in the progeny).

1.4.4 Fluctuating selection hypothesis

The fluctuating selection hypothesis, which has received relatively little attention compared to the two other hypotheses, may also play a role in pathogen-mediated selection (Hedrick, 2002; Spurgin & Richardson, 2010). It proposes that the spatial and temporal fluctuation in pathogen types and abundance may maintain the diversity at the MHC. This pathogen fluctuation may lead to fluctuations in the intensity of directional selection at MHC genes; hence different MHC alleles are selected for at different time or spatial points.

The hypotheses explaining pathogen-mediated selection outlined above are not mutually exclusive (Spurgin & Richardson, 2010). Pathogens may maintain high MHC diversity through a combination of mechanisms. In addition, the evolution of MHC-dependent mating preferences also plays a role in

maintaining MHC diversity. Sexual selection mechanisms such as MHC-disassortative mating and post-copulatory mate choice, which generate heterozygous progeny or a moving target against rapidly evolving pathogens, may also maintain MHC diversity.

1.5 Other MHC evolution hypotheses

1.5.1 Associative Balancing Complex (ABC) evolution

The ABC evolution hypothesis (vanOosterhout, 2009) proposes that recessive deleterious mutations accumulate as a 'sheltered load' (Stone, 2004) near MHC genes. Natural selection does not purge these mutations because they are rarely expressed as homozygotes, given the high MHC gene diversity. These accumulated recessive deleterious mutations increase selection against individuals that are homozygous for the same haplotype. Therefore, if a haplotype becomes very common, it has a greater chance of occurring as a homozygote, at which point purifying selection will reduce its frequency (vanOosterhout, 2009). Selection against homozygotes can thus maintain MHC polymorphisms, even in the absence of strong pathogen-mediated selection. These accumulated recessive deleterious mutations can also lead to the reinforcement of linkage disequilibrium in haploblocks by epistatic selection, which reduces the effective recombination rates (vanOosterhout, 2009). Recombinant haplotypes are selected against because they express part of their sheltered load when paired with either parental haplotype. The deleterious mutations continue to accumulate, which further increases balancing selection and reinforces linkage by epistatic selection against recombinants. ABC evolution is particularly efficient in maintaining MHC polymorphisms because purifying selection acts in all generations, independent of temporal fluctuations in selective pressure by pathogens. ABC evolution is not mutually exclusive to other evolutionary mechanisms and it can operate alongside balancing selection on the actual immune genes. Pre- and post-copulatory sexual selection, such as the abortion of MHC homozygous embryos, may be mechanisms that facilitate ABC evolution.

1.5.2 Gene conversion

In addition to point mutation, MHC sequence variation can be generated by gene conversion (Parham & Ohta, 1996; Jakobsen *et al.*, 1998; Richman *et al.*, 2003). Gene conversion can generate new haplotypes by transferring sections of DNA sequences within and across loci (Ohta, 1991; Chen *et al.*, 2007). Gene conversion may be the primary mechanism generating MHC haplotype variation in bottlenecked populations (Spurgin *et al.*, 2011), because the rate at which it can do so may greatly exceed that of point mutation (Parham & Ohta, 1996). Although gene conversion can re-shuffle existing polymorphisms within and across loci to create new haplotypes, variation is only generated at the haplotype level; variation at the sequence level will become homogenized over time (Spurgin *et al.*, 2011).

1.6 European badgers:

A review of mating systems and disease susceptibility

1.6.1 European badgers

The European badger (*Meles meles*) is a semi-fossorial nocturnal carnivore that lives in underground burrow systems, termed setts (Thornton, 1988), by day, and emerges to forage independently at dusk (Roper, 2010). In high-density populations in Southwest England, badgers can form large social groups of up to 30 individuals (Rogers *et al.*, 1997; Macdonald *et al.*, 2004). The badger population at Wytham Woods, Oxfordshire, United Kingdom, has one of the highest recorded badger population densities in the world (Macdonald & Newman, 2002; Macdonald *et al.*, 2008; Macdonald *et al.*, 2009). Wytham Woods is a 424 ha mixed deciduous / coniferous woodland located 5 km northwest of Oxford (51°46′N, 01°19′W), which is surrounded by permanent pasture and mixed arable land (Kruuk, 1978b; Macdonald *et al.*, 2004). The numbers of badgers in Wytham Woods rose from 60 to 228 adults between 1987 and 1996, that is from 10 adult badgers/km^{2,} to 38 badgers/km^2 (Macdonald & Newman, 2002) and have continued to maintain this high density henceforth (Macdonald *et al.*, 2008; Macdonald *et al.*, 2009). Badgers have been studied continuously at Wytham Woods since the 1970s

(Kruuk, 1978b, a) and the population has been trapped 4 times per annum since 1987 (for detailed methods see Macdonald & Newman, 2002). Extensive and systematic studies have been undertaken for this study population on different areas (e.g. Macdonald & Newman, 2002; Dugdale *et al.*, 2007; Dugdale *et al.*, 2008; Macdonald *et al.*, 2009), and a rich database will enable me to explore new questions regarding MHC-based mate choice and pathogen resistance.

1.6.2 Mating system

European badgers can mate during any month of the year in Britain (Neal & Cheeseman, 1996; Roper, 2010), but with the primary mating peak occurring immediately postpartum (Cresswell *et al.*, 1992; Yamaguchi *et al.*, 2006). They exhibit a polygynandrous mating system (Carpenter *et al.*, 2005; Dugdale *et al.*, 2007), that is, individuals of both sexes may mate with more than one individual of the opposite sex. Only 31% of adult females and 27% of adult males bred each year (Dugdale *et al.*, 2007). After mating, fertilized ova enter embryonic diapause at the blastocyst stage, which lasts up to eleven months before implantation of the embryo in December or January (Canivenc & Bonnin, 1979; Yamaguchi *et al.*, 2006). The oestrous cycle may continue during embryonic diapause, female badgers are potentially capable of superfoetation (Canivenc & Bonnin, 1979; Yamaguchi *et al.*, 2006) and embryo reabsorption (Yamaguchi *et al.*, 2006). Badgers give birth once a year around February after this period of delayed implantation. This reproductive strategy facilitates multiple paternity (Evans *et al.*, 1989; Dugdale *et al.*, 2007) and may lead to post-copulatory sexual selection.

Field observations of extra-group mounting behaviour are very scarce (Christian, 1995), but recent studies have revealed that nearly half of assigned paternity is attributed to extra-group males, mainly from neighbouring social groups (Carpenter *et al.*, 2005; Dugdale *et al.*, 2007). In addition, Dugdale et al. (2007) showed that within-group and extra-group males can sire cubs within the same multiple-paternity litter. Within-group relatedness has been demonstrated to be high ($R = 0.20$) and approximately one-third of within-group dyads were more likely to represent first-order kin than unrelated pairs (Dugdale *et al.*, 2008); this may facilitate inbreeding avoidance and favour extra-group mating behaviours. MHC-based recognition has been proposed as a mechanism for inbreeding avoidance (Potts *et al.*, 1994). In addition, MHC-based choice can be involved in extra-pair mating,

for example, females may seek extra-pair matings if their social partners have low MHC diversity or similar MHC (Richardson *et al.*, 2005; Schwensow *et al.*, 2008b).

Due to these sociological and reproductive system characteristics, the European badger is an exemplarily species with which to explore how MHC selection may be modified and regulated by mate choice. Attributes such as a long mating period (Cresswell *et al.*, 1992), polygynandrous mating system (Dugdale *et al.*, 2007), delayed implantation (Yamaguchi *et al.*, 2006), putative superfoetation (Yamaguchi *et al.*, 2006) and embryo reabsorption (Yamaguchi *et al.*, 2006), high-level of extra-group paternity (Carpenter *et al.*, 2005; Dugdale *et al.*, 2007), low fecundity (Macdonald *et al.*, 2009), and lack of paternal care (Dugdale *et al.*, 2010), all may underscore pre-copulatory and post-copulatory MHC-based mate choice.

1.6.3 Information transmission by odour

MHC genotypes might be recognizable through olfactory cues (Yamazaki *et al.*, 1982; Ziegler *et al.*, 2005; Brennan & Kendrick, 2006), where MHC variation is known to be related to sexually selected ornaments (vonSchantz *et al.*, 1996; vonSchantz *et al.*, 1997; Dunn *et al.*, 2012). Therefore olfactory cues may provide a basis for individual recognition and mate choice (Penn, 2002).

European badgers exchange information using faeces, urine and secretions from the anal, subcaudal and interdigital glands, often at latrines (Kruuk, 1978b; Stewart *et al.*, 2001; Buesching *et al.*, 2003) along group-home range boundaries. They also use subcaudal gland secretion to mark both the environment (i.e. object-marking) and conspecifics (i.e. allo-marking) (Buesching *et al.*, 2002b). The anal secretions deposited with faeces are individual-specific and identifiable by group (Davies *et al.*, 1988; Roper *et al.*, 1993; Christian, 1995) while subcaudal secretions confer information about the sex and reproductive status of individuals (Buesching *et al.*, 2002c, b; Buesching *et al.*, 2003).

To create a shared group-odour and maintain group cohesion, badgers perform mutual allo-marking using their subcaudal gland, i.e. two badgers pressing their subcaudal pouch against each other

simultaneously (Buesching *et al.*, 2003). Sequential allo-marking, when one badger marks the body of another, is more frequent than mutual marking. It is proposed that the specific odour is generated partly by bacterial flora that metabolize primary gland products in the pouch (Albone & Perry, 1975; Buesching *et al.*, 2003), consequently mutual marking facilitates the exchange of pouch bacteria and thus generates a common group-smell in addition to the individual-specific odour, while sequential allo-marking serves to distribute the common group smell, and to advertise individual-specific, fitness-related information, e.g. body condition and reproductive status. Individual-specific olfactory signal may also be used as a cue for pre-copulatory mate choice based on MHC, and it has been shown to influence mate choice in mammals and fish (Egid & Brown, 1989; Potts *et al.*, 1991; Reusch *et al.*, 2001; Leinders-Zufall *et al.*, 2004; Milinski *et al.*, 2005; Spehr *et al.*, 2006),.

1.6.4 Pathogens and badgers

Female mate choice is proposed to be based on phenotypic cues whose expression is dependent on their genetically based resistance to pathogens (Hamilton & Zuk, 1982). The MHC plays an important role in pathogen recognition and is a good candidate for testing the relationships between immune genes and both phenotypic cues expression and pathogen resistance (Dunn *et al.*, 2012).

Pathogens play a role in badger sociality, through the solicitation of allo-grooming (Johnson *et al.*, 2004) and in communal behaviour pertaining to sett chamber use (Butler & Roper, 1996). Hancox (1980) summarizes a wide variety of ectoparasites and endoparasites that have been identified to infect badgers. Living underground socially accounts for many aspects of the ecology and species composition of badger pathogens, particularly by facilitating the spread of infection (Hancox, 1980), such as badger lice (*Trichodectes melis*) and fleas (*Paraceris melis*) in bedding (Butler & Roper, 1996; Cox *et al.*, 1999).

In addition to ectoparasites (Butler & Roper, 1996; Cox *et al.*, 1999; Ogden *et al.*, 2000; Johnson *et al.*, 2004; Newman *et al.*, 2004; San, 2007), badgers are susceptible to, and implicated in the transmission of, zoonotic disease, e.g. bovine tuberculosis (bTB) (Cheeseman *et al.*, 1989; Clifton-

Hadley *et al.*, 1993; Courtenay *et al.*, 2006; Macdonald *et al.*, 2006). Although studies have shown that the prevalence of infectious individuals carrying *Mycobacterium bovis* among common farmland wildlife is very low (Mathews *et al.*, 2006a). Fatal epidemic diseases have also been reported in badgers, e.g. canine distemper virus (Burtscher & Url, 2007) and rabies (Smith & Wilkinson, 2002).

Both blood and gut parasites have been studied in the Wytham badger population (Macdonald *et al.*, 1999; Anwar *et al.*, 2000; Newman *et al.*, 2001; Lizundia *et al.*, 2011); the prevalence of *Trypanosoma pestanai* and *Babesia missirolii* being 7.7% and 77.2 % respectively (Macdonald *et al.*, 1999). Prevalence was highest in younger age classes, suggesting immature badgers are more susceptible to infection. Two coccidia species, *Eimeria melis* and *Isospora melis*, have been identified in faecal samples collected from the same population (Newman *et al.*, 2001). Cubs had significantly higher intensity and prevalence of infection, which diminished markedly in adult badgers (Anwar *et al.*, 2000). Infantile coccidiosis in badgers may increase mortality in the cub population or reduce fitness in their future life. By acting together with climatic fluctuation (Newman *et al.*, 2001; Macdonald *et al.*, 2010), coccidia are one of the many pathogens controlling cub survival (Nouvellet *et al.*, 2010).

As badgers are host to such a broad variety of pathogens, they provide a good opportunity to examine the association of MHC and multiple pathogens to infer possible pathogen-mediated selection in the wild (Hughes & Nei, 1992; McClelland *et al.*, 2003; Spurgin & Richardson, 2010). In addition, given their reproductive biology and mating system, we postulate that MHC diversity is maintained through pathogen-mediated selection and MHC-based sexual selection (Brown & Eklund, 1994; Jordan & Bruford, 1998; Penn & Potts, 1999).

1.7 Thesis aims and organization

This thesis is comprised of five inter-linked data chapters that together investigate how the MHC genes of an animal may provide indirect benefit and associate with the behavioral and evolutionary levels.

Chapter 2 (Sin *et al.*, 2012b) and 3 (Sin *et al.*, 2012c) consider the characterization and evolution of MHC class II and class I genes respectively, from both gDNA and cDNA. The evidence for selection across the loci is also examined, for example by determining the rates of non-synonymous and synonymous nucleotide substitutions (d_N:d_S) (Hill & Hastie, 1987; Hughes & Nei, 1988) for antigen-binding site (ABS) and non-ABS. Phylogenetic analyses are also performed to investigate whether trans-species polymorphism and orthology have happened. This is a crucial first step before I investigate the relationship between MHC, mate choice and pathogen resistance.

In **Chapter 4 (Sin *et al.*, 2012a)**, I examine the subcaudal gland secretion microbiota and investigate its correlation with different individual-specific parameters.

In **Chapter 5**, I investigate whether MHC-based mate choice occurred in this badger population. I also compare the MHC with neutral microsatellite markers to distinguish MHC-specific effects from genome-wide effects. I specifically test whether the assigned parent pairs were more similar or dissimilar in their MHC compared to random pairing of breeding females and candidate males.

In **Chapter 6**, I investigate the association between MHC alleles and pathogen burden of a wide variety of taxa. Infection intensity and pathogens presence/absence is examined for multiple species, and the co-infection status is determined. I also test whether there is less co-infecting pathogens in individuals that are MHC-heterozygous.

Finally, in **Chapter 7**, a unifying synthesis is made in the context of how MHC link all these findings together, with regard to broadening and enhancing our understanding of immuno-competence inheritance paradigms. Directions for future research are also discussed.

Chapter 2

MHC class II genes in the European badger (*Meles meles*): Characterization, patterns of variation and transcription analysis[1]

[1] This chapter has been published as: Sin, Y.W., Dugdale, H.L., Newman, C., Macdonald, D.W. & Burke, T. 2012. MHC class II genes in the European badger (*Meles meles*): characterization, patterns of variation, and transcription analysis. *Immunogenetics* **64:** 313–327. doi: 310.1007/s00251-00011-00578-00259.

2.1 Abstract

The major histocompatibility complex (MHC) comprises many genes, some of which are polymorphic with numerous alleles. Sequence variation among alleles is most pronounced in exon 2 of the class II genes, which encodes the α1 and β1 domains that form the antigen-binding site (ABS) for the presentation of peptides. The MHC thus plays an important role in pathogen defense. European badgers (*Meles meles*) are a good species in which to study the MHC, as they harbor a variety of pathogens. We present the first characterization of MHC class II genes, isolated from genomic DNA (gDNA) and complementary DNA (cDNA), in the European badger. Examination of seven individuals revealed four *DRB*, two *DQB*, two *DQA* and two *DRA* putatively functional gDNA sequences. All of these sequences, except *DRA*, exhibited high variability in exon 2; *DRB* had the highest variability. The ABS codons demonstrated high variability, due potentially to balancing selection, while non-ABS codons had lower variability. Positively-selected sites were detected in *DRB* and *DQA*. Phylogenetic analysis demonstrated trans-species polymorphism of class II genes. Comparison with cDNA from whole blood revealed that only *DRB* had a transcription pattern reflecting the alleles that were present in the gDNA, while the other three genes had disparities between gDNA and cDNA. Only one sequence was transcribed, even though two gDNA sequences were present, from each of both *DQB* and *DRA*. Our characterization of badger MHC sequences forms a basis for further studies of MHC variability, mate choice and pathogen resistance in this, and other, species.

2.2 Introduction

Major histocompatibility complex (MHC) genes play an important role in the vertebrate immune system and auto-immunity. Cell-surface glycoproteins encoded by the MHC genes are important in both humoral and cell-mediated immune responses, as they bind and present antigens to T-cells and trigger an immune cascade (Swain, 1983). The mammalian MHC is organized into functional regions including class I and class II. The genes in each region are subjected to birth-and-death evolution (Takahashi *et al.*, 2000; Piontkivska & Nei, 2003), in which new genes are created by gene duplication, and some genes become nonfunctional through deleterious mutations. MHC class I molecules are primarily responsible for intracellular antigen binding and are expressed on the surface of all nucleated somatic cells (Hughes & Yeager, 1998). MHC class II molecules, on the other hand, principally bind peptides from the extracellular environment and are only expressed on antigen-presenting cells, such as B-cells and macrophages (Hughes & Yeager, 1998). If a helper T-cell recognizes a presented antigen, the antigen-presenting cell is activated to trigger an immune response.

At the molecular level, MHC class II molecules are heterodimers that consists of two trans-membrane proteins: an α and a β chain, which are encoded by class II genes *A* and *B*, respectively. The α1 and β1 domains are regions in the α and β chains that contain specific sites forming the antigen-binding site (ABS; Furlong & Yang, 2008). In placental mammals, the class II region is divided into sub-regions, or gene clusters, known in humans as DR, DQ, DP and DM (Takahashi *et al.*, 2000; Kelley *et al.*, 2005). Each cluster contains a functional α chain gene and one or more functional β chain gene(s). A similar organization of class II genes has been found in the canine MHC, known as the dog leukocyte antigen (DLA) region (Wagner, 2003). Some gene clusters have apparently been lost in some evolutionary lineages. For example, the feline MHC lacks the DQ sub-region and does not transcribe DP molecules; however, the feline MHC has an expanded DR sub-region (Yuhki *et al.*, 2003). DR and DQ molecules present different peptides and are expressed differentially, according to cell type, for example, DQ is expressed at a lower frequency than DR on peripheral blood monocytes (Alonso *et al.*, 1985; Edwards *et al.*, 1985; Rees *et al.*, 2003).

The high number of MHC alleles and high degree of sequence diversity among alleles are principally due to the numerous and diverse exons that encode the domains forming the ABS (Hughes & Yeager, 1998). The nucleotide diversity within MHC genes is generally attributed to balancing selection (Bergström & Gyllensten, 1995), which encompasses many specific forms of selection that act to maintain large numbers of alleles in populations. Balancing selection greatly enhances the persistence of allelic diversity over long periods of time relative to neutral genetic variation (Richman, 2000). Balancing selection, driven principally by ever-evolving pathogens (Spurgin & Richardson, 2010), leads to trans-species polymorphism (Klein, 1987), resulting in orthologous alleles among related species.

There is currently a paucity of studies on the relationships among MHC genotype, parasite burden, and mate choice in wild mammal populations (Bernatchez & Landry, 2003). In this study, we characterize the MHC class II genes of the European badger (*Meles meles*), a nocturnal and fossorial carnivore that rests by day in underground burrow systems, termed setts, and emerges to forage independently at dusk (Roper, 2010). *M. meles* has an extensive geographical range, stretching from Europe to Japan and from the Arctic Circle to the Mediterranean, and exhibits considerable socio-spatial variability (Rogers *et al.*, 1997; Macdonald *et al.*, 2004; Rosalino *et al.*, 2004; Roper, 2010). *M. meles* has a long mating-season (Buesching *et al.*, 2009), delayed implantation (Thom *et al.*, 2004), putative superfoetation (Yamaguchi *et al.*, 2006), and a sensory predisposition towards olfaction (Buesching *et al.*, 2002a). In high-density populations *M. meles* has a polygynandrous mating system (Dugdale *et al.*, 2007; Dugdale *et al.*, 2011b) with high levels of extra-group paternity (Dugdale *et al.*, 2007) and low fecundity (Macdonald *et al.*, 2009). Additionally, it has been the subject of a diverse range of parasitic disease studies (Hancox, 1980; Macdonald *et al.*, 1999; Newman *et al.*, 2001; Anwar *et al.*, 2006; Rosalino *et al.*, 2006; San, 2007; Lizundia *et al.*, 2011). *M. meles* is thus potentially well suited for investigating how MHC selection, and conferred immunological advantages (e.g. parasite resistance), is modified and regulated by mate choice.

In this study we: (1) characterized the MHC class II genes of *M. meles* from a high-density population, and tested for evidence of selection; (2) identified the transcription pattern of alleles by comparing

gDNA and cDNA sequences from whole blood samples, which is important as MHC genes identified using gDNA may be nonfunctional (Knapp, 2007); and (3) performed phylogenetic analyses to investigate whether *M. meles* MHC sequences belong in monophyletic groups, or whether trans-species polymorphism occurs.

2.3 Materials and methods

Sample collection and nucleic acid isolation

Blood samples were collected *ad hoc* from seven badgers associated with five different social groups (Macdonald *et al.*, 2008), as part of an ongoing long-term study (see Macdonald *et al.*, 2009) in Wytham Woods, Oxfordshire, UK (global positioning system reference 51:46:26N, 1:19:19W; all trapping and handling protocols were subject to ethical review and performed under government licence). The methods are detailed in Macdonald and Newman (2002). The mean of the pairwise Queller and Goodnight (1989) index of relatedness (*R*) of these seven individuals was 0.37 (95% CI = 0.29–0.44); slightly higher than half-siblings (0.31 (95% CI = 0.30–0.33); Dugdale *et al.*, 2008), but unfortunately the wild nature of this population restricted sample availability. Nevertheless, no additional *DRB* sequences have been found from genotyping of the entire population (Y. W. Sin, unpublished data). Approximately 3 ml of blood was taken by jugular venipuncture and collected in a vacutainer containing EDTA. Samples were stored at -20°C until DNA isolation was performed. A single sample from an American badger *Taxidea taxus jacksoni* (sampling locality: Wisconsin, USA) was also examined and the sequences generated were included in the phylogenetic analysis. gDNA was isolated using the GFX Genomic Blood DNA Purification Kit (Amersham Biosciences, Little Chalfont, UK), following the scalable method in the manufacturer's protocol. In order to validate whether the identified alleles were transcribed, a 500-μL blood sample, from each of the seven individuals, was also transferred into RNAprotect Animal Blood Tubes (Qiagen, Hilden, Germany) and stored immediately at -20°C for less than a month before RNA isolation. Total cellular RNA was isolated from each blood sample using an RNeasy Protect Animal Blood Kit (Qiagen, Hilden, Germany).

cDNA synthesis

A standard cDNA synthesis was performed on 2 μg of RNA, quantified using a NanoDrop 8000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA), from each individual. The reverse-transcription reaction components included 4 units of reverse transcriptase (Omniscript®, Qiagen), 0.5 mM of each dNTP and 1 μM oligo-dT primer, in 1x RT buffer. The reverse transcription reactions were incubated at 37°C for 60 min and stored at -20°C.

Primer design and PCR amplification

To amplify the near full-length transcribed class II genes from cDNA and class II genes in exon 2 from gDNA, published primers of other canivores - the domestic dog *Canis lupus familiaris* (Wagner *et al.*, 1996; Kennedy *et al.*, 2006), DRB exon 2 of the European mink *Mustela lutreola* (Becker *et al.*, 2009) and transcribed class II genes of the California sea lion *Zalophus californianus* (Bowen *et al.*, 2002; Bowen *et al.*, 2004), were tested. Oligonucleotide primers, recognizing highly conserved regions of each of four MHC class II genes (*DRB, DQB, DQA,* and *DRA*), were also designed using OligoAnalyzer 3.1 (http://eu.idtdna.com/analyzer/applications/oligoanalyzer), based on alignments with GenBank's nucleotide sequences from mustelids (*DRB*: EU121854–EU121855, EU263550–EU263558; *DQB*: EU121856–EU121857), canines (*DRB*: AF016910–AF016912, CFU44777–CFU44778, M29611, NC006594, NM001014768; *DQB*: AF016908–AF016909, M90802, NC006594; *DQA*: NC006594, Y07942–Y07944; *DRA*: L37332, NC006594), pinnipeds (*DRB*: AY491456–AY491467; *DQB*: AF503397–AF503401, AF503406–AF503407; *DQA*: AF093799, AF502560–AF502564; *DRA*: AY491450–AY491455), porcines (*DQA*: NM001130224) and humans (*DQA*: NM002122; *DRA*: NM019111).

PCR amplification, using these primers on 10–30 ng of cDNA/gDNA, was performed in a 20-μl reaction mix that also contained: 0.5 μM of each primer (Table 2.1), 200 μM of each dNTP, 1 X PCR buffer (contains $MgCl_2$; Qiagen), and 2 units HotStarTaq (Qiagen). The PCR cycle began with incubation at 94°C for 15 min, followed by 30 incubation cycles at 94°C for 30 s, annealing

Table 2.1 MHC class II-specific primers for European badgers. F = forward, R = reverse. T_a = annealing temperature (^{o}C). bp = base pair

	Primer name	Primer sequence	Product size (bp): Region amplified	T_a	Species specificity	Source reference
Amplification of transcribed genes	F: DRBU116	GAGGCTCCTGGATGACAGCTYTG	776 bp; exon 1 – 3' UTR	60	Mammalian *DRB*	Bowen et al. 2004
	R: DRB3'UTR	AGCTGGGAAGGAAGGTTCTTCCT				Bowen et al. 2004
	F: ZcDQBUEX1D	TGTTCCCAATGATGCCTGGGAAG	654 bp: 5′UTR – 5′exon 4	60	Mammalian *DQB*	Bowen et al. 2002
	R: DQB-L769	CTCTGGGMAGATTCAGACTG				Bowen et al. 2002
	F: Meme-DQAex1F	ACAGAGTTCTGATTCTGGGG	592 bp; exon 1 – exon 4	55	Badger *DQA*	This study
	R: Meme-DQAex4R	CAGCTCTGACATAGGGGTTGGAA				This study
	F: DRAU102	CATAAGTGGAGTCCCWGTGCTAG	647 bp; exon 1 – exon 4	55	Badger *DRA*	Bowen et al. 2004
	R: Meme-DRAex4R	AATGATGCCAACCAGACCCAC				This study
Amplification of exon 2	F: Meme-DRBex2F	CGTCCCCACAGGACATTTC	248 bp; exon 2	59	Badger *DRB*	This study
	R: Meme-DRBex2R	CTCGCCGCTGCACC				This study
	F: Meme-DQBex2F	TCCCCGCAGAGGATTTCGTG	248 bp; exon 2	55	Badger *DQB*	This study
	R: Meme-DQBex2R	GCTCACCTCGCCTCTGC				This study
	F: Meme-DQAex2F	TACCAGTCTTATGGTCCC	176 bp; exon 2	48	Badger *DQA*	This study
	R: Meme-DQAex2R	CATTGGTAGCAGCGGTATAGTTG				This study
	F: Meme-DRAex2F	AGGACCATGTGATCATCCAG	209 bp; exon 2	55	Badger *DRA*	This study
	R: Meme-DRAex2R	ACGTACCATTGGTGTTTGGGGTG				This study

temperature (Table 2.1) for 30 s, and 72°C for 1–2 min according to amplicon length, ending with an extension step at 72°C for 10 min. The PCR products were electrophoresed on a 1.5% agarose gel and visualised by UV light and ethidium bromide staining. A 100-bp DNA ladder (New England Biolabs, Herts, UK) was used to size the DNA fragments. Bands of expected size were excised from the gel and purified using QIAquick Gel Extraction Kits (Qiagen). PCR products that gave rise to relatively bright bands of the expected size were cloned and sequenced.

Cloning and DNA sequencing

Purified PCR fragments were ligated into a TA cloning vector (pGEM-T Easy Vector Systems, Promega, Madison, WI, USA). Transformation, growth and blue-white selection were carried out in *Escherichia coli* competent cells (JM109). PCRs were performed on positive clones using M13 forward and reverse primers (Messing, 1983). Consensus sequences were created. All amplified fragments were sequenced in both complementary directions using dideoxynucleotide chain-termination and loaded on an automated DNA Sequencer (48-well capillary ABI3730, Applied Biosystem, Foster City, CA, USA). Between six and twenty-four clones were sequenced for each individual. Identical sequences were derived from a minimum of two badgers or from independent PCR reactions from the same individual, in compliance with DLA nomenclature rules (Kennedy *et al.*, 1999). Single unique sequences (possible chimeras) were excluded. Nucleotide sequences were analyzed using CodonCode Aligner 3.7.1 (CodonCode Corporation, Dedham, MA, USA), and were compared with known MHC class II sequences using the NCBI BLAST program (Altschul *et al.*, 1990). The DNA sequences from the European and American badger were assigned GenBank accession numbers: HQ908092–HQ908108.

Data analyses

Phylogenetic analyses were performed on the consensus alignments of European and American badger *DRB, DQB, DQA* and *DRA* exon 2 sequences, against sequences from other Carnivore species available in GenBank (e.g. sea otter *Enhydra lutris,* European mink *M. lutreola, Mustela* sp.,

California sea lion *Z. californianus,* giant panda *Ailuropoda melanoleuca* and dog *C. lupus familiaris*; accession numbers are provided in the legends of Fig. 2.1–2.4). Domain borders were assigned based on sequence homology to different HLA genes respectively (IMGT/HLA Database; Robinson *et al.*, 2001). Bayesian phylogenetic inference was performed using MrBayes 3.1.2 (Ronquist & Huelsenbeck, 2003). Data were divided into three partitions based on gene codon positions. We used the web-based application FindModel (http://www.hiv.lanl.gov/content/sequence/findmodel/findmodel.html) to find the best-fit model of nucleotide substitution, which was the general time-reversible model with gamma-distributed rate variation (GTR+Γ) for both datasets. A Markov chain Monte Carlo (MCMC) search was initiated with random trees and run for 2 000 000 generations, with a sampling frequency of every 100 generations. The standard deviation of split frequencies converged to a value of 0.005 and 0.007 for the α and β genes, respectively. We checked for convergence by plotting the likelihood scores against generation, and discarded the first 25% of the generations as 'burn-in'. Two separate analyses and four independent chains were then executed to check for convergence of topology.

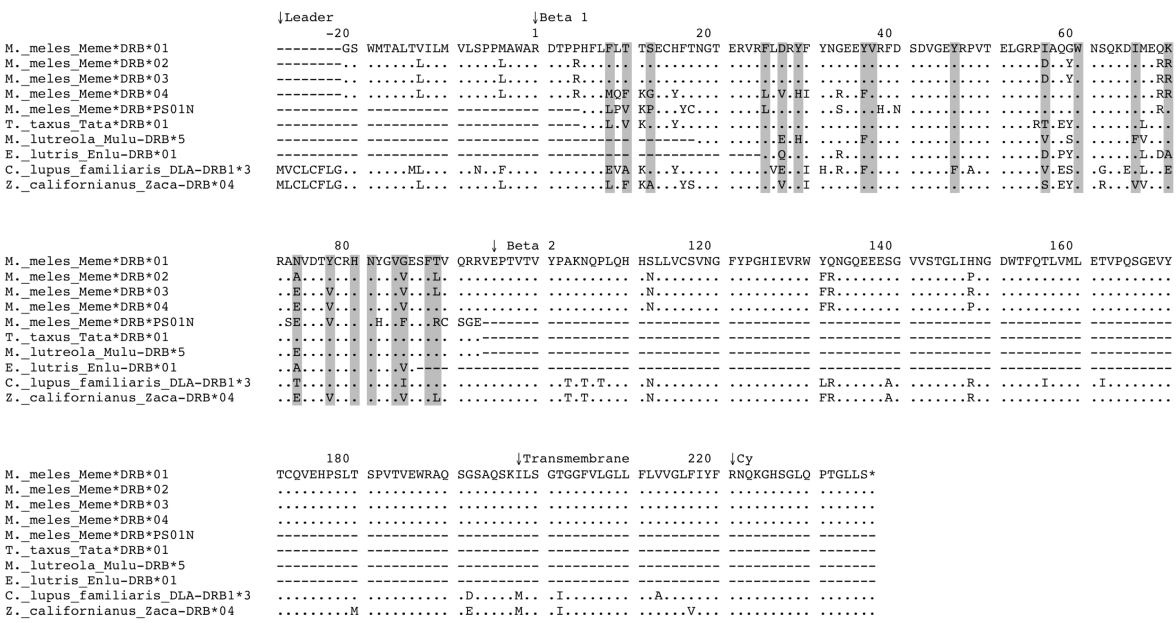


Fig. 2.1 Amino acid sequence identity for the European badger *Meles meles* class II MHC clones, mustelid, phocine, and canine DRβ sequences. The GenBank accession numbers for *DRB* sequences from other mammals are EU263554 (*Mustela lutreola*), EU121855 (*Enhydra lutris*), NM001014768 (*Canis lupus familiaris*) and AY491459 (*Zalophus californianus*). The complete amino acid sequence of Meme-DRB*01 is shown; numbers above the sequence indicate the codon position in the β domain. Single letters and dots represent amino acids that are distinct from or identical to Meme-DRB*01, respectively. Dashes indicate missing sequences. Putative antigen-binding sites were defined according to Reche & Reinherz (2003) and are marked with grey boxes

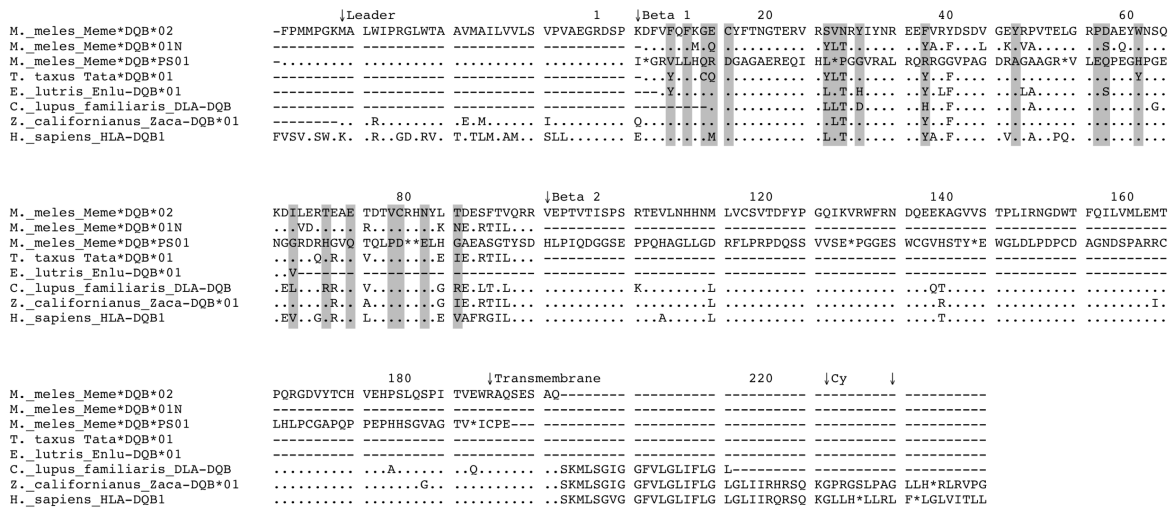


Fig. 2.2 Amino acid sequence identity for the European badger class II MHC clones, mustelid, phocine, canine, and human DQβ sequences. The GenBank accession numbers for *DQB* sequences from other mammals are EU121857 (*Enhydra lutris*), M90802 (*Canis lupus familiaris*), AF503397 (*Zalophus californianus*) and NM002123 (*Homo sapiens*). The complete amino acid sequence of Meme-DQB*02 is shown; numbers above the sequence indicate the codon position in the β domain. Single letters and dots represent amino acids that are distinct from or identical to Meme-DQB*02 respectively. Dashes indicate missing sequences. Putative antigen-binding sites were defined according to Reche & Reinherz (2003) and are marked with grey boxes

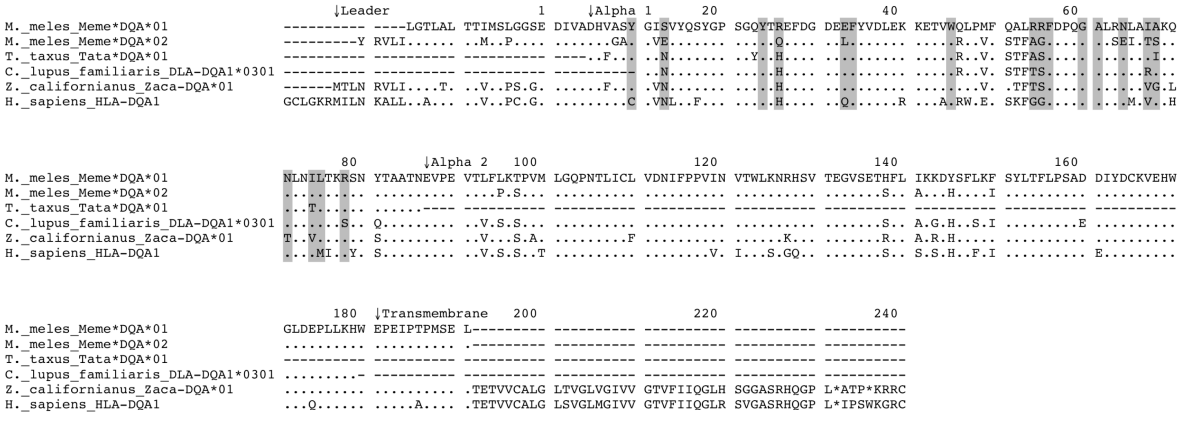


Fig. 2.3 Amino acid sequence identity for the European badger class II MHC clones, mustelid, phocine, canine, and human DQα sequences. The GenBank accession numbers for *DQA* sequences from other mammals are Y07944 (*Canis lupus familiaris*), AF502560 (*Zalophus californianus*) and NM002122 (*Homo sapiensi*). The complete amino acid sequence of Meme-DQA*01 is shown; numbers above the sequence indicate the codon position in the α domain. Single letters and dots represent amino acids that are distinct from or identical to Meme-DQA*01 respectively. Dashes indicate missing sequences. Putative antigen-binding sites were defined according to Reche & Reinherz (2003) and are marked with grey boxes

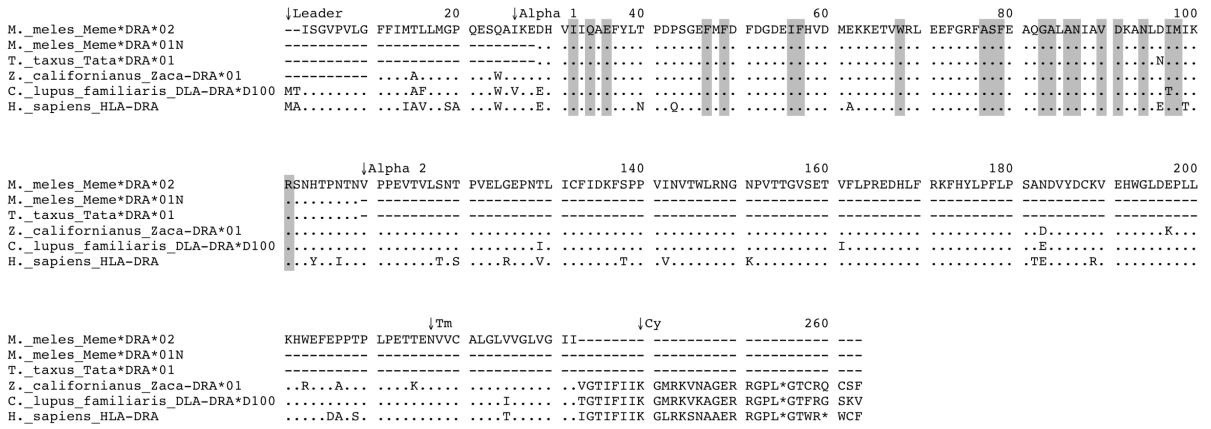


Fig. 2.4 Amino acid sequence identity for the European badger class II MHC clones, mustelid, phocine, canine, and human DRα sequences. The GenBank accession numbers for *DRA* sequences from other mammals are L37332 (*Canis lupus familiaris*), AY491450 (*Zalophus californianus*) and NM019111 (*Homo sapiens*). The complete amino acid sequence of Meme-DRA*02 is shown; numbers above the sequence indicate the codon position in the α domain. Single letters and dots represent amino acid that are distinct from or identical to Meme-DRA*02 respectively. Dashes indicate missing sequences. Putative antigen-binding sites were defined according to Reche & Reinherz (2003) and are marked with grey boxes

Rates of non-synonymous (d_N) and synonymous (d_S) substitutions per site were estimated in DnaSP 4.0 (Rozas *et al.*, 2003) and MEGA 4 (Tamura *et al.*, 2007) according to the method of Nei & Gojobori (1986), with Jukes & Cantor (1969) correction. Standard errors were obtained with 1000 bootstrap replicates. Synonymous and non-synonymous substitutions were calculated separately for the ABS and non-ABS as determined by Reche & Reinherz (2003). CODEML in PAML 4.4b (Yang, 2007) was used to check for positive selection sites (PSS) in MHC class II DR and DQ α1 and β1. PSS are indicated by a ratio ω (non-synonymous over synonymous substitutions, d_N/d_S) larger than 1. Three null models of neutral evolution (M0 – one ratio, M1a – nearly neutral, and M7 – beta) were applied and compared against their nested models allowing for positive selection (M3 – discrete, M2a – positive selection, M8 – beta & ω, respectively). Model M0 (one ratio) assumes a constant ω ratio for all codons. M1a (nearly neutral) assumes two site classes (conserved ($0 < \omega_0 < 1$) and neutral ($\omega_1 = 1$). M2a (positive selection) adds a third site class ($\omega_2 > 1$) to M1a. M3 (discrete) estimates the proportion of conserved, neutral and unrestricted codons from the data. M7 (beta) and M8 (beta & ω) assume a ω ratio with a beta distribution among sites, where M8 adds an extra site class ($\omega > 1$). We determined whether the alternative model provided a significantly improved fit, versus the null model, using a likelihood ratio test (LRT), by comparing twice the difference of the log likelihood ratios

(2ΔInL) to a χ^2 distribution. At each codon CODEML was used to calculate Bayes Empirical Bayes (BEB) posterior probabilities (Yang *et al.*, 2005) for comparisons of M1a versus M2a and M7 versus M8, to identify codons under positive selection (posterior probabilities >0.95).

2.4 Results

MHC class II sequence variability and selection patterns

Five different sequences were isolated for *DRB* genes, three for *DQB*, two for *DQA* and two for *DRA* from gDNA and cDNA using the primers detailed in Table 2.1. Two to three putatively functional *DRB* sequences were isolated from each individual, indicating the presence of at least two *DRB* loci. No more than two putatively functional sequences were isolated from each individual for *DQB*, *DQA* and *DRA* genes, signifying the presence of a single locus. In addition, we inferred the presence of one *DRB* and one *DQB* pseudogene loci (next paragraph). Among the four genes, variation between *DRB* sequences was highest, while that between *DRA* sequences was the lowest (Table 2.2). The majority of variable sites were in exon 2 (Table 2.2), where most mutations represented a non-synonymous nucleotide substitution. Hence, DRβ, DQβ, and DQα exhibit more polymorphic amino acid residues among the β1 or α1 domain (Fig. 2.1–2.3, Table 2.3); in contrast the DRα had the least polymorphic amino acid residues (Fig. 2.4, Table 2.3).

Pseudogene features (Fig. 2.1–2.2) were found in two sequences: nucleotide deletions caused a frame-shift that occurred at position 355 for *Meme-DRB*PS01N*, and a deletion of 17 nucleotides occurred at the beginning of exon 2 for *Meme-DQB*PS01*, which caused premature stop codons. The deletion in *Meme-DQB*PS01* occurred at the 5' primer annealing site for exon 2 amplification from gDNA, so no sequence was detected from gDNA. Accordingly, these were regarded as pseudogenes and were not included in the nucleotide substitution calculations.

The ratios of non-synonymous to synonymous substitutions for *DRB* and *DQA* were higher in the ABS (Table 2.4), whereas synonymous substitutions were higher than non-synonymous substitutions in the

Table 2.2 Sequence polymorphism of MHC class II genes delineated by exon. The number of nucleotide and derived amino acid sequences isolated from seven European badgers *Meles meles* were compared and the numbers of synonymous and non-synonymous nucleotide substitutions are shown

Exons	1	2	3	4	5	6
DRB						
Sequences for comparison	4	4	4	4	4	4
Variable sites	4	43	7	1	0	0
Mutations	4	43	8	1	0	0
Synonymous	2	13	3	1	0	0
Non-synonymous	2	26	5	0	0	0
DQB						
Sequences for comparison	1	2	1	1	−	−
Variable sites	−	39	−	−	−	−
Mutations	−	39	−	−	−	−
Synonymous	−	11	−	−	−	−
Non-synonymous	−	27	−	−	−	−
DQA						
Sequences for comparison	2	2	2	2	−	−
Variable sites	2	38	11	0	−	−
Mutations	2	38	11	0	−	−
Synonymous	0	15	3	0	−	−
Non-synonymous	2	23	8	0	−	−
DRA						
Sequences for comparison	1	2	1	1	−	−
Variable sites	−	3	−	−	−	−
Mutations	−	3	−	−	−	−
Synonymous	−	3	−	−	−	−
Non-synonymous	−	0	−	−	−	−

non-ABS. We inferred PSS for the α1 and β1 domain of MHC DR and DQ using PAML. Models M3 (discrete), M2a (positive selection) and M8 (beta&ω), which allow for positive selection at a subset of sites, gave a significantly better fit than the models without positive selection (p<0.01 in LRTs) in *DRB* and *DQA* but not *DQB* and *DRA* (Table S2.1 in supplementary material). That is, positive selection signals were detected in *DRB* and *DQA*. Parameter estimates (Table S2.1 in Electronic supplementary material) suggest that 11.9% (under M2a) and 11.7% (under M8) of *DRB* codon sites are under positive selection with ω = 9.2 (M2a) and ω = 9.4 (M8). Parameter estimates from both M2a and M8 suggest that 27.9% of DQA codon sites are under positive selection, with ω = 5 (both M2a and M8). Six *DRB* codon sites (11th, 51st, 60th, 74th, 78th and 90th) and three *DQA* codon sites (9th,

Table 2.3 Polymorphisms of four MHC class II molecules according to their domains. The derived amino acid sequences (isolated from seven European badgers *Meles meles*) were compared and the numbers of polymorphic amino acid residues are shown

Domains	Leader	β1/α1 domain	β2/α2 domain	Transmembrane domain	Cytoplasmic domain	Total
DRβ						
No. of amino acids	21 [a]	95	103	23	16	258
Polymorphic amino acid residues [b]	2	21	4	0	0	27
DQβ						
No. of amino acids	32	90	94	8 [a]	–	224
Polymorphic amino acid residues [b]	–	22	–	–	–	22
DQα						
No. of amino acids	20 [a]	82	94	11 [a]	–	207
Polymorphic amino acid residues [b]	2	18	6	0	–	26
DRα						
No. of amino acids	23 [a]	84	107	16 [a]	–	230
Polymorphic amino acid residues [b]	–	0	–	–	–	0

[a] only part of the domain is included in this study
[b] refer to Table 2.2 for the number of sequences that were compared

29th and 47th) were identified as being under positive selection, with high values under the BEB analysis (Table 2.5). Four out of six positively selected sites (11th, 74th, 78th and 90th) in DRβ1 and two out of three positively selected sites (9th and 29th) in DQα1 were within the ABS (grey boxes in Fig. 2.1 & 2.3).

Transcription analyses

Eight of 11 sequences detected in the gDNA were also amplified from the cDNA isolated from whole blood; two were detected inconsistently (Table 2.6). Pseudogene *Meme-DRB*PS01N* was detected only in gDNA, but *Meme-DQB*PS01* was detected in the cDNA, indicating the presence of a transcribed non-functional pseudogene (Mayer *et al.*, 1993; Fernandez-Soria *et al.*, 1998). The transcription of *Meme-DRB*01* and *Meme-DQA*01* was only detected in one out of seven individuals for the *DRB* and *DQA* genes (Table 2.6), although these alleles were present in their gDNA. In both cases, only one in twenty sequenced clones from cDNA was identified as *Meme-DRB*01* and *Meme-DQA*01*, with the remaining 19 clones determined to be *Meme-DRB*02* and *Meme-DQA*02*,

Table 2.4 Rates (± standard error) of non-synonymous (d_N) and synonymous (d_S) substitutions for antigen-binding site (ABS), non-ABS, and combined (ABS + non-ABS) at the European badger *Meles meles* MHC loci

Position	Number of codons	d_N	d_S	ω
DRB				
ABS	21	0.315 ± 0.096	0.181 ± 0.093	1.74
Non-ABS	68	0.029 ± 0.012	0.092 ± 0.035	0.32
Combined	90	0.087 ± 0.021	0.109 ± 0.031	0.80
DQB				
ABS	21	0.208 ± 0.080	0.264 ± 0.182	0.79
Non-ABS	68	0.147 ± 0.042	0.192 ±0.083	0.77
Combined	89	0.162 ± 0.037	0.208 ± 0.074	0.78
DQA				
ABS	19	0.395 ± 0.171	0.134 ± 0.108	2.95
Non-ABS	63	0.090 ± 0.030	0.287 ± 0.092	0.31
Combined	82	0.151 ± 0.037	0.249 ± 0.075	0.61
DRA				
ABS	21	0.000 ± 0.000	0.104 ± 0.132	0.00
Non-ABS	60	0.000 ± 0.000	0.065 ± 0.049	0.00
Combined	81	0.000 ± 0.000	0.074 ± 0.046	0.00

ω = ratio of non-synonymous to synonymous nucleotide substitution

respectively (badger 1, Table 2.6, Table S2.2 in Electronic supplementary material). This implies that some alleles have low levels of transcription in whole blood. *Meme-DRB*04* was transcribed and identified in only a single individual in this study. More than 1100 individuals from the same population have since been genotyped and more of this sequence has been found (Chapter 5).

For both *DQB* and *DRA*, only a single sequence was transcribed, in all individuals examined, even though two sequences were present in the gDNA of each gene (Table 2.6). This consistent transcription pattern of the same sequences among all individuals examined was likely also the case for *DQA*; however, *Meme-DQA*01* could be transcribed at such a low level in whole blood that it was not easy to detect by cloning and sequencing. Only *DRB* showed a high degree of variation in

Table 2.5 CODEML detection of amino acid sites under positive selection

	Codon number	Amino acid	[a] Probability $\omega > 1$ (M1a v M2a)	Mean $\omega \pm$ SE (M1a v M2a)	[a] Probability $\omega > 1$ (M7 v M8)	Mean $\omega \pm$ SE (M7 v M8)
DRB	11	T	0.904	7.75 ± 2.70	0.955*	8.02 ± 2.26
	51	I	0.937	8.03 ± 2.41	0.968*	8.13 ± 2.11
	60	G	0.961*	8.22 ± 2.15	0.982*	8.24 ± 1.93
	74	N	0.997**	8.49 ± 1.67	0.999**	8.37 ± 1.67
	78	Y	0.999**	8.51 ± 1.64	1.000**	8.38 ± 1.66
	90	T	0.962*	8.22 ± 2.15	0.980*	8.23 ± 1.95
DQA	9	S	0.938	6.99 ± 2.58	0.969*	6.96 ± 2.39
	29	E	0.947	7.07 ± 2.53	0.969*	6.98 ± 2.40
	47	Q	0.971*	7.21 ± 2.36	0.986*	7.08 ± 2.26

[a] Positively selected sites were identified in models M2a and M8 by Bayes Empirical Bayes (BEB) posterior probabilities (Yang *et al.* 2005).
* $p > 95\%$, ** $p > 99\%$. ω = ratio of non-synonymous to synonymous nucleotide substitution

Table 2.6 Presence of MHC class II sequences in genomic DNA (gDNA) and complementary DNA (cDNA) from seven European badgers *Meles meles*. **X** represents sequence from cDNA. Blank represents no detected sequence. Gray shading indicates sequence from gDNA

Gene	Sequence	Individual						
		1	2	3	4	5 [b]	6 [b]	7 [b]
DRB	*Meme-DRB*01*	X						
	*Meme-DRB*02*	X	X	X	X	X	X	X
	*Meme-DRB*03*		X				X	X
	*Meme-DRB*04*				X			
	*Meme-DRB*PS01N* [a,c]							
DQB	*Meme-DQB*01N* [c]							
	*Meme-DQB*02*	X	X	X	X	X	X	X
	*Meme-DQB*PS01* [a]	X	X			X		
DQA	*Meme-DQA*01*	X						
	*Meme-DQA*02*	X	X	X	X	X	X	X
DRA	*Meme-DRA*01N* [c]							
	*Meme-DRA*02*	X	X	X	X	X	X	X

[a] *PS* signifies pseudogenes with a frameshift (in *Meme-DQB*PS01*: a deletion occurred at the 5' primer annealing site for exon 2 amplification from gDNA; as a result, no sequence was detected from gDNA)
[b] sequences were not determined from gDNA for DQB, DQA and DRA
[c] *N* signifies the presence of genomic sequences, where no cDNA sequences were found

transcribed sequences, reflecting the sequences found in the gDNA. From the transcription analysis, it is thus probable that not all MHC class II sequences are transcribed in the blood of *M. meles*.

Phylogenetic analyses

The phylogenetic trees of the class II MHC exon 2 highlight that each gene coding for the β chain (*DRB* and *DQB*; Fig. 2.5) and the α chain (*DQA* and *DRA*; Fig. 2.6) formed a distinct and highly supported clade. For the *DRB* gene (Fig. 2.5), mustelid sequences, except for *Meme-DRB*04*, formed a distinct clade (posterior probability (PP) support = 95%). The *M. meles DRB* sequences within this clade did not form a monophyletic group, rather they were interspersed with sequences from *T. taxus* and *E. lutris*. For the *DQB* gene (Fig. 2.5) a distinct clade was formed by *M. meles* sequences: *Meme-DQB*02* and *Meme-DQB*PS01* (PP value = 100%), however this did not cluster with *Meme-DQB*01N*. For the sequences coding for the α chain (Fig. 2.6), two *M. meles DQA* sequences, and two *M. meles DRA* sequences did not form a monophyletic group and were interspersed with other mammalian *DQA* and *DRA* sequences. Although more than one *DRB* sequence was found in *T. taxus*, only one of them was identified in more than one PCR. In compliance with DLA nomenclature rules (Kennedy *et al.*, 1999), we therefore acknowledged and included only a single sequence for this gene. Only a single sequence was identified for the other three genes.

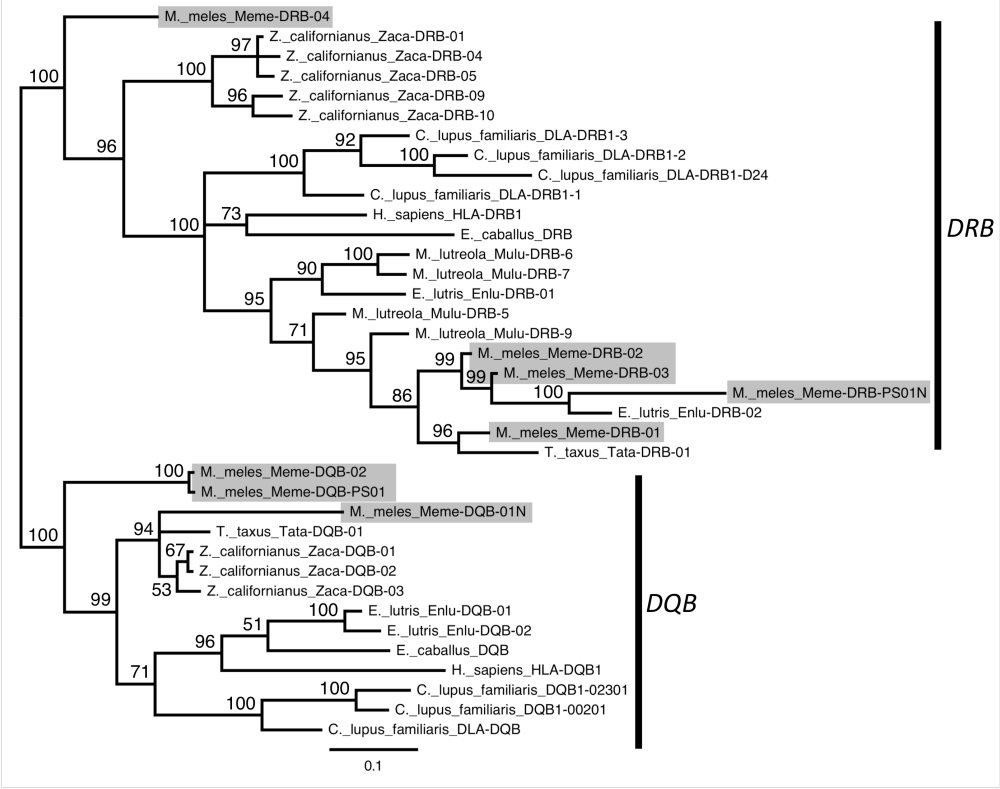


Fig. 2.5 Phylogenetic tree of MHC class II DRB and DQB exon 2 sequences from *Meles meles*, *Taxidea taxus*, and other mammal species including mustelid, phocine, canine, equine, and human (GenBank accession numbers are provided in the materials and methods), based on the 50% majority rule tree from the Bayesian analysis. Bayesian posterior probabilities above 50% are shown above the branches. *M. meles* class II sequences are marked with grey boxes

2.5 Discussion

MHC variation and selection

This is the first study to characterize class II MHC genes in European badgers and, importantly, this was performed using both the genome and transcriptome. The *DRB* genes showed higher diversity compared to the *DQB, DQA* and *DRA* genes, which among individuals from this Wytham population exhibited a nearly uniform distribution of alleles. Similarly, in other mammals, the *DRB* gene has the highest reported diversity among all class II MHC genes (Wagner *et al.*, 1999; Doxiadis *et al.*, 2001; Bowen *et al.*, 2006b). The other mammalian species tested also exhibited lower diversities for the *DQB* and *DQA* genes (Doxiadis *et al.*, 2001; Bowen *et al.*, 2002; Bowen *et al.*, 2006b; Wagner, 2003), with the *DLA-DRA* locus appearing monomorphic (Wagner, 2003).

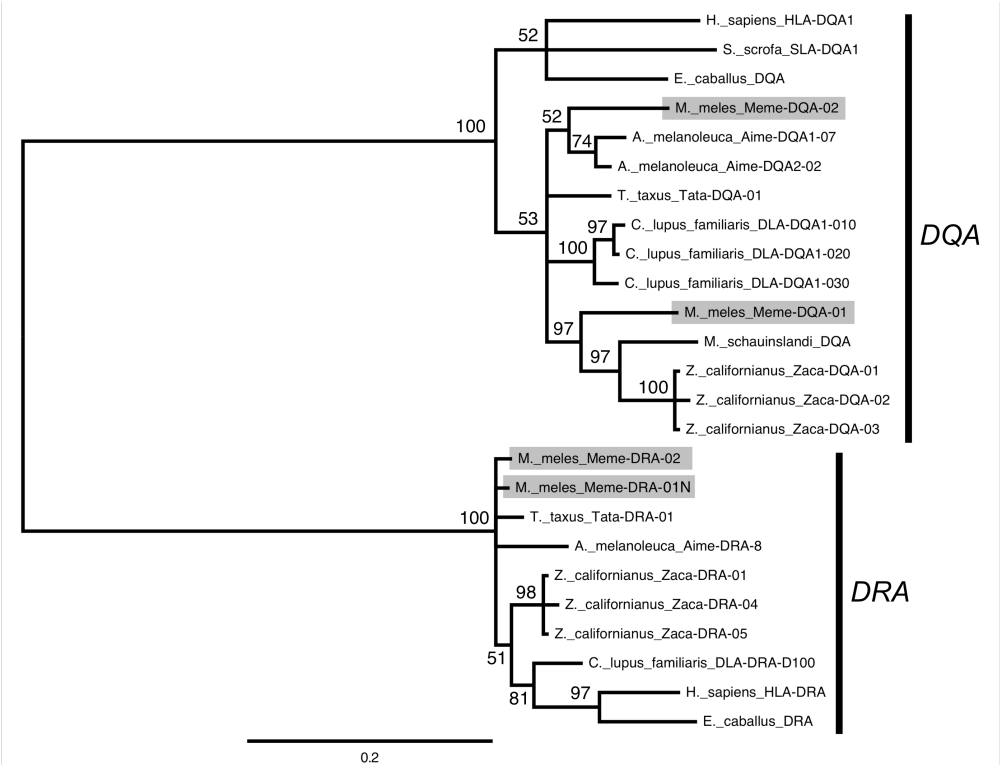


Fig. 2.6 Phylogenetic tree of MHC class II DQA and DRA exon 2 sequences from *Meles meles*, *Taxidea taxus*, and other mammal species including mustelid, phocine, canine, equine, and human (GenBank accession numbers are provided in the materials and methods), based on the 50% majority rule tree from the Bayesian analysis. Bayesian posterior probabilities above 50% are shown above the branches. *M. meles* class II sequences are marked with grey boxes

In comparison to other mustelid species (Bowen *et al.*, 2006b; Becker *et al.*, 2009), we observed moderate variability in the number of *DRB* sequences for *M. meles*. When compared to closely related species in the Carnivora, the suggested number of loci for the four class II genes that we describe here is identical to reports from a mustelid, *E. lutris* (Bowen *et al.*, 2006b), and a pinniped, *M. angustirostris* (Weber *et al.*, 2004). In addition, one of the *DRB* loci in *M. angustirostris* is monomorphic, while the second is polymorphic with two alleles, which could be similar in *M. meles*. The number of *DRB* loci in different mammalian species ranges from one to in excess of three (Takahashi *et al.*, 2000; Doxiadis *et al.*, 2001; Yuhki *et al.*, 2003; Bowen *et al.*, 2004; Becker *et al.*, 2009). The MHC diversity in these mammalian species usually arises from a high level of allelic variation at a single locus and in some species this locus may be paired with one or two relatively monomorphic loci (Robbins *et al.*, 1997). A few species with multiple *DRB* loci have low polymorphism, but haplotypes are highly variable in their gene content (see Doxiadis *et al.*, 2001;

Bowen *et al.*, 2004). In contrast to the *DRB*, the *DQB*, *DQA* and *DRA* have only one locus in many species (Takahashi *et al.*, 2000). Considering that the four *DRB* alleles that we have characterized are based on seven individuals, which come from a minuscule part of the European badger's range, this represents a high possibility to detect more alleles in other populations.

The majority of the polymorphic sites among *M. meles* sequences were found in exon 2, and there were more non-synonymous than synonymous nucleotide substitutions within this region, contributing to a higher nucleotide and amino acid sequence allelic diversity than within other exons. Exon 2 is responsible for coding the α1 or β1 domains, which play a role in presenting antigenic peptides to the immune system (Hughes & Yeager, 1998). The high variability in the exon 2 sequences of different alleles allows a population to present a wider repertoire of antigens, increasing their ability to combat parasitic infection (Hughes & Nei, 1992).

Within exon 2, some amino acid residues form the ABS (Reche & Reinherz, 2003), which is responsible for binding peptides. In the ABS, the rate of non-synonymous substitution (d_N) was higher than the rate of synonymous substitution (d_S) in *DRB* and *DQA*. By contrast, d_S was higher than d_N in the non-ABS. This higher level of d_N in the ABS, which is similar in other species (Cheng *et al.*, 2009; deBellocq *et al.*, 2009), indicates that this is the target of balancing selection, which would maintain the high levels of genetic diversity that we observed (Richman, 2000). Sites outside of the ABS showed a higher d_S than d_N and appeared to have undergone purifying selection, through which non-synonymous substitutions that affect normal gene function will usually be eliminated (Hughes & Yeager, 1998).

In our maximum likelihood models, model M0 (one ratio), which assumes a single ω over all codons in the sequence, gave results of $\omega < 1$, revealing that there has been overall purifying selection acting over *M. meles* MHC class II genes. Positive selection most likely affects only a few sites at a few time points (Yang, 2002; Nielsen, 2005) and so interpreting the signs of selection from ω averaged over a whole genetic region could be misleading. We detected sites exhibiting positive selection and an excess of non-synonymous substitutions in amino acid residues in *M. meles* MHC II *DRB* and *DQA*

genes, which demonstrates that balancing selection plays a role in maintaining these MHC sequence polymorphisms. Most of the sites of positive selection in DRβ1 and DQα1 were within the ABS. This is consistent with the theory that balancing selection is mainly act on the ABS and drives the high polymorphism in exon 2 of the MHC class II genes. In five of the six mammal species studied by (Furlong & Yang, 2008), DRβ1 sites 11L, 13F, 57D, 74A, and 86G were under positive selection. We confirm here that the 11th and 74th of these sites are also under positive selection in the European badger, implying that the selective pressures shaping variation at these sites may be driven similarly in different mammals.

Transcription pattern

Many MHC sequences can be detected at the genomic level, but not at the cDNA level (deGroot *et al.*, 2004). This could be because these sequences are pseudogenes, or due to fitness advantages conferred by suppressing some inherited genes (DuPasquier *et al.*, 1989): an over-extensive diversity of MHC molecules would minimize the T-cell repertoire of an individual during thymic selection (Nowak *et al.*, 1992; but see Borghans *et al.*, 2003). In terms of how MHC is related to mate choice and pathogen resistance, nonfunctional MHC sequences (i.e. pseudogenes) should be identified before conducting any analysis, however only a few studies (e.g. Babik *et al.*, 2009) have demonstrated the transcription of MHC class II alleles. We found that not all sequences detected in gDNA were detectable in the sequences transcribed from whole blood. Only one *DQB* and *DRA* sequence was found to be transcribed, although two sequences were detected at the genomic level for each of them.

Caution is required when interpreting transcription data, as some of the alleles found in gDNA may be transcribed at very low background levels that are hard to detect (e.g. *Meme-DRB*01* and *Meme-DQA*01* in this study). Alternatively, some alleles may be transcribed at a level high enough to be detected only if the individual is challenged with pathogen invasion, or the transcription of some alleles may vary throughout the life of an individual. In addition, some putative functional sequences may not transcribe at all. Our results suggest that some of the sequences exhibit no transcription, or lower transcription levels than the others, which would be possible if they are situated on different

loci, or if there is allele-specific transcription within a locus with transcription controlled by different promoters (Villard *et al.*, 1997; LeibundGut-Landmann *et al.*, 2004; Handunnetthi *et al.*, 2010). Interestingly, the two sequences for which the results suggest a low level of transcriprtion (*Meme-DRB*01* and *Meme-DQA*01*) were both detected in the same individual, potentially because this individual was subjected to a particular immune challenge at the time of sampling. The expression of different MHC genes can vary considerably (Handunnetthi *et al.*, 2010) due to pathogenic challenge (Chen *et al.*, 2006), or differential developmental of the immune system (Handunnetthi *et al.*, 2010). Differential MHC transcription was demonstrated in peripheral blood leukocytes of Steller sea lions (*Eumetopias jubatus*) by using quantitative PCR (Bowen *et al.*, 2006a), which identified extensive variation in the transcription of *DRB* genes even though sequences in the gDNA were identical in all individuals.

Despite the caveats above, we found that *DRB* and *DQA*, where all the detected putative functional sequences were transcribed, showed signals of positive selection. Conversely, models of neutral evolution could not be rejected in *DQB* and *DRA*, where only one of the two putative functional sequences was transcribed. The transcription pattern could be used to explain why no positive selection was detected on *DQB* and *DRA*, as in all probability one of the two sequences was not transcribed and thus not subject to positive selection. The possibility also exists that sequences were missed and it is not possible to prove that certain sequences were not transcribed in other tissues, given that we only tested blood samples. More individuals from different geographical locations would need to be examined to elucidate these transcription patterns, and the diversity of MHC class II genes fully in *M. meles*.

The *DRB* gene exhibited higher variability than the other class II genes among the sampled badgers. Furthermore, all four *DRB* gDNA sequences were transcribed. Although *Meme-DRB*01* was not detected from the cDNA of some individuals, it should still be considered to be a transcribed sequence, as it may potentially have low transcription levels. The *DRB* gene also has the highest variability among all of the class II MHC genes in a variety of other species (Wagner *et al.*, 1999; Kennedy *et al.*, 2002; Garrigan & Hedrick, 2003). The *DRB* gene in European badgers is thus a better

candidate class II MHC gene, than *DQB*, *DQA* and *DRA*, for pursuing further studies on MHC and mate choice / pathogen resistance.

Trans-specific evolution

The European badger is classified within the subfamily Melinae, which diverged from the subfamily Taxidiinae somewhere between 16–17 (Marmi *et al.*, 2004) and 20 (Yonezawa *et al.*, 2007) million years ago (Mya). Taxidiinae now only contains the extant American badger, *Taxidea taxus*, which forms the basal group in the family Mustelidae (Yonezawa *et al.*, 2007; Koepfli *et al.*, 2008). The divergence time of Melinae to other Mustelidae subfamilies, i.e. Lutrinae (otters) and Mustelinae, is estimated as 20–12 Mya (Marmi *et al.*, 2004; Yonezawa *et al.*, 2007). The superfamily Musteloidea is a sister taxon to Pinnipedia, and together with Ursidae and Canidae they form the suborder Caniformia (Flynn *et al.*, 2005).

In contrast to the broader phylogenetic relationship resolved by other studies, the MHC sequences of *M. meles* reported here do not form a monophyletic group, or cluster, with closely related taxa; rather a mixture of MHC alleles among the Pinnipedia, Ursidae and Canidae was established. This phenomenon has been observed in a number of other MHC studies (Edwards & Hedrick, 1998), and is termed trans-species polymorphism (Klein, 1987). The variability of the MHC genes in the ancestral species may have been retained even when the species diverged, thus an allele may be more closely related to an allele from another species rather than from another allele within the same species. Balancing selection (Penn & Potts, 1999), indicated by higher non-synonymous substitution in the ABS, is the proposed mechanism that maintains this ancestral variation, even though *M. meles* diverged from the other species included in this study more than 10 Mya.

Conclusions and implications

This is the first published study to describe MHC sequences in the European badger. The gene-specific primers developed in this study will facilitate future research into the genetic diversity at the

DRB, DQB, DQA, and *DRA* loci in badgers and other closely related mustelid species, and the correlation between these MHC genes and fitness-related traits. Balancing selection acts to maintain genetic diversity, evidenced here by the high diversity of the *DRB* sequences, the detection of positive selection, and trans-species polymorphism. Informatively, transcription analysis showed that the gDNA and cDNA sequence patterns were identical in *DRB*, which was moderately variable between individuals. Conversely, there were large disparities between gDNA and cDNA sequences obtained from whole blood in *DQB*, *DQA* and *DRA*. This highlights the need for future MHC studies to examine transcription patterns before investigating any further questions.

Supplementary material

Table S2.1 Positively selected sites, parameter estimates, and log-likelihood values of different models of codon evolution for the α1 and β1-domain of both MHC class II DR and DQ in European badgers

	Model	Log-likelihood	Parameter estimate(s)	Positive selected sites
DRB	M0 (one ratio)	-566.95	ω=0.320	None
	M3 (discrete)	-546.23	p_0=0.315 ω_0=0.000 p_1=0.569 ω_1=0.214 p_2=0.116 ω_2=9.412	Not allowed
	M1a (nearly neutral)	-552.55	p_0=0.700 ω_0=0.000 p_1=0.300 ω_1=1.000	Not allowed
	M2a (positive selection)	-546.25	p_0=0.880 ω_0=0.128 p_1=0.001 ω_1=1.000 p_2=0.119 ω_2=9.225	60G, **74N, 78Y**, 90T
	M7 (beta)	-552.55	p=0.005 q=0.012	Not allowed
	M8 (beta&ω)	-546.23	p_0=0.883 p=1.304 q=8.090 p_1=0.117 ω=9.384	11T, 57I, 60G, **74N**, **78Y**, 90T
DQB	M0 (one ratio)	-484.14	ω=0.292	None
	M3 (discrete)	-480.05	p_0=0.316 ω_0=0.035 p_1=0.620 ω_1=0.390 p_2=0.064 ω_2=999.0	Not allowed
	M1a (nearly neutral)	-481.28	p_0=0.687 ω_0=0.075 p_1=0.313 ω_1=1.000	Not allowed
	M2a (positive selection)	-480.10	p_0=0.927 ω_0=0.245 p_1=0.000 ω_1=1.000 p_2=0.073 ω_2=351.870	Not significant
	M7 (beta)	-481.37	p=0.106 q=0.172	Not allowed
	M8 (beta&ω)	-480.05	p_0=0.936 p=1.220 q=3.227 p_1=0.064 ω=382.0	Not significant
DQA	M0 (one ratio)	-475.93	ω= 0.731	None
	M3 (discrete)	-466.20	p_0=0.556 ω_0=0.000 p_1=0.164 ω_1=0.000 p_2=0.279 ω_2=5.037	Not allowed
	M1a (nearly neutral)	-470.86	p_0=0.576 ω_0=0.000 p_1=0.424 ω_1=1.000	Not allowed
	M2a (positive selection)	-466.20	p_0=0.721 ω_0=0.000 p_1=0.000 ω_1=1.000 p_2=0.279 ω_2=5.037	47Q
	M7 (beta)	-470.88	p=0.005 q=0.008	Not allowed
	M8 (beta&ω)	-466.20	p_0=0.721 p=0.005 q=0.783 p_1=0.279 ω=5.037	9S, 22E, 47Q
DRA	M0 (one ratio)	-335.35	ω=0.000	None
	M3 (discrete)	-335.35	p_0=0.389 ω_0=0.000 p_1=0.398 ω_1=0.000 p_2=0.214 ω_2=0.000	Not allowed
	M1a (nearly neutral)	-335.35	p_0=1.000 ω_0=0.000 p_1=0.000 ω_1=1.000	Not allowed
	M2a (positive selection)	-335.35	p_0=1.000 ω_0=0.000 p_1=0.000 ω_1=1.000 p_2=0.000 ω_2=1.000	None
	M7 (beta)	-335.35	p=0.005 q=99.0	Not allowed
	M8 (beta&ω)	-335.35	p_0=1.000 p=0.005 q=3.725 p_1=0.000 ω=1.000	None

Log-likelihood values and estimated parameters for each model were calculated using CODEML in PAML 4.4b (Yang, 2007). ω=dN/dS ratio; p_n is the proportion of amino acids in the ω_n site class. p and q are the statistical parameters of the beta distribution. Positively selected sites were identified in models M2a and M8 by Bayes Empirical Bayes (BEB) posterior probabilities (Yang *et al.*, 2005). Sites under positive selection are identified at the cutoff p > 95%, with those with p > 99% shown in bold

Table S2.2 The number of clones with different MHC class II sequences found in the seven European badger individuals. *Numbers in brackets* are clones of PCR products from cDNA. *Numbers out of brackets* are clones of PCR products from gDNA

Gene	Sequence	Individual							
		1	2	3	4	5 [c]	6 [c]	7 [c]	Total
DRB	*Meme-DRB*01*	3 (1)	4 (0)	6 (0)	4 (0)	7 (0)	2 (0)	3 (0)	29 (1)
	*Meme-DRB*02*	3 (19)	5 (7)	2 (6)	4 (1)	6 (12)	2 (1)	2 (5)	24 (51)
	*Meme-DRB*03*		2 (5)				1 (5)	3 (1)	6 (11)
	*Meme-DRB*04*				2 (5)				2 (5)
	*Meme-DRB*PS01N* [a,b]		1 (0)				1 (0)	1 (0)	3 (0)
	Total	6 (20)	12 (12)	8 (6)	10 (6)	13 (12)	6 (6)	9 (6)	64 (68)
DQB	*Meme-DQB*01N* [b]	1 (0)	4 (0)	2 (0)					7 (0)
	*Meme-DQB*02*	5 (16)	1 (11)	5 (6)	6 (6)	(11)	(6)	(6)	17 (62)
	*Meme-DQB*PS01* [a]	(4)	(1)			(1)			(6)
	Total	6 (20)	5 (12)	7 (6)	6 (6)	(12)	(6)	(6)	24 (68)
DQA	*Meme-DQA*01*	3 (1)	3 (0)	3 (0)	1 (0)				10 (1)
	*Meme-DQA*02*	3 (19)	3 (12)	4 (6)	5 (6)	(12)	(6)	(6)	15 (67)
	Total	6 (20)	6 (12)	7 (6)	6 (6)	(12)	(6)	(6)	25 (68)
DRA	*Meme-DRA*01N* [b]	4 (0)	1 (0)	2 (0)	3 (0)				10 (0)
	*Meme-DRA*02*	2 (24)	3 (16)	6 (6)	5 (6)	(12)	(6)	(6)	16 (76)
	Total	6 (24)	4 (16)	8 (6)	8 (6)	(12)	(6)	(6)	26 (76)

[a] *PS* signifies pseudogenes with a frameshift (in *Meme-DQB*PS01*: deletion occurred at the 5' primer annealing site for amplifying exon 2 from gDNA, leading to no sequence was detected from gDNA)

[b] *N* signifies the presence of genomic sequences but no cDNA sequences were found

[c] sequences were not determined from gDNA for DQB, DQA and DRA

Chapter 3

Evolution of MHC class I genes in the European badger (*Meles meles*) [2]

[2] This chapter has been published as: Sin, Y.W., Dugdale, H.L., Newman, C., Macdonald, D.W. & Burke, T. 2012. Evolution of MHC class I genes in the European badger (*Meles meles*). *Ecology and Evolution* **3:** 285. doi: 210.1002/ece1003.1285.

3.1 Abstract

The major histocompatibility complex (MHC) plays a central role in the adaptive immune system and provides a good model with which to understand the evolutionary processes underlying functional genes. Trans-species polymorphism and orthology are both commonly found in MHC genes; however, mammalian MHC class I genes tend to cluster by species. Concerted evolution has the potential to homogenise different loci, whereas birth-and-death evolution can lead to the loss of orthologs; both processes result in monophyletic groups within species. Studies investigating the evolution of MHC class I genes have been biased towards a few particular taxa and model species. We present the first study of MHC class I genes in a species from the superfamily Musteloidea. The European badger (*Meles meles*) exhibits moderate variation in MHC class I sequences when compared to other carnivores. We identified seven putatively functional sequences and nine pseudogenes from genomic (gDNA) and complementary (cDNA) DNA, signifying at least two functional class I loci. We found evidence for separate evolutionary histories of the α1 and α2/α3 domains. In the α1 domain, several sequences from different species were more closely related to each other than to sequences from the same species, resembling orthology or trans-species polymorphism. Balancing selection and probable recombination maintain genetic diversity in the α1 domain, evidenced by the detection of positive selection and a recombination event. By comparison, two recombination breakpoints indicate that the α2/α3 domains have most likely undergone concerted evolution, where recombination has homogenised the α2/α3 domains between genes, leading to species-specific clusters of sequences. Our findings highlight the importance of analysing MHC domains separately.

3.2 Introduction

The major histocompatibility complex (MHC) is of particular importance to the study of evolutionary genetics owing to its pattern of molecular evolution. MHC is a diverse gene family that plays a crucial role in the vertebrate adaptive immune system and in auto-immunity. Cell-surface glycoproteins, encoded by the MHC genes, are vital in both humoral and cell-mediated immune responses, as they bind and present antigens to T-cells and trigger an immune cascade (Swain, 1983). MHC genes are classified into groups including class I and class II. MHC class II molecules principally bind peptides from the extracellular environment and are only expressed on antigen-presenting cells, such as B-cells and macrophages (Hughes & Yeager, 1998). MHC class I genes comprise classical (class Ia) and non-classical (class Ib) loci that differ in polymorphism, structure, function and expression pattern (Parham & Ohta, 1996; Rodgers & Cook, 2005). MHC class Ia molecules are responsible primarily for intracellular antigen binding and are expressed on the surface of all nucleated somatic cells (Bjorkman & Parham, 1990). Because of this crucial role in the immune system, MHC genes are under constant selective pressures due to challenges from parasites and pathogens (Jeffrey & Bangham, 2000; Piertney & Oliver, 2006). This arms race between pathogens and hosts is posited to be the driving force for the extreme diversity in MHC genes (e.g. class Ia genes such as HLA-A, HLA-B, and HLA-C in humans; Piertney & Oliver, 2006).

The high degree of diversity among MHC alleles arises from the diverse exons that encode the domains forming the antigen-binding site (ABS; Hughes & Yeager, 1998). The nucleotide diversity within the MHC genes has been attributed to balancing selection (Bergström & Gyllensten, 1995), such as over-dominance and frequency-dependent selection, which act to maintain large numbers of alleles in populations. The persistence of ancestral allelic diversity over long periods of time, relative to neutral genetic variation (Richman, 2000), is enhanced substantially by balancing selection, which leads to high levels of allelic diversity within species (Hughes & Yeager, 1998; Penn *et al.*, 2002). Some mammalian MHC allelic lineages are more than a million years old and are maintained after speciation (Figueroa *et al.*, 1988). Phylogenetic reconstructions therefore reveal trans-species polymorphism (Klein, 1987; Klein *et al.*, 1998; Klein *et al.*, 2007), where alleles between species are more closely related (or even identical) than alleles within species. Phylogenetic reconstructions can

also reveal orthologous relationships, by which sequences group by gene rather than by species, forming orthologous gene clusters (Nei *et al.*, 1997; Nei & Rooney, 2005). In orthologous clusters the diveregnce pattern of genes reflects species phylogeny (when trans-species polymorhism does not also occur) because genes have diverged from a common ancestor due to a speciation event (Fitch, 2000).

In contrast to class II genes, which usually exhibit orthology and trans-species polymorphism in mammals, class I genes tend to form monophyletic groups within species and therefore it is difficult to establish orthologous relationships among closely related mammalian species (Takahashi *et al.*, 2000). One plausible explanation is that class I loci undergo a faster rate of birth-and-death evolution than do class II loci (Nei *et al.*, 1997; Vogel *et al.*, 1999; Takahashi *et al.*, 2000; Piontkivska & Nei, 2003), whereby new genes are created by gene duplication and some genes become non-functional through deleterious mutations, or become deleted from the genome. The orthologous relationship may therefore become lost due to birth-and-death evolution. The high divergence of class I genes between closely related species has also been posited to be due to frequent genetic exchange by recombination or gene conversion termed 'concerted evolution' (Rada *et al.*, 1990; Hess & Edwards, 2002; Nei & Rooney, 2005). This yields genes within a species that are more similar to each other than they are to the orthologs in closely related species, i.e. intra-specific paralogs are more similar than inter-specific orthologs. The homogenising effect of recombination can mask orthology and duplicaiton history and impede the reconstruction of the phylogeny. With higher rates of concerted evolution, divergence between orthologs increases, which makes the signal for orthology harder to detect. In the concerted evolution model, different loci are homogenised constantly by gene conversion (Rada *et al.*, 1990; Nei & Rooney, 2005). This tends to eliminate variation and generates intra-specific similarity while reducing inter-specific similarity, thus facilitating diversification. Exchange of sequence segments by gene conversion, and recombination, play a prominent role in the evolution of the MHC genes (Parham & Ohta, 1996; Jakobsen *et al.*, 1998; Richman *et al.*, 2003). In addition to its homogenization effect, it has been proposed that recombination creates and maintains diversity in the MHC genes (Holmes & Parham, 1985; Jakobsen *et al.*, 1998). The consequent phylogenetic inconsistencies between different regions of the same gene, due to concerted evolution, has been observed in the

MHC of many species (e.g. Shum *et al.*, 2001; Bos & Waldman, 2006; Burri *et al.*, 2008), including humans (Houlden *et al.*, 1996; Jakobsen *et al.*, 1998; Bos & Waldman, 2006).

To date, the domestic dog, *Canis lupus familiars,* has provided the model for MHC research in carnivores. In the Dog Leukocyte Antigen (DLA; Kennedy *et al.*, 2001; Wagner, 2003), one class Ia gene is present, with three class Ib genes and two pseudogenes. This class Ia gene is highly polymorphic, with more than 50 alleles identified (Wagner, 2003). Studies of the MHC class I genes in many carnivores however, such as those in the superfamily Musteloidea, are lacking. Here we characterise the MHC class I genes of the European badger (*Meles meles*). *M. meles* is well suited to investigate how MHC selection and conferred immunological advantages (e.g. pathogen and parasite resistance) are regulated by mate choice in the wild. *M. meles* has a long mating-season (Buesching *et al*., 2009), delayed implantation (Thom *et al*., 2004), putative superfoetation (Yamaguchi *et al*., 2006), and a sensory predisposition towards olfaction (Buesching *et al.*, 2002c). In high-density populations *M. meles* has a polygynandrous mating system (Dugdale *et al.*, 2007; Dugdale *et al.*, 2011b) with high levels of extra-group paternity (Dugdale *et al*., 2007) and low fecundity (Macdonald *et al*., 2009). Additionally, it has been the subject of a diverse range of endo-parasitic disease studies (Macdonald *et al.*, 1999; Anwar *et al.*, 2000; Newman *et al.*, 2001; Anwar *et al.*, 2006; Rosalino *et al.*, 2006; Lizundia *et al.*, 2011). In particular, *M. meles* is a wildlife reservoir of *Mycobacterium bovis* (Delahay *et al.*, 2001; Mathews *et al.*, 2006b; Riordan *et al.*, 2011); an intracellular bacteria that is the cause of bovine tuberculosis (bTB) in cattle and wildlife. MHC class I-dependent immunity is known to play an important role in the eradication of *M. bovis* in mice (Ladel *et al*., 1995). *M. meles* with different MHC genotypes, or the presence/absence of certain MHC alleles, may therefore have differential susceptibilities to *M. bovis*, which could contribute to genetic-based bTB control strategies in badgers.

In this study we: (1) characterised the MHC class I genes of *M. meles* from a high-density population and tested for evidence of selection and recombination; (2) identified the transcription pattern by comparing gDNA and cDNA sequences from whole blood samples, which is important as MHC genes identified using gDNA may be nonfunctional; and (3) performed phylogenetic analyses to investigate whether *M. meles* sequences belong to monophyletic groups, or whether sequences transcend species

boundaries. Characterisation of MHC class I genes in *M. meles* will clarify whether these have a more rapid turnover rate than class II loci (Sin *et al.*, 2012b) and facilitate elucidation of the underlying evolutionary processes within different regions. Moreover, the development of MHC markers will facilitate studies on the relationship between genetics and disease in this controversial animal, as well as other closely related species.

3.3 Materials and methods

Sample collection and nucleic acid isolation

Blood samples were collected from 11 badgers that resided in eight different social groups (Sin *et al.*, 2012b) in Wytham Woods, Oxfordshire, UK (global positioning system reference 51°46'26N, 1°19'19W). All trapping and handling protocols are detailed in Macdonald & Newman (2002). These protocols were subject to ethical review and were performed under Natural England Licence (currently 20104655) and UK Home Office Licence (PPL 30/2835). Approximately 3 ml of blood was taken by jugular venipuncture and collected in a vacutainer containing EDTA. Samples were stored at -20°C until DNA isolation was performed. gDNA was isolated using the GFX Genomic Blood DNA Purification Kit (Amersham Biosciences, Little Chalfont, UK), following the scalable method in the manufacturer's protocol. In order to validate whether the identified alleles were transcribed, a 500-μL blood sample, from each of the 11 individuals, was also transferred into RNAprotect Animal Blood Tubes (Qiagen, Hilden, Germany) and stored immediately at -20°C for less than a month before RNA isolation. Total cellular RNA was isolated from each blood sample using an RNeasy Protect Animal Blood Kit (Qiagen, Hilden, Germany). Methods of cDNA synthesis are detailed in Sin *et al*. (2012).

Primer design and PCR amplification

MHC class I molecules are heterodimers, that consist of an α chain and a β_2-microglobulin (β_2m) molecule. The α chain is composed of three extracellular domains α1, α2 and α3 (encoded by exon 2, 3 and 4 respectively), a trans-membrane domain (exon 5) and a cytoplasmic domain (exon 6; Fig. 1; Bjorkman & Parham, 1990). The α1 and α2 domains are α chain regions that comprise specific sites forming the ABS.

To amplify the class I genes from *M. meles*, we tested published primers used successfully in other carnivores (Zhong *et al.*, 1998; Aldridge *et al.*, 2006). Oligonucleotide primers, which recognise highly conserved regions of the MHC class I genes, were also designed using OligoAnalyzer 3.1 (Owczarzy *et al.*, 2008), based on alignments with GenBank's nucleotide sequences from domestic dog (*C. lupus familiaris*; AF218297, AF218299, AF218301, AF218303, DQ056267, DQ056268 and M32283), harbour seal (*Phoca vitulina*; U88874), domestic cat (*Felis catus*; M26318, U07670, U07672 and U07674), horse (*Equus caballus*; NM001123381) and human (*Homo sapiens*; AF287959, U03907, NM002117 and NM002127).

Using these primers on 10–30 ng of cDNA/gDNA, PCR amplification was performed in a 20-μl reaction mix that also contained: 0.5 μM of each primer (Table 3.1), 200 μM of each dNTP, 1 X PCR buffer (containing $MgCl_2$; Qiagen), and 2 units of HotStarTaq (Qiagen). The PCR cycle began with incubation at 94°C for 15 min, followed by 35 incubation cycles at 94°C for 30 s, annealing temperature (Table 3.1) for 30 s, and 72°C for 60–90 s according to amplicon length, ending with an extension step at 72°C for 10 min. The PCR products were electrophoresed on a 1.5% agarose gel and visualised using UV light and ethidium bromide staining. A 100-bp DNA ladder (New England Biolabs, Herts, UK) was used to size the DNA fragments. Bands of expected size were excised from the gel and purified using QIAquick Gel Extraction Kits (Qiagen). PCR products that gave rise to relatively bright bands of the expected size were cloned and sequenced.

Cloning and DNA sequencing

Purified PCR fragments were cloned using the pGEM-T Easy Vector Systems (Promega, Madison, WI, USA). The cloning and sequencing procedure is detailed in Sin *et al.* (2012). Between 9 and 72 clones were sequenced for each individual. Identical sequences were derived from a minimum of two badgers or from independent PCR reactions from the same individual, in compliance with dog leukocyte antigen (DLA) nomenclature rules (Kennedy *et al.*, 1999). Single unique sequences (possible chimeras) were excluded. Nucleotide sequences were analysed using CodonCode Aligner 3.7.1 (CodonCode Corporation, Dedham, MA, USA) and were compared with known MHC class I

Table 3.1 MHC class I-specific primers for *M. meles*

Primer name	Primer sequence	Product size (bp): Region amplified	T_a (°C)	Source reference
F: Meme-MHCIex1F	GGCCCTGGCCGTGACC	1012 bp; exon 1 – exon 6	59	This study
R: Meme-MHCI-ex6R	ATCAGAGCCCTGGGCACTGTC			This study
F: Meme-MHCIex2F	GGCTCCCACTCCCTGAGG	543 bp: exon 2 – exon 3 on cDNA	59	This study
R: Meme-MHCIex3R	GCGCAGCAGCGACTCCTT	731-784 bp: exon 2 – exon 3 on gDNA		This study
R: PpLAa1L250	GGCCTCGCTCTGGTTGTAG	270 bp; exon 2	55	Aldridge *et al*. 2006
F: Meme-MHCIex3F	GGGTCTCACACCATCCAG	(Pairs with Meme-MHCIex3R) 273 bp; exon 3	59	This study

F = forward, R = reverse. bp = base pair. T_a = annealing temperature.

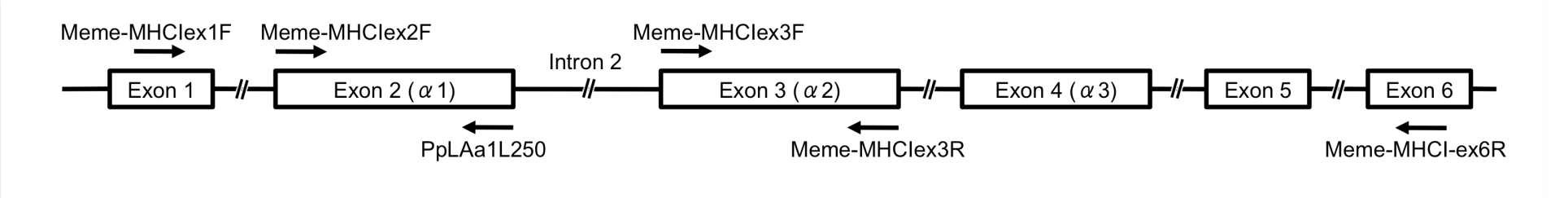


Fig. 3.1 Schematic representation of the positions of the primers used for amplification of MHC class I sequences from cDNA/gDNA. α1−3 domains are labeled at the exons encoding them.

sequences using the NCBI BLAST program (Altschul *et al*., 1990). The DNA sequences from *M. meles* were assigned the GenBank accession numbers JQ425427−JQ425447.

Data analyses

Selection and recombination

Selection at the amino acid level was measured as the rates of non-synonymous (d_N) and synonymous (d_S) substitutions per codon site, estimated in DnaSP 4.0 (Rozas *et al*., 2003) and MEGA 4 (Tamura *et al*., 2007) according to the method of Nei and Gojobori (1986), with Jukes and Cantor (1969) correction. Standard errors were derived from 1000 bootstrap replicates. Synonymous and non-synonymous substitutions were calculated separately for the ABS and non-ABS, as determined by Bjorkman *et al*., (1987). CODEML in PAML 4.4b (Yang, 2007) was used to check for positively selected sites (PSS) in the α1 and α2 domains, which are indicated where the ratio ω (non-synonymous/synonymous substitution rate ratio, d_N/d_S) exceeds 1, meaning that nucleotide mutations that alter the amino acid sequence of a protein occur more frequently than nucleotide mutations that do not alter amino acids. Thus, $\omega > 1$ is indicative of positive selection whereby beneficial amino acid changes are fixed. Codon-based likelihood analysis was used to test for evidence of positive selection, using several models: M1a (nearly neutral), M2a (positive selection), M7 (beta) and M8 (beta & ω). The assumptions of these models are detailed in Yang *et al.* (2000; 2005). Two null models of neutral evolution (M1a and M7) were applied and compared against their nested models, which allow stringent testing for positive selection (M2a and M8, respectively, which assume a different distribution of mutations; Anisimova *et al.*, 2001, 2002). We determined whether the alternative models (M2a and M8) provided a significantly improved fit, versus their null models (M1a and M7, repectively), using a likelihood ratio test (LRT), which compares twice the difference of the log likelihood ratios (2ΔInL) to a χ^2 distribution. To identify codons under positive selection (posterior probabilities >0.95), through comparisons of M1a versus M2a and M7 versus M8, CODEML was used to calculate the Bayes Empirical Bayes (BEB) posterior probabilities (Yang *et al*., 2005) at each codon. Different domains were analysed separately, as their evolutionary histories could be different due to recombination events in the intervening intron.

Recombination analyses were performed on the nucleotide alignment spanning exon 2, intron 2 and exon 3 using RDP3 alpha 44 (Martin *et al*., 2010). Four methods, i.e. RDP (Martin & Rybicki, 2000), GENECONV (Padidam *et al*., 1999), MaxChi (Smith, 1992), and Bootscan (Martin *et al*., 2005), were applied in the first run to detect recombination events. Default settings were applied with a maximum *p*-value of 0.05, applying Bonferroni correction for multiple comparisons. Any recombination signals, which were detected by at least three methods, were then re-checked with all available methods (Martin *et al*., 2010). Any putative recombination events detected were verified further by examination of MaxChi plots and matrices and Neighbour-Joining trees from the inferred fragments, to assess recombinant designation and breakpoint placement. Only the recombination events that were confirmed after these procedures were considered significant. The effects of recombination and gene conversion on sequence evolution are similar in small sequence fragments, therefore we did not differentiate between them, and we refer to them as recombination *sensu lato* hereafter (Richman *et al.*, 2003; Burri *et al.*, 2008).

Phylogenetic analyses

Phylogenetic analyses were performed on the consensus alignments of *M. meles* MHC class I exon 2, exon 3 and exon 4 sequences, against sequences from other species available in GenBank [e.g. *C. lupus familiaris* (accession number: M32283, NM001014378, NM001014379, NM001014767, NM001020810, U55029), *P. vitulina* (PVU88874), *F. catus* (U07672, U07674), *E. caballus* (LOC100056062), *H. sapiens* (AF287959, NM002117.4, NM002127.5, U03907), *Ailuropoda melanoleuca* (EU162658, EU162659), *Leopardus pardalis* (U07678), *Acinonyx jubatus* (U07666), *Bos taurus* (AB245424) and *Monachus schauinslandi* (Aldridge *et al.* 2006)]. Tree reconstruction was performed separately on three domains, in order to maximise the detection of different evolutionary histories due to genetic exchange. Domain borders were assigned according to Koller and Orr (1985).

Phylogenetic networks allow visualisation of reticulate phylogenetic signals (Huson & Bryant, 2006) whereas phylogenetics trees may poorly describe complex evolutionary scenarios; thus, phylogenetic networks provide an effective way of evaluating evolutionary relationships involving gene duplication and recombination. We used the Neighbor-Net algorithm in SplitsTree 4.12.3 (Huson & Bryant, 2006)

to analyse the phylogenetic relationships for exon 2, exon 3 and intron 2. We constructed a Neighbor-Net network based on uncorrected *p*-distances and conducted 1000 bootstrap replicates to estimate their support.

Bayesian phylogenetic inference was performed using MrBayes 3.1.2 (Ronquist & Huelsenbeck, 2003) to analyse the phylogenetic relationship of exon 4. We used the web-based application FindModel (http://www.hiv.lanl.gov/content/sequence/findmodel/findmodel.html) to ascertain the best-fit model of nucleotide substitution, which was identified as the Hasegawa-Kishino-Yano plus gamma model (HKY+Γ). A Markov chain Monte Carlo (MCMC) search was initiated with random trees and run for 3 000 000 generations, sampling every 100 generations. Data were partition by gene codon positions. Two separate analyses and four independent chains were executed. Convergence was indicated when the average standard deviation of split frequencies was less than 0.01 (Ronquist *et al.*, 2005). We also checked for convergence by plotting the likelihood scores against generation and discarded the first 25% of the generations as "burn-in".

3.4 Results

Diversity and transcription of MHC class I sequences

Sixteen different sequences were isolated from gDNA and cDNA, using the primers detailed in Table 3.1, Fig. 3.1 and Table S3.1. A maximum of four putatively functional sequences was derived from a single individual, indicating the presence of at least two loci. Transcription analysis showed that nine out of 16 sequences detected in the gDNA were also amplified from the cDNA, which was isolated from whole blood (Fig. 3.2; Table 3.2), while four sequences were detected only in the gDNA (Fig. 3.2).

Pseudogene features (Fig. 3.2) were found in four sequences: nucleotide deletions caused a frame-shift for *Meme-MHC I*PS01*, *Meme-MHC I*PS02*, *Meme-MHC I*PS03* and *Meme-MHC I*PS04N*, which caused premature stop codons (Fig. 3.2; *PS* signifies pseudogenes with a frame-shift, *N* signifies the

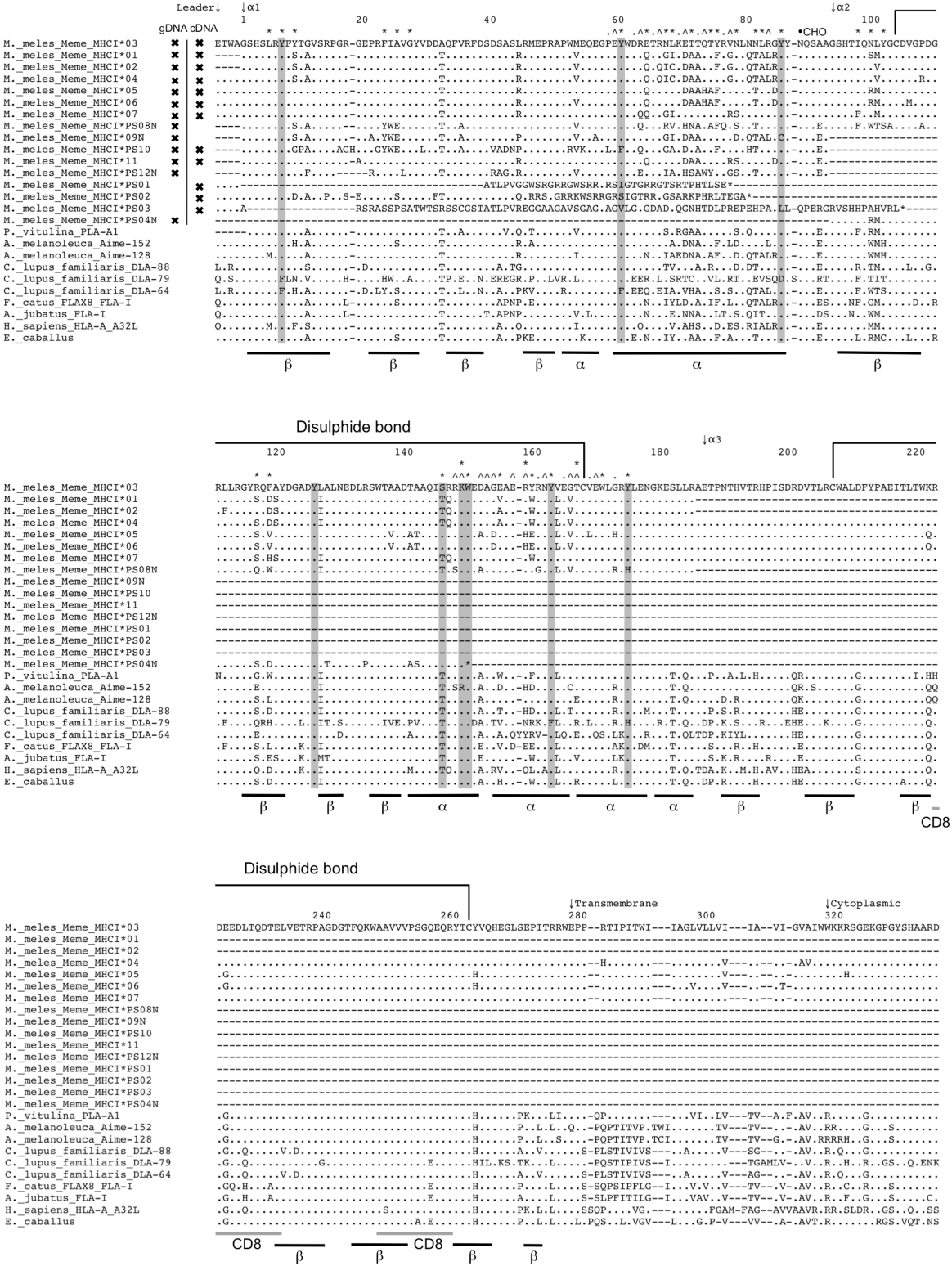


Fig. 3.2 Amino acid sequence identity for class I sequences of *M. meles* and seven other mammals (*P. vitulina*, *A. melanoleuca*, *C. lupus familiaris*, *F. catus*, *A. jubatus*, *E. caballus* and *H. sapiens*; GenBank accession numbers are provided in the materials and methods). The complete amino acid sequence of *Meme-MHC I*03* is shown. *PS* signifies pseudogenes with a frame-shift. *N* signifies the presence of genomic sequences, where no cDNA sequences were found. *Single letters* and *dots* within the alignment/sequence represent amino acids that are distinct from or identical to *Meme-MHC I*03*, (**Fig. 3.2 continued**) respectively. *Dashes* (-) indicate missing sequences. *Numbers above the sequence* indicate the codon position. *Arrows above the sequence* label the beginning of a domain. *Asterisks* (*) indicate amino acid residues pointing towards the postulated antigen binding site, which

were defined according to Bjorkman *et al*. (1987). *Carets* (^) indicate residues pointing up on an alpha-helix, postulated to interact with peptides and/or T cell receptors (TCRs), and *dots above the sequence* (.) indicate residues on an alpha-helix that is pointing away from the antigen binding site, postulated to interact with TCRs (Bjorkman *et al.*, 1987). Conserved sites that bind the peptide N- and C-termini are marked with *grey boxes*. The location of the N-linked glycosylation site is at position 88 (•CHO). Disulphide bonds formed between cysteine residues are shown with a line spanning the two cysteine residues. Residues that form the β-sheet or α-helix, and residues that influence the binding of the CD8 glycoprotein, are marked under the alignment. The columns on the left of the sequences indicate whether the sequences were found in gDNA and/or cDNA.

presence of genomic sequences, where no cDNA sequences were found). Pseudogene *Meme-MHC I*PS04N* was detected only in gDNA, whereas *Meme-MHC I*PS01*, *Meme-MHC I*PS02* and *Meme-MHC I*PS03* were only detected in the cDNA, indicating the presence of a transcribed non-functional pseudogene (Mayer *et al.*, 1993; Fernandez-Soria *et al.*, 1998). The nucleotide deletion in *Meme-MHC I*PS01 and *Meme-MHC I*PS03* occurred at the 5' primer annealing site for exon 2 amplification from gDNA, thus no sequence was detected from gDNA. Only exon 2 was amplified for four sequences (*Meme-MHC I*09N*, *Meme-MHC I*PS10*, *Meme-MHC I*11* and *Meme-MHC I*PS12N*), in which the insertion or deletion of nucleotides was found in two of these (Fig. 3.2). The intron 2 of *Meme-MHC I*PS08N* is 45-56 bp longer than other putatively functional sequences; a longer intron was also found in a DLA pseudogene (DLA-53; Wagner, 2003; Fig. 3.3 & 3.4c). Accordingly, the above sequences were regarded as pseudogenes and were not included in the nucleotide substitution calculations. Only sequences without pseudogene features, and with both exon 2 and exon 3 amplified from both gDNA and cDNA, were regarded as functional sequences and included in further analyses of nucleotide substitution and selection (i.e. *Meme-MHC I*01−Meme-MHC I*07*; Table 3.2; Fig. 3.2).

Of the seven putatively functional sequences, all were detected in both gDNA and cDNA (Table 3.2; Fig. 3.2). *Meme-MHC I*05* was detected in all individuals. By contrast, *Meme-MHC I*01* and *Meme-MHC I*02* were identified and expressed in only one individual in this study (Table 3.2). 1153 individuals from the same population have been genotyped subsequently and these sequences have been identified in more individuals (330 for *Meme-MHC I*01* and 132 for *Meme-MHC I*02*; Chapter 5).

Table 3.2 Presence of MHC class I sequences in genomic DNA (gDNA) and complementary DNA (cDNA) from 11 *M. meles*. The total number of clones with different sequences found is given. Total number of clones: numbers out of brackets are clones of PCR products from gDNA (numbers in brackets are clones of PCR products from cDNA)

	Individual												Total no.
Sequence	1	2	3	4	5	6 *	7 *	8 *	9 †	10 †	11 †	All	of clones
*Meme-MHC I*01*					X							X	3 (2)
*Meme-MHC I*02*				X								X	8 (5)
*Meme-MHC I*03*			X				X	X				X	3 (27)
*Meme-MHC I*04*	X	X	X	X	X	X	X					X	7 (72)
*Meme-MHC I*05*	X	X	X	X	X	X	X	X				X	88 (49)
*Meme-MHC I*06*	X							X				X	12 (14)
*Meme-MHC I*07*	X							X				X	5 (26)

X represents sequence from cDNA. Blank represents no detected sequence. Grey shading indicates sequence from gDNA.
* gDNA was not sequenced. † cDNA was not sequenced.

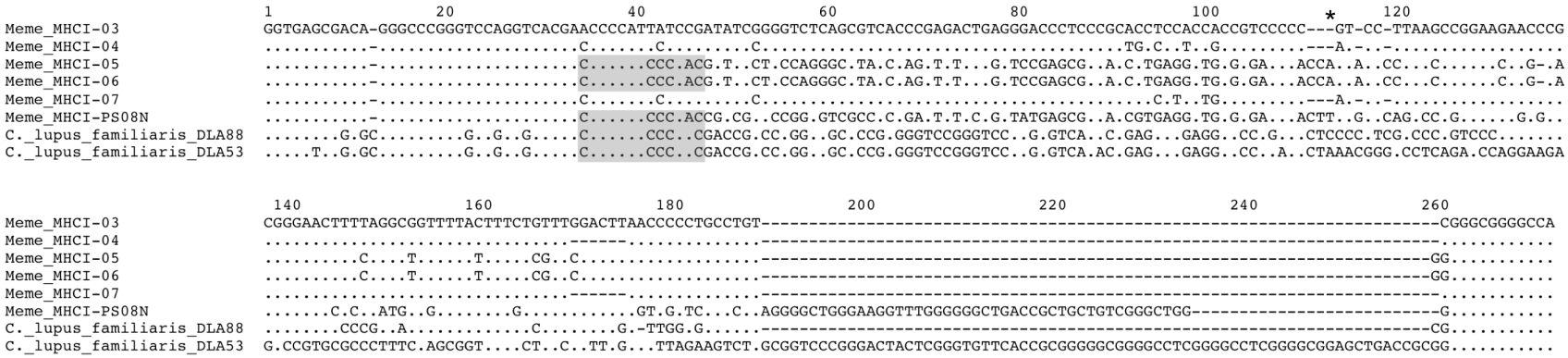


Fig. 3.3 Nucleotide sequence identity for the MHC class I intron 2 of *M. meles* clones and *C. lupus familiaris*. The GenBank accession numbers for sequences from *C. lupus familiaris* are NW876254.1 and U55029.1. The complete nucleotide sequence of *Meme-MHC I*03* is shown; *numbers above the sequence* indicate the nucleotide position. *PS* signifies pseudogenes with a frame-shift. *N* signifies the presence of genomic sequences, where no cDNA sequences were found. *Single letters* and *dots* (.) represent nucleotides that are distinct from or identical to *Meme-MHC I*03*, respectively. *Dashes* (-) indicate missing sequences. The degenerate 13−bp sequence motif (CCNCCNTNNCCNC) that is crucial in crossover events at human recombination hotspots (Myers *et al.*, 2008) is marked with *grey boxes*. The nucleotide breakpoint for *Meme-MHC I*05* and *Meme-MHC I*06* is marked with an *asterisk* (*) above the alignment.

Table 3.3 Sequence polymorphism of MHC class I genes/molecules delineated by exon/domain (leader, α1, α2, α3, transmembrane and cytoplasmic domains encoded by exon 1, 2, 3, 4, 5 and 6, respectively). The number of nucleotide and derived amino acid sequences isolated from 11 *M. meles* were compared and the numbers of synonymous and non-synonymous nucleotide substitutions and polymorphic amino acid residues are shown

Exons	1 *	2	3	4	5	6 *
Sequences for comparison	5	7	7	5	5	5
Variable sites	0	37	44	5	7	2
Mutations	0	39	47	5	7	2
Synonymous	0	7	14	2	1	1
Non-synonymous	0	28	26	3	6	1
No. of amino acids	4	90	92	92	29	18
Polymorphic amino acid residues	0	23	24	3	6	1

* only part of the exon is included in this study

The majority of variable sites in the identified sequences were in exon 2 and exon 3, where most mutations represented a non-synonymous nucleotide substitution (Table 3.3). As a consequence, there were more polymorphic amino acid residues among the α1 and α2 domains (Fig. 3.2, Table 3.3), which are involved in antigen binding. The average sequence divergence was 6.5% (SE = 0.62%); 7.0% (SE = 0.84%) and 1.0% (SE = 0.25%) in exon 2, exon 3 and exon 4 respectively. Two of the sequences, *Meme-MHC I*02* and *04*, were highly similar, differing from each other at only one nucleotide position in exon 2 (Fig. 3.4a) and five positions in exon 3 (Fig. 3.4b). *Meme-MHC I*03* was highly divergent (9.3−9.6%) from *Meme-MHC I*01*, *02* and *04*; *Meme-MHC I*07* was highly divergent (9.3−9.6%) from *Meme-MHC I*02* and *04*, in the exon 2 region. In the exon 3 region *Meme-MHC I*05* and *06* were highly divergent from the other 5 sequences (8.8−12.3%). The intron 2 region of *Meme-MHC I*05* and *06* was identical (Fig. 3.3 & 3.4c) yet both of them were highly divergent from other putatively functional sequences (28.3−30.1%).

Nine amino acid residues (three in the α1 domain and six in the α2 domain) are located at the two ends of the binding groove of the MHC molecule, where these residues bind the N- and C- termini of the

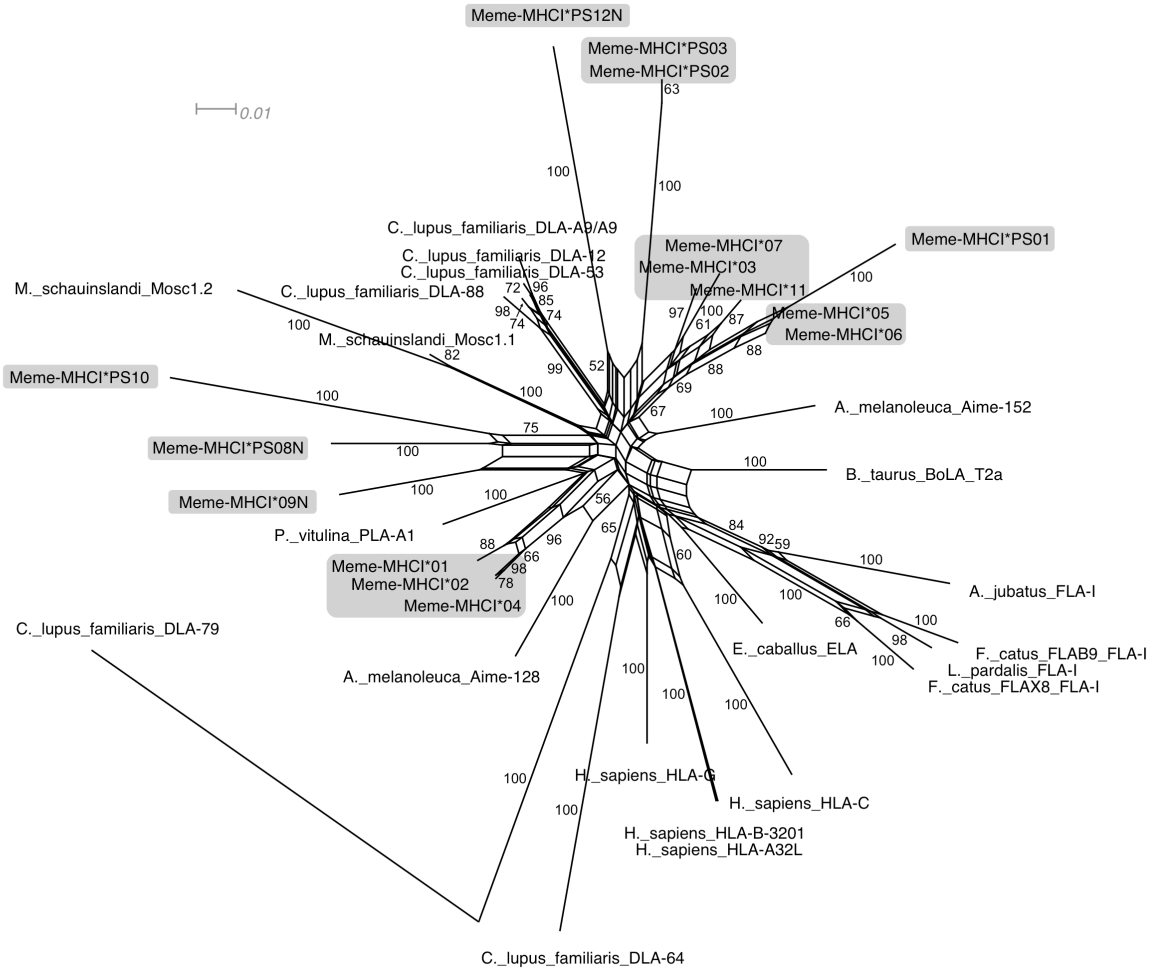


Fig. 3.4a

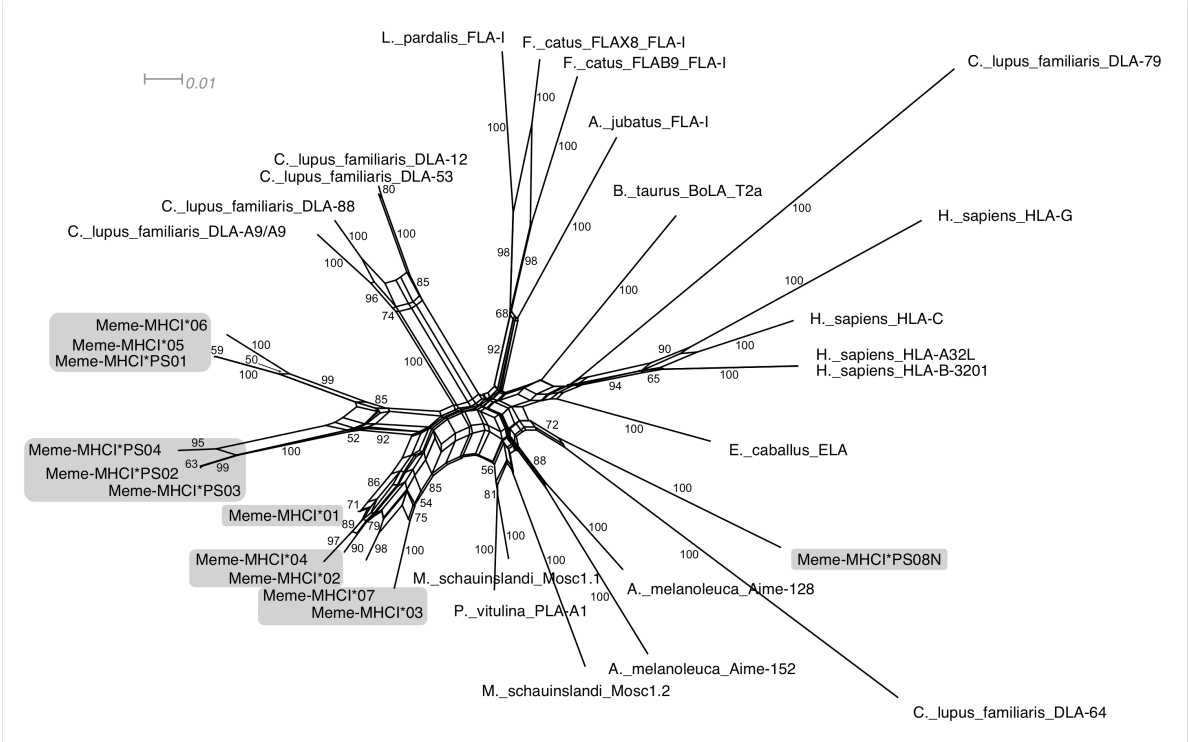


Fig. 3.4b

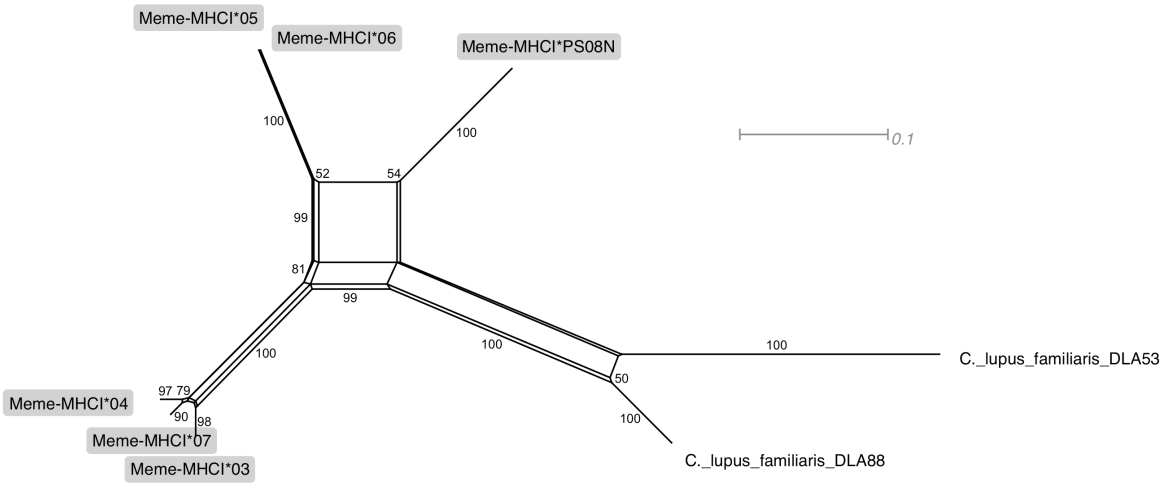


Fig. 3.4c

Fig. 3.4 Neighbour-Net of MHC class I (a) exon 2/α1-domain, (b) exon 3/α2-domain, and (c) intron 2 sequences from *M. meles* and other mammalian species including phocine, ursine, canine, feline, equine, and human (GenBank accession numbers are provided in the materials and methods). *M. meles* class I sequences are marked with grey boxes. *PS* signifies pseudogenes with a frame-shift. *N* signifies the presence of genomic sequences, where no cDNA sequences were found. Only bootstrap support values above 50 are shown; the central webbing indicates loops that have conflicting support for alternative branching events. Inclusion of 45 DLA-88 alleles (i.e. different major allelic types, Kennedy *et al.*, 1999; Kennedy *et al.*, 2001) formed a distinct clade comprising DLA-88, 12, 53 & A9, as shown in Fig. 3.4, therefore only one allele for DLA-88 is shown.

presented peptide; these residues are highly conserved across vertebrates (Y7, Y59, Y84, Y123, T143, K146, W147, Y159 and Y171; Kaufman *et al.*, 1994; Shum *et al.*, 1999; Mesa *et al.*, 2004). In *M. meles* all of these residues except one were conserved (Fig. 3.2). Residue T143 is polymorphic in *M. meles* and four sequences have the common threonine (T), which is the same amino acid seen in six out of seven species included in the sequence alignment (Fig. 3.2). The seventh species, *E. caballus*, has a serine (S) instead, which is the same as in the other three *M. meles* sequences.

The *M. meles* sequences also coded for other conserved residues, characteristic of class I molecules, which are important for the molecular structure and peptide binding. These include four cysteine (C) residues (positions 103, 167, 206 and 262; Fig. 3.2) that form the intra-domain disulphide bonds, an N-linked glycosylation site in the α1 domain (position 88; Fig. 3.2); and a threonine (T) residue at position 136 (Fig. 3.2) that is critical for interactions with T-cell antigen processing (TAP) complexes and endogenous peptide loading (Peace-Brewer *et al*., 1996). Two regions in the α3 domain (position 222-232 and 248-259; Fig. 3.2), bearing many negatively charged residues, interact with the positively

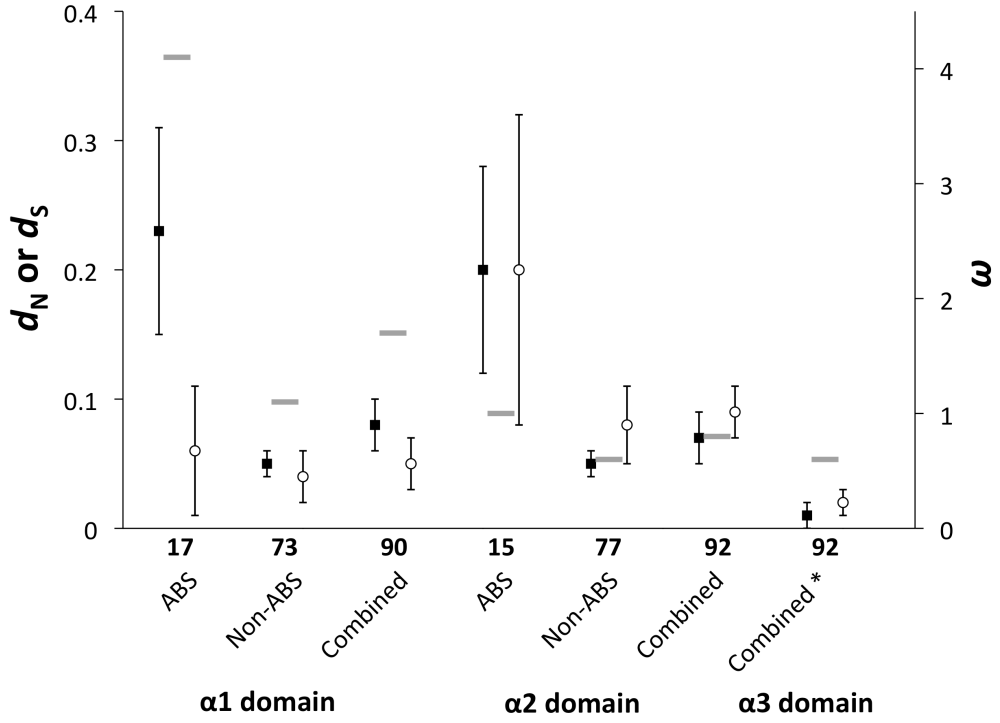


Fig. 3.5 Rates (± standard error) of non-synonymous (d_N; solid marker) and synonymous (d_S; open marker) substitutions and ratio of d_N to d_S (ω; grey bar) for antigen-binding site (ABS), non-ABS, and combined (ABS + non-ABS) at the *M. meles* MHC class I loci. The number of codons for each region is given under the x-axis. *Asterisk* (*) indicates no ABS in α3 domain.

charged side-chains of the CD8 glycoprotein on the T-cells (Kaufman *et al*., 1994). Twenty of the 23 residues in these two regions are conserved in the *M. meles* sequences, except that glutamic acid (E224) replaces glycine (G) in *Meme-MHC I*03*, *Meme-MHC I*04* and *Meme-MHC I*07*, leucine (L227) replaces glutamine (Q) in all sequences, and glutamine (Q256) replaces glutamic acid (E) in all sequences. The glutamine (Q229) residue (Fig. 3.2), which is important for CD8 accessory functions (Durairaj *et al*., 2003), is conserved here and in other mammalian, avian and reptilian sequences (Kaufman *et al*., 1994).

Selection and recombination

Rates of non-synonymous substitutions were higher than synonymous substitutions in the ABS in both the α1 and α2 domains and the non-ABS of the α1 domain (Fig. 3.5). Synonymous substitutions, however, were higher than non-synonymous substitutions in the α3 domain (which has no ABS) and the non-ABS in the α2 domain. PAML models M2a (positive selection) and M8 (beta & ω), which permit positive selection at a subset of codon sites, gave a significantly better fit than did the models

without positive selection in the α1 domain (Table 3.4). That is, positive selection signals were detected in the α1 domain. In the α2 domain, however, models that allow for positive selection did not provide a significantly better fit than the models of neutral evolution (Table 3.4). Parameter estimates (Table 3.4) indicate that 3.8% (under both M2a and M8) of α1 amino acid sites are under positive selection with ω = 13.7 (M2a) and ω = 14.4 (M8). Two sites were identified as being under positive selection (Table 3.4; Fig. 3.6); both PSS were within the ABS (Fig. 6; Kaufman *et al.*, 1994).

One significant recombination event was detected in *M. meles* class I sequences, indicating that both *Meme-MHC I*05* and *Meme-MHC I*06* were likely to have originated from *Meme-MHC I*PS08N* and *Meme-MHC I*04*. The recombination signals were significant for RDP, GENECONV, BootScan, SiScan (all $p < 0.001$), MaxChi and Chimera (both $p < 0.01$). Two recombination breakpoints were determined of which one was situated in exon 2 (nucleotide position 216 from the beginning of exon 2; MaxChi: $p < 0.001$) and one in intron 2 (nucleotide position 382; MaxChi: $p < 0.001$; Fig. 3.3).

Phylogenetic analyses

The phylogenetic tree of α1 (Fig. 3.4a) highlights that sequences from *M. meles* did not form a monophyletic clade, rather they were intermingled with phocine, canine and ursine sequences. No sequences from the other species included in the analysis formed a monophyletic clade, except for those from *H. sapiens* and *M. schauinslandi*. But the support of clades with mixed sequences from difference species was not strong (e.g. bootstrap value = 56 for the clustering of *Aime*-128 and *Meme-MHC I*01, 02* and *04*). Some *M. meles* pseudogenes, however, formed highly or moderately supported clades without grouping with any putatively functional sequences (e.g. *Meme-MHC I*PS02* and *Meme-MHC I*PS03*; *Meme-MHC I*PS08N*, *Meme-MHC I*09N* and *Meme-MHC I*PS10*). For the sequences encoding the α2 domain (Fig. 3.4b) all *M. meles* sequences, except for *Meme-MHC I*PS08N*, formed a monophyletic clade. Within this clade, pseudogenes *Meme-MHC I*PS04N*, *Meme-MHC I*PS02* and *Meme-MHC I*PS03* grouped together. Sequences from *A. melanoleuca* and *H. sapiens* also formed monophyletic clades. For α3 sequences, *M. meles* formed a distinct and highly supported clade [posterior probability (PP) support = 1.0; Fig. 3.7] that grouped together with

Table 3.4 Positively selected sites (PSS), parameter estimates, log-likelihood values of different models of codon evolution and summary of test statistics for the likelihood-ratio test (LRT) of the α1 and α2-domain of MHC class I genes in *M. meles*

Domain	Model name		Log-likelihood	Parameter estimate(s)						PSS	LRT	Test statistic	*p* value
α1	M1a	- nearly neutral	-588.31	p_0=0.61	ω_0=0.00	p_1=0.39	ω_1=1.00			Not allowed	M1a vs M2a	12.78	<0.002
	M2a	- positive selection	-581.92	p_0=0.51	ω_0=0.00	p_1=0.45	ω_1=1.00	p_2=0.04	ω_2=13.74	**66N**, 67L			
	M7	- beta	-588.32	p=0.01	q=0.01					Not allowed	M7 vs M8	12.74	<0.002
	M8	- beta & ω	-581.95	p_0=0.96	p=0.01	q=0.01	p_1=0.04	ω=14.36		**66N**, 67L			
α2	M1a	- nearly neutral	-623.67	p_0=0.66	ω_0=0.00	p_1=0.34	ω_1=1.00			Not allowed	M1a vs M2a	0	1
	M2a	- positive selection	-623.67	p_0=0.66	ω_0=0.00	p_1=0.26	ω_1=1.00	p_2=0.08	ω_2=1.00	*NS*			
	M7	- beta	-623.79	p=0.01	q=0.03					Not allowed	M7 vs M8	0.24	0.89
	M8	- beta & ω	-623.67	p_0=0.66	p=0.01	q=70.51	p_1=0.34	ω=1.00		*NS*			

Log-likelihood values and estimated parameters for each model were calculated using CODEML in PAML 4.4b (Yang, 2007). $\omega = d_N/d_S$ ratio; p_n is the proportion of amino acids in the ω_n site class. *p* and *q* are the statistical parameters of the beta distribution. PSS were identified in models M2a and M8 by Bayes Empirical Bayes (BEB) with posterior probabilities (PP) > 0.95 (Yang *et al.*, 2005); sites with with PP > 0.99 are shown in bold. *NS* not significant. Test statistics = twice the difference of the log likelihood ratios (2ΔInL). Degree of freedom is two for all LRTs. *p* value determined by comparing the test statistic to a χ^2 distribution

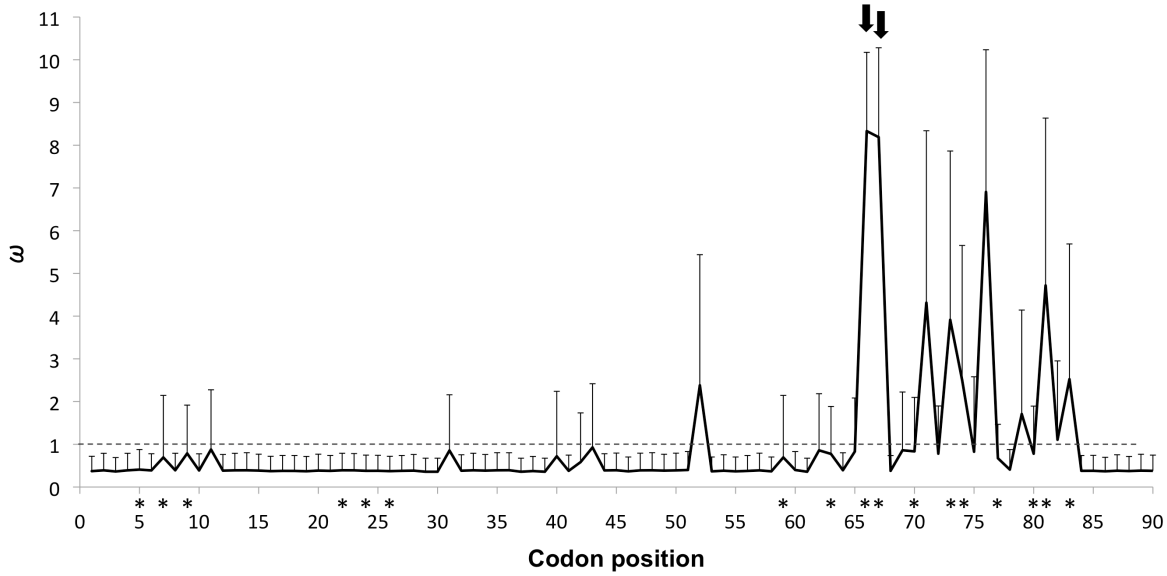


Fig. 3.6a

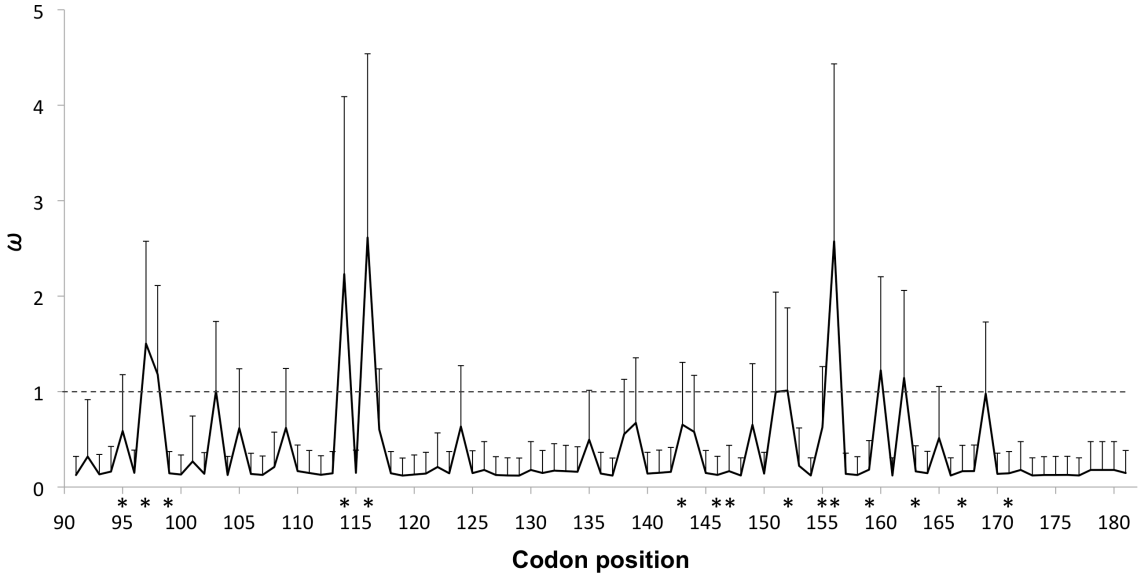


Fig. 3.6b

Fig. 3.6 Mean ω at each codon position across (a) exon 2 and (b) exon 3 of *M. meles* class I genes. Means of ω were calculated as the average of ω over the 11 site classes, weighted by the posterior probabilities under M8 computed by CODEML. *Asterisks* (*) indicate codons coding for putatively antigen-binding sites that were defined according to Bjorkman *et al*. (1987). The dotted line indicates ω=1. *Error bars* indicate the S.E. of the mean. *Arrows* indicated significant positively selected sites identified by the BEB procedure ($p > 0.95$).

P. vitulina and *A. melanoleuca* (PP = 0.67). This clade further grouped with all carnivores in our comparison including canine and feline sequences (PP = 0.86), separating it from equine, bovine and human sequences.

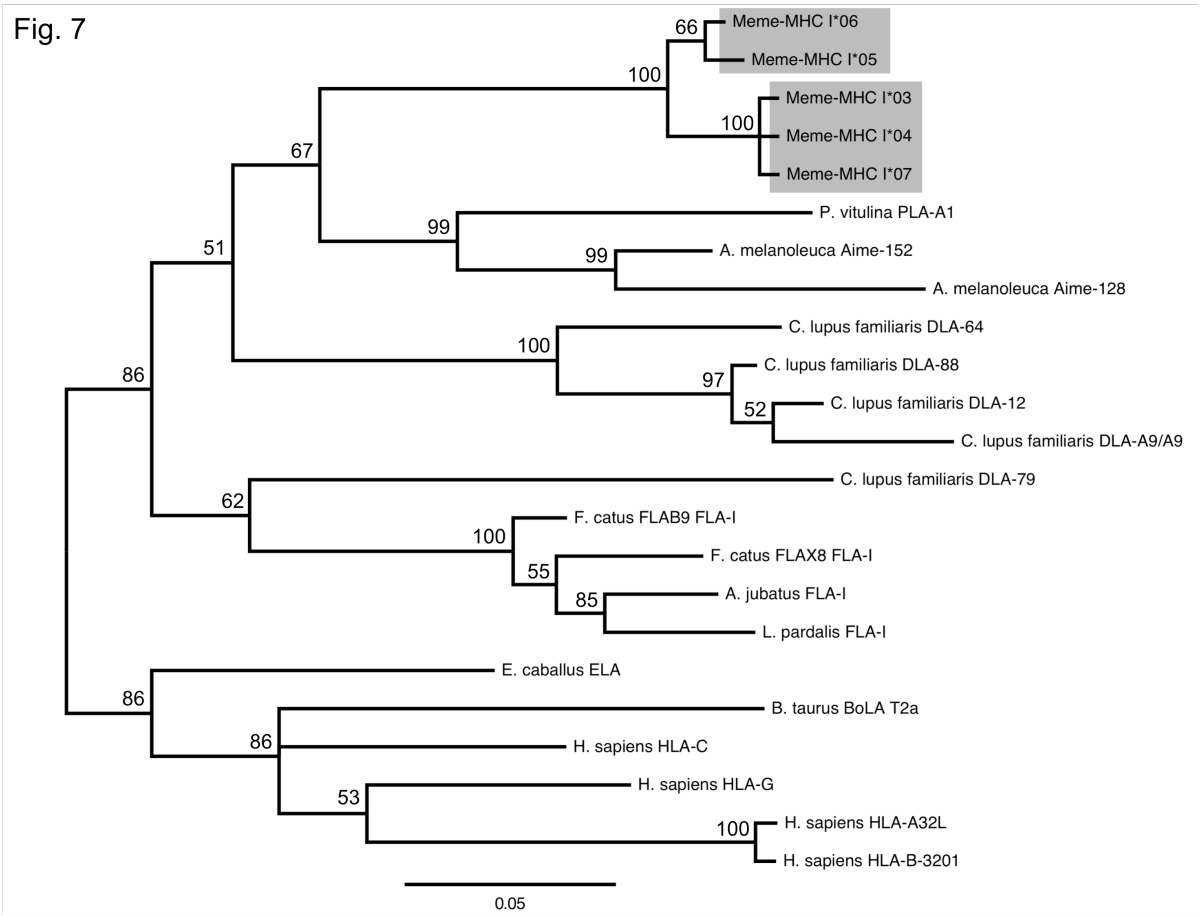


Fig. 3.7 Phylogenetic tree of MHC class I exon 4/α3-domain sequences from *M. meles* and other mammalian species (GenBank accession numbers are provided in the materials and methods), based on the 50% majority rule tree from the Bayesian analysis. Bayesian posterior probabilities above 50% are shown above the branches. *M. meles* class I sequences are marked with *grey boxes*. *PS* signifies pseudogenes with a frame-shift. *N* signifies the presence of genomic sequences, where no cDNA sequences were found.

3.5 Discussion

Diversity and transcription of MHC class I sequences

This is the first study to characterise MHC class I genes in *M. meles*. Moreover, this work was performed using both the genome and transcriptome. The sequences identified included pseudogenes and putatively functional sequences that encode domains necessary to form functional class I molecules. The major structural features (Kaufman *et al.*, 1994) that distinguish class Ia molecules are all present in these putatively functional sequences, including highly conserved amino acid residues that bind the N- and C- termini of the peptide, cysteine (C) residues that form the disulphide bonds, an

N-linked glycosylation site, a threonine (T) residue for interaction with TAP complex, and regions to interact with CD8 glycoproteins on the T-cells. In addition, we demonstrate that all putatively functional sequences were expressed in the RNA level in whole blood. These features, together with the detected polymorphisms, indicate that these sequences belong to at least two class Ia loci, in contrast to class Ib genes that are typically monomorphic (Rodgers & Cook, 2005).

The variability in the number of class I sequences in *M. meles* is intermediate compared to other carnivores. The most closely related species for which class I genes have been characterised is the Hawaiian monk seal (*M. schauinslandi*; Aldridge *et al*. 2006), which has at least two loci. No variability was found, however, within more than 80 individuals of this endangered seal species. Within the infraorder Arctoidea, the giant panda (*A. melanoleuca*) has two classical genes (*Aime*-152 and *Aime*-128; Pan *et al*. 2008), plus a non-classical gene closely related to DLA-79. *Aime*-152 appears to be monomorphic, while nine *Aime*-128 sequences were found in five individuals (Pan *et al*., 2008). The carnivores that, to date, have had their MHC organisation characterised most thoroughly are *C. lupus familiaris* and *F. catus*. In the feline leukocyte antigen (FLA) there are at least three class Ia genes, in common with human and murine MHCs (Yuhki *et al*., 2007). In DLA (Burnett *et al.*, 1997; Kennedy *et al.*, 1999; Wagner *et al.*, 1999; Wagner, 2003), there is one class Ia gene (DLA-88), three class Ib genes (DLA-79, -12 and -64) and two pseudogenes (DLA-53 and -12a). Among these, DLA-88 is the most polymorphic gene, with more than 50 alleles identified (Wagner, 2003). As discussed above, the number of class I genes and their organisation can differ between species greatly (Kelley *et al*., 2005). Mammals usually possess 1−3 class I genes (Nei *et al*., 1997) and have a variable range of polymorphism (Parham & Ohta, 1996; Wagner, 2003). Considering the extensive geographical range of *M. meles* and their considerable socio-spatial variability (Macdonald *et al.*, 2004; Rosalino *et al.*, 2004; Newman *et al.*, 2011), it is highly likley that more sequences will be detected as and when other populations are examined.

Our transcription analysis demonstrated that not all the detected sequences were expressed in whole blood. In addition to the seven putatively functional sequences, for which the sequences detected from gDNA and cDNA were identical; nine sequences were identified as pseudogenes. The phylogenetic

analyses indicate that *Meme-MHC I*PS02*, *Meme-MHC I*PS03* and *Meme-MHC I*PS04N*, which form a strongly supported clade, belong to the same pseudogene locus, with *Meme-MHC I*PS02* and *Meme-MHC I*PS03* both present in two individuals and *Meme-MHC I*PS04N* detected in another two individuals. Another grouping of *Meme-MHC I*PS08N*, *Meme-MHC I*09N* and *Meme-MHC I*PS10* indicated that these belong to two closely related pseudogene loci, as 1−3 sequences of this group were detected from seven individuals separately. Other studies have shown that many MHC sequences can be detected at the genomic level, but not at the cDNA level (deGroot *et al*., 2004). The presence of expressed non-functional MHC pseudogenes has also been reported (Mayer *et al.*, 1993; Fernandez-Soria *et al.*, 1998), even in the class II genes of *M. meles* (Sin *et al.*, 2012b). This is concordant with the finding that the MHC class I and class II regions have duplicated many times, generating many pseudogenes in addition to novel functional genes (Beck *et al*., 1999).

Evolution of the MHC class I genes

Our phylogenetic analyses reveal the evolutionary histories of the class I domains we examined to be very different from each other. Within the α1 domain there were more non-synonymous than synonymous nucleotide substitutions in both the ABS and non-ABS, contributing to a higher nucleotide and amino acid sequence diversity. Evidence of positive selection was detected in the α1 domain, in which positively selected sites were all within the ABS (Fig. 3.6a). *M. meles* sequences of the α1 domain were not monophyletic within this species but intermingled with sequences from other species (e.g. *P. vitulina*); a phenomenon characteristic of MHC genes. This could be due to trans-species polymorphism (Klein, 1987; Hughes & Yeager, 1998; Klein *et al.*, 1998), whereby balancing selection maintains this ancestral variation over a long period of time (many generations), even after species divergence (Penn & Potts, 1999; Bernatchez & Landry, 2003), leading to differences between gene- and species- trees. As the α1 domain is responsible for antigen binding the diversifying and balancing selection that drives and maintains this ABS polymorphism permits a population to present a wider repertoire of antigens, thus increasing its ability to combat pathogenic and parasitic infections (Hughes & Nei, 1992; Hughes & Yeager, 1998). The clustering of sequences among species could also be due to orthology, whereby sequences from an orthologous gene will cluster together, which may produce a similar clustering pattern as that observed with trans-species polymorphism. Without

the availability of information about loci identity it was not possible to disentangle between signals for orthlogy and trans-species polymorphism. Recombination between exons that encode ABSs can also increase allelic diversity in MHC genes (Ohta, 1991; Jakobsen *et al.*, 1998; Shum *et al.*, 2001; Bos & Waldman, 2006). We detected a recombination event at the 3' end of exon 2 that may have increased variation in *M. meles* class I genes. Nevertheless, point mutation and selection have been proposed as the major reason for high diversity in the MHC (Nei *et al.*, 1997; Piontkivska & Nei, 2003; Nei & Rooney, 2005).

As is the case for α1 domain, the α2 domain functions as an antigen-binding domain and showed a higher non-synonymous/synonymous rate ratio (ω) within the ABS than in the non-ABS (Fig. 3.6b). Comparisons of maximum likelihood models, which allow for positive selection (M2a and M8) relative to their corresponding null models, detected no significant positive selection acting on this region, however. A possible explanation for less intense diversifying selection on the α2 domain would be that it has more conserved sites than the α1 domain and so purifying selection may act to maintain the structural features of the MHC molecule. Our phylogenetic analyses also showed that all *M. meles* α2 domain sequences were grouped together except for one pseudogene [*Meme-MHC I*PS08N* clustering with DLA-64 but with low bootstrap support (72.4); Fig. 3.4b]; a very different pattern to the pattern that we observed in the α1 domain. The birth-and-death model of evolution predicts that this type of clustering by species (Nei *et al.*, 1997; Piontkivska & Nei, 2003) will result in less orthologous relationships, due to higher rates of gene duplication and deletion. Mechanistically, a differential rate of birth-and-death evolution explains the more rapid turnover rate of class I genes than class II genes in mammals (Hughes & Nei, 1989; Piontkivska & Nei, 2003) and possibly reveals why sequences are more closely related within *M. meles* than to sequences from other species. The α1 sequences we observed, however, showed that ancestral polymorphism or orthology at multiple lineages was maintained after the analysed species (Fig. 3.4a) diverged from one another, which indicates that different selective pressures act on different class I gene regions. Under this birth-and-death model, balancing selection at the α1 domain should maintain ancestral polymorphism even after species divergence, while strong purifying selection and recent duplication are required to produce a species-specific clustering at the α2 domain.

Alternatively, and more plausibly, species-specific gene clusters may result from concerted evolution (Hess & Edwards, 2002; Joly & Rouillon, 2006), where genes are homogenised by recombination. Recombination in the α2 coding region could produce *M. meles* sequences that are more similar to one another than to the sequences from other species. Under this scenario, a recombination breakpoint is needed to separate selection on the α1 and α2 regions, given that they exhibit different evolutionary histories. Our recombination analysis exposed that there are indeed recombination breakpoints located in the middle of the α-helix of the α1 domain and in the middle of intron 2, which is between the α1 and α2 coding exons. These recombination breakpoints indicate the probable crossover location and they most likely contributed to the independent evolutionary histories of the α1 and α2 domains by recombination in the α2 encoding region. Recombination within intron 2 has also been reported in other species (Holmes & Parham, 1985; Shum *et al.*, 2001; Bos & Waldman, 2006), which gives greater support to the hypothesis of concerted evolution through shuffling of the α1 and α2 domains. These recombination breakpoints are shared among some *M. meles* sequences, demonstrating that breakpoints are not random and that these recombination events are due to *in vivo* recombination but not *in vitro* chimera formation. Coincidentally, the degenerate 13−bp sequence motif, which is crucial in the crossover event of human recombination hotspots (Myers *et al*., 2008), was also found in the *M. meles* intron 2 sequences in which the recombination breakpoints were identified (Fig. 3.3). Variation in the zinc-finger protein PRDM9, however, affects species-specific binding to the sequence motif (Myers *et al*., 2010), hence more studies are needed to elucidate the recombination mechanism in *M. meles*.

Phylogenetic relationships of the α3 encoding sequences were more similar to that of the α2 than the α1 domain, demonstrating species-specific clustering of *M. meles* α3 sequences. This indicates that a recombination event most plausibly separated the α1 domain and the remaining 3' segments of the gene, leading to inconsistencies between the α1 gene tree and the α2 and α3 gene trees. However, the α3 domain is not antigen binding and it shows a very low number of mutations and polymorphic amino acid residues. Different from the antigen binding domains that have positive selection to increase diversity, strong purifying selection could eliminate mutations leading to unfavourable

changes in the molecular structure or the regions that interact with T-cells (Kaufman *et al*., 1994). The relatively conserved α3 domain is thus used to reconstruct evolutionary relationships more frequently than the α1 and α2 domains (e.g. Glaberman & Caccone, 2008). Here, the phylogenetic relationships of the α3 sequences follow the molecular phylogeny of the extant Carnivora (Flynn *et al.*, 2005; Fulton & Strobeck, 2006), in which Musteloidea, Pinnipedia and Ursoidea together form the infraorder Arctoidea. This infraorder then groups with Canidae to form the suborder Caniformia and together with Feliformia comprise the order Carnivora. An exception is *P. vitulina* that forms a clade with *A. melanoleuca* instead of *M. meles*. Musteloidea diverged from Ursoidea and Pinnipedia around 36 and 35.5 million years ago, respectively (Bininda-Emonds *et al*., 1999). The proximity of these divergence times might lead to incomplete lineage sorting (Maddison & Knowle, 2006), which is common in highly polymorphic genes (Lu, 2001).

Conclusions

Our findings highlight the importance of examining gene regions separately in order to obtain a more comprehensive understanding of MHC evolution. While the MHC class II genes (e.g. *DRB* and *DQB*; Sin *et al*., 2012) in *M. meles* did not exhibit extensive trans-species polymorphism, it nevertheless was observed. The class II genes are therefore likely to be undergoing balancing selection, whereas phylogenetic inconsistency between the α1 and α2/α3 domains of the class I genes indicates that these domains have different evolutionary histories. Concerted evolution provides a more plausible hypothesis than birth-and-death-evolution in the α2 and α3 domains, given the intron 2 breakpoint and separate evolutionary history of the α1 domain. Probable gene conversion has homogenised the α2 and α3 encoding exons, after ancestral gene duplication. As the α2 domain is antigen binding, whereas α3 is not, these domains were subject to different selective pressures leading to the observed difference in the degree of sequence divergence over evolutionary time. Separated from the α2 and α3 encoding exons, the α1 domain is likely to be subject to balancing selection (Nei *et al*., 1997). The high number of pseudogenes we found is typical in class I loci, which often show a high turnover rate, generating an abundant number of pseudogenes with various degrees of divergence from their functional counterparts (Hughes, 1995; Cadavid *et al.*, 1996; Beck *et al.*, 1999; Piontkivska & Nei, 2003). Non-functional sequences could serve as a reservoir for genetic exchange (Ohta, 1991; deGroot *et al.*,

2004), which would be the case here if the recombination event happened after the parental sequence *Meme-MHC I*PS08N* lost its function.

This is the first study to characterise MHC class I genes in a species within the superfamily Musteloidea. Given the polygynandrous mating system of *M. meles*, where high levels of extra-group paternity are observed (Dugdale *et al.*, 2007), further studies of mate choice and MHC would be informative. Examination of the association between MHC genotypes and pathogens could also have significant implications for research into bTB epidemiology in badgers (Allen *et al.*, 2010).

Supplementary material

Table S3.1 Different *M. meles* MHC class I alleles could be obtained by PCR amplification using the primers marked with *ticks*. Primer information and amplified region are given in Table 3.1 and Figure 3.1

	Primers					
	Meme-MHCIex1F	Meme-MHCIex2F	PpLAa1L250	Meme-MHCIex3F	Meme-MHCIex3R	Meme-MHCIex6R
*Meme-MHC I*01*		✔		✔	✔	
*Meme-MHC I*02*		✔		✔	✔	
*Meme-MHC I*03*	✔	✔	✔	✔	✔	✔
*Meme-MHC I*04*	✔	✔	✔	✔	✔	✔
*Meme-MHC I*05*	✔	✔	✔	✔	✔	✔
*Meme-MHC I*06*	✔	✔		✔	✔	✔
*Meme-MHC I*07*	✔	✔		✔	✔	✔
*Meme-MHC I*PS08N*		✔		✔	✔	
*Meme-MHC I*09N*		✔	✔			
*Meme-MHC I*PS10*		✔	✔			
*Meme-MHC I*11*		✔	✔			
*Meme-MHC I*PS12N*		✔	✔			
*Meme-MHC I*PS01*	✔					✔
*Meme-MHC I*PS02*	✔					✔
*Meme-MHC I*PS03*	✔					✔
*Meme-MHC I*PS04N*				✔	✔	

Chapter 4

Molecular characterization of the microbial communities in the subcaudal gland secretion of the European badger [3]

[3] This chapter has been published as: Sin, Y.W., Buesching, C.D., Burke, T. & Macdonald, D.W. 2012. Molecular characterization of the microbial communities in the subcaudal gland secretion of the European badger (*Meles meles*). *FEMS Microbiology Ecology* **81:** 648–659.

4.1 Abstract

Many mammals possess specialized scent glands, which convey information about the marking individual. As the chemical profile of scent-marks is likely to be affected by bacteria metabolizing the primary gland products, the variation in bacterial communities between different individuals has been proposed to underpin olfactory communication. However, few studies have investigated the dependency of microbiota residing in scent organs on the host's individual-specific parameters. Here, we used terminal restriction fragment length polymorphism analysis of the PCR-amplified 16S rRNA gene and clone library construction to investigate the microbial communities in the subcaudal gland secretion of the European badger (*Meles meles*). As the secretion has been shown to encode individual-specific information, we investigated the correlation of the microbiota with different individual-specific parameters (age, sex, body condition, reproductive status and season). We discovered a high number of bacterial species (56 OTUs from four phyla: Actinobacteria, Firmicutes, Proteobacteria and Bacteroidetes), dominated by Actinobacteria (76.0%). The bacterial communities of cubs and adults differed significantly. Cubs possessed considerably more diverse communities dominated by Firmicutes, while in adults the communities were less diverse and dominated by Actinobacteria, suggesting that the acquisition of a 'mature bacterial community' is an ontogenetic process related to physiological changes during maturation.

4.2 Introduction

Scent marking is common in mammals, and urine, faeces, or secretions from specialized scent glands can convey information about the marking individual (Brown & Macdonald, 1985). Scent glands typically contain sebaceous and apocrine glands that secrete fatty and/or organic substances into scent pouches or dense hair tufts on the skin surface (Albone & Shirley, 1984; Brown & Macdonald, 1985; Thody & Shuster, 1989). These places are generally warm, moist, nutrient rich, and ideal for the proliferation of symbiotic bacteria, which are proposed to play an important role in scent marking (Albone *et al.*, 1974; Albone & Perry, 1975) as they are likely to metabolize the primary scent gland products. The fermentation hypothesis for chemical recognition (Albone *et al.*, 1974; Albone & Perry, 1975; Gorman, 1976) proposes that symbiotic bacteria living in scent glands produce volatile odorants that contribute to cues used for individual recognition, and that these cues vary with the composition of the bacterial community. Mammalian scent glands often harbor diverse bacterial communities (reviewed in Archie & Theis, 2011), which include fermentative anaerobic bacteria that produce short-chain fatty acids (SCFAs) amongst other odorants. Depending on pH, available primary gland products and other factors, bacterial communities may vary with individual-specific parameters, such as the age, sex, individual identity, reproductive status, health status, diet and competitive ability of their hosts.

Although odour generation and olfactory communication in mammals generally involve microbiota, few studies have investigated the microflora residing in scent organs or their potential dependency on individual-specific parameters. Here, we investigated the bacterial communities in the subcaudal gland of the European badger (*Meles meles*). As a nocturnal carnivore with a well-developed sense of smell, *M. meles* constitutes a good model species to study olfactory communication (Buesching & Macdonald, 2001). Badgers possess a unique skin gland, the subcaudal gland, which consists of a well-developed subcaudal pouch and is lined with sebaceous and apocrine tissues secreting paste-like secretion into a common lumen (Stubbe, 1971). The pouch opens into the environment through a 20−80 mm wide horizontal slit (Buesching *et al.*, in prep.), situated between the base of the tail and the anus. The secretion is transferred from the pouch onto the substrate by a quick squatting motion

during which the pouch opening is pressed briefly against objects (Buesching & Macdonald, 2004) or conspecifics (Buesching *et al.*, 2003).

The subcaudal gland secretion has been shown to encode information about fitness-related parameters in addition to group membership and individuality (Gorman *et al.*, 1984; Buesching *et al.*, 2002c). The lower molecular weight components of the odour are thought to be by-products from bacteria metabolizing the primary gland products (Kruuk *et al.*, 1984), and the time-related changes in the chemical composition of the secretion are believed to be the results of chemical and microbial degradation (Buesching *et al.*, 2002b). The microflora inhabiting the subcaudal pouch of *M. meles* is thus likely to play an important role in generating the factual odour of the subcaudal gland secretion. Therefore, we hypothesized that the bacterial pouch communities should vary with individual-specific parameters such as age, sex, body condition and reproductive status. Badger cubs, however, do not start to produce subcaudal gland secretion until approximately four months of age (Buesching *et al.*, 2003; Fell *et al.*, 2006), and increase in secretion during their first year is age-related (Buesching *et al.*, 2003). The slow raise in the amount of secretion stored in the subcaudal pouch should thus boost bacterial colonization (Nordstrom & Noble, 1984), which is likely to happen as a succession of communities (Favier *et al.*, 2002). Hence, we anticipated the bacterial communities of cubs and adults to differ.

The objective of this study was to investigate the microbial community using a culture-independent method: terminal restriction fragment length polymorphism (T-RFLP; Liu *et al.*, 1997) in combination with cloning and sequencing of PCR-amplified 16S rRNA genes. In the past, most studies on the microbial communities of mammalian scent organs used traditional culture-dependent techniques to isolate and identify bacteria from cultures. However, only a very small portion of bacteria (<1%) from the environment can be cultivated (Amann *et al.*, 1995; Pace, 1997; Hugenholtz *et al.*, 1998), whereas culture-independent molecular methods, such as T-RFLP, can overcome the limitations of the traditional cultivation approach. Although all fingerprinting techniques have limitations, such as not detecting rare species (Bent *et al.*, 2007; Orcutt *et al.*, 2009), T-RFLP provides a convenient way to determine the diversity of bacterial communities. More importantly, it also allows comparison of the

relative proportions of bacteria among sampled communities, which would be impossible with culture techniques. In this study, we aimed to (1) characterize the microbial community of the subcaudal gland secretion; and (2) investigate the correlation of the microbiota with different individual-specific parameters (age, sex, season, body condition and reproductive status). According to the fermentation hypothesis, we predicted bacterial communities to show differences between some, if not all, of the factors, especially between age categories.

4.3 Materials and methods

Study animals and sample collection

Subcaudal gland secretion samples were collected in spring (27th May−9th Jun) and summer (6th−18th Sep) of 2010 from a population of *M. meles* in Wytham Woods, Oxfordshire, UK (global positioning system reference 51:46:26N, 1:19:19W; trapping and handling was subject to the University of Oxford's ethical review and performed under government licences). As part of an ongoing long-term study, this badger population has been trapped 3−5 times per year since 1987 (for details see Macdonald & Newman, 2002). A unique number, tattooed in the left inguinal region at first capture (usually as cubs), permits individual identification and determination of age. After capture, badgers were sedated with 0.2 ml ketamine hydrochloride/kg body weight. Tattoo, sex and body condition [categorised as 1(=emaciated) − 3 (=very good condition); following Buesching *et al.*, 2009; Macdonald *et al.*, 2009] were recorded. Length and diameter of teats were measured to establish recent lactation in females (Dugdale *et al.*, 2011a), and the degree to which the testes were descended into the scrotum was examined in males (Buesching *et al.*, 2009). As badgers are induced ovulators (Yamaguchi *et al.*, 2006), it is unlikely that the microbial community would mirror such short-term endocrinological changes, and thus no attempt was made to correlate pouch bacteria with female oestrus.

Thirty-nine secretion samples were collected in spring and 40 in summer (Table 4.1). Eighteen animals (eight males and ten females) were sampled in both trapping events. Subcaudal gland secretion was scooped out of the subcaudal pouch using a rounded stainless steel spatula

Table 4.1 Details of analyzed *M. meles* subcaudal gland secretion samples

		Season		
		Spring	Summer	Total
Cubs	Male	1	2	3
	Female	1	2	3
Adults	Male	10	16	26
	Female	27	20	47
Total		39	40	79

(Buesching *et al.*, 2002c) and stored inside a microcentrifuge tube. Samples were frozen immediately and stored at -20ºC until DNA extraction. Between sampling, the spatula was wiped clean and sterilized with absolute ethanol and a flame to avoid contamination of samples. Contact of the spatula with body parts other than the inside of the subcaudal pouch was strictly avoided.

Only 16.7% (total number=42) of the cubs had detectable amounts of secretion in their subcaudal pouches in spring, and 57.9% (total number=19) in summer. However, the volume was usually too small to collect for further analyses, and thus only six samples were derived from cubs in this study.

DNA extraction

Pre-treatment for Gram-positive bacteria was carried out to lyse the bacterial cell walls before DNA purification. Secretions were transferred to microcentrifuge tubes and resuspended in 180 µl enzymatic lysis buffer [20 mM Tris·Cl (pH 8.0), 2 mM EDTA, 1.2% Triton X-100] containing 20 mg/ml lysozyme, and incubated at 37°C for at least 1 hr in a shaking incubator. Samples were then incubated with proteinase K at 56°C for at least 1 hr, followed by the addition of 200 µl ethanol. DNA isolation was continued with pipetting the mixture into the spin column according to the manufacturer's instructions for animal tissue samples using the DNeasy Blood & Tissue kit (Qiagen, Hilden, Germany).

T-RFLP analysis

Internal fragments of 16S rRNA genes were amplified from the isolated DNA using universal bacterial primers 341f (5'-CCTACGGGAGGCAGCAG-3'; Muyzer *et al.*, 1993) and 926r (5'-CCGTCAATTCMTTTRAGTTT-3'; Muyzer *et al.*, 1995). Primer 341f was labeled at its 5' end with the dye 6-carboxyfluorescein (6-FAM) and 926r labeled with hexachloro-6-carboxyfluorescein

(HEX). PCR mixtures comprised 1X PCR buffer, 1 mM $MgCl_2$, each dNTP at a concentration of 0.2 mM, each primer at a concentration of 0.6 µM and 0.5 U Taq polymerase (Bioline). Amplification was carried out in two separate 25 µl reaction mixtures under conditions: 95°C for 3 min, amplified for 35 cycles of 95°C, 30 sec; 55°C, 1 min; 72°C, 1 min 15 sec; and a final extension at 72°C for 10 min. For each sample, duplicate fluorescently labeled PCR products were pooled after PCR. PCR product sizes were confirmed by agarose gel electrophoresis using 100 bp DNA ladders as size markers and by staining with ethidium bromide. The amplified products (around 586-bp) were purified using the QIAquick gel extraction kit (Qiagen).

Fluorescently labeled PCR products were digested with MspI restriction enzymes [Applied Biosystems (ABI)] for 6 h at 37°C, followed by treatment for 20 min at 80°C for enzyme inactivation. The digested fragments were separated on the ABI 3730 Genetic Analyzer (ABI). The sizes of the fluorescently labeled terminal restriction fragments (TRFs) were determined by comparison with the GeneScan 500 ROX size standard (ABI). Before injection, 0.5 µL of the DNA sample was denatured in the presence of 9.5 µL Hi-Di formamide and 0.05 µL GS 500 ROX size standard (ABI) at 95°C for 5 min. Injection was performed electrokinetically at 2 kV for 10 s, and electrophoresis was run at 15 kV for 30 min. After electrophoresis, T-RFLP electropherograms were imaged using the GENEMAPPER software (v. 3.7; ABI). The lengths of TRFs were determined by comparison with the internal standard, within the lower threshold at 50 bp and upper threshold at 500 bp. Only peaks with heights exceeding 50 fluorescence units were evaluated.

T-RFLP profiles were aligned using T-Align (Smith *et al.*, 2005), which identified all fragments unequivocally with ±0.5 bp in all profiles generated, and determined the presence and absence of TRFs and their relative fluorescence in all samples. Log (n+1) transformation was performed in order to normalize the data.

Cloning and sequencing

The identities of bacterial species corresponding to the TRFs were confirmed by cloning and sequencing. The 16S rRNA genes were amplified from genomic DNA using the unlabeled primers 341f and 926r using the cycling conditions as described above. DNA samples from nine individuals,

which comprised >80% of the TRFs were selected. Duplicate PCR products from each individual were pooled after PCR and the around 586-bp amplified products were purified using the QIAquick gel extraction kit (Qiagen). Purified PCR fragments were ligated into a TA cloning vector (pGEM-T Easy Vector Systems, Promega, Madison, WI, USA). Transformation, growth and blue-white selection were carried out in *Escherichia coli* competent cells (JM109). PCRs were performed on 384 positive clones using M13 forward and reverse primers (Messing, 1983). The PCR products were digested with MspI (ABI) as above and restriction fragments were separated in a 2.5% agarose gel. Unique RFLP patterns were identified and a total of 96 corresponding PCR products were purified with ExoSAP-IT (GE Healthcare, Buckinghamshire, UK) and sequenced.

All amplified fragments were sequenced with M13 primers using dideoxynucleotide chain-termination and loaded on an automated DNA Sequencer (48-well capillary ABI3730, ABI, Foster City, CA). Nucleotide sequences were analyzed using CodonCode Aligner 3.7.1 (CodonCode Corporation, Dedham, MA), and were compared with GenBank sequences using the NCBI BLAST program (Altschul *et al.*, 1990). The DNA sequences from this study were assigned GenBank accession numbers JN806160-JN806215.

Phylogenetic tree reconstruction and statistical analyses

Phylogenetic analyses were performed on the consensus alignments of the sequenced 16S rRNA sequences. Bayesian phylogenetic inference was performed using MrBayes 3.1.2 (Ronquist & Huelsenbeck, 2003). The web-based application FindModel (http://www.hiv.lanl.gov/content/sequence/findmodel/findmodel.html) was employed to find the best-fit model of nucleotide substitution. The general time-reversible model with gamma-distributed rate variation (GTR+Γ) was determined to be the best fit for the dataset. A Markov chain Monte Carlo (MCMC) search was initiated with random trees and run for 3 000 000 generations, with a sampling frequency of every 100 generations. The standard deviation of split frequencies converged to a value <0.01 for the 16S rRNA gene analyses. After convergence was checked by plotting the likelihood scores against generation, the first 25% of the generations were discarded as the burn-in of the chain (Ronquist & Huelsenbeck, 2003).

Discriminant function analysis (DFA) was used to test the hypothesis that the relative abundance of bacteria groups in the secretion was different between different classes of individuals. The samples were categorized by age, sex, body condition and reproductive status of the individual and season of sampling. Nineteen males (with fully descended testes) were classified as in breeding condition, while only one male (with completely retracted testes) was classified as not in breeding condition and six males were in an intermediate state so that reproductive activity could not be determined unambiguously.

DFA generates discriminant functions that provide the best discrimination between the defined categories. Here, the statistical significance of category differences was tested using Wilks' λ and χ^2. The effectiveness of the discriminant function to correctly assign cases to categories was tested using classification statistics. The "leave-one-out" cross-validation analysis was reported because it produced a more reliable function as the case to be classified was not used to develop the discriminant function.

An analysis of similarity (ANOSIM, Clarke, 1993) was conducted to verify the significance of the grouping and examine the degree of similarity between bacterial communities. A Bray-Curtis similarity matrix computed from the relative abundance of all TRFs was used to generate one-way ANOSIM statistics with 999 permutations. It calculated a test statistic (R) that is equal to 1 if all individuals within a category were more similar to each other than to any individual in another category, and 0 if there was no difference between categories. The Similarity percentage (SIMPER; Clarke & Gorley, 2001) procedure was used to determine the percentage contribution of each TRF to the overall difference between categories.

The bacterial complexity was described using two diversity indices. Richness was defined as the total number of TRF (bacterial groups) in the TRF profile. The Shannon index (H') took into account the evenness in addition to the species number, and was thus higher if the diversity is more even. The diversity indices were compared using Kruskal-Wallis test.

DFA and Kruskal-Wallis tests were performed using PASW 18.0 (SPSS, Chicago, IL, USA). The ANOSIM and SIMPER analyses were performed with PRIMER 6 (Primer-E Ltd; Clarke, 1993).

4.4 Results

T-RFLP analysis

The structure of bacterial communities associated with the subcaudal gland secretion of European badgers was determined by T-RFLP analysis of the 16S rRNA gene. This approach enabled us to determine differences in relative abundances of some phyla. Analysis of the T-RFLP data revealed a total of 50 TRFs in the subcaudal gland secretion of the 79 sampled badgers. The mean number of TRFs found in an individual was 5.58 (range 2–15, SD±1.58). None of the TRFs were recorded in all individuals. The most common TRF was 140.4 (Table S4.1 and S4.2; supplementary material), which was present in 91.1% (72/79) of individuals. The corresponding reverse primer TRFs (62.2 and 368) were present in 19.0% and 97.5% of individuals respectively. This TRF (140.4) accounted for an average of 72.9% (SD±16.3%) of the total fluorescence area of the T-RFLP profiles from all badgers.

Analysis of clone libraries

To provide information regarding the identity of the bacterial species, 16S rRNA gene clone libraries were constructed from DNA isolated from the subcaudal gland secretion. Clone libraries were generated from nine badgers and a total of 384 clones were screened for their individual RFLP types. All unique and redundant RFLP types were selected for sequencing.

Out of the 96 clones sequenced, 56 operational taxonomic units (OTUs) were found. We assigned a species to an OTU with ≥97% identity, while those with ≥95% identity were assigned to the same genus (Schloss & Handelsman, 2005; Santelli *et al.*, 2008). The OTUs clustered into four bacterial phyla: Actinobacteria, Firmicutes, Proteobacteria and Bacteroidetes (Fig. 4.1). The Proteobacteria present belonged to the Alpha, Beta and Gamma classes, with the most OTUs (eight OTUs) belonging to the Alphaproteobacteria (Fig. 4.1), while members of Gammaproteobacteria were not abundant [relative abundance (i.e. abundance of corresponding TRFs within T-RFLP profiles) = 0.06%].

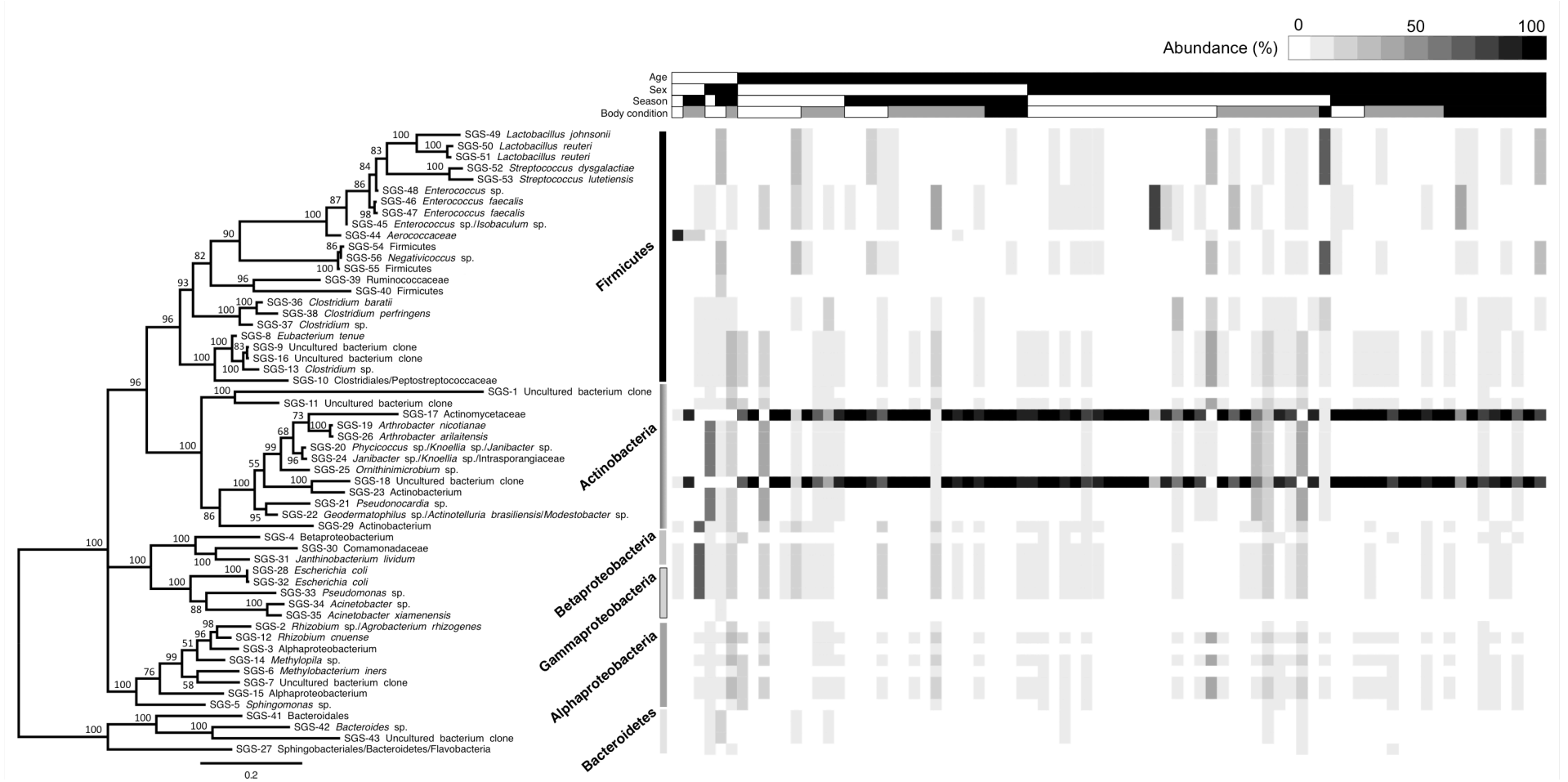


Fig. 4.1 Phylogenetic tree showing different OTUs found in the subcaudal gland secretions of nine *M. meles* based on 16S rRNA gene sequences. Bayesian posterior probabilities above 50% are shown above branches. Sequences were assigned to the same species with ≥97% identity, whereas sequences with ≥95% identity were assigned the genus name followed by 'sp.'. The right side of the figure shows the relative abundance of forward TRFs per sample in gray scale values (white: 0% present; black: 90−100%). Bars above the abundance graph indicate age (white = cubs; black = adults), sex (white = male; black = females), season of sampling (white = spring; black = summer) and body condition (white = emaciated – black = very good condition) of each individual sampled. Attention should be paid that some OTUs correspond to the same TRFs (see Table S4.1 for details) and thus have the same abundance patterns.

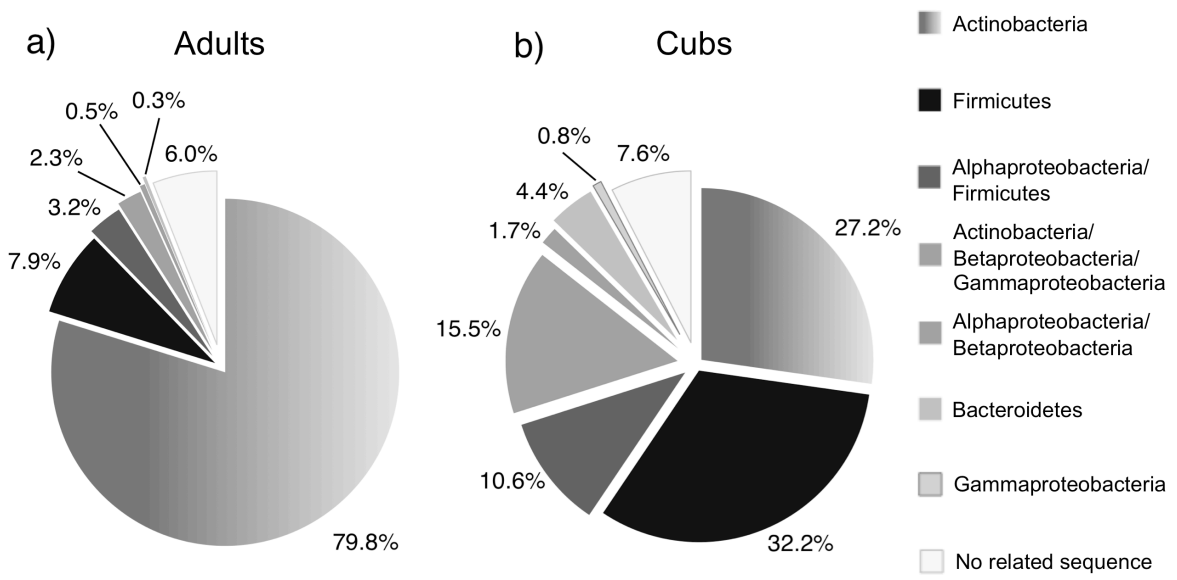


Fig. 4.2 Pie chart showing the relative abundance of different bacterial phyla/classes found in subcaudal gland secretion of *M. meles* (a) adults (n=73) and (b) cubs (n=6). The proportions are based on T-RFLP data. TRFs with no matches in the clone libraries were classified as no related sequence (white area).

Members of the Actinobacteria were the most abundant (relative abundance = 76.0%), with the majority corresponding to the order Actinomycetales (SGS-17, 19, 20, 21, 22, 24, 25 and 26). The phylum Firmicutes was the most diverse group with 23 OTUs identified, which were closely related to *Enterococcus*, *Aerococcaceae*, *Lactobacillus*, *Streptococcus*, *Eubacterium*, *Clostridium* and *Negativicoccus*. Only four OTUs of Bacteroidetes were found, which were not abundant and diverse compared to other phyla (relative abundance = 0.6%).

Identification of TRFs from 16S rRNA sequenced genes

The nucleotide sequence data obtained were used to identify the corresponding TRFs based on the predicted sizes for the TRFs generated following restriction digestion (Table S4.1; supplementary material). 28% of TRFs were identified, which contributed to 93.9% of the total peak area in all individuals. However, for many of the TRFs, more than one species/genus could be matched with the particular fragment size. Differences existed between the observed and predicted TRFs, which have been reported frequently in other studies (Kaplan *et al.*, 2001).

By linkage of TRFs to the sequence data, we found that the abundance of TRFs corresponding to Actinobacteria represented more than 76 % of all bacterial communities, followed by Firmicutes and Proteobacteria with relative abundances of more than 9 % and 3 % respectively. The most dominant TRF (140.4), found in the majority of the individuals, were in phylum Actinobacteria (Fig. 4.1). Approximately 6.1% of the T-RFLP abundance data, mainly TRFs with low relative abundances, could not be correlated with any sequence data (Fig. 4.2).

Effects of age, sex, season, body condition and reproductive status

The bacterial communities of cubs and adults differed significantly. DFA based on age revealed a discriminant function that differentiated the T-RFLP profile between cubs and adults (Wilks' λ=0.003, p<0.001, Table 4.2). The function explained 99.7% of the variance in the grouping variable. Using the computed discriminant function, 98.7% of secretions were assigned to the correct category (Table 4.2), with 100% of cubs (6/6) and 98.6% of adults (72/73) classified correctly. The ANOSIM analysis also illustrated a significant overall difference (Global R statistic=0.71, p=0.001, Table 4.2) of the bacterial community in cubs and adults. This was not due to pseudoreplication as there was also a significant difference when only one sample per individual was included (Table 4.2). When age was categorized as cubs (age 0; n=6), yearlings (age=1; n=12), young adults (age 2-5; n=33) and old adults (age>5; n=17), only cubs were significantly different from other age categories (Wilks' λ=0.0, p<0.001, Fig. 4.3).

The T-RFLP profile of adults of different body condition was significantly different in spring, shown by DFA (Wilks' λ=0.205, p=0.044, Table 4.2), but not ANOSIM. However, this difference between body conditions was not apparent in summer. We found no significant differences between T-RFLP profiles of males and females, nor amongst females of different reproductive status, and bacterial communities did not differ with seasons (Table 4.2). DFA, but not ANOSIM, showed a significant difference between males in breeding condition (n=19) and males in intermediate reproductive condition (n=6) when all samples were included. However, if only one sample per male was included,

Table 4.2 DFA and ANOSIM comparing T-RFLP profiles of different categories (age, sex, body condition, reproductive status and season of sampling) of *M. meles* subcaudal gland secretion samples. Analyses were performed on different data sets, i.e. either all or only one sample per individual was included. Results of functions 1 and 2 from the DFA were shown for analyses with three factor levels

Test	Data set	λ	χ^2	df	*P*	% Cross-validation	ANOSIM Global R	ANOSIM *P*
Sex	All data	0.44	47.03	41	0.24		0.00	0.43
	One sample per ID	0.38	40.58	35	0.24		-0.02	0.61
	All adults	0.47	40.14	36	0.29		-0.01	0.49
	Spring, adult only	0.37	23.56	23	0.43		0.08	0.21
	Summer, adult only	0.34	23.10	25	0.57		-0.01	0.66
Age	All data	**0.003**	**327.44**	**47**	**<.001**	**98.7**	**0.71**	**0.001**
	One sample per ID	**0.001**	**271.53**	**35**	**<.001**	**96.7**	**0.71**	**0.001**
Season	All data	0.45	45.59	41	0.29		0.02	0.03
	One sample per ID	0.44	33.62	36	0.58		0.01	0.15
	All adult	0.48	38.58	36	0.35		0.02	0.07
Reproductive status	All adult females [a]	0.24	39.47	29	0.09		0.20	0.01
	One sample per female	0.27	27.18	24	0.30		0.10	0.07
Body condition	All data						-0.03	0.80
	Function 1	0.18	95.59	82	0.15			
	Function 2	0.47	42.80	40	0.35			
	One sample per ID						-0.03	0.80
	Function 1	0.12	85.72	72	0.13			
	Function 2	0.39	37.79	35	0.34			
	All adults						-0.01	0.54
	Function 1	0.20	84.37	72	0.15			
	Function 2	0.48	38.42	35	0.32			
	Spring, adult only [b]	**0.21**	**35.71**	**23**	**0.044**	**80.6**	0.12	0.07
	Summer, adult only [c]							
	Function 1	0.06	60.85	50	0.14		0.01	0.36
	Function 2	0.37	20.92	24	0.64			

We report classification results only for significant analyses. Significant DFA analyses and significant ANOSIM with Global R>0.5 are in bold. df degrees of freedom.
[a] Non-breeding females: n=14; breeding females: n=30.
[b] Only body condition 1 (n=24) and 2 (n=13) are included, as body condition 3 occurred only once.
[c] Body condition 1: n=7; body condition 2: n=16; body condition 3: n=13.

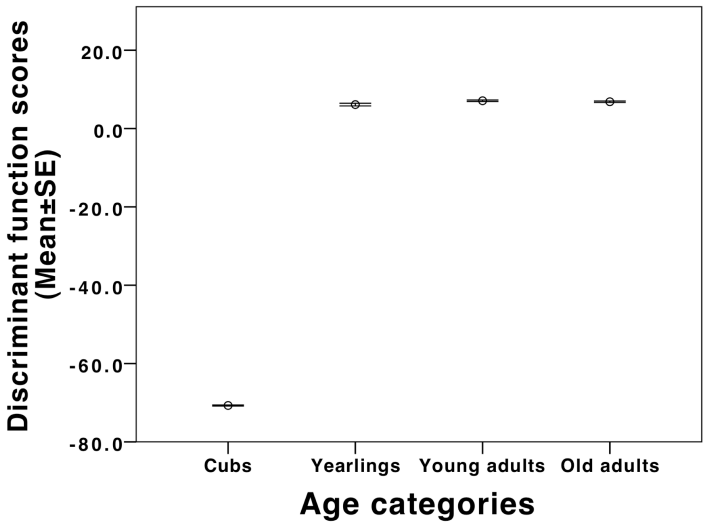


Fig. 4.3 Discriminant function scores (mean ± SE) from DFA of 79 subcaudal gland secretion samples differentiating T-RFLP profiles of cubs (age 0; n=6), yearlings (age=1; n=12), young adults (age 2-5; n=33) and old adults (age>5; n=17). The function explained 100% of the variance in the grouping variable.

analyses showed no significant difference (data not shown), indicating that the significant result might have been caused by pseudoreplication.

The diversity of the microbiota was estimated based on the T-RFLP data using the Shannon index of diversity and richness. On average there were 8.67 (SD±1.69) TRFs in cub and 5.33 (SD±1.51) TRFs in adult subcaudal gland secretions. The bacterial community in cub secretions was significantly more diverse (p=0.014) and rich (p=0.019) than in the secretions of adults (Fig. 4.4). In addition, the diversity (p=0.014) and richness (p=0.023) of recently breeding and non-breeding females differed significantly (Fig. 4.4), although no other life-history parameters appeared to affect the composition of the subcaudal gland microbiota.

The proportions of different bacterial phyla varied significantly between the bacterial pouch communities of adults and cubs (Fig. 4.2; see Fig. 4.1 for inter-individual differences). SIMPER analysis showed that TRF 140.4 bp, 212.0 bp, and 160.6 bp contributed the most to the differences between bacterial communities of cubs and adults (16.4%, 11.5%, and 8.79% respectively, Table 4.3). TRF 140.4 bp, which was identified to be in phylum Actinobacteria (SGS-17, 18), was more common in adults (present in 95.9% of individuals) than in cubs (present in 33.3% of individuals) and also had

Fig. 4.4a

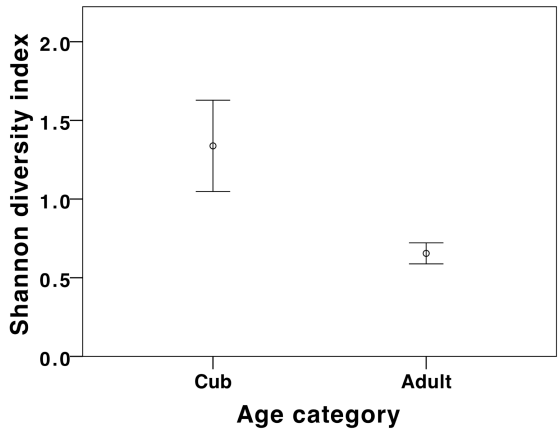


Fig. 4.4b

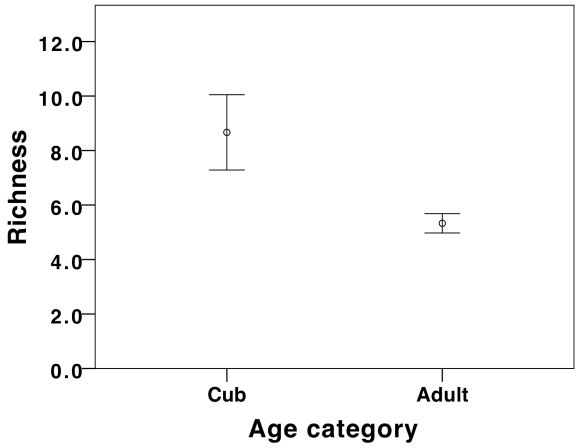


Fig. 4.4c



Fig. 4.4d

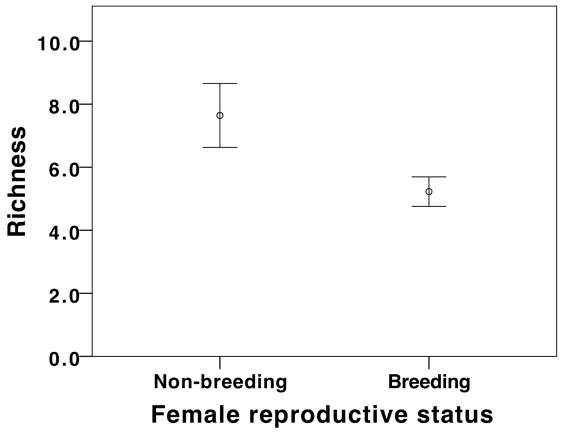


Fig. 4.4 Mean (± SE) difference of (a) Shannon index of diversity of cubs (n=6) and adults (n=73); (b) richness of cubs and adults; (c) Shannon index of diversity of non-breeding (n=14) and breeding females (n=30); and (d) richness of non-breeding and breeding females. Richness signified the total number of TRF and the Shannon index of diversity was calculated from the relative abundance of TRFs of badger subcaudal gland secretion bacterial communities.

higher relative abundance in adults (average: 77.7%, SD±16.5%) than in cubs (average: 14.0%, SD±13.8%). Conversely, TRF 212.0 bp, which was identified as a member of the Firmicutes, was more common and abundant in cubs (66.6% presence) than in adults (6.8% presence). TRF 160.6, which mostly comprised Gammaproteobacteria, was only found in cubs (relative abundance = 0.8%; Fig. 4.2).

4.5 Discussion

This study provides the first detailed molecular characterization of the microbiota inhabiting the subcaudal pouch of *M. meles*. We discovered 56 OTUs in the subcaudal gland secretion of *M. meles*.

Table 4.3 The percentage contribution of the top 10 TRFs to the overall difference between cubs and adults

TRF	Cum. % [a]	Clone
140.4	16.43	SGS-17, 18
212.0	27.93	SGS-44
160.6	36.73	SGS-28 to 33
138.9	43.06	SGS-19 to 26
133.7	49.37	SGS-7 to 16
194.2	54.84	NA [b]
212.8	60.10	SGS-45 to 48
189.3	64.42	SGS-36 to 38
214.1	68.08	SGS-49 to 56
208.2	71.13	SGS-41 to 43

[a] Cumulative percentage
[b] No related sequence

The high diversity in the bacterial community was unusual when compared to the scent organs of 13 other mammalian species reviewed in Archie & Theis (2011). However, only very few studies (e.g. Roze *et al.*, 2010) had used culture-independent methods, limiting the diversity of identified species.

The bacterial community in the subcaudal pouch was dominated by Actinobacteria, which co-existed with numerous minor OTUs classified under Firmicutes, Bacteroidetes, Alphaproteobacteria, Betaproteobacteria and Gammaproteobacteria. Some Actinobacteria have been shown to play a major role in the transformation of odorless steroids into odorous derivatives. For example, in the human axilla, different bacteria such as *Corynebacterium* metabolize the odorless sebaceous and apocrine gland secretions (e.g. steroids and sebum) into odorous compounds (Gower *et al.*, 1986; Kohl *et al.*, 2001), generating the typical armpit smell. Moreover, studies showed humans to have individually distinctive axillary bacterial communities, and thus resulting odour profiles, which may serve as a cue for individual (Natsch *et al.*, 2006; Xu *et al.*, 2007; Kuhn & Natsch, 2009) and kin (Weisfeld *et al.*, 2003) recognition. Badger subcaudal gland secretion has been shown to be rich in short- and medium-chain fatty acids (Buesching *et al.*, 2002c, b) that are likely to be pheromonally active end products of Actinobacteria metabolizing long-chain fatty acids (Albone & Perry, 1975; Gorman, 1976), as has been suggested for the perineal glands of *E. dorsatum,* whose bacterial community is also dominated by Actinobacteria (Roze *et al.*, 2010).

Other less abundant species are also likely to contribute to odour generation. For example, *Lactobacillus*, *Streptococcus* and *Enterococcus* produce lactic acid as their major fermentation product, and were also identified in the scent organ of other mammals [e.g. the perineal glands of *E. dorsatum* (Roze *et al.*, 2010) and tarsal tufts of white-tailed deer *Odocoileus virginianus* (Alexy *et al.*, 2003)]. *Clostridium*, which is also found in the anal sac of the red fox *Vulpes vulpes* (Ware & Gosden, 1980), generates an unpleasant odour during anaerobic decomposition of proteins due to the production of fatty acids, ammonia, hydrogen sulfide and amines. *Bacteroides* and *Escherichia*, identified in the anal glands of beavers (*Castor canadensis*) as well as in the scent organs of other species (for details see Archie & Theis, 2011), are able to produce a mixture of organic acids as fermentation end products. They are widespread in mammalian digestive systems, and can easily be transferred from the anus to the subcaudal pouch. Thus, their ability to produce odorants may have resulted in the evolution of a long-term mutualistic association between a variety of bacteria and *M. meles*. However, our study focused on a single population and it is theoretically possible for the bacterial taxa to be population-specific. Thus, additional studies on other populations are needed to understand whether the bacterial communities found here are typical for badgers throughout their geographic range.

Interestingly, cubs and adults differed highly significantly in the bacterial communities populating their subcaudal pouches, indicating that the acquisition of a 'mature bacterial community' is an ontogenetic process related to physiological changes during maturation. As cubs start to produce subcaudal gland secretion at approximately four months of age (Buesching *et al.*, 2002a; Fell *et al.*, 2006), different species of bacteria may start to colonize the subcaudal pouch at this stage. It is highly possible that some opportunistic bacteria would take advantage of this habitat and achieve high abundances before a 'mature bacterial community' is established. Succession of symbiotic bacterial communities is common (Deplancke *et al.*, 2000; Favier *et al.*, 2002; Lu *et al.*, 2003), e.g. the colonization of bacteria on human skin has been related to maturity and increase in skin substrate availability (Nordstrom & Noble, 1984).

In badgers, mutual allo-marking is commonly assumed to facilitate the exchange, and thus gradual assimilation, of pouch bacteria between group members (Kruuk *et al.*, 1984; Buesching *et al.*, 2003). While very young cubs have been shown to anoint themselves with the scent of adult group members by brushing past their subcaudal pouch region (a behavior termed scent-theft: Fell *et al*. 2006), older cubs producing their own subcaudal gland secretion have to obtain the typical bacterial community from other badgers that harbor the specific 'mature bacterial community' to enable them to produce typical adult scent profiles. Hence, the subcaudal pouch microflora of cubs is likely to undergo a succession from the more diverse and variable pioneering communities dominated by Firmicutes to less diverse and more specific communities dominated by Actinobacteria found in older individuals (yearlings and adults).

In addition, the composition of the microbial community is affected by the primary gland products. Some Actinobacteria are capable of metabolizing sex steroids (Gower *et al.*, 1986; Kohl *et al.*, 2001), which will only be present in the primary gland secretion in measurable amounts after the host reaches sexual maturity. The chemical composition of the subcaudal gland secretion changes significantly with maturity (Buesching *et al.*, 2002c). Hence the shift from Firmicutes in cubs to Actinobacteria as the most prevalent bacteria in adults is likely to be influenced by the onset of puberty (such as in humans; Nordstrom & Noble, 1984), making the resources and environment progressively more suitable for Actinobacteria. On the other hand, there was no significant difference between age categories in badgers older than one year (i.e. yearlings, young and old adults), indicating that bacterial communities stay relatively stable once established during the first year of life and are not liable to major changes through senescence (Buesching *et al.*, 2009; Dugdale *et al.*, 2011c).

Previous studies on humans have shown that the levels of blood androgens contribute to the differences in axillary bacterial community in men and women (Kohl *et al.*, 2001). Similarly, the difference in bacterial diversity between recently breeding and non-breeding female badgers is likely due to diverging hormone levels, and thus the composition of their primary gland products, resulting in the differences in the chemical composition (e.g. dodecanoic acid, tetradecanoic acid and heptadecanoic acid) reported previously (Buesching *et al.*, 2002c). The accuracy of species richness

estimation from T-RFLP has been criticized because of its limited ability to detect rare taxa (Bent *et al.*, 2007), particularly in highly diverse communities (Orcutt *et al.*, 2009). However, it is generally considered to be robust for the purpose of relative comparisons, identifying dominant species and examining richness of low-diversity communities (Engebretson & Moyer, 2003; Orcutt *et al.*, 2009).

Bacterial communities differed with badger body condition in spring but not in summer. As some of the chemicals in the subcaudal gland secretion are correlated with badger body condition in some seasons, a possible explanation for this observation is that the composition and/or amounts of primary gland products serving as food sources for different bacteria also vary with body condition (Buesching *et al.*, 2002a; Buesching *et al.*, 2002c), although this variation is not ubiquitous throughout the year (Buesching *et al.*, 2002c). However, the difference between body conditions was not as obvious as the difference between cubs and adults. As the difference between body conditions was only found in spring but not in summer, the small sample size in one season may limit the power of the analyses. Analysis involves more samples should be performed to clarify the difference between body conditions in future studies.

The chemical composition of *M. meles* subcaudal gland secretion encodes a diverse array of information about the scent owner (Buesching *et al.*, 2002c, b). Surprisingly, our results showed that highly significant differences in the bacterial communities existed only between cubs and adults, whereas other factors (except spring body condition) did not affect the microflora significantly. However, it is possible that factors such as kinship (Weisfeld *et al.*, 2003), social group membership (Buesching *et al.*, 2003), and host genotypes (Kuhn & Natsch, 2009), particularly the major histocompatibility complex (Scott & Dobson, 1989; Lanyon *et al.*, 2007; Zomer *et al.*, 2009; Sin *et al.*, 2012b; Sin *et al.*, 2012c), have a greater influence on shaping the community of the symbiotic bacteria. Future studies examining bacterial communities in badger subcaudal gland secretion should thus include the above key factors together with analyses of the semio-chemical profiles to determine the role of symbiotic bacteria in the olfactory communication of badgers.

Supplementary material

Table S4.1 16S rRNA gene sequencing results from *M. meles* subcaudal gland secretion. Identification of terminal restriction fragments (TRFs) generated from samples analyzed by terminal restriction fragment length polymorphism (T-RFLP). Shown are the predicted sizes of the TRFs based on the nucleotide sequence of individual clones compared to the detected TRFs in the T-RFLP profile

	Best matched sequence				Forward primer		Reverse primer	
Clone number	Name	% identity	Accession number	Division	Predicted TRF (bp)	Detected TRF (bp)	Predicted TRF (bp)	Detected TRF (bp)
SGS-1	Uncultured bacterium clone HWB2224-3-70	99	HM243988.1		102	97.6	245	241.4
SGS-2	*Rhizobium* sp./ *Agrobacterium rhizogenes*	100	HQ698283.1/ EU420078.1	Alphaproteobacteria	102	97.6	420	416.4
SGS-3	Alpha proteobacterium/ *Bacillus* sp.	97	AB302355.1/ EF489025.1	Alphaproteobacteria / Firmicutes	102	97.6	298	295.5
SGS-4	Uncultured beta proteobacterium clone GASP-MB2S3_G01	96	EF665258.1	Betaproteobacteria	103	99.2	311	308.1
SGS-5	Uncultured *Sphingomonas* sp. clone DWIIB03	99	HQ711900.1	Alphaproteobacteria	103	99.2	90	85.0
SGS-6	*Methylobacterium iners* strain 5317S-33	99	EF174497.1	Alphaproteobacteria	103	99.2	179	177.2
SGS-7	Uncultured bacterium clone ncd985d08c1	99	HM332406.1		138	133.7	179	177.2
SGS-8	*Eubacterium tenue*	100	FR749984.1	Firmicutes	138	133.7	367	364.6
SGS-9	Uncultured bacterium clone PB2_aai24h05	99	EU460716.1		138	133.7	367	364.6
SGS-10	Uncultured Clostridiales bacterium clone 8837-D0-A-9C/ Uncultured Peptostreptococcaceae bacterium clone 8817-D4-C-9A	97/ 97	EU289123.1 / EU289092.1	Firmicutes/ Firmicutes	138	133.7	204	201.8
SGS-11	Uncultured bacterium clone ncd1676f01c1	99	JF147747.1		138	133.7	204	201.8
SGS-12	*Rhizobium cnuense*	100	AM498743.1	Alphaproteobacteria	138	133.7	311	308.1
SGS-13	Uncultured *Clostridium* sp. clone B255_G12	99	EF092228.1	Firmicutes	138	133.7	367	364.6
SGS-14	Uncultured bacterium/ *Methylopila* sp. JZL-4	98/ 97	AB461120.1 /FJ502233.3	/ Alphaproteobacteria	138	133.7	420	416.4
SGS-15	Uncultured alpha proteobacterium	100	FM866279.1	Alphaproteobacteria	138	133.7	204	201.8
SGS-16	Uncultured bacterium clone	100	EU460716.1		138	133.7	367	364.6
SGS-17	Uncultured Actinomycetaceae bacterium clone D12_40	93	EU266839.1	Actinobacteria	143	140.4	68	62.2
SGS-18	Uncultured bacterium clone 8837-D60-PY-G-6C *Propionibacterium avidum*	97 94	GQ358900.1 FN824488	Actinobacteria	143	140.4	371	368.0
SGS-19	*Arthrobacter nicotianae* strain GJA819	99	HM209738.1	Actinobacteria	143	138.9	371	368.0
SGS-20	Uncultured *Phycicoccus* sp. clone P8s-320/ *Knoellia* sp. DMZ1/ *Janibacter* sp. 0705C10-1	99	GQ329496.1/ HQ171909.1/ HM222661.1	Actinobacteria/ Actinobacteria/ Actinobacteria	143	138.9	301	298.5
SGS-21	Uncultured bacterium clone RT_7/ *Pseudonocardia* sp. CNS139 PL04/ uncultured Actinobacteridae bacterium	99/ 98/ 98	EU644220.1/ DQ448726.1/ AJ555207.1	/ Actinobacteria/ Actinobacteria	143	138.9	207	204.5
SGS-22	*Geodermatophilus* sp. 73NP15/ *Actinotelluria brasiliensis* strain Tu6233/ *Modestobacter* sp. R-36506	99	AB242705.1/ DQ029102.1/ FR682686.1	Actinobacteria/ Actinobacteria/ Actinobacteria	143	138.9	207	204.5
SGS-23	Uncultured actinobacterium clone O_2B10	98	HM480650.1	Actinobacteria	143	138.9	207	204.5
SGS-24	*Janibacter* sp. M3-5/ Intrasporangiaceae bacterium VUG-A40a/	99	HQ425307.1/ GQ454846.1/	Actinobacteria/ Actinobacteria/	143	138.9	302	299.4

	Knoellia sp. R-43101		FR682681.1	Actinobacteria				
SGS-25	Uncultured *Ornithinimicrobium* sp. clone De253	99	HQ183913.1	Actinobacteria	143	138.9	301	298.5
SGS-26	Uncultured bacterium clone Luq_GN490_032/	100/	HQ445817.1/	/	143	138.9	371	368.0
	Arthrobacter arilaitensis	99	FQ311875.1	Actinobacteria				
SGS-27	Uncultured Sphingobacteriales bacterium clone YL117/	98/	HM856483.1/	Bacteroidetes/	158	156.6	366	365.6
	Uncultured Bacteroidetes bacterium/	98/	FN811193.1/	Bacteroidetes/				
	Uncultured Flavobacteria bacterium clone GASP-WB2W1_F09	98	EF074147.1	Bacteroidetes				
SGS-28	*Escherichia coli* KO11	100	CP002516.1	Gammaproteobacteria	162	160.6	68	62.2
SGS-29	Uncultured actinobacterium clone GASP-WC2S3_B11	99	EF075075.1	Actinobacteria	162	160.6	267	264.2
SGS-30	Uncultured Comamonadaceae bacterium clone Elev_16S_1456	99	EF020052.1	Betaproteobacteria	163	160.6	310	308.1
SGS-31	*Janthinobacterium lividum* strain NBGD31	99	HQ003440.1	Betaproteobacteria	163	160.6	310	308.1
SGS-32	Uncultured bacterium clone J2Q401_43C06/	100/	FJ833807.1/	/	163	160.6	68	62.2
	Escherichia coli KO11	99	CP002516.1	Gammaproteobacteria				
SGS-33	*Pseudomonas* sp. W15Feb31	99	EU681014.1	Gammaproteobacteria	163	160.6	311	308.1
SGS-34	*Acinetobacter* sp. 4087	99	FJ405317.1	Gammaproteobacteria	164	163.0	421	418.0
SGS-35	*Acinetobacter xiamenensis*	97	EF030545.1	Gammaproteobacteria	164	163.0	421	418.0
SGS-36	*Clostridium baratii* strain IP 2227/	99/	NR_029229.1/	Firmicutes	190	189.3	310	308.1
	Uncultured Firmicutes bacterium	99	CU919549.1					
SGS-37	*Clostridium baratii* strain IP 2227	96	NR_029229.1	Firmicutes	190	189.3	310	308.1
SGS-38	*Clostridium perfringens* str. 13	99	BA000016.3	Firmicutes	190	189.3	310	308.1
SGS-39	Uncultured Ruminococcaceae bacterium clone EMP_N10	99	EU794265.1	Firmicutes	192	191.4	299	296.7
SGS-40	Uncultured bacterium clone TU1_aaa01f06/	99/	EU779135.1/	/	193	191.4	68	62.2
	Uncultured Firmicutes bacterium clone CF2-133	98	GU959006.1	Firmicutes				
SGS-41	Uncultured Bacteroidales bacterium clone EMP_W5	96	EU794146.1	Bacteroidetes	209	208.2	366	365.6
SGS-42	Uncultured *Bacteroides* sp. clone EMP_Z12	99	EU794298.1	Bacteroidetes	210	208.2	29	NA[a]
SGS-43	Uncultured bacterium clone OK2_d08	90	EU468645.1		210	208.2	29	NA[a]
SGS-44	Uncultured Aerococcaceae bacterium clone 8837-D0-C-9H	95	EU289060.1	Firmicutes	214	212.0	370	366.3
SGS-45	Uncultured *Enterococcus* sp. clone 1-4/	96/	AY830399.1/	Firmicutes/	214	212.8	302	299.4
	Isobaculum melis strain M577-94	96	NR_025445.1	Firmicutes				
SGS-46	Uncultured bacterium clone PCD-8/	100/	EF608536.1/	/	215	212.8	302	299.4
	Enterococcus faecalis strain LW88	99	HQ641405.1	Firmicutes				
SGS-47	*Enterococcus faecalis* strain LW88	99	HQ641405.1	Firmicutes	215	212.8	302	299.4
SGS-48	*Enterococcus durans* strain 41D/	99	HQ677826.1/	Firmicutes	215	212.8	302	298.5
	Enterococcus faecium/		AB596997.1/					
	Enterococcus faecalis strain R16		HQ228219.1					
SGS-49	*Lactobacillus johnsonii* strain 466	100	HM162410.1	Firmicutes	215	214.1	302	299.4
SGS-50	*Lactobacillus reuteri* DSM 20016	99	CP000705.1	Firmicutes	215	214.1	68	62.2
SGS-51	*Lactobacillus reuteri*	99	AB596995.1	Firmicutes	215	214.1	68	62.2
SGS-52	*Streptococcus dysgalactiae* subsp. equisimilis strain zsg	99	HQ684022.1	Firmicutes	216	214.1	68	63.5
SGS-53	*Streptococcus lutetiensis* strain NWL68	100	HQ293091.1	Firmicutes	216	214.1	81	75.7
SGS-54	Uncultured bacterium clone nbw159g11c1 /	95/	GQ073954.1/	/	216	214.1	68	63.5
	Negativicoccus succinicivorans strain BI001	94	HQ264056.1	Firmicutes				
SGS-55	*Negativicoccus succinicivorans* strain ADV 07/08/06-B-1388	94	FJ715930.1	Firmicutes	216	214.1	68	63.5
SGS-56	*Negativicoccus succinicivorans* strain ADV 07/08/06-B-1388	95	FJ715930.1	Firmicutes	216	214.1	68	63.5

[a] TRF too short to be determined by T-RFLP

Table S4.2 Relative abundance of forward and reverse TRFs in T-RFLP data of all individuals (n = 79)

Forward primer		Reverse primer	
Detected TRF (bp)	Relative abundance (%)	Detected TRF (bp)	Relative abundance (%)
97.6	0.46	62.2	1.23
99.2	0.57	63.5	0.12
114.3	0.01	64.2	0.06
119.3	0.02	75.7	0.68
125.6	0.05	76.5	0.05
132.4	0.52	85.0	0.16
133.7	3.30	124.0	0.45
134.3	0.09	140.7	0.01
135.3	0.09	143.7	0.02
136.0	0.17	177.2	0.37
136.5	0.13	179.8	0.02
138.9	3.09	200.7	0.02
140.4	72.74	201.8	0.25
141.7	0.34	203.1	0.22
142.5	0.20	204.5	1.61
143.8	0.06	221.0	0.02
144.6	0.21	241.4	0.15
149.5	0.01	243.3	0.39
152.5	0.05	250.8	0.02
153.7	0.01	264.2	0.08
156.6	0.09	295.5	0.12
157.8	0.01	296.7	0.12
159.5	0.06	298.5	1.83
160.6	3.30	299.4	1.42
161.7	0.72	306.6	0.06
163.0	0.06	308.1	2.03
186.8	0.17	309.0	0.30
188.0	0.04	310.4	0.17
189.3	1.63	311.9	0.08
191.4	0.13	346.6	0.02
192.9	0.27	356.0	0.02
194.2	2.46	362.0	0.26
204.3	0.02	363.1	0.16
208.2	0.52	364.6	1.47
212.0	1.77	365.6	0.28
212.8	3.60	366.3	2.14
214.1	2.64	368.0	78.76
215.2	0.01	369.0	0.07
216.4	0.01	369.7	0.03
220.9	0.06	370.4	0.01
246.1	0.02	371.1	0.01
250.6	0.01	416.4	1.97
269.2	0.03	417.2	0.36
271.1	0.07	418.0	0.21
279.6	0.01	419.1	2.08
280.2	0.01	419.6	0.10
298.7	0.03	426.6	0.01
366.6	0.08		
375.2	0.02		
376.7	0.04		

Chapter 5

MHC-based patterns:

Female preference for MHC-similar male

European badgers (*Meles meles*)

5.1 Abstract

The major histocompatibility complex (MHC) plays a crucial role in the immune system, and has been shown in some species as a target by which individuals choose mates in order to optimize the fitness of their offspring, potentially mediated by olfactory cues. Under the genetic compatibility hypothesis, females are predicted to choose mates with dissimilar MHC, to enhance the resistance of their offspring to pathogens. Studies on MHC-based mate choice in wild mammals are under-represented currently. We investigated whether mate choice based on compatibility of the MHC class I genes and MHC class II DRB genes occurred in a population of wild European badgers (*Meles meles*). We also compared the MHC with neutral microsatellite markers to distinguish MHC-specific effects from genome-wide effects. We found evidence for MHC-assortative mating based on DRB genes for neighbouring-group parent pairs; there was an overall dissimilarity on DRB amino acid distance for both the whole exon 2 sequences and antigen-binding sites (ABS) only. We also found negative correlation between MHC class I genes amino acid distance and relatedness at microsatellite loci, indicating possible correlation between inbreeding avoidance and MHC class I genes-based mating. We found considerable annual fluctuations in the occurrence of MHC-based patterns. MHC class II DRB similarity was the apparent actual target, as no genome-wide based inbreeding, based on 35 microsatellite loci, was found in those mating randomizations.

5.2 Introduction

Indirect genetic benefits have been proposed as the mechanism underpinning mate selection, especially in case where direct benefits have not been detected (Zelano & Edwards, 2002). In theory, females should select mates based on 'good genes' or genetic compatibility (Trivers, 1972; Neff & Pitcher, 2005; Kempenaers, 2007). The good genes hypothesis predicts that all breeding females in the population show consistent patterns of choice, irrespective of their own genotypes (Neff & Pitcher, 2005). While compatibility, which contributes non-additive genetic variance (Puurtinen *et al.*, 2005), requires individuals to assess the genotype of potential mates by self-reference to their own genotype.

The major histocompatibility complex (MHC) has the potential to be the target of this compatibility assessment (Penn & Potts, 1998a; Mays & Hill, 2004). Evidence for a role of the MHC in mate choice has been accumulating (Bernatchez & Landry, 2003) since it was discovered in mice (Yamazaki *et al.*, 1976). The MHC is a diverse gene family that plays a crucial role in the adaptive immune system, as it encodes cell surface glycoproteins that bind and present antigens to T cells and trigger an immune cascade (Swain, 1983). The MHC may therefore provide the basis on which individuals discriminate mating partners, in order to optimize the fitness of their offspring (Penn & Potts, 1999; Penn *et al.*, 2002; Milinski, 2006). MHC genes are under constant selective pressures as a consequence of challenges from parasites and pathogens (Jeffrey & Bangham, 2000; Piertney & Oliver, 2006). This arms race between pathogens and hosts maintains the extreme diversity in MHC genes. MHC-related reproductive mechanisms, such as mating preferences, selective fertilization and abortion, could also contribute to generating MHC diversity (Ziegler *et al.*, 2005). MHC genotypes have been posited to be detectable as olfactory cues (Penn, 2002; Leinders-Zufall *et al.*, 2004; Milinski *et al.*, 2005), or through other traits (e.g. ornamentation; vonSchantz *et al.*, 1996; vonSchantz *et al.*, 1997; Dunn *et al.*, 2012), thus allowing the MHC to be an important target for mate choice through it's indirect genetic benefits.

The evolution of MHC-based mate choice is expected to favour the combination of MHC that confers the highest resistance to pathogenic invasions. Under the genetic compatibility hypothesis, females are predicted to exhibit preferences for mates possessing complementary genotypes to their own (Trivers,

1972; Tregenza & Wedell, 2000; Neff & Pitcher, 2005). Because of the codominant nature of MHC expression, individuals with a greater MHC diversity may have an advantage when facing a wide array of pathogens (Hughes & Nei, 1992; McClelland *et al.*, 2003). Choosing mates with dissimilar MHC can maximize the MHC diversity, and consequently enhance disease resistance among offspring. MHC-disassortative mating has been found in a variety of animals (Potts *et al.*, 1991; Landry *et al.*, 2001; Olsson *et al.*, 2003; Consuegra & Garcia-de-Leaniz, 2008; Neff *et al.*, 2008; Schwensow *et al.*, 2008a; Havlicek & Roberts, 2009; Miller *et al.*, 2009; Setchell *et al.*, 2010). In contrast, a preference for MHC-similar males has also been found in other species (Yamazaki *et al.*, 1978; Roberts *et al.*, 2005; Bonneaud *et al.*, 2006a; Bos *et al.*, 2009; Bollmer *et al.*, 2012). Less attention has been devoted to the risk of mating with a partner with highly divergent MHC genes, despite the fact that disruption of co-adapted genes may reduce fitness (Kaufman, 1999; Hendry *et al.*, 2000; Neff, 2004).

The European badger (*Meles meles*) is well suited to investigate MHC-based mate choice in the wild given their mating system and reproductive biology. In high-density populations, *M. meles* is group living (Newman *et al.*, 2011) and has a polygynandrous mating system (i.e. they do not have one exclusive social mate; Dugdale *et al.*, 2007; Dugdale *et al.*, 2011b) with high levels of extra-group paternity (Carpenter *et al.*, 2005; Dugdale *et al.*, 2007) and low fecundity (i.e. 1-2 cubs once a year; Macdonald & Newman, 2002; Carpenter *et al.*, 2005; Dugdale *et al.*, 2007). European badgers have delayed implantation, which uncouples mating and parturition (Thom *et al.*, 2004), and a mating season with a post-parturition peak in late-February to March, but with some females ovulating subsequently later into the summer (Woodroffe, 1995; Yamaguchi *et al.*, 2006; Roper, 2010). As female badgers are potentially capable of superfoetation (conception during pregnancy; Yamaguchi *et al.*, 2006) and embryo reabsorption (Yamaguchi *et al.*, 2006), in combination this extends the opportunity for females to select the most suitable mates. These components may underscore MHC-based mate choice. Furthermore, badgers are under selective pressure from a wide array of pathogens (Jeffrey & Bangham, 2000; Piertney & Oliver, 2006). Given this mating system, together with their life-history characteristics (Macdonald *et al.*, 2009) and lack of paternal care (Dugdale *et al.*, 2010), we predict that badgers are strong candidates to have evolved mate choice for indirect genetic benefits.

MHC-based choice can be used to select extra-pair mates, for example, females may seek extra-pair matings if their social partners have low MHC diversity or similar MHC (Richardson *et al.*, 2005; Schwensow *et al.*, 2008b) where this results in offspring with greater MHC heterozygosity (Potts *et al.*, 1991). In this badger population, approximately 50% of paternities were assigned to extra-group males, the majority of which were neighbouring males (Carpenter *et al.*, 2005; Dugdale *et al.*, 2007). Given that badgers are more related to within-group than to neighbouring-group members (Dugdale *et al.*, 2008), MHC-similarity is likely to vary with respect to group membership, we therefore analysed within-group and extra-group parent pairs separately.

We investigated the genetic compatibility hypothesis by examining whether badgers choose mates based on MHC similarity (class II DRB and class I), taking into account both the number of shared MHC alleles and functional differences between alleles. To distinguish MHC-specific effects from genome-wide effects, we compared MHC similarity with that of potentially neutral microsatellite markers. Mating associated with MHC similarity but not microsatellite similarity would indicate that the MHC is the actual target of mate choice.

5.3 Materials and methods

Study site and sample collection

A high-density (44 badgers/km^2; Macdonald & Newman 2002) population of badgers was studied in Wytham Woods, a 4 km^2 deciduous woodland in Oxfordshire, UK (51°46'26N, 1°19'19W). Trapping events have usually been undertaken four times a year since 1987 (Macdonald & Newman, 2002), generally over two weeks in June, September and November, with one week of trapping in January of some years (Macdonald *et al.*, 2009). Badgers were caught in mesh-traps baited with peanuts placed near the entrances of active setts (Macdonald & Newman, 2002; Macdonald *et al.*, 2009). Captured badgers were transferred to holding cages and transported to a central handling facility. Badgers were then sedated by an intra-muscular injection using ketamine hydrochloride at 0.2 ml/kg body weight (McLaren *et al.*, 2005). Upon their first capture all badgers were tattooed with a unique individual

number on the left inguinal region for permanent individual identification. The sex, age group/maturity (cub or adult based on body size) and location (sett name) of each badger were recorded. These protocols were subject to ethical review and were performed under Natural England Licence (currently 20104655) and UK Home Office Licence (PPL 30/2835). During trapping DNA samples were collected: guard hairs were plucked and stored in 80% ethanol, and approximately 3 ml of blood was taken by jugular venipuncture and collected in a vacutainer containing EDTA. Blood samples were stored at -20°C, and hair samples in 80% ethanol at room temperature, until DNA isolation was performed.

Social group ranges were established using bait-marking techniques every two years (Delahay *et al.*, 2000; Macdonald *et al.*, 2008; Kilshaw *et al.*, 2009), the mean number of social groups in which badgers were present each year was 19 ± 2 (range = 14–26; Dugdale *et al.*, 2008). The majority of badgers in this population exhibit high group-fidelity; only 19% of individuals dispersed to other social groups (Macdonald *et al.*, 2008). Temporary group-movements do however occur at quite a high rate (16.4%, Macdonald *et al.* 2008). Taking into account both permanent dispersal and temporary movements, we defined the social group of residence of an individual using the following criteria: (a) badgers were assigned to the social group (natal social group) in which they were first trapped; or (b) a badger was resident within a social group if the 2 most recent captures were made within the same social group.

DNA extraction

Genomic DNA was isolated using the GFX Genomic Blood DNA Purification Kit (Amersham Biosciences, Little Chalfont, UK), following the scalable method in the manufacturer's protocol, or from a minimum of 20 hairs with visible follicles using a Chelex protocol (Walsh *et al.*, 1991).

MHC gene amplification

The DRB gene is the most variable among the four MHC class II genes studied in badgers, with four putatively functional sequences identified from both gDNA and cDNA (Sin *et al.*, 2012b). Seven putatively functional MHC class I sequences have also been identified from both gDNA and cDNA in

the same population (Sin *et al.*, 2012c). The variable DRB and class I genes thus are good targets for investigation of MHC-based mate choice in this species. Their intermediate, but not too high, variability also allows inference of the numbers of loci and haplotypes. We used published primers to amplify the regions that encode the antigen-binding domain in class I (Meme-MHCIex3, for the exon 3; Sin *et al.* 2012b) and class II DRB (Meme-DRBex2, for the exon 2; Sin *et al.* 2012c) genes. Previous studies indicated the presence of at least two DRB loci and two class I loci (Sin *et al.*, 2012b; Sin *et al.*, 2012c). Using these primers on 10–30 ng of gDNA, PCR amplification was performed in a 10-μl reaction mix that also contained: 0.5 μM of each primer, 200 μM of each dNTP, 1 X PCR buffer (containing $MgCl_2$; Qiagen), and 1 units of HotStarTaq (Qiagen). The PCR cycle began with incubation at 94°C for 15 min, followed by 35 incubation cycles at 94°C for 30 s, 59°C for 30 s, and 72°C for 60 s, ending with an extension step at 72°C for 10 min.

MHC genotyping

The MHC was genotyped using reference strand-mediated conformation analysis (RSCA) following Arguello *et al.* (1998). Fluorescent-labeled reference strands (FLRs) were generated using alleles from closely related species (American badger *Taxidea taxus*, polecat *Mustela putorius*, stoat *Mustela erminea* and mink *Neovison vison*) and based on phylogenetic analyses. FLRs were generated by PCR using cloned alleles as templates and a 5'-FAM-labelled primer. Cloning and sequencing details are described in Sin *et al.* (2012b, 2012c). The same PCR protocol, detailed in the MHC gene-amplification section, was used to generate the FLRs, except that the primer proportion was altered to 0.5 μM FAM-labelled primer (Meme-DRBex2R or Meme-MHCIex3F) and 0.1 μM unlabeled primer (Meme-DRBex2F or Meme-MHCIex3R). Nine FLRs were trialled for each of the class II DRB and the class I genes to determine a subset that could best resolve all sequences in the control samples. A final sets of three FLRs derived from the American badger, polecat and stoat clones and three FLRs derived from American badger and mink were used for the class II DRB genes and the class I genes, respectively.

All the resulting FLRs were diluted 1:5 in ddH_2O before use in the hybridization reactions. To form heteroduplexes, 2 μl diluted FLR and 2 μl test sample PCR product were mixed and incubated in a

thermal cycler at 95°C for 10 min, cooled down to 55°C at 1°C/s, hybridized at 55°C for 20 min, cooled down to15°C for 1 min and put on ice for 30 min. The plate was stored at 4°C until required. Subsequently, 3 µl of hybridization product was mixed with 7.82 µl water and 0.18 µl Genescan Rox-500 size standard [Applied Biosystems (ABI), Foster City, CA, USA] in a 96-well plate. The samples were then run on an ABI 3100 Genetic Analyzer, 4% Genescan non-denaturing polymer (ABI), and data collected using matrix Dye set D. The running protocol was an injection voltage of 8 kV, injection time 15 s, run voltage of 15 kV, run temperature of 18°C. The heteroduplex peaks were identified using GENEMAPPER 3.7 (ABI) and their motilities were estimated relative to the internal ROX size standard. Control alleles from cloned plasmids were included in each run to control for variation between runs. Peaks with the same motility across different FLRs were designated as identical putative alleles. All seven class I alleles and four class II DRB alleles identified were sequenced (Sin *et al.*, 2012b; Sin *et al.*, 2012c).

Microsatellite typing and parentage analyses

Individuals were genotyped at 35 microsatellites loci (amplification conditions described in Annavi *et al.*, 2011). Details of candidate parent rules, genotyping procedures and parentage analysis are described elsewhere (Dugdale *et al.*, 2007; Annavi *et al.*, subm.). Briefly, parentage and sibships were assigned with 80% confidence (accounting for genotyping error and unsampled individuals: see Dugdale *et al.*, 2007; Annavi *et al.*, subm.) using MasterBayes 2.47 (Hadfield *et al.*, 2006) implemented in R 2.12.2 (R Development Core Team 2010), and Colony 2.0 (Wang & Santure, 2009), respectively. Candidate fathers were all sexually mature males (older than one year old) present in the year before the cub was born (due to delayed implantation; Yamaguchi *et al.*, 2006), and candidate mothers were females aged two or more in the year the cub was born that were resident in the cub's natal group.

Presence of individuals were based on trapping history according to the following rules: (a) individuals trapped in that year; (b) individuals were included for an extra 2 years after the their last capture if their death date was unknown, to minimize error from badgers that were present but not caught (Dugdale *et al.*, 2007); (c) and individuals were included for the whole period between two

trappings. Only six years [1993 (class I data set: $N = 36$; class II data set: $N = 47$), 2004 (class I: $N = 32$; class II: $N = 33$), 2005 (class I: $N = 52$; class II: $N = 55$), 2008 (class I: $N = 27$; class II: $N = 27$), 2009 (class I: $N = 30$; class II: $N = 30$) and 2010 (class I: $N = 23$; class II: $N = 23$)] with high number of cubs that were assigned parents were used in this study. The loci were in Hardy–Weinberg equilibrium and linkage equilibrium (see Dugdale *et al.*, 2007; Annavi *et al.*, subm.). Parents that were from the same social group were categorised as within-group pairs, whereas parents that were from the different social group were categorised as extra-group pairs.

In this paper we use mate choice/preference in a broad sense such that it comprises both pre- and post-copulatory mate choice. Assigned parent pair here means a male and female badger that were genetically assigned as the mother and father of the same cub.

Data analysis

MHC similarity parameters

To examine whether assigned parent pairs have more (or less) similar MHC genes than would be expected under random mating, we examined MHC compatibility using both the extent of allele-sharing and the magnitude of the differences between genotypes. According to the heterozygote advantage hypothesis, it is beneficial to have alleles that are different (Potts & Wakeland, 1990). On the other hand, two alleles differing by more amino acids than two less different alleles could potentially recognize and bind a wider range of antigens. The allele-sharing value is the number of alleles shared between assigned parents. To incorporate the functional similarity between alleles in the analysis, we also calculated the amino acid distance from pairwise combinations of alleles in assigned parent pairs (Landry *et al.*, 2001). We used the average amino acid distance (Landry *et al.*, 2001; Forsberg *et al.*, 2007; Miller *et al.*, 2009) because individuals might carry different numbers of alleles and hence the number of pairwise comparisons could be different between parent pairs. We calculated the amino acid distance for both the whole exon sequences (exon 2 for DRB and exon 2+3 for class I; Sin *et al.* 2012b, 2012c) and just the antigen-binding sites (ABS) in these exons only (Sin *et al.*, 2012b; Sin *et al.*, 2012c). The ABS is the basis of antigen recognition, so it may only be variation in

this region that is functionally important. We analysed MHC class I and class II genes separately as they have different structure, function and expression patterns (Hughes & Yeager, 1998).

Randomization test

Near half of the parentage were assigned to neighbouring-group candidate fathers (six years data set: 98/230 = 42.6%) and since most extra-group mating occurred between neighbouring groups (six years data set: 89/98 = 90.8%), the analyses for extra-group mating were restricted to assigned parent pairs from neighbouring groups only and random pairs were assigned from neighbouring groups only during randomization. By performing a randomization test (Landry *et al.*, 2001), we compared the mean allele-sharing value and amino acid distance value for the assigned parent pairs with the frequency distribution of mean values generated from 1000 simulations of the same number of random parent pairs. Specifically, all parent pairs assigned in parentage analyses were disassociated and all assigned mothers were reassembled to candidate fathers (see below) at random with replacement and the mean values were calculated 1000 times in order to obtain a distribution of values under random mating. The *p*-value was calculated as the proportion of the total number of iterations greater or smaller than the observed mean (Fisher, 1935) and a 2-tailed test with $\alpha = 0.05$ was applied, with any values that fall out of this 97.5-2.5% CI are significant. All statistical analyses were performed in R v. 2.15.0 (R Core Development Team 2012).

We examined the MHC similarity of the assigned fathers and other candidate fathers to assigned mothers in two sets of comparison, given that badgers are more related to within-group than to neighbouring-group members (Dugdale *et al.*, 2008):

(a) First, we compared the assigned within-group parents with random within-group pairs, which included all within-group candidate fathers including the males that were not assigned paternity in addition to the assigned fathers.

(b) Second, we compared MHC similarity of the assigned neighbouring-group parents with random neighbouring-group pairs, which included all neighbouring candidate fathers that were not assigned paternity in addition to the assigned fathers.

To ensure that any observed MHC-based mating was not as a consequence of selection on wider genetic diversity, we also examined mating patterns at neutral markers. We calculated the genetic similarity of assigned parent pairs based on 35 microsatellite loci (Annavi *et al.*, 2011) using a pairwise relatedness measure *R* (Queller & Goodnight, 1989). The mean pairwise relatedness of assigned parent pairs was compared with 1000 simulations of randomly chosen pairs.

Combining probabilities

We assessed overall significance, across years, using Fisher's method of combining probabilities (Sokal & Rohlf, 1994) for independent tests. The data points were different reproductive events with different candidate males and females presented each year. Furthermore, females are able to prevent males from mounting them (59/257 = 23% of mounting events are failed mountings; Dugdale *et al.*, 2011b) so they won't be forced to reproduce with the same males each year. 116 females that were assigned maternity were examined, 16 of these females were assigned the same male in >1 year and 8 of them were also assigned to other males in the same or different year. The repeated assigned pairs that appeared in >1 year were excluded randomly to follow the assumption of combining *p*-values from independent tests using Fisher's method. Fisher's method was carried out in left or right-sided *p*-values ($\alpha = 0.025$) for detecting if there were lower or higher mean assigned values respectively.

Correlation between variables

As amino acid distance could be correlated with allele sharing, relatedness, within/neighbouring-group pairing or year, we tested for associations between them and amino acid distance. We used linear mixed model, included four fixed effects: two continuous (allele sharing and relatedness) and two categorical effects (pairing [within- or neighbouring-group] and year), with amino acid distance (all sites) as the response variable. Amino acid distance (ABS only) was not included as it was highly correlated with amino acid distance (all sites) (correlation coefficient is 1.0 [class II] and 0.97 [class I]). We controlled for females with maternity assigned in more than one year and social group of the assigned mother, by including individual identity and social group identity of assigned mothers as random effects, respectively. Model selection was based on Akaike's information criterion corrected for sample size (AICc; Akaike 1973). Models that are better supported by the data result in lower

AICc values. Multi-model inference (Burnham & Anderson, 2002) was performed for models with ΔAICc < 7 (Burnham *et al.*, 2011). Model averaged parameter estimates, unconditional standard errors and 95% confidence intervals (Anderson, 2008) of parameter estimates are reported.

5.4 Results

Major histocompatibility complex screening

MHC class II DRB genotypes were determined for 520 assigned parent pairs and MHC class I genotypes for 501 assigned parent pairs. Among them, six class II DRB and 12 class I genotypes were identified. Four class II DRB and seven class I genes were identified and shown to be putatively functional sequences (Sin *et al.* 2012b, 2012c). Three DRB and five class I haplotypes (Fig. 5.1) were inferred from parentage assignments, based on the assumption of Mendelian inheritance.

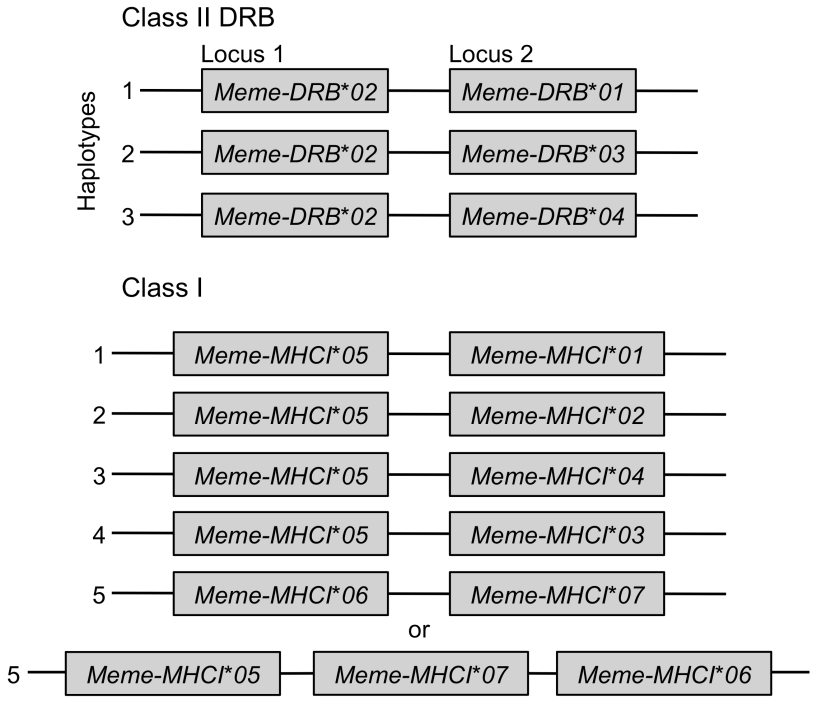


Fig. 5.1 Three MHC class II DRB haplotypes and five MHC class I haplotypes in the study population of European badger (*Meles meles*). For MHC class I haplotype 5, it was not certain whether it comprises *Meme-MHCI*05* or not, because *Meme-MHCI*05* was present in all examined individuals and the frequency of haplotype 5 was low in the population and no homozygote was identified.

(a) *Within-group parents*

There was no overall effect of assigned within-group parent pairs shared different number of class I and class II alleles from randomised within-group pairs (Fisher's method of combining probabilities: class I: $\chi^2_{12} = 20.54$, $p > 0.05$; class II: $\chi^2_{12} = 13.04$, $p > 0.05$; Table 5.1; Fig. 5.2 & 5.3). Assigned within-group pairs overall did not have different amino acid distances from random within-group pairs [Fisher's method of combining probabilities: class I (all sites): $\chi^2_{12} = 21.12$, $p > 0.05$; class II (all sites): $\chi^2_{12} = 21.9$, $p > 0.05$; class I (ABS only): $\chi^2_{12} = 20.97$, $p > 0.05$; class II (ABS only): $\chi^2_{12} = 21.35$, $p > 0.05$; Table 5.1]. There was an overall effect of assigned within-group pairs less related than random within-group pairs (Fisher's method of combining probabilities: class I data set: $\chi^2_{12} = 44.81$, $p < 0.001$; DRB data set: $\chi^2_{12} = 42.12$, $p < 0.001$; Table 5.1).

(b) *Neighbouring-group parents*

Assigned neighbouring-group pairs had more similar MHC class II genes than random neighbouring-group pairs (Table 5.1; Fig. 5.3). There was no difference in allele sharing between assigned neighbouring-group pairs and random neighbouring-group pairs (Fisher's method of combining probabilities: class I: $\chi^2_{12} = 16.27$, $p > 0.05$; class II: $\chi^2_{12} = 12.23$, $p > 0.05$; Table 5.1), but assigned neighbouring-group pairs had overall smaller amino acid distance than random neighbouring-group pairs for class II genes [Fisher's method of combining probabilities: class II (all sites): $\chi^2_{12} = 32.32$, $p < 0.002$; class II (ABS only): $\chi^2_{12} = 39.0$, $p < 0.001$; Table 5.1]. There was no difference in relatedness between assigned neighbouring-group pairs and random neighbouring-group pairs (Fisher's method of combining probabilities: class I data set: $\chi^2_{12} = 10.72$, $p > 0.05$; DRB data set: $\chi^2_{12} = 9.86$, $p > 0.05$; Table 5.1).

Correlation between variables

No significant association was found between DRB amino acid distance and allele sharing, relatedness, within/neighbouring-group pairing or year (Table 5.2). MHC class I gene amino acid distance was associated negatively with allele sharing and relatedness (Table 5.3). Class I gene amino acid distance in year 2010 was larger than in 1993 (Table 5.3).

Table 5.1 Overall significance of MHC-based mating and inbreeding avoidance across six years (1993, 2004, 2005, 2008, 2009 and 2010), calculated using Fisher's method of combining probabilities (Sokal & Rohlf 1994). *p*-values are combined for lower (or higher) allele sharing, smaller amino acid (a.a.) distance (all site and ABS only) and lower Queller & Goodnight's index of pairwise relatedness (R) of assigned parent pairs to random pairs. Significant p-values are shown in bold.

Parent pairs	MHC gene	Allele sharing	Small a.a. distance (all)	Small a.a. distance (ABS)	Low R
Within-group	Class I	(low)	$\chi^2_{12} = 21.12$	$\chi^2_{12} = 20.97$	**$\chi^2_{12} = 44.81$**
		$\chi^2_{12} = 20.54$, p>0.05	p>0.05	p>0.05	**p<0.001**
	Class II	(low)	$\chi^2_{12} = 21.90$	$\chi^2_{12} = 21.35$	**$\chi^2_{12} = 42.12$**
		$\chi^2_{12} = 13.04$, p>0.05	p>0.05	p>0.05	**p<0.001**
Neighbouring-group	Class I	(high)	$\chi^2_{12} = 20.70$	$\chi^2_{12} = 20.79$	$\chi^2_{12} = 10.72$
		$\chi^2_{12} = 16.27$, p>0.05	p>0.05	p>0.05	p>0.05
	Class II	(high)	**$\chi^2_{12} = 32.32$**	**$\chi^2_{12} = 39.00$**	$\chi^2_{12} = 9.86$
		$\chi^2_{12} = 12.23$, p>0.05	**p<0.002**	**p<0.001**	p>0.05

Table 5.2 The association between MHC class II DRB amino acid distance (all amino acid sites) and DRB allele sharing, Q&G's index of pairwise relatedness, within/neighbouring-group pairing and year: Parameter estimates with their unconditional standard errors (SE), 95% confidence intervals and relative importance after model averaging (models with ΔAICc < 7). Estimates where the 95% confidence intervals do not overlap zero are in bold.

Parameter	Estimate	Unconditional SE	95% Confidence interval		Relative importance^
(Intercept)	0.058	0.005	**0.049**	**0.067**	
DRB allele sharing	-0.001	0.002	-0.005	0.004	0.25
Relatedness	0.005	0.003	-0.002	0.011	0.48
Within-group pairing [a]	-0.001	0.002	-0.004	0.002	0.28
Year 2004 [b]	-0.004	0.005	-0.013	0.006	1.00
Year 2005 [b]	-0.004	0.005	-0.013	0.006	"
Year 2008 [b]	0.000	0.005	-0.009	0.010	"
Year 2009 [b]	0.000	0.005	-0.010	0.009	"
Year 2010 [b]	0.006	0.005	-0.004	0.016	"

^ Relative importance is the sum of Akaike weights for all models (ΔAICc < 7) containing the parameter

[a] neighbouring-group pairing was the reference category

[b] Year 1993 was the reference category

Table 5.3 The association between MHC class I gene amino acid distance (all amino acid sites) and MHC class I allele sharing, Q&G's index of pairwise relatedness, within/neighbouring-group pairing and year: Parameter estimates with their unconditional standard errors (SE), 95% confidence intervals and relative importance after model averaging (models with ΔAICc < 7). Estimates where the 95% confidence intervals do not overlap zero are in bold.

Parameter	Estimate	Unconditional SE	95% Confidence interval		Relative importance^
(Intercept)	0.124	0.004	**0.121**	**0.134**	
Allele sharing	**-0.011**	**0.002**	**-0.015**	**-0.009**	**1.00**
Relatedness	**-0.010**	**0.004**	**-0.017**	**-0.002**	**0.95**
Within-group pairing [a]	-0.003	0.002	-0.007	0.000	0.65
Year 2004 [b]	0.004	0.003	-0.002	0.010	0.83
Year 2005 [b]	0.003	0.003	-0.003	0.009	"
Year 2008 [b]	0.004	0.003	-0.001	0.011	"
Year 2009 [b]	0.001	0.003	-0.006	0.006	"
Year 2010 [b]	**0.007**	**0.003**	**0.001**	**0.013**	**"**

^ Relative importance is the sum of Akaike weights for all models (ΔAICc < 7) containing the parameter

[a] neighbouring-group pairing was the reference category

[b] Year 1993 was the reference category

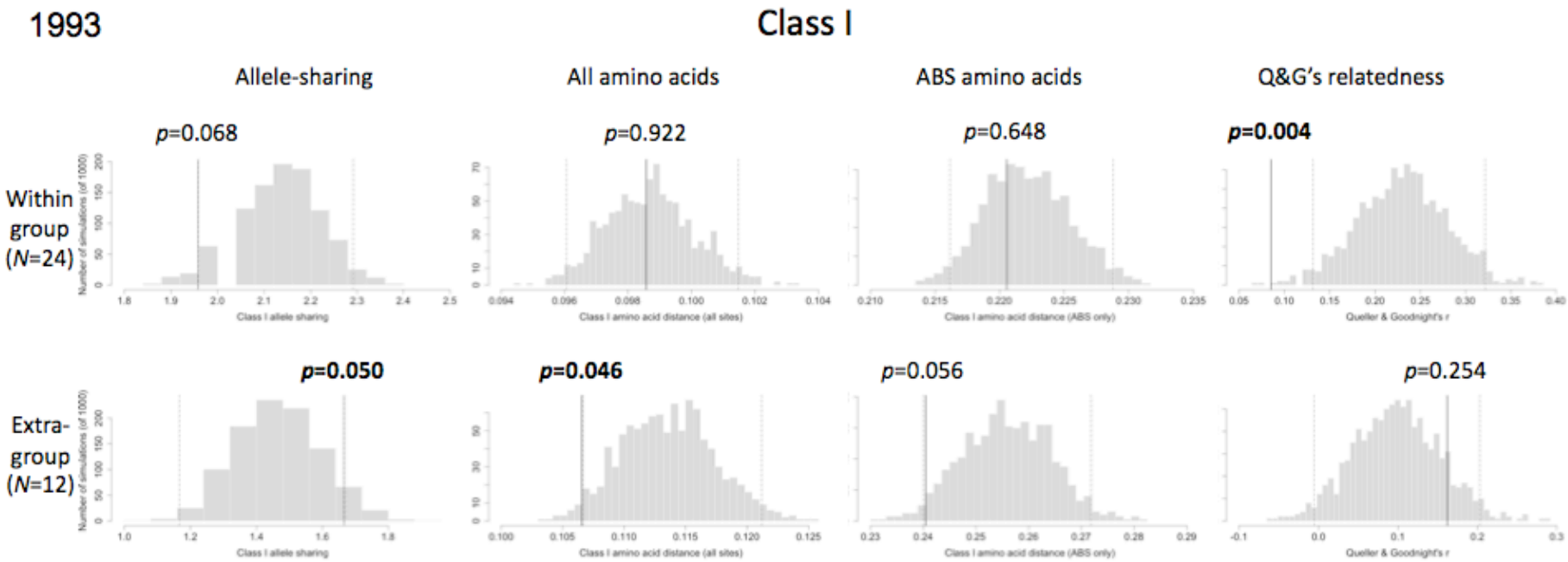


Fig. 5.2a

Fig. 5.2 Mean MHC class I allele sharing, mean amino acid distance (all amino acid sites and ABS only) and mean Queller & Goodnight's index of pairwise relatedness (*R*) of assigned parent-pairs (solid line) compared with the frequency distribution of mean values generated from 1000 simulations of random pairings between breeding females and all of their within-group or neighbouring-group candidate fathers (grey bars), with two-tailed 95% CI (dashed lines) indicating cut-offs for significant departures from random mating. The data were analyzed separately for each year: (a) 1993, number of candidate fathers (N_C)=84, (b) 2004, N_C=102, (c) 2005, N_C=101, (d) 2008, N_C=136, (e) 2009, N_C=110, and (f) 2010, N_C=130. *N* indicates number of assigned parent pairs. Significant p-values are shown in bold.

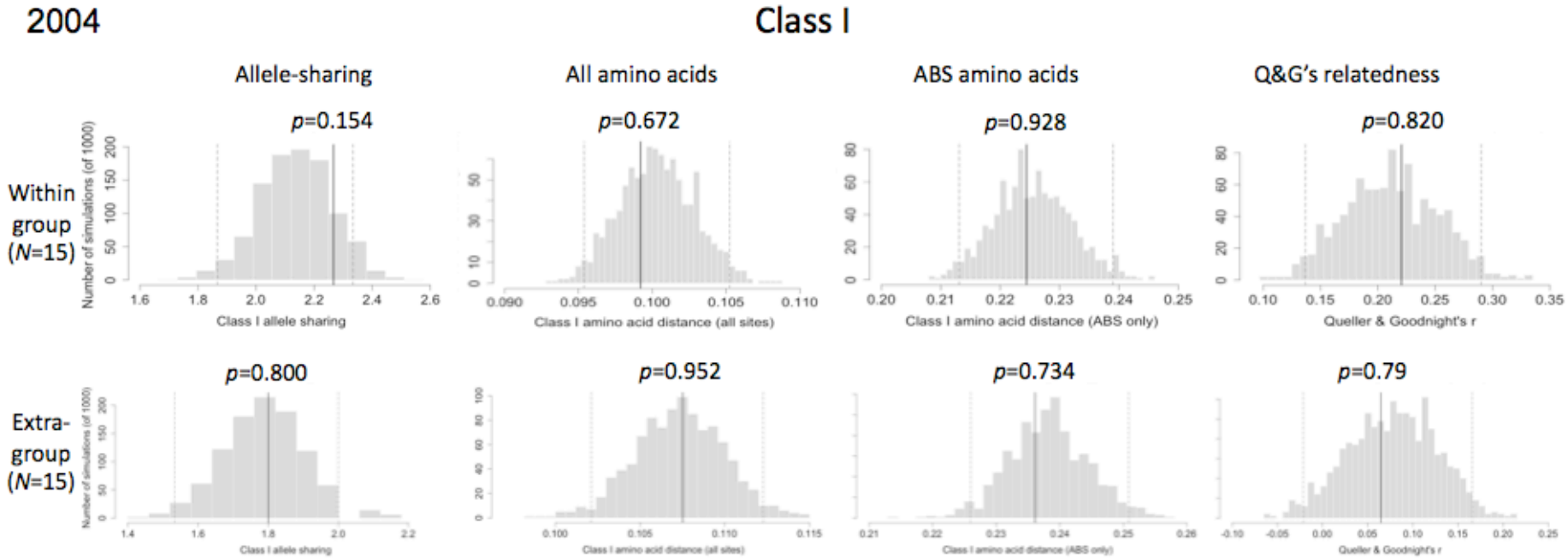


Fig. 5.2b

Fig. 5.2 Mean MHC class I allele sharing, mean amino acid distance (all amino acid sites and ABS only) and mean Queller & Goodnight's index of pairwise relatedness (*R*) of assigned parent-pairs (solid line) compared with the frequency distribution of mean values generated from 1000 simulations of random pairings between breeding females and all of their within-group or neighbouring-group candidate fathers (grey bars), with two-tailed 95% CI (dashed lines) indicating cut-offs for significant departures from random mating. The data were analyzed separately for each year: (a) 1993, number of candidate fathers (N_C)=84, (b) 2004, N_C=102, (c) 2005, N_C=101, (d) 2008, N_C=136, (e) 2009, N_C=110, and (f) 2010, N_C=130. *N* indicates number of assigned parent pairs. Significant p-values are shown in bold.

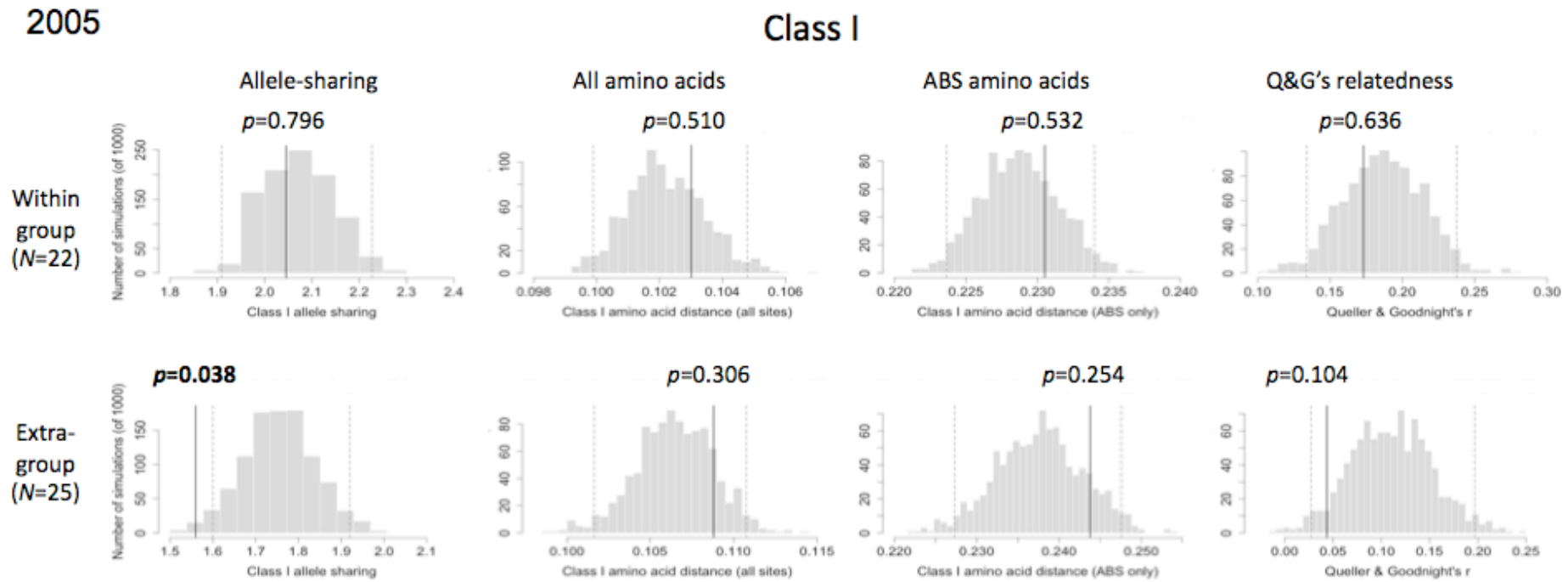


Fig. 5.2c

Fig. 5.2 Mean MHC class I allele sharing, mean amino acid distance (all amino acid sites and ABS only) and mean Queller & Goodnight's index of pairwise relatedness (*R*) of assigned parent-pairs (solid line) compared with the frequency distribution of mean values generated from 1000 simulations of random pairings between breeding females and all of their within-group or neighbouring-group candidate fathers (grey bars), with two-tailed 95% CI (dashed lines) indicating cut-offs for significant departures from random mating. The data were analyzed separately for each year: (a) 1993, number of candidate fathers (N_C)=84, (b) 2004, N_C=102, (c) 2005, N_C=101, (d) 2008, N_C=136, (e) 2009, N_C=110, and (f) 2010, N_C=130. *N* indicates number of assigned parent pairs. Significant p-values are shown in bold.

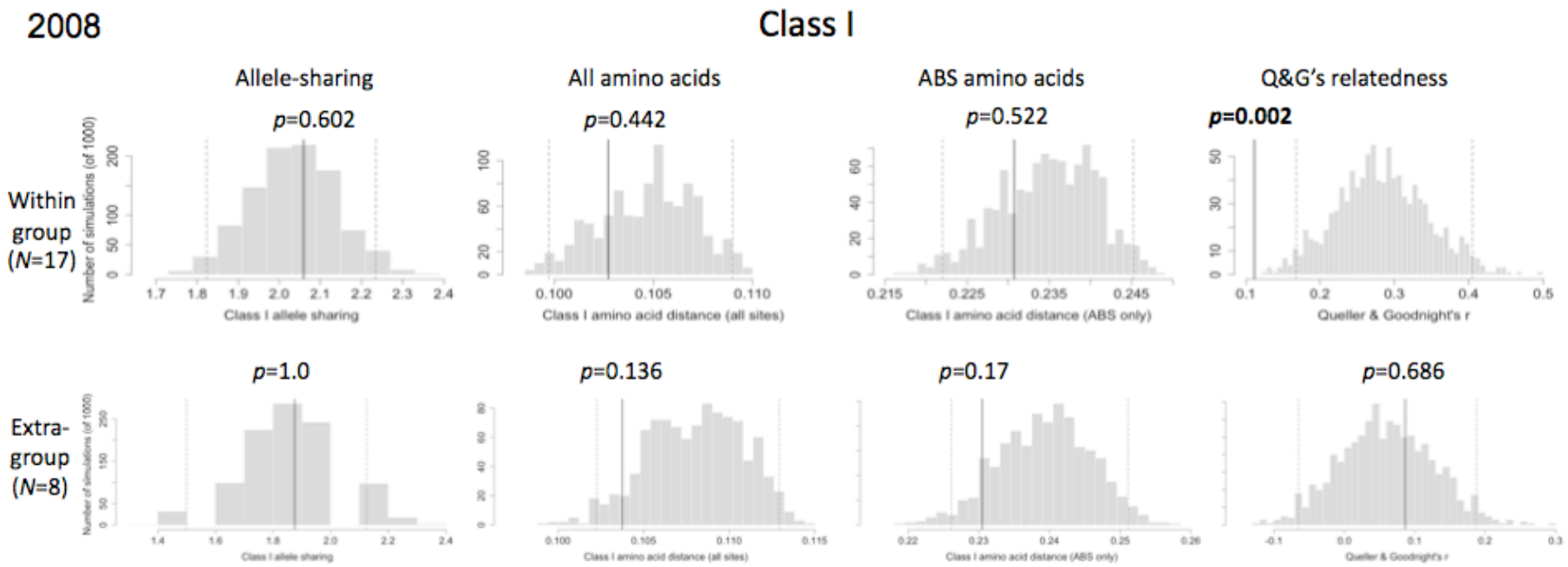


Fig. 5.2d

Fig. 5.2 Mean MHC class I allele sharing, mean amino acid distance (all amino acid sites and ABS only) and mean Queller & Goodnight's index of pairwise relatedness (*R*) of assigned parent-pairs (solid line) compared with the frequency distribution of mean values generated from 1000 simulations of random pairings between breeding females and all of their within-group or neighbouring-group candidate fathers (grey bars), with two-tailed 95% CI (dashed lines) indicating cut-offs for significant departures from random mating. The data were analyzed separately for each year: (a) 1993, number of candidate fathers (N_C)=84, (b) 2004, N_C=102, (c) 2005, N_C=101, (d) 2008, N_C=136, (e) 2009, N_C=110, and (f) 2010, N_C=130. *N* indicates number of assigned parent pairs. Significant p-values are shown in bold.

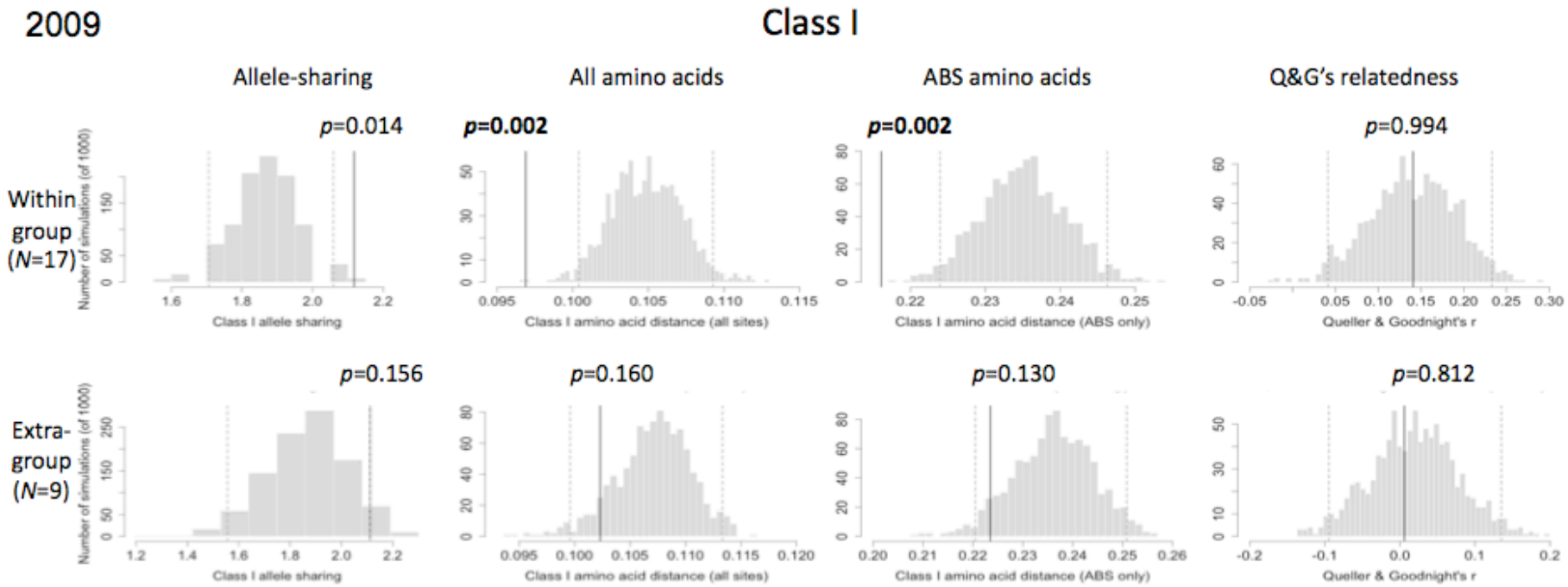


Fig. 5.2e

Fig. 5.2 Mean MHC class I allele sharing, mean amino acid distance (all amino acid sites and ABS only) and mean Queller & Goodnight's index of pairwise relatedness (*R*) of assigned parent-pairs (solid line) compared with the frequency distribution of mean values generated from 1000 simulations of random pairings between breeding females and all of their within-group or neighbouring-group candidate fathers (grey bars), with two-tailed 95% CI (dashed lines) indicating cut-offs for significant departures from random mating. The data were analyzed separately for each year: (a) 1993, number of candidate fathers (N_C)=84, (b) 2004, N_C=102, (c) 2005, N_C=101, (d) 2008, N_C=136, (e) 2009, N_C=110, and (f) 2010, N_C=130. *N* indicates number of assigned parent pairs. Significant p-values are shown in bold.

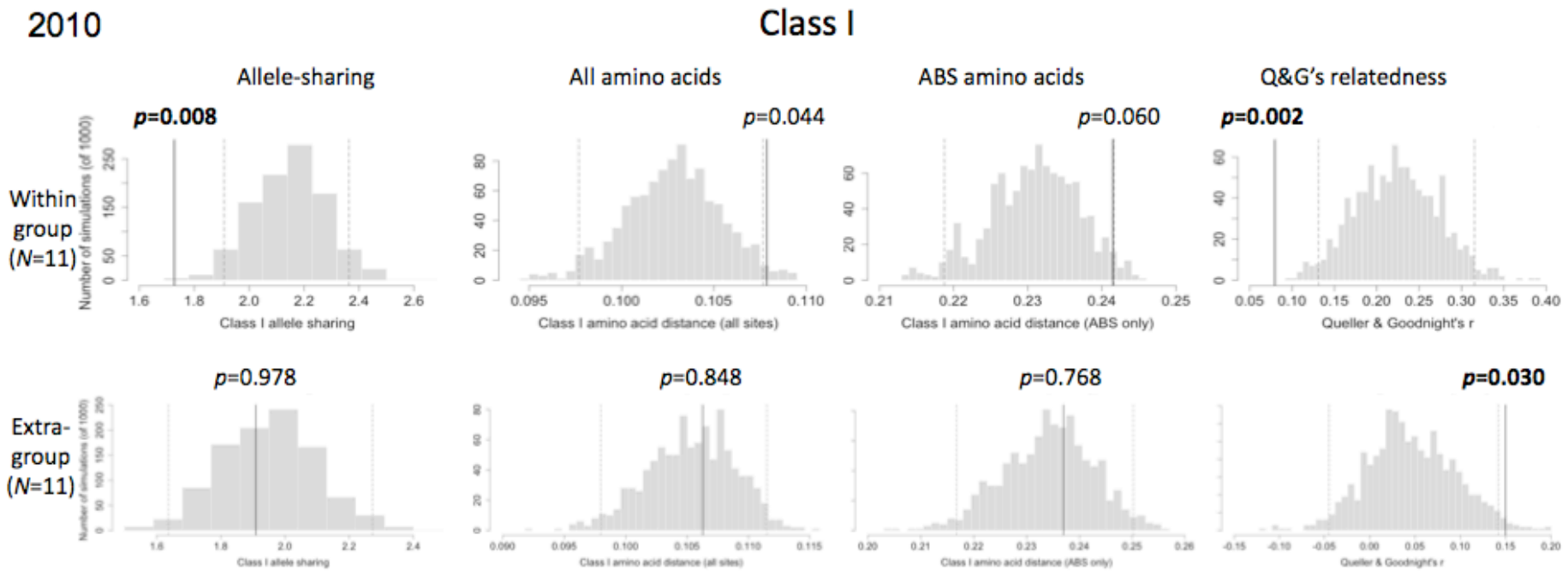


Fig. 5.2f

Fig. 5.2 Mean MHC class I allele sharing, mean amino acid distance (all amino acid sites and ABS only) and mean Queller & Goodnight's index of pairwise relatedness (*R*) of assigned parent-pairs (solid line) compared with the frequency distribution of mean values generated from 1000 simulations of random pairings between breeding females and all of their within-group or neighbouring-group candidate fathers (grey bars), with two-tailed 95% CI (dashed lines) indicating cut-offs for significant departures from random mating. The data were analyzed separately for each year: (a) 1993, number of candidate fathers (N_C)=84, (b) 2004, N_C=102, (c) 2005, N_C=101, (d) 2008, N_C=136, (e) 2009, N_C=110, and (f) 2010, N_C=130. *N* indicates number of assigned parent pairs. Significant p-values are shown in bold.

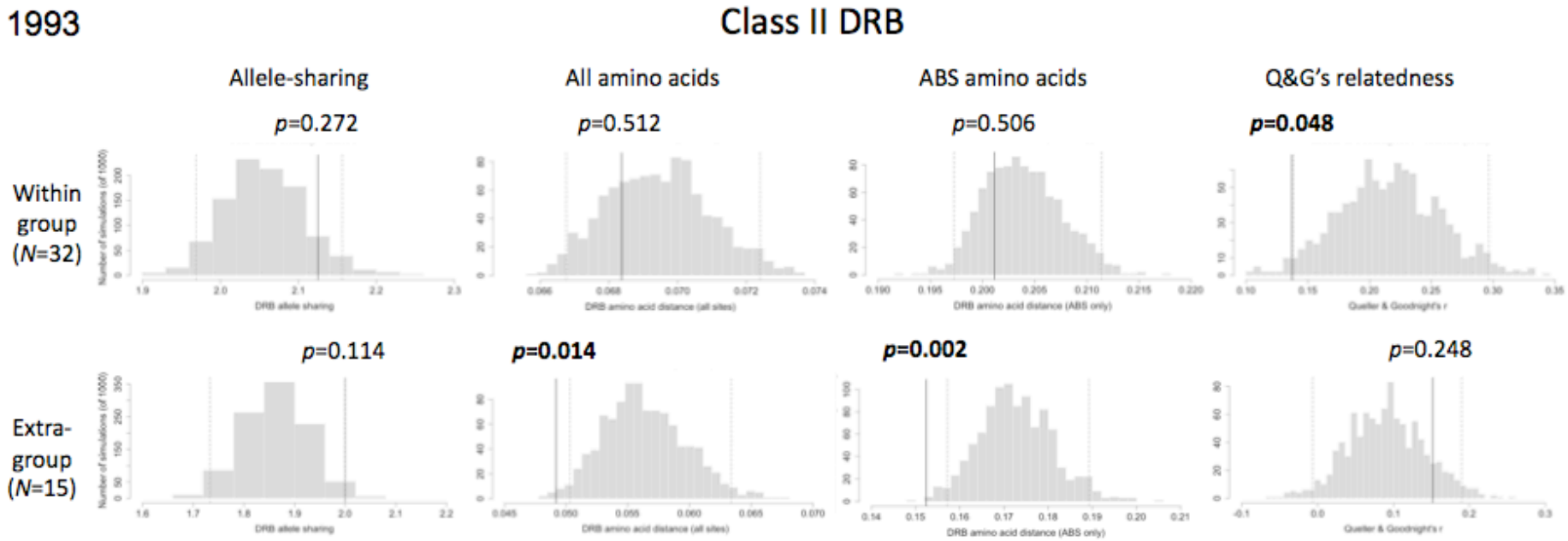


Fig. 5.3a

Fig. 5.3 Mean MHC class II DRB allele sharing, mean amino acid distance (all amino acid sites and ABS only) and mean Queller & Goodnight's index of pairwise relatedness (*R*) of assigned parent-pairs (solid line) compared with the frequency distribution of mean values generated from 1000 simulations of random pairings between breeding females and all of their within-group or neighbouring-group candidate fathers (grey bars), with two-tailed 95% CI (dashed lines) indicating cut-offs for significant departures from random mating. The data were analyzed separately for each year: (a) 1993, number of candidate fathers (N_C)=87, (b) 2004, N_C=101, (c) 2005, N_C=100, (d) 2008, N_C=136, (e) 2009, N_C=109, and (f) 2010, N_C=128. *N* indicates number of assigned parent pairs. Significant p-values are shown in bold.

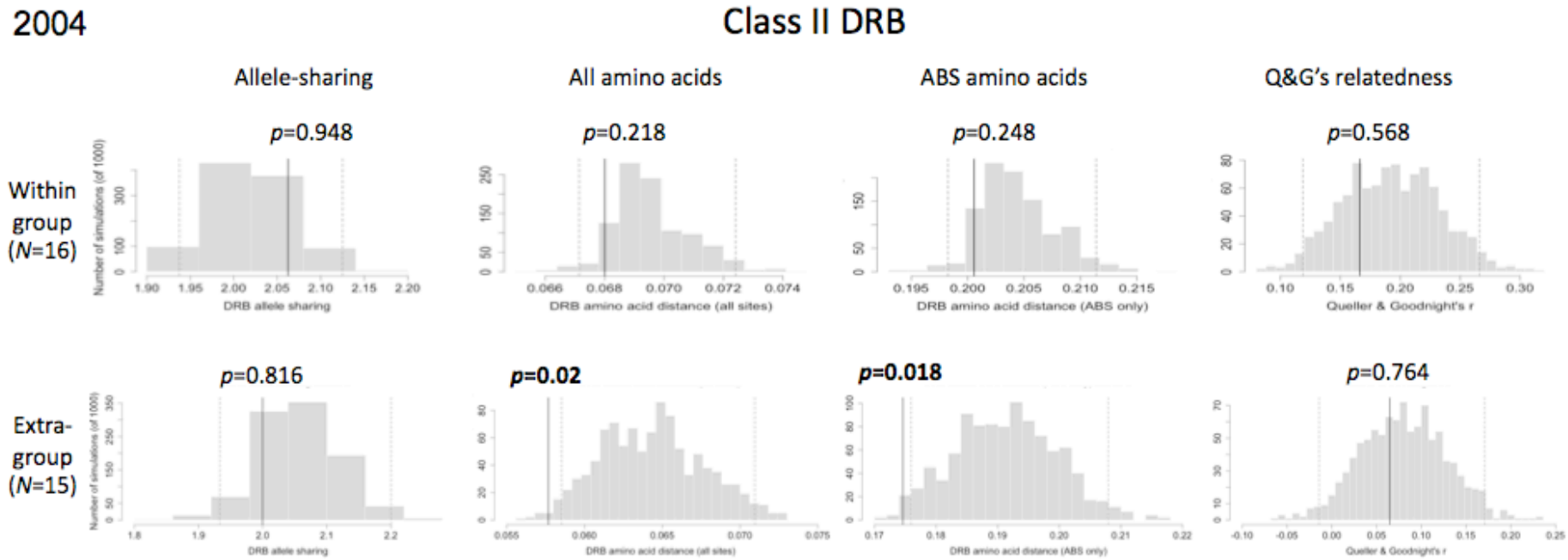


Fig. 5.3b

Fig. 5.3 Mean MHC class II DRB allele sharing, mean amino acid distance (all amino acid sites and ABS only) and mean Queller & Goodnight's index of pairwise relatedness (*R*) of assigned parent-pairs (solid line) compared with the frequency distribution of mean values generated from 1000 simulations of random pairings between breeding females and all of their within-group or neighbouring-group candidate fathers (grey bars), with two-tailed 95% CI (dashed lines) indicating cut-offs for significant departures from random mating. The data were analyzed separately for each year: (a) 1993, number of candidate fathers (N_C)=87, (b) 2004, N_C=101, (c) 2005, N_C=100, (d) 2008, N_C=136, (e) 2009, N_C=109, and (f) 2010, N_C=128. *N* indicates number of assigned parent pairs. Significant p-values are shown in bold.

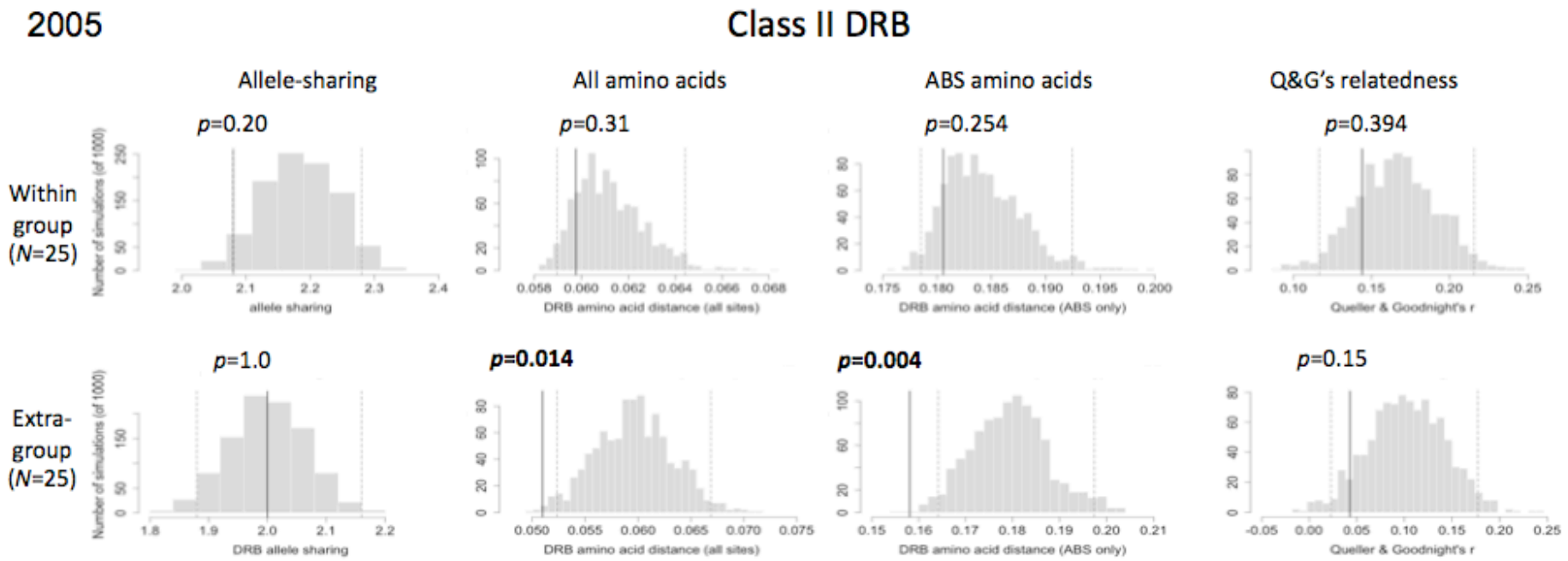


Fig. 5.3c

Fig. 5.3 Mean MHC class II DRB allele sharing, mean amino acid distance (all amino acid sites and ABS only) and mean Queller & Goodnight's index of pairwise relatedness (*R*) of assigned parent-pairs (solid line) compared with the frequency distribution of mean values generated from 1000 simulations of random pairings between breeding females and all of their within-group or neighbouring-group candidate fathers (grey bars), with two-tailed 95% CI (dashed lines) indicating cut-offs for significant departures from random mating. The data were analyzed separately for each year: (a) 1993, number of candidate fathers (N_C)=87, (b) 2004, N_C=101, (c) 2005, N_C=100, (d) 2008, N_C=136, (e) 2009, N_C=109, and (f) 2010, N_C=128. *N* indicates number of assigned parent pairs. Significant p-values are shown in bold.

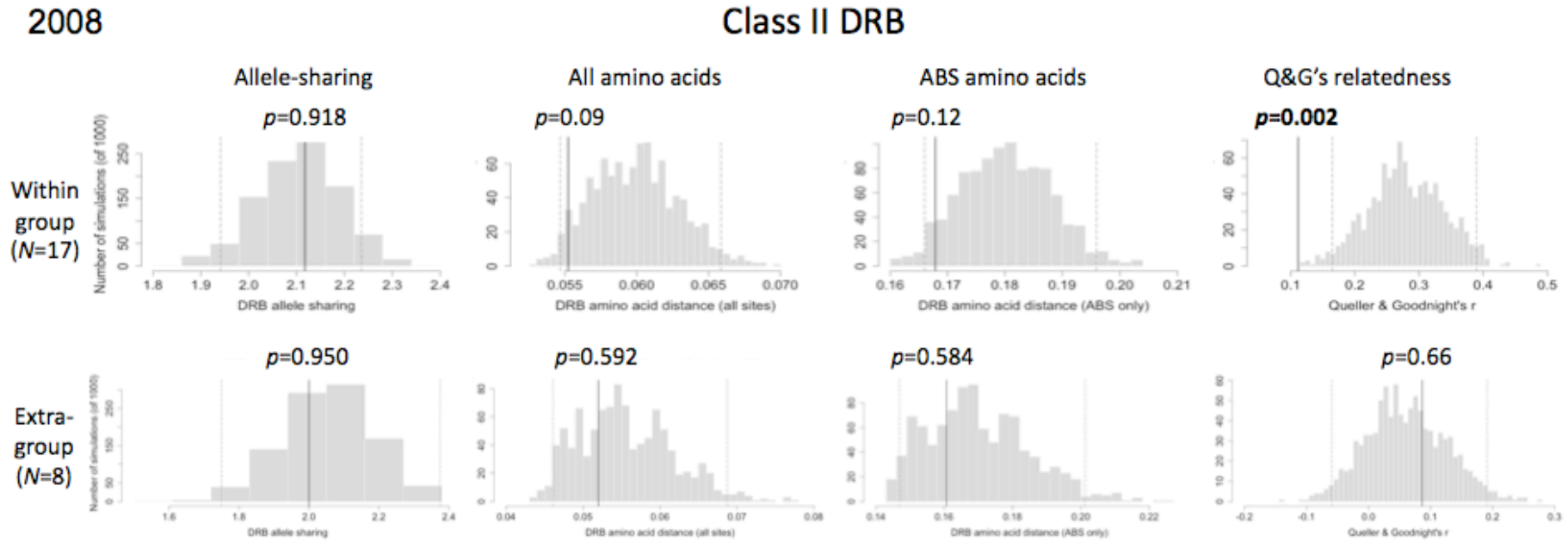


Fig. 5.3d

Fig. 5.3 Mean MHC class II DRB allele sharing, mean amino acid distance (all amino acid sites and ABS only) and mean Queller & Goodnight's index of pairwise relatedness (*R*) of assigned parent-pairs (solid line) compared with the frequency distribution of mean values generated from 1000 simulations of random pairings between breeding females and all of their within-group or neighbouring-group candidate fathers (grey bars), with two-tailed 95% CI (dashed lines) indicating cut-offs for significant departures from random mating. The data were analyzed separately for each year: (a) 1993, number of candidate fathers (N_C)=87, (b) 2004, N_C=101, (c) 2005, N_C=100, (d) 2008, N_C=136, (e) 2009, N_C=109, and (f) 2010, N_C=128. *N* indicates number of assigned parent pairs. Significant p-values are shown in bold.

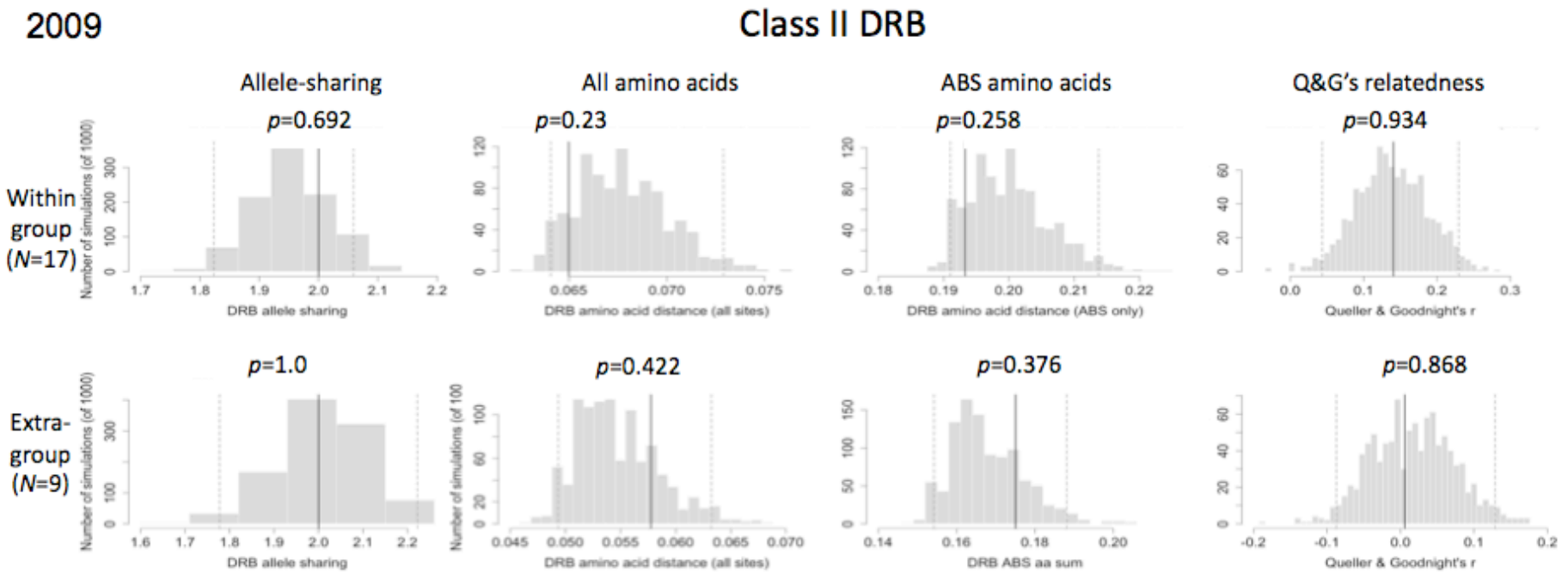


Fig. 5.3e

Fig. 5.3 Mean MHC class II DRB allele sharing, mean amino acid distance (all amino acid sites and ABS only) and mean Queller & Goodnight's index of pairwise relatedness (*R*) of assigned parent-pairs (solid line) compared with the frequency distribution of mean values generated from 1000 simulations of random pairings between breeding females and all of their within-group or neighbouring-group candidate fathers (grey bars), with two-tailed 95% CI (dashed lines) indicating cut-offs for significant departures from random mating. The data were analyzed separately for each year: (a) 1993, number of candidate fathers (N_C)=87, (b) 2004, N_C=101, (c) 2005, N_C=100, (d) 2008, N_C=136, (e) 2009, N_C=109, and (f) 2010, N_C=128. *N* indicates number of assigned parent pairs. Significant p-values are shown in bold.

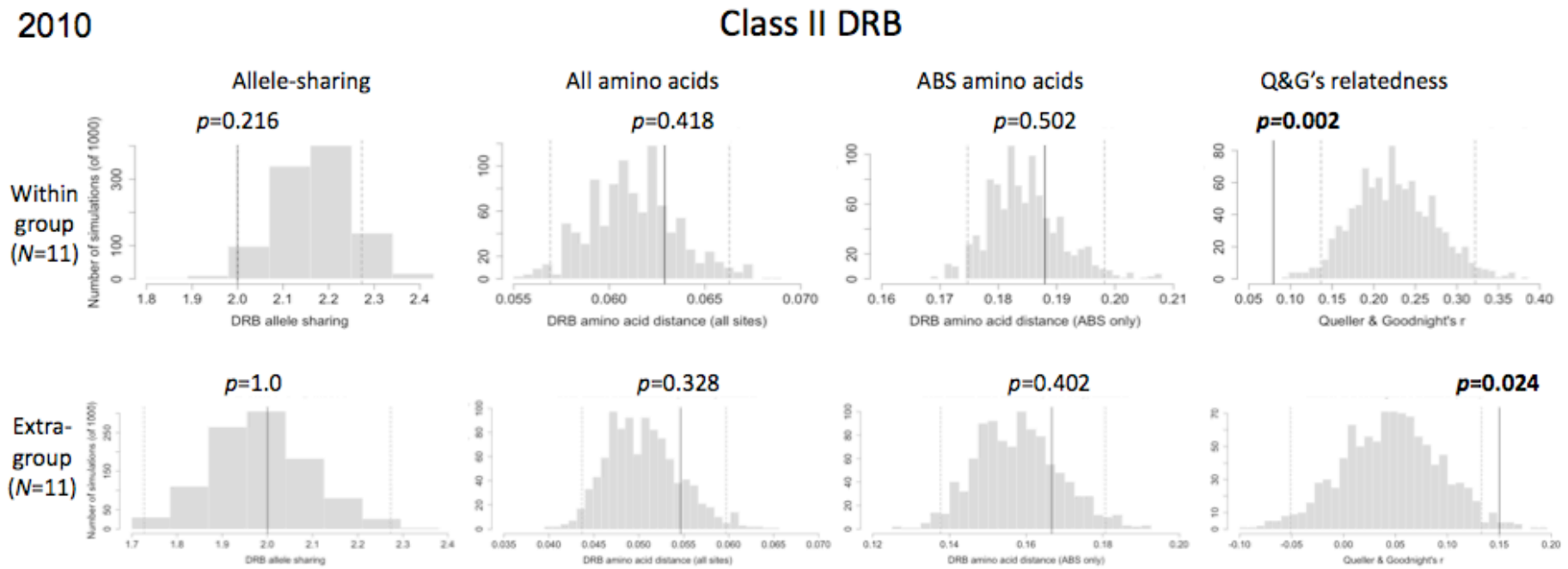


Fig. 5.3f

Fig. 5.3 Mean MHC class II DRB allele sharing, mean amino acid distance (all amino acid sites and ABS only) and mean Queller & Goodnight's index of pairwise relatedness (*R*) of assigned parent-pairs (solid line) compared with the frequency distribution of mean values generated from 1000 simulations of random pairings between breeding females and all of their within-group or neighbouring-group candidate fathers (grey bars), with two-tailed 95% CI (dashed lines) indicating cut-offs for significant departures from random mating. The data were analyzed separately for each year: (a) 1993, number of candidate fathers (N_C)=87, (b) 2004, N_C=101, (c) 2005, N_C=100, (d) 2008, N_C=136, (e) 2009, N_C=109, and (f) 2010, N_C=128. *N* indicates number of assigned parent pairs. Significant p-values are shown in bold.

5.5 Discussion

Our study provides evidence for a relationship between MHC similarity and mate choice. In general, neighbouring-group parent pairs were more similar in their MHC class II DRB genes than randomized neighbouring-group parent pairs. These MHC-assortative patterns found for class II DRB genes are consistent with mate choice for optimal levels of MHC variation and maintenance of co-adapted gene complexes (Kaufman, 1999; Hendry *et al.*, 2000; Neff, 2004). The results based on allele sharing did not show significant difference such as those based on amino acid distance. This is not unexpected as amino acid distance incorporates functional difference between allele. The divergence between these parameters is also evident in other studies, which have shown the amino acid distance to be more informative than allele sharing in identifying MHC-based mate choice (Landry *et al.*, 2001; Miller *et al.*, 2009). There is also a growing number of studies that group MHC alleles into supertypes, based on similar peptide specificities of the ABS (Trachtenberg *et al.*, 2003; Huchard *et al.*, 2010), which is useful when a large number of MHC alleles present in the population (Sepil *et al.*, 2013). However, only a small number of MHC alleles present in this badger population, thus amino acid distance is enough to incorporate functional differences between alleles. The small DRB amino acid distance of assigned neighbouring-group parent pairs was not due to inbreeding, which was only found in assigned neighbouring-group pairs in 2010 and no MHC-assortative mating was found in that year, and there was no correlation of DRB amino acid distance and relatedness.

Female badgers prefer neighbouring-group males with similar MHC class II DRB genes, This is different to MHC-disassortative mating found in other species (Landry *et al.*, 2001; Gillingham *et al.*, 2009; Miller *et al.*, 2009; Juola & Dearborn, 2012), as badgers do not maximize MHC differences between mates. Only a handful of other studies have found MHC-assortative mating pattern, i.e. preference for MHC-similar individuals, which was reported in mice (Yamazaki *et al.*, 1978), humans (Roberts *et al.*, 2005), Malagasy giant jumping rat *Hypogeomys antimena* (Sommer, 2005a), house sparrows *Passer domesticus* (Bonneaud *et al.*, 2006a), tiger salamanders *Ambystoma tigrinum* (Bos *et al.*, 2009) and common yellowthroat *Geothlypis trichas* (Bollmer *et al.*, 2012). Possible explanations for these results are that females are choosing mates to avoid disadvantages arise from having highly divergent MHC genes. Potential disadvantages include increased loss of T-cells (Lawlor *et al.*, 1990;

Nowak *et al.*, 1992), increased risk of autoimmune diseases, and disruption of local adaptations or co-adapted gene complexes (Kaufman, 1999; Hendry *et al.*, 2000; Neff, 2004), and thus having offspring of lower fitness. Furthermore, males with MHC alleles that are more similar to a female over males exhibiting more divergent MHC alleles may have reproductive advantage (Bos *et al.*, 2009).

MHC-disassortative mating has been proposed as a mechanism for inbreeding avoidance (Potts *et al.*, 1994). Analyses showed negative correlation between MHC class I genes amino acid distance and relatedness at microsatellite loci, indicating possible correlation between inbreeding avoidance and MHC class I genes-based mating, with class I genes-based cues potentially used as a reference for inbreeding avoidance. Inbreeding avoidance was only identified in within-group parent pairs but not neighbouring-group parent pairs. Possible explanation is that inbreeding avoidance happened only if most candidate fathers were close relatives (i.e. with high relatedness), which was the case in within-group pairing (Fig. 5.2 & 5.3; Dugdale *et al.*, 2008). Whereas the candidate neighbouring-group fathers were less related (i.e. with lower relatedness), no inbreeding avoidance was found for assigned neighbouring-group parent pairs. Furthermore, MHC-assortative mating based on DRB genes was only identified in neighbouring-group pairs but not within-group pairs, which indicates the effect of DRB gene-assortative mating may be overpowered by inbreeding avoidance when candidate fathers are highly related. Alternative explanations are that inbreeding depression happened and females were able to seek MHC class II DRB genes-similar neighbouring-group males more actively than within-group males.

There were considerable annual fluctuations in the occurrence of MHC-based mate choice based on class II DRB genes. This marked fluctuation in the level of MHC similarity between assigned parent pairs suggests that females may actively choose males on the basis on MHC genotypes in some years. We cannot provide a definitive answer to which conditions favoured MHC-based mating in those years. Fluctuation in population relatedness structure is proposed to be affected by ecological constraints such as predation and disease risk (Jaatinen *et al.*, 2012). We speculate that MHC-based mate choice may only pay for females when ecological constraints such as disease risk are severe

enough, and the specific conditions needed for MHC-based mate choice to be favoured may vary over time.

Our study identified MHC class II DRB genes-assortative mate choice and also possible correlation between MHC class I genes and inbreeding avoidance. An underlying assumption to mate choice based on MHC genes is that MHC genotypes can be recognised as cues by choosy partners. Studies on MHC-based mate choice indicate that olfactory cues are generally used to discriminate among MHC genotypes (Yamazaki *et al.*, 1982; Reusch *et al.*, 2001; Brennan & Kendrick, 2006), as the MHC molecules and/or peptides derived from them can contribute to individual olfactory profiles. For example, MHC genes are involved in the generation and recognition of odour-types in mice (Leinders-Zufall *et al.*, 2004; but see Hurst *et al.*, 2005; Ziegler *et al.*, 2005). Badgers potentially exchange information about individual specific parameters through subcaudal gland scent (Buesching *et al.*, 2002c). This secretion may include information on the MHC genotype of the badger that deposited it. Another possibility is MHC genes may influence microbiota (e.g. inside the subcaudal pouch; Sin *et al.*, 2012a) that generate secondary metabolites and contribute to the actual odour of the host (Singh *et al.*, 1990). The odour of badgers is thus likely to play an important role if pre-copulatory MHC-based mate choice happens in this species. An alternative, but not mutually exclusive, explanation is that MHC-based mate choice operates after copulation, through cryptic female choice such as selective fertilization and selective abortion (Wedekind *et al.*, 1996; Ziegler *et al.*, 2002; Ziegler *et al.*, 2005). MHC-based post-copulatory choice occurs in mammals such as grey mouse lemurs (*Microcebus murinus*; Schwensow *et al.*, 2008a). Putatively superfoetation and embryo reabsorption in badgers (Canivenc & Bonnin, 1979; Yamaguchi *et al.*, 2006) is likely to facilitate multiple paternity (Dugdale *et al.*, 2007) and post-copulatory sexual selection.

In addition to pre- and post-copulatory mate choice, the MHC-dependent pattern identified could be caused by alternative mechanisms, such as male competition, sperm competition and cub mortality before emergence from below the ground. However, male competition in badgers is unlikely, as male mounting frequency was not related to dominance rank or body condition index, and male mounting frequency did not correlate with paternity success (Dugdale *et al.*, 2011b). The promiscuous and

repeated mounting of female badgers (Carpenter *et al.*, 2005; Dugdale *et al.*, 2007; Dugdale *et al.*, 2011b) may promote sperm competition, or MHC-dependent fertilization according to genetic compatibility or diversity (Skarstein *et al.*, 2005). As cubs could only be trapped after they have emerged from below the ground at around 12-week old (Macdonald & Newman, 2002), the possibility that the MHC-dependent pattern identified was due to cub mortality before emergence could not be ruled out.

There is growing evidence indicating a trend of mate preference towards minimal or intermediate MHC divergence (Bonneaud *et al.*, 2006a; Milinski, 2006; Forsberg *et al.*, 2007; Bos *et al.*, 2009; Kalbe *et al.*, 2009; Woelfing *et al.*, 2009; Bollmer *et al.*, 2012) in addition to maximal divergence (Landry *et al.*, 2001; Miller *et al.*, 2009; Juola & Dearborn, 2012) − i.e. optimal divergence. There is also growing evidence showing a complex and context-dependent pattern of MHC-based reproduction (Richardson *et al.*, 2005; Bos *et al.*, 2009) rather than simple MHC-disassortative mating. Mate choice based on the MHC but not on genome-wide background has been reported for a number of species (e.g. Landry *et al.*, 2001; Juola & Dearborn, 2012), but there is a lack of study in wild mammals currently. This is the first published study of MHC-based mate choice in the order Carnivora, of evidence for MHC-assortative mating, which is contrast to other studies that found a preference for MHC dissimilar mates (e.g. Landry *et al.*, 2001; Neff *et al.*, 2008; Radwan *et al.*, 2008; Schwensow *et al.*, 2008b; Gillingham *et al.*, 2009; Miller *et al.*, 2009; Juola & Dearborn, 2012) or no preference (e.g. Paterson & Pemberton, 1997; Westerdahl, 2004). Our study thus adds to the support for the evidence-lacking MHC-assortative mating and the changing view of MHC-based mate choice.

Chapter 6

Pathogen burden, co-infection and major histocompatibility complex variability in the European badger (*Meles meles*)

6.1 Abstract

Pathogen-mediated selection is believed to maintain the extreme diversity in the major histocompatibility complex (MHC) genes through a heterozygous advantage, rare-allele advantage and fluctuating selection. Nevertheless, the relative importance of these non-mutually exclusive mechanisms has been controversial. Here, we investigate whether badger MHC diversity is driven by pathogen-mediated selection. Heterozygote advantages, of recognizing and binding a wider range of antigens than homozygotes, may only be realized when multiple pathogens are considered simultaneously. Therefore, we investigated the co-infection status of 13 pathogens from a range of taxa, and examined their relationships with MHC diversity in a wild population of European badgers (*Meles meles*). This population has a relatively low number of MHC alleles and thus provides great potential for studying MHC-pathogen relationships. We examined pathogen prevalence (infected or not) and infection intensity to determine qualitative and quantitative resistance/susceptibility conferred by MHC alleles. We found associations between specific MHC alleles and both Mustelid herpes virus and *Yersinia* infection. MHC heterozygosity was not correlated with co-infection status, but there was weak evidence of a heterozygote advantage for individual pathogen infections (trypanosomes and *Campylobacter*), which may be due to the presence of specific alleles. Our results thus indicate that rare-allele advantages and/or fluctuating selection are likely to be the major selective force shaping MHC diversity in this species, considering the pathogens investigated. We conclude that examining both pathogen prevalence and infection intensity, and co-infection of a wide range of pathogens is crucial to enable a more complete picture of the mechanisms driving MHC diversity.

6.2 Introduction

The MHC is a diverse gene family that plays a crucial role in the adaptive immune system (Hedrick, 1994). MHC genes encode cell surface glycoproteins that are vital in both humoral and cell-mediated immune responses, as MHC molecules bind and present antigens to T cells and trigger an immune cascade (Swain, 1983). Because of this critical role in the immune system, MHC genes are under constant selective pressures due to challenges from parasites and pathogens (hereafter, both are referred to as "pathogens") (Jeffrey & Bangham, 2000; Piertney & Oliver, 2006). It is generally believed that balancing and diversifying selection, through an arms race between pathogens and hosts, maintains the extreme diversity in MHC genes. This pathogen-mediated selection is proposed to operate through a heterozygote advantage, rare-allele advantage and fluctuating selection (Spurgin & Richardson, 2010).

The rare-allele advantage hypothesis proposes that rare alleles within the population are likely to offer greater protection to pathogens than common alleles, and so have a selective advantage (Takahata & Nei, 1990; Slade & McCallum, 1992). Through a rare-allele advantage, specific pathogens drive a cyclic change in the frequency of MHC alleles, known as negative frequency-dependent selection. The heterozygote advantage hypothesis (Doherty & Zinkernagel, 1975; Penn *et al.*, 2002) assumes that MHC heterozygous individuals are able to recognize and bind more antigens than homozygous individuals, and therefore trigger immune responses more effectively against both individual pathogen (Carrington *et al.*, 1999; Penn *et al.*, 2002; Worley *et al.*, 2010) and co-infection (Hughes & Nei, 1992; McClelland *et al.*, 2003). We use "heterozygote advantage" in the broad sense, where on average heterozygotes exhibit higher fitness than the average fitness of homozygotes due to dominance, or heterozygotes have a fitness advantage over the fittest homozygotes due to over-dominance (Hughes & Nei, 1988; Penn *et al.*, 2002). The fluctuating selection hypothesis proposes that spatial and temporal fluctuation in pathogen types, and their abundances, may maintain MHC diversity (Hedrick, 2002; Spurgin & Richardson, 2010). Pathogen fluctuations may alter the intensity of directional selection on MHC genes, hence different MHC alleles may be selected for at different times or spatial points.

The majority of studies of MHC-pathogen relationships have reported associations between specific resistant/susceptible MHC alleles and pathogens, e.g., in fish (Langefors *et al.*, 2001), mammals (Harf & Sommer, 2005; Meyer-lucht & Sommer, 2005) and birds (Bonneaud *et al.*, 2006b; Loiseau *et al.*, 2011), while a heterozygote advantage has limited support (but see Thursz *et al.*, 1997; Penn *et al.*, 2002; Froeschke & Sommer, 2005). The association between specific alleles and pathogens could be due to frequency-dependent selection (Takahata & Nei, 1990); however, there may also be a frequency-dependent component to a heterozygote advantage (Apanius *et al.*, 1997). Given that rare alleles occur mostly in heterozygous genotypes (Apanius *et al.*, 1997), both rare-allele advantages and heterozygote advantages favour rare alleles and also maintain large numbers of alleles, making it difficult to distinguish between them. The advantage of heterozygotes being able to recognize and bind a wider range of antigens may be beneficial in the occurrence of both single pathogen infections (Carrington *et al.*, 1999; Worley *et al.*, 2010) and co-infection with multiple pathogens (Hughes & Nei, 1992; McClelland *et al.*, 2003). MHC heterozygotes may have an advantage of being co-infected by fewer pathogens than homozygotes, rather than an advantage due specifically to rare alleles responding to particular pathogens (McClelland *et al.*, 2003). The limited evidence reported for a heterozygote advantage may therefore be due to studies that only focused on single or a few pathogen taxa (Paterson *et al.*, 1998; Harf & Sommer, 2005; Meyer-lucht & Sommer, 2005; Schad *et al.*, 2005), limiting examination of the overall selection pressure to which animals are exposed. The relationship between the MHC and pathogens can thus only be tested properly by examining multiple pathogens.

The interaction between the MHC and pathogens may be context-dependent and only emerge under particular environmental conditions (Bernatchez & Landry, 2003; Sommer, 2005b; Piertney & Oliver, 2006). To understand MHC evolution, it is therefore important to examine the association between MHC and pathogens in natural populations. Here, we examine the associations between MHC class I and class II gene variation and multiple pathogens in a population of wild European badgers (*Meles meles*). Because of the analytical complications that arise from high allelic diversity (Richardson *et al.*, 2005), a population with a relatively low number of MHC alleles offers great potential for identifying the ecological consequences underlying MHC variation (Oliver *et al.*, 2009). In this badger population, only one DRB locus was polymorphic with three putatively functional sequences (Sin *et*

al., 2012b; Chapter 5), while seven putatively functional MHC class I sequences were also identified (Sin *et al.*, 2012c). This system, with relatively low MHC diversity, thus affords an ideal opportunity to study MHC-pathogen relationships.

This high-density badger population form social groups (Newman *et al.*, 2011) that use communal underground setts where conditions are likely to facilitate pathogen transmission. European badgers are infected by a wide taxonomic range of pathogens, e.g., viral [mustelid herpes virus (MHV); (King *et al.*, 2004)], bacterial [*Salmonella*; (Wilson *et al.*, 2003)], protozoan [*Trypanosoma pestanai* (Macdonald *et al.*, 1999); *Eimeria melis* and *Isospora melis* (Newman *et al.*, 2001)], helminth (*Capellaria*, *Strongyloides*, other nematodes), and invertebrate ectoparasites [badger fleas (*Paraceras melis*), badger lice (*Trichodectes melis*) and ticks (*Ixodes hexagonus*); (San, 2007)]. Different pathogens exert different extend of disadvantage to the host fitness. Some pathogens, such as *E. melis* (Newman *et al.*, 2001), cause host morbidity or mortality. Other pathogens, causing diagnostically only mild symptoms, may decrease host fitness by using up the host's resource and/or rendering the host to spend resource on resistance or tolerance to them (Råberg *et al.*, 2007), and thus potentially compromising to other mortality factors (e.g. oxidative stress; vandeCrommenacker *et al.*, 2012; Bilham *et al.*, 2013). Studying such a broad range of pathogens in a wild animal population facilitates a more comprehensive understanding of the MHC-pathogen interaction.

This study investigates both the pathogen prevalence (presence/absence of pathogen) and infection intensity (pathogen quantity) to achieve a more complete picture of the pathogen-mediated selection (Westerdahl *et al.*, 2012) that might drive badger MHC diversity. We first examine for any patterns in infectious status among hosts (e.g. by sex, age, weight/length ratio, standardized microsatellite heterozygosity etc.). We then test whether:

(1) particular MHC alleles were associated with lower or higher pathogen burden (prevalence and infection intensity); and

(2) MHC heterozygotes exhibited less individual pathogen infection and/or lower number of co-infecting pathogens than homozygotes.

It is difficult to disentangle different pathogen-mediated selection mechanisms (Spurgin & Richardson, 2010). If associations between specific pathogens and particular MHC alleles are found this would indicate a more substantial role for either rare-allele advantages or fluctuating selection. If associations between MHC heterozygosity and lower individual pathogen burden and/or number of co-infecting pathogens are found, this would indicate a possible role for heterozygote advantages (Spurgin & Richardson, 2010).

6.3 Materials and methods

Study population and sample collection

This study was conducted on a high-density badger population (36.4 badgers/km^2, *SE* = 2.55 badgers/km^2; Macdonald and Newman 2009) in Wytham Woods (a 6 km^2 deciduous woodland in Oxfordshire, UK; 51°46'26N, 1°19'19W). Trapping events have usually been undertaken four times a year in different seasons (spring, summer, autumn and winter) since 1987 (Macdonald & Newman, 2002), generally over two weeks in June (spring), September (summer) and November (autumn), with one week of trapping in January (winter) of some years (Macdonald *et al.*, 2009). Badgers were caught in mesh-traps baited with peanuts (but without pre-baiting, making the overall effect of food supplentation inconsequential – see Macdonald et al. 2009) , placed near the entrances of active setts (Macdonald & Newman, 2002; Macdonald *et al.*, 2009). Captured badgers were transferred to holding cages and transported to a central handling facility. Badgers were then sedated by an intra-muscular injection using ketamine hydrochloride at 0.2 ml/kg body weight (McLaren *et al.*, 2005). Upon their first capture all badgers were tattooed with a unique individual number on the left inguinal region for permanent individual identification. The sex, age class (cub [<1 years old] or adult, based on body size, refer to as age hereafter), weight (to the nearest 0.1 kg), body length (mm), and trapping location (sett name) of each badger were recorded. Weight and body length were used to calcualte the body condition index (weight/length ratio) (Macdonald *et al.*, 2002). These protocols were subject to ethical review and were performed under Natural England Licence (currently 20104655) and UK Home Office Licence (PPL 30/2835).

During trapping DNA samples were collected: ~100 guard hairs were plucked, and approximately 3 ml of blood was taken by jugular venipuncture using a vacutainer containing EDTA. Blood samples were stored at -20°C, and hair samples in 80% ethanol at room temperature, until DNA isolation was performed. Faecal samples were collected following administration of an enema consisting of 7.5 ml warm soapy water per kg bodyweight (Newman *et al.*, 2001). Faecal samples were preserved individually as two sub-samples. The first sub-sample was preserved using 2.5% aqueous potassium dichromate ($K_2Cr_2O_7$) at 4^{o}C for later examination using faecal flotation technique (Foreyt, 2001), and the second sub-sample was stored at -20°C until DNA isolation was performed. The blood and faecal samples for pathogen screening were collected from individuals trapped June 2009 – January 2010.

Pathogen screening

Trypanosomes and Mustelid Herpes Virus (MHV)

DNA from 200 µl whole blood samples was isolated using QIAamp DNA Blood kit (QIAGEN, Hilden, Germany) and eluted in 100 µl ddH_2O. A quantitative real-time PCR (qPCR) approach was used to determine the prevalence and intensity of trypanosomes infection (*Trypanosoma pestanai*; number of badgers [n_b] = 217; number of samples [n_s] = 360) and MHV (n_b = 218; n_s = 361). qPCR primers were designed for regions in the 18S rRNA gene from *T. pestanai* (forward: 5'-GTCCAGCGAATGAACGAAATTAA; reverse: 5'-AGGGCAGTTGTTCGTCAGAAG) and the DNA polymerase gene from MHV (forward: 5'-GGAGAGTGCTGACCGATGGA; reverse: 5'-AAAAGCCTGGAATTGGATCAATAA; 150bp), using the software StepOne 2.1 [Applied Biosystems (ABI), Foster City, CA, USA]. qPCR was performed in a 20 µl reaction mix consisted of: 10 µl SYBR Green PCR Master Mix (ABI), 0.4 µl (for trypanosomes) or 0.1 µl (for MHV) of each forward and reverse primer [200nM, trypanosmes; 50 nM, MHV], 4.2 µl (trypanosomes) or 4.8 µl (MHV) of RNase-free water, and 5 µl DNA sample. Amplification and real-time fluorescence detection were performed with StepOnePlus PCR Systems (ABI). Each real-time PCR assay contained serial dilutions of 10^7 to 10^1 plasmid standards, which contained the PCR products amplified by each primer set (cloning methods are described in Sin *et al*. 2012a), to produce the calibration curve. All seven standards, samples and one negative control were run in triplicate on the same plate. The

thermal cycling conditions were: 95°C for 10 min, followed by 40 cycles of 95°C for 15 sec, and 60°C for 1 min. A melting curve analysis was added at the end to confirm the identity of the product. The number of pathogen DNA copies was calculated from the Ct values and standard curves and expressed as copies per 1 µl of extracted DNA.

Enteric bacteria (Salmonella, Yersinia and Campylobacter)

DNA from faecal samples was isolated using QIAamp DNA Stool kit (Qiagen), according to the manufacturer's instructions. A semi-quantitative PCR approach was used to determine the prevalence and infection intensity of three bacteria: *Salmonella*, *Yersinia* and *Campylobacter* ($n_b = 99$; $n_s = 150$). PCR primers were designed for regions in the *ompC* gene from *Salmonella* [forward: 5'-ATCGCTGACTTATGCAATCG; reverse: 5'-GTTGCTGATGTCCTTACCTTTAG; (355bp)], *pla* gene from *Yersinia* (forward: 5'-GCTTTATGACGCAGAAACAGGA; reverse: 5'-AACCAGCCTTTCACATTGAGGT; 270bp) and 16S rRNA gene from *Campylobacter* (forward: 5'-GGATGACACTTTTCGGAG; reverse: 5'-AATTCCATCTGCCTCTCC; 246bp). PCR was performed in a 10 µl reaction mix consisted of 0.6 µl of forward and reverse primer (100 µM), 2.7 µl of ddH_2O, 0.1 µl BSA (ABI), 200 µM of each dNTP, 1× PCR buffer (Qiagen), 2 units of HotStarTaq (Qiagen) and 1 µl of DNA. The thermal cycling conditions were: initial denaturation at 95°C for 15 min, followed by 38 cycles at 94°C for 30 s, 63°C (*Salmonella*) or 52°C (*Campylobacter*) for 30 s, and 72°C for 30 s, with a final extension at 72°C for 10 min. Touchdown PCR was performed for *Yersinia*; cycles started at 65°C and dropped by 1°C per cycle until the remaining 29 cycles continued at 55°C. The reactions were performed in triplicate. PCR products were resolved by electrophoresis on 1.5% agarose gel, visualized by ethidium bromide staining, and analyzed using GeneTools (SynGene).

Enteric parasites (Coccidia and helminths)

The faecal flotation technique (Foreyt, 2001) was used to assess the prevalence and intensity of gastro-intestinal parasites present in faecal samples ($n_b = 185$; $n_s = 275$), including coccidia (*Eimeria melis* and *Isospora melis*) and helminths (any nematodes e.g., *Strongyloides*, *Capellaria*). Following the quantitative methods of Newman *et al*. (2001), two slides per faecal sample were screened using microscopy for coccidia oocysts, helminth eggs and larvae (MAFF, 1986; Parker & Duszynski, 1986;

Lloyd & Smith, 1997) and the mean intensity of infection (e.g. coccidian oocysts per gram of pelletted faecal material; Anwar *et al.*, 2000) was then calculated. While absolute rate of infection can only be established by necropsy, assessing gastro-intestinal parasitoses from faecal oocyst / egg counts, as an indirect method, provides a reliable analogue (Seivwright *et al.*, 2004). This technique allows estimation of the parasite burden in "live hosts" (Hall, 1982; Michael & Bundy, 1989). *Isospora*, *Capellaria*, *Strongyloides* and other nematodes were only included in the co-infection analysis as their prevalence and infection intensity were very low.

Ecto-parasites (fleas, lice, ticks)

Ecto-parasites were counted when handling sedated badgers ($n_b = 226$; $n_s = 418$). The number of badger fleas (*Paraceras melis*) was counted during a routine examination of 20 sec duration, involving parting the fur and inspecting the animal's back and underside (Cox *et al.*, 1999). The total number of ticks (*Ixodes hexagonus*) on the body, and the number of badger lice (*Trichodectes melis*) in 4 cm^2 of the inguinal skin region were also counted.

Co-infection

Co-infection status was estimated for 120 time-points from 90 individuals. The number of pathogens co-infecting each badger simultaneously was evaluated for badgers in which screening was performed for all 13 pathogens (trypanosomes, MHV, *Salmonella*, *Yersinia* and *Campylobacter*, *Eimeria*, *Isospora*, *Capellaria*, *Strongyloides*, other nematodes, badger fleas, badger lice and ticks).

MHC and microsatellite genotyping

Genomic DNA was isolated using the GFX Genomic Blood DNA Purification Kit (Amersham Biosciences, Little Chalfont, UK), following the scalable method in the manufacturer's protocol, or from a minimum of 20 hairs with visible follicles, using a Chelex protocol (Walsh *et al.*, 1991). We used published primers to amplify exon 3 and exon 2 regions that encode the antigen-binding domain in MHC class I and class II DRB genes, respectively (Sin *et al.*, 2012b; Sin *et al.*, 2012c). Previous studies of the badger MHC indicated the presence of at least two class I loci and two class II DRB loci and, with four and seven putatively functional sequences identified by transcription analysis,

respectively (Sin *et al.*, 2012b; Sin *et al.*, 2012c). The MHC sequences were separated by reference strand-mediated conformation analysis (RSCA) following Chapter 5. We used the number of alleles within an individual as a measure of MHC heterozygosity across multiple loci (Richardson *et al.*, 2005; Westerdahl *et al.*, 2005). The number of class II DRB alleles is indicative of the heterozygosity of the polymorphic locus; one of the two DRB loci was monomorphic for *Meme-DRB*02* (see Fig. 5.1 in Chapter 5). 'Heterozygosity' hereafter refers to this and to the allele diversity exhibited in class I genes.

In order to measure neutral variation to control for potential confounding effects of demographic processes on MHC structure (Spurgin & Richardson, 2010), we genotyped all individuals at 35 microsatellite loci (detailed in Annavi *et al.*, 2011) and estimated the standardized multilocus heterozygosity of each individual (Coltman *et al.*, 1999; Annavi et al. submitted).

Data analysis

Multi-model inference

Because a multitude of factors influence pathogenic infection in the wild, we employed linear mixed models, which allow the inclusion of multiple explanatory variables and random effects (Paterson *et al.*, 1998; Oliver *et al.*, 2009). We used multi-model inference to establish which explanatory variables had an effect, averaged over all plausible models (Burnham & Anderson, 2002; Anderson, 2008; Symonds & Moussalli, 2011). Analyses were performed using the packages lme4 0.999375-42 (Bates & Maechler, 2009), MuMIn v1.7.7 (Barton, 2009) and AICcmodavg v1.25 (Mazerolle, 2011) in R 2.15.0 (R Core Development Team 2012). We conducted an initial exploration of our data to ascertain the distribution and spread of the data, identify outliers and examine relationships between variables (Zuur *et al.*, 2009). The infection intensities of trypanosomes, MHV, *Eimeria*, *Salmonella*, *Yersinia*, *Campylobacter* and lice were $\log_{10}$(intensity + 1) transformed, to correct for heterogeneity of variance.

To reduce the number of factors included in one model, we conducted multi-model inference in two steps. In step one, we tested whether different life-history factors affected the: (1) prevalence

(presence/absence) of a particular pathogen, (2) pathogen infection intensity (infected individuals only, as non-infection can be due to non-exposure or resistance), or (3) number of species co-infecting a badger simultaneously. We built linear mixed models (LMM: non-count intensity data) and generalized linear mixed models [GLMM: prevalence (binomial error, logit link), and co-infection or ecto-parasite counts (Poisson error, log link)]. The starting models included five fixed effects: three categorical [age (cub or adult), sex and season] and two continuous effects [weight/length ratio and standardized microsatellite heterozygosity]. The number of fleas was also included in the trypanosomes infection model, as *P. melis* is the vector of *T. pestanai* (Lizundia *et al.*, 2011). We controlled for individuals with multiple samples and individuals from the same group, by including individual identity and social group identity as random effects, respectively. Model selection was based on Akaike's information criterion corrected for sample size (AICc; Akaike 1973). Models that are more plausible have lower AICc value. Multi-model inference (Burnham & Anderson, 2002) was performed for models with ΔAICc < 7 (Burnham *et al.*, 2011). Significant variables were retained in a reduced model for use in step 2.

In step two, we investigated whether pathogen prevalence, pathogen infection intensity and co-infection were related to (1) the presence of specific MHC alleles and (2) MHC heterozygosity. We included allele presence or heterozygosity as fixed effects in the reduced model. Eight alleles (three class II DRB: *Meme-DRB*01*, *-DRB*03*, and *-DRB*04*, Sin *et al*., 2012b; five class I: *Meme-MHCI*01*, *-MHCI*02*, *-MHCI*03*, *-MHCI*04*, and *-MHCI*07*, Sin *et al*., 2012a) were included in the reduced model; *Meme-DRB*02* and *Meme-MHCI*05* were present in all individuals and *Meme-MHCI*06* was only present with *Meme-MHCI*07* (see Chapter 5), and were thus excluded from the analyses. Model averaged parameter estimates and parameter estimates with shrinkage (i.e. parameter estimates are set to zeros in models that do not contain the parameter) are reported. The unconditional standard errors (a conservative measure as it accounts for model uncertainty) and 95% confidence intervals (Anderson, 2008) of parameter estimates are also reported. These allow model uncertainty to be included in both model evaluation and derivation of parameter estimates. The relative importance of a parameter was defined as the sum of Akaike weights (where the Akaike weight of each model is it's relative likelihood (exp[-0.5*ΔAICc]) divided by the sum of Akaike weights of all models) for all

models (ΔAICc < 7) containing the predictor (Burnham & Anderson, 2002). The parameter with the largest sum was inferred to be the most influential.

In order to explore whether identified effect of MHC heterozygosity or allele presence on pathogen burden was potentially due to association between MHC heterozygosity and presence of specific allele, we present the Kendall rank correlation between MHC heterozygosity and allele presence when significant MHC effect on pathogen prevalence or infection intensity has been found.

6.4 Results

Trypanosomes

There was no significant association between prevalence and specific allele presence (Fig. 6.1), but badgers with higher class I genes heterozygosity exhibited lower trypanosome prevalence (Fig. 6.1; Table S6.1). The correlations between class I genes heterozygosity and allele presence was highest for *Meme-MHCI*03* (0.42), and correlations for *Meme-MHCI*01*, *02*, *04* and *07* were 0.17, 0.19, 0.15 and 0.38 respectively in the trypanosome data set. The prevalence of trypanosomes was higher in adults than in cubs, and lower in females than in males (Table S6.1). Flea load was associated positively with trypanosomes prevalence.

Trypanosomes infection intensity was lower in adults and females than in cubs and males, respectively. Badgers with lower standardised microsatellite heterozygosity had lower infection intensity (Table S6.2). The intensity of infection by trypanosomes was not associated with the MHC alleles or MHC heterozygosity analysed (Fig. 6.2).

MHV

MHV was prevalent in samples from nearly all individuals (354/361 = 98.1%), thus only infection intensity was analysed for MHV. Adults had lower MHV intensity than cubs, and weight/length ratio was negatively associated with MHV infection intensity (Table S6.3). Seasonably, badgers had lower MHV intensity in winter than in spring. The presence of *Meme-DRB*01* and *Meme-MHCI*03* were correlated with the intensity of MHV infection (Fig. 6.2; Table S6.3). Individuals with *Meme-DRB*01*

had higher MHV intensity and individuals with *Meme-MHCI*03* had lower MHV intensity. The correlation between *Meme-MHCI*03* and class I genes heterozygosity was 0.42 in the MHV data set. The correlation between *Meme-DRB*01* and class II genes heterozygosity was 0.08. The infection intensity by MHV was not associated with the MHC heterozygosity (Fig. 6.2).

Enteric parasites

Adults had lower prevalence and infection intensity of *Eimeria* than did cubs (Table S6.4 and S6.5). Badgers in summer and autumn had lower *Eimeria* infection intensity than those in spring (Table S6.5). *Eimeria* infection was not associated with specific MHC alleles or MHC heterozygosity (Fig. 6.1 and 6.2).

Enteric bacteria

The prevalence of *Salmonella* was lower in adults than in cubs (Table S6.6). More individuals were infected by *Salmonella* in summer and autumn than in spring (Table S6.6). Infection intensity of *Salmonella* was associated negatively with standardized microsatellite heterozygosity (Table S6.7). *Salmonella* infection was not associated with specific MHC alleles or MHC heterozygosity (Fig. 6.1 and 6.2).

The prevalence of *Yersinia* was higher in summer and autumn than in spring (Table S6.8). Individuals that carried *Meme-MHCI*04* were more likely to be infected by *Yersinia* (Fig. 6.1). The correlation between *Meme-MHCI*04* and class I genes heterozygosity was 0.28 in this data set. The prevalence of infection by *Yersinia* was not associated with MHC heterozygosity (Fig. 6.1). The infection intensity of *Yersinia* was higher for individuals with *Meme-DRB*01* and lower for individuals with *Meme-DRB*04* (Fig. 6.2; Table S6.9). The correlation between *Meme-DRB*01* and class II genes heterozygosity was 0.01 in the *Yersinia* data set. The correlation between *Meme-DRB*04* and class II genes heterozygosity was 0.46 in the *Yersinia* data set. The intensity of infection by *Yersinia* was not associated with MHC heterozygosity (Fig. 6.1 and 6.2).

None of the tested explanatory variables correlated with *Campylobacter* prevalence (Fig. 6.1; Table S6.10). Infection intensity of *Campylobacter* was lower in summer than spring (Table S6.11). Badgers with high heterozygosity in class II DRB genes exhibited lower infection intensity (Fig. 6.2; Table S6.11). The correlations of class II DRB genes heterozygosity with *Meme-DRB*01, 03* and *04* were 0.14, 0.62 and 0.37 respectively in the *Campylobacter* data set.

Ectoparasites

The prevalence of fleas was greater for adults than for cubs, and negatively associated with weight/length ratio (Table S6.12). Flea prevalence was higher in summer and autumn than in spring. Flea intensity was higher in summer and lower in winter than in spring (Table S6.13). Both flea prevalence and intensity was not associated with specific MHC alleles or MHC heterozygosity (Fig. 6.1 and 6.2).

The prevalence of lice was associated negatively with weight/length ratio (Table S6.14). Lice prevalence was higher in summer than in spring and higher for males than for females. The infection intensity of lice was associated negatively with weight/length ratio (Table S6.15). Females had lower lice intensity than males. Lice infection was not associated with specific MHC alleles or MHC heterozygosity (Fig. 6.1 and 6.2).

Co-infection

Individual badgers were simultaneously co-infected with between two and eleven different pathogens (mean = 6.0, SE ± 0.075). Co-infection was not associated with the presence/absence of individual MHC alleles or MHC heterozygosity (Fig. 6.3), and there was no association between the number of co-infecting species and any of the tested explanatory variables (Table S6.16). By excluding ectoparasites, which may subject to direct behavioural intervention by the host, we examined the number of co-infecting internal pathogens (range = 2-9, mean = 4.4, SE ± 0.065), and there was no association with the explanatory variables.

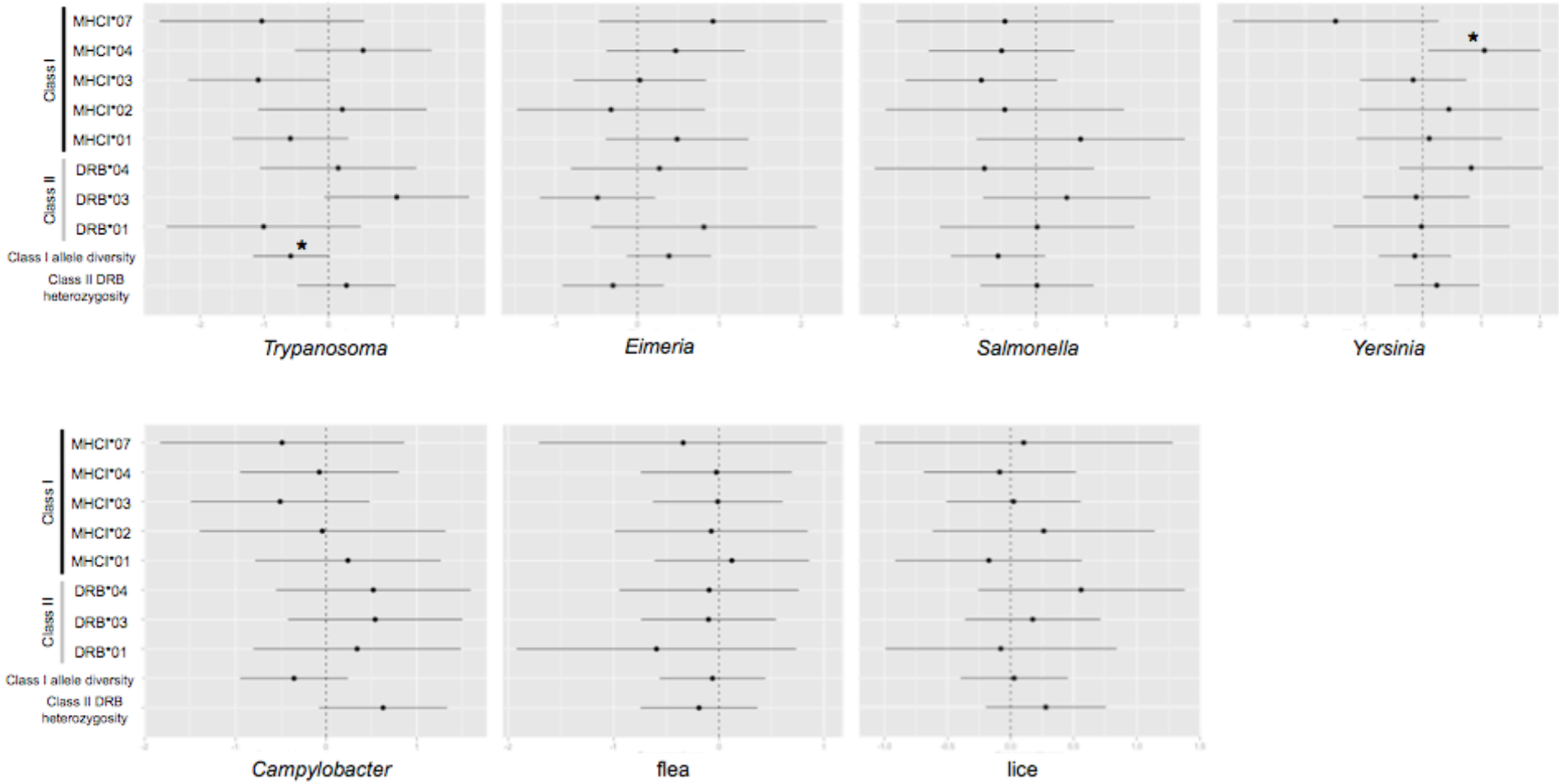


Fig. 6.1 Model averaged parameter estimates and their 95% confidence intervals for the eight MHC predictors (MHC class I and II genes) associated with the prevalence of seven pathogens. MHV was present in nearly all samples (354/361 = 98.1%), thus only infection intensity was analysed for MHV. * indicates parameter that shows a significant effect.

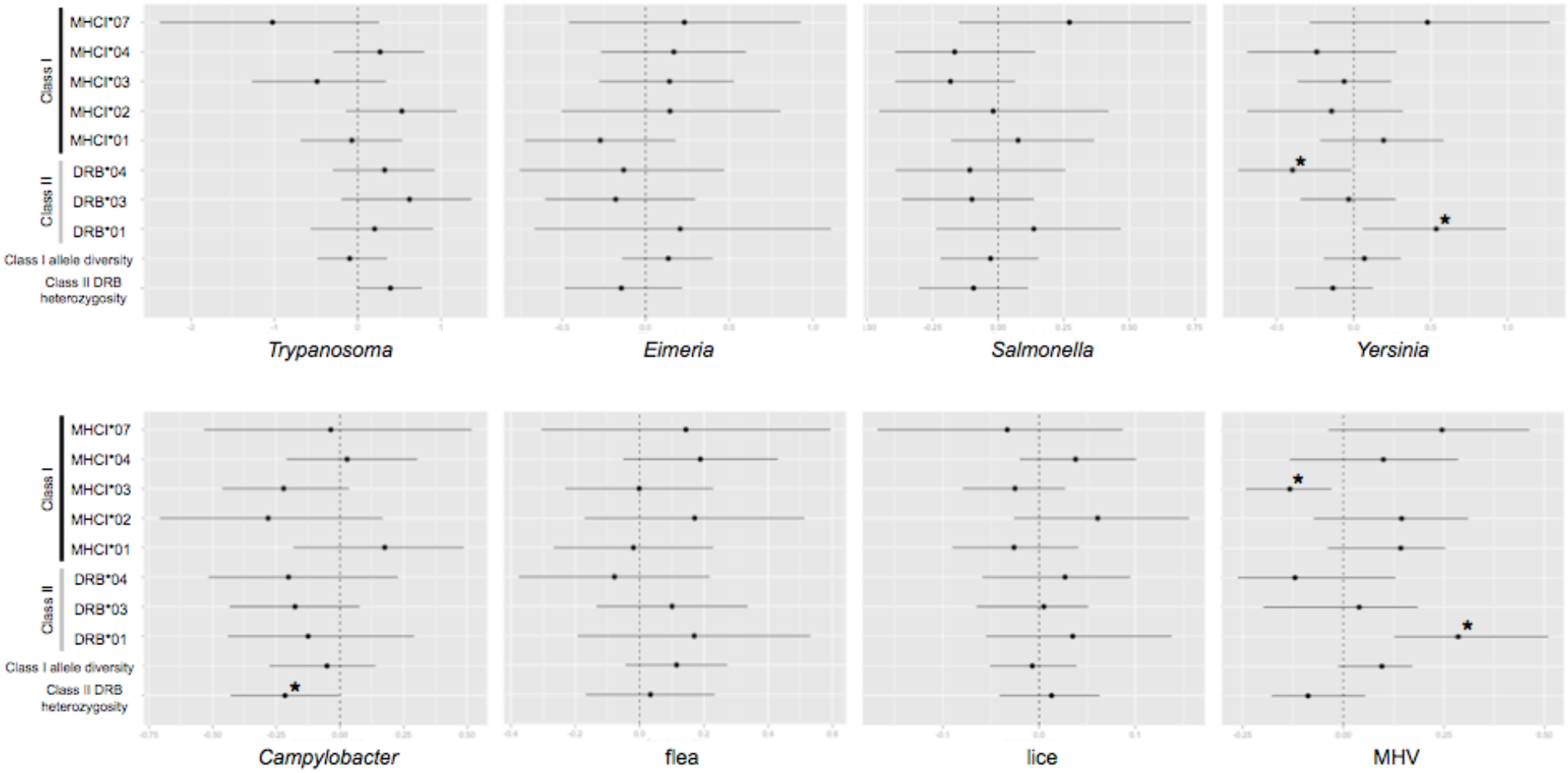


Fig. 6.2 Model averaged parameter estimates and their 95% confidence intervals for the eight MHC predictors (MHC class I and II genes) associated with the infection intensity of eight pathogens. * indicates parameter that shows a significant effect.

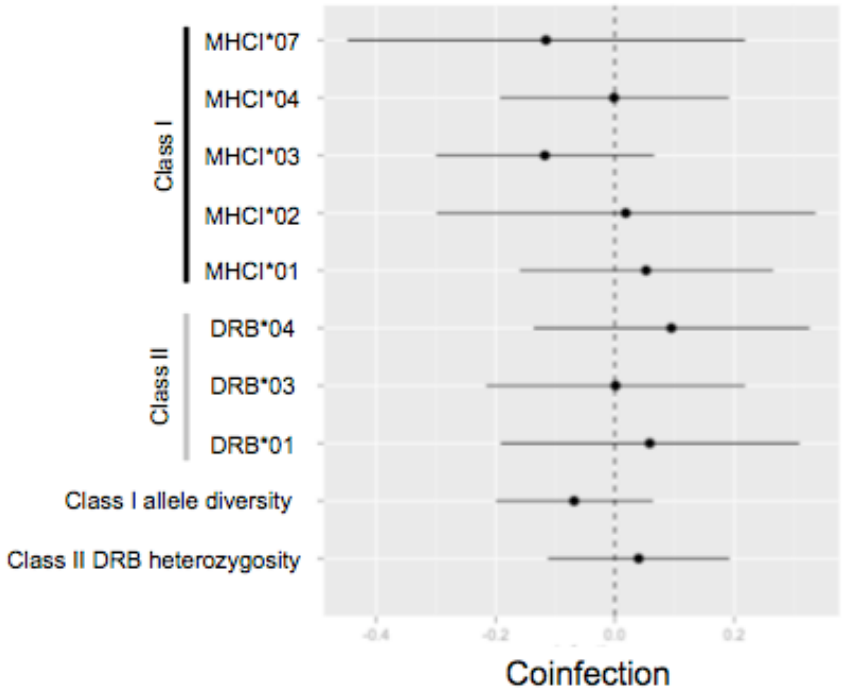


Fig. 6.3 Model averaged parameter estimates and their 95% confidence intervals for the ten MHC predictors (MHC class I and II genes, MHC genes heterozygosity) associated with the co-infection by 13 pathogens. * indicates parameter that shows a significant effect.

6.5 Discussion

Associations between specific pathogens and particular MHC alleles were found in this population. Certain allele conferred susceptibility to two different pathogens, i.e. *Meme-DRB*01* was associated with higher MHV and *Yersinia* infection intensity. By examining infection intensity, alleles that provided quantitative resistance (e.g. *Meme-DRB*04*) were identified in addition to qualitative resistance alleles (Westerdahl *et al.*, 2012), thus providing a more complete picture of MHC-pathogen associations. These associations between certain MHC alleles and pathogen resistance or susceptibility are in concordance with the prediction of rare-allele advantage hypothesis (Takahata & Nei, 1990), whereby MHC alleles that have been the target of recent pathogen evasion will be susceptible and become less frequent, and rare or new alleles will become resistant to newly mutated pathogens and increase in frequency. In addition to the host-pathogen co-evolution and cyclic change in allele frequencies, the presence of susceptible alleles could be due to simultaneous resistance to a pathogen and susceptibility to others (Penn & Potts, 1999).

Conversely, fluctuating selection can also produce the same effect on MHC-pathogen associations. Both rare-allele advantage and fluctuating selection can lead to associations between pathogens and specific MHC alleles, and it is very difficult to separate the two non-exclusive mechanisms (Spurgin & Richardson, 2010). These pathogen-MHC allele associations were identified while controlling for genome-wide heterozygosity, therefore these were not by-products of confounding effects of demographic processes on MHC structure (Spurgin & Richardson, 2010). Associations with microsatellite heterozygosity were only identified for trypanosomes and *Salmonella* infection intensities. Although Salmonella intensity was lower for individuals with higher microsatellite heterozygosity, in concordance with heterozygosity–fitness correlations (Hansson & Westerberg, 2002), individuals with lower microsatellite heterozygosity had lower trypanosomes intensity.

Both class I and class II alleles were associated with particular pathogens in this study. MHC class II molecules principally bind exogenous antigens and are only expressed on antigen-presenting cells, such as B cells and macrophages (Hughes & Yeager, 1998). MHC class I molecules are responsible primarily for intracellular antigen binding and are expressed on the surface of all nucleated somatic

cells (Bjorkman & Parham, 1990), while cross-presentation also allows class I molecules to process exogenous antigens (Heath & Carbone, 2001; Ackerman & Cresswell, 2004). Our investigation found that class I alleles were associated with both an intracellular pathogen (e.g. MHV) and an extracellular pathogen (e.g. *Yersinia*). Class II DRB alleles also associated with an extracellular pathogen (e.g. *Yersinia*) and an intracellular pathogen (e.g. MHV).

The majority of MHC-pathogen studies have considered only one or a few pathogens, and it might be possible that evidence of heterozygous advantages may only be exposed by including more pathogens (Langefors *et al.*, 2001). We examined thirteen pathogenic species from a wide variety of taxa to establish the co-infection status, but still found no support for MHC heterozygosity being associated with fewer co-infecting pathogens. Some studies have also investigated and found no association between MHC heterozygosity and pathogen resistance (e.g. Langefors *et al.*, 2001; Harf & Sommer, 2005; Meyer-lucht & Sommer, 2005; Schad *et al.*, 2005; Tollenaere *et al.*, 2008) but demonstrated associations between individual MHC alleles and specific pathogens. However, MHC heterozygote advantage can also work against single pathogen infection if heterozygous individuals recognise and present more antigens originating from that individual pathogen. We found a heterozygote advantage of class I genes for trypanosomes prevalence and of class II DRB genes for *Campylobacter* infection intensity. However, there was also evidence for a possible association between MHC heterozygosity and the presence of specific allele (for trypanosomes prevalence, correlation between class I heterozygosity and *Meme-MHCI*03* was 0.42; for *Campylobacter* intensity, correlation between class II DRB heterozygosity and *Meme-DRB*03* was 0.62), which also had the same direction of effect though marginally not significant. This suggests that the identified heterozygote advantage could potentially be due to the effect of specific allele presence. This highlights the difficulties of disentangling different pathogen-mediated selection mechanism (Spurgin & Richardson, 2010), as there is a rare allele advantage component to heterozygote advantage (Apanius *et al.*, 1997), i.e. different mechanisms are not mutually exclusive.

Factors other than the presence of MHC alleles were also associated with the burden of different pathogen species. Flea (*P. melis*) number correlated positively with trypanosomes (*T. pestanai*)

prevalence, consistent with the finding that *P. melis* is the vector of *T. pestanai* in European badgers (Lizundia *et al.*, 2011). Adults exhibited a higher prevalence of fleas and this may be the reason that they had a higher prevalence of trypanosomes than did cubs, with continuing life-time exposure to the risk of infection. Age also correlated with certain other pathogens: compared with cubs adults exhibited lower MHV intensity; prevalence of *Salmonella*; and prevalence and intensity of *Eimeria*. Cubs are still under-going immunological development, and are consequently more susceptible to pathogens. This is consistent with Anwar *et al.* (2000) and Newman *et al.* (2001) who report that cubs have a much higher burden of *Eimeria*. Similarly, MHV intensity was also higher in cubs than in adults, where cubs may be first exposed to MHV by vertical transmission (Jones, 2001; Corey & Wald, 2009), rather than through sexual transmission. Due to acquired immunological memory, the immune systems of adults should be more capable of resolving these infections than that of cubs.

Badgers with a higher body condition index (weight/length ratio), i.e., notionally sturdier badgers, exhibited lower flea and lice prevalence and lower MHV and lice intensity. Individuals in higher body-condition may be better able to allocate resources to combating pathogens (Macdonald *et al.*, 2002; Montes *et al.*, 2011). Alternatively, individuals with higher severity of infection may suffer compromised metabolic budgets, and thus loose condition as a consequence. Flea intensity affects growth and body size in great tits (*Parus major*) (Fitze *et al.*, 2004), while juvenile gerbils (*Gerbillus andersoni*) parasitised by natural levels of fleas lost body mass faster and gained body mass slower than uninfected individuals (Hawlena *et al.*, 2006b), leading to compromised survival probability (Hawlena *et al.*, 2006a). Variation in pathogen burdens can thus impact on host health and survival traits, which highlights the importance of having advantageous alleles to mount the most effective immune response, enhancing host fitness (Brouwer *et al.*, 2010).

Conclusion

Our study provides evidence for the association between specific MHC alleles and certain pathogens in a population of wild mammals. We found no evidence for a MHC heterozygote advantage under co-infection status and weak evidence of heterozygote advantage to single pathogen infection, which may potentially be due to correlation with specific allele presence. This suggests rare-allele advantage

and/or fluctuating selection are likely to be the major selective force shaping the MHC diversity, where heterozygote advantage may play a lesser role in this species. We stress, however, that these findings do not preclude other selection mechanisms, such as mating preferences or post-copulatory mate choice, and apply only to those pathogens investigated here. Future studies will be needed to evaluate the importance of these mechanisms in MHC evolution in this and other species.

Supplementary material

Table S6.1 The association of (a) presence or absence of alleles or (b) MHC genes heterozygosity and prevalence (presence or absence) of trypanosomes (*Trypanosoma pestanai*) infection in badgers: Parameter estimates with their unconditional standard errors (SE), 95% confidence intervals and relative importance after model averaging (models with ΔAICc < 7). Estimates where the 95% confidence intervals do not overlap zero are in bold.

Parameter	Estimate (with shrinkage)	Estimate	Unconditional SE	95% Confidence interval		Relative importance^
(a)						
(Intercept)	0.907	0.907	0.785	-0.632	2.446	
Fleas	0.053	0.053	0.028	-0.001	0.107	1.00*
Age (adult)	**1.368**	**1.368**	**0.358**	**0.666**	**2.07**	**1.00***
Sex (female)	**-1.202**	**-1.202**	**0.341**	**-1.87**	**-0.534**	**1.00***
DRB*03 (present)	0.696	1.062	0.576	-0.067	2.191	0.65
MHCI*03 (present)	-0.829	-1.092	0.559	-2.187	0.003	0.76
DRB*01 (present)	-0.474	-1.009	0.772	-2.522	0.504	0.47
MHCI*01 (present)	-0.245	-0.592	0.461	-1.495	0.311	0.41
MHCI*07 (present)	-0.432	-1.036	0.812	-2.628	0.555	0.42
MHCI*04 (present)	0.205	0.540	0.544	-0.526	1.607	0.38
DRB*04 (present)	0.037	0.151	0.621	-1.067	1.368	0.25
MHCI*02 (present)	0.053	0.216	0.669	-1.095	1.528	0.25
(b)						
(Intercept)	1.210	1.210	0.994	-0.739	3.159	
Fleas	**0.054**	**0.054**	**0.027**	**0**	**0.108**	**1.00***
Age (adult)	**1.394**	**1.394**	**0.346**	**0.715**	**2.073**	**1.00***
Sex (female)	**-1.111**	**-1.111**	**0.328**	**-1.754**	**-0.469**	**1.00***
Class I allele diversity	**-0.416**	**-0.587**	**0.298**	**-1.171**	**-0.004**	**0.71**
DRB heterozygosity	0.093	0.281	0.392	-0.486	1.048	0.33

* Included in all models

^ Relative importance is the sum of Akaike weights for all models (ΔAICc < 7) containing the parameter

Table S6.2 The association of (a) presence or absence of alleles or (b) MHC genes heterozygosity and infection intensity of trypanosomes (*Trypanosoma pestanai*) in badgers: Parameter estimates with their unconditional standard errors (SE), 95% confidence intervals and relative importance after model averaging (models with ΔAICc < 7). Estimates where the 95% confidence intervals do not overlap zero are in bold.

Parameter	Estimate (with shrinkage)	Estimate	Unconditional SE	Confidence interval		Relative importance^
(a)						
(Intercept)	**2.180**	**2.180**	**0.635**	**0.975**	**3.411**	
Age (adult)	**-0.559**	**-0.559**	**0.255**	**-1.056**	**-0.059**	1.00*
Sex (female)	**-0.430**	**-0.430**	**0.190**	**-0.808**	**-0.061**	1.00*
SH[a]	**1.051**	**1.051**	**0.503**	**0.111**	**2.092**	1.00*
DRB*03 (present)	0.388	0.621	0.399	-0.198	1.366	0.63
MHCI*02 (present)	0.274	0.529	0.342	-0.143	1.186	0.52
MHCI*07 (present)	-0.534	-1.023	0.674	-2.375	0.257	0.52
MHCI*03 (present)	-0.227	-0.488	0.407	-1.273	0.342	0.47
DRB*04 (present)	0.118	0.323	0.317	-0.300	0.926	0.36
MHCI*04 (present)	0.093	0.269	0.280	-0.293	0.797	0.35
DRB*01 (present)	0.058	0.202	0.383	-0.567	0.905	0.29
MHCI*01 (present)	-0.018	-0.071	0.322	-0.689	0.536	0.25
(b)						
(Intercept)	**1.880**	**1.880**	**0.778**	**0.507**	**3.54**	
Age (adult)	**-0.576**	**-0.576**	**0.251**	**-1.073**	**-0.086**	1.00*
Sex (female)	**-0.374**	**-0.374**	**0.185**	**-0.739**	**-0.012**	1.00*
SH [a]	**0.947**	**0.947**	**0.472**	**0.025**	**1.878**	1.00*
DRB heterozygosity	0.276	0.392	0.204	-0.008	0.772	0.7
Class I allele diversity	-0.028	-0.097	0.212	-0.485	0.352	0.29

* Included in all models

^ Relative importance is the sum of Akaike weights for all models (ΔAICc < 7) containing the parameter

[a] Standardised microsatellite heterozygosity (Coltman *et al*., 1999; Annavi *et al.* submitted)

Table S6.3 The association of (a) presence or absence of alleles or (b) MHC genes heterozygosity and infection intensity of Mustelid herpes virus in badgers: Parameter estimates with their unconditional standard errors (SE), 95% confidence intervals and relative importance after model averaging (models with ΔAICc < 7). Estimates where the 95% confidence intervals do not overlap zero are in bold.

Parameter	Estimate (with shrinkage)	Estimate	Unconditional SE	Confidence interval		Relative importance^
(a)						
(Intercept)	1.844	1.844	0.153	**1.529**	**2.132**	
Age (adult)	-0.634	-0.634	0.078	**-0.794**	**-0.490**	1.00*
Season (Summer) [a]	-0.022	-0.022	0.048	-0.115	0.076	1.00*
Season (Autumn) [a]	-0.001	-0.001	0.058	-0.111	0.116	"
Season (Winter) [a]	-0.355	-0.355	0.118	**-0.582**	**-0.118**	"
Weight/length ratio	-0.042	-0.042	0.014	**-0.069**	**-0.014**	1.00*
DRB*01 (present)	0.263	0.285	0.108	**0.126**	**0.509**	0.92
MHCI*03 (present)	-0.106	-0.133	0.067	**-0.242**	**-0.029**	0.80
MHCI*07 (present)	0.170	0.245	0.132	-0.037	0.462	0.69
MHCI*01 (present)	0.088	0.143	0.092	-0.039	0.253	0.61
MHCI*02 (present)	0.071	0.145	0.105	-0.074	0.310	0.49
DRB*04 (present)	-0.049	-0.120	0.109	-0.262	0.129	0.41
MHCI*04 (present)	0.039	0.099	0.103	-0.133	0.285	0.40
DRB*03 (present)	0.013	0.039	0.091	-0.198	0.185	0.32
(b)						
(Intercept)	**2.002**	**2.002**	**0.183**	**1.757**	**2.318**	
Age (adult)	**-0.706**	**-0.706**	**0.077**	**-0.86**	**-0.557**	1.00*
Season (Summer) [a]	-0.012	-0.012	0.049	-0.107	0.086	1.00*
Season (Autumn) [a]	-0.005	-0.005	0.059	-0.118	0.113	"
Season (Winter) [a]	**-0.337**	**-0.337**	**0.121**	**-0.572**	**-0.098**	"
Weight/length ratio	**-0.034**	**-0.034**	**0.014**	**-0.061**	**-0.006**	1.00*
Class I allele diversity	0.066	0.095	0.050	-0.013	0.172	0.70
DRB heterozygosity	-0.046	-0.088	0.060	-0.178	0.054	0.53

[a] Spring was the reference category

* Included in all models

^ Relative importance is the sum of Akaike weights for all models (ΔAICc < 7) containing the parameter

Table S6.4 The association of (a) presence or absence of alleles or (b) MHC genes heterozygosity and prevalence (presence or absence) of *Eimeria* infection in badgers: Parameter estimates with their unconditional standard errors (SE), 95% confidence intervals and relative importance after model averaging (models with ΔAICc < 7). Estimates where the 95% confidence intervals do not overlap zero are in bold.

Parameter	Estimate (with shrinkage)	Estimate	Unconditional SE	95% Confidence interval		Relative importance^
(a)						
(Intercept)	1.032	1.032	0.719	-0.377	2.441	
Age (adult)	**-2.067**	**-2.067**	**0.376**	**-2.803**	**-1.331**	1.00*
DRB*03 (present)	-0.231	-0.487	0.360	-1.192	0.219	0.47
MHCI*07 (present)	0.414	0.924	0.711	-0.470	2.317	0.45
DRB*01 (present)	0.343	0.813	0.703	-0.564	2.190	0.42
MHCI*04 (present)	0.188	0.468	0.431	-0.378	1.313	0.40
MHCI*01 (present)	0.188	0.485	0.445	-0.388	1.357	0.39
MHCI*02 (present)	-0.090	-0.322	0.587	-1.472	0.828	0.28
DRB*04 (present)	0.076	0.269	0.552	-0.813	1.350	0.28
MHCI*03 (present)	0.008	0.029	0.412	-0.779	0.838	0.28
(b)						
(Intercept)	1.275	1.275	0.780	-0.254	2.804	
Age (adult)	**-2.143**	**-2.143**	**0.371**	**-2.871**	**-1.415**	1.00*
Class I allele diversity	0.197	0.386	0.263	-0.130	0.902	0.51
DRB heterozygosity	-0.107	-0.298	0.317	-0.918	0.323	0.36

* Included in all models

^ Relative importance is the sum of Akaike weights for all models (ΔAICc < 7) containing the parameter

Table S6.5 The association of (a) presence or absence of alleles or (b) MHC genes heterozygosity and infection intensity of *Eimeria* in badgers: Parameter estimates with their unconditional standard errors (SE), 95% confidence intervals and relative importance after model averaging (models with ΔAICc < 7). Estimates where the 95% confidence intervals do not overlap zero are in bold.

Parameter	Estimate (with shrinkage)	Estimate	Unconditional SE	95% Confidence interval		Relative importance^
(a)						
(Intercept)	**2.481**	**2.481**	**0.306**	**1.873**	**3.073**	
Age (adult)	**-1.049**	**-1.049**	**0.165**	**-1.364**	**-0.717**	1.00*
Season (Summer)[a]	**-0.558**	**-0.559**	**0.165**	**-0.884**	**-0.238**	1.00*
Season (Autumn)[a]	**-0.605**	**-0.605**	**0.223**	**-1.048**	**-0.173**	"
MHCI*01 (present)	-0.090	-0.271	0.231	-0.727	0.18	0.33
MHCI*04 (present)	0.043	0.169	0.222	-0.268	0.601	0.25
MHCI*07 (present)	0.055	0.232	0.354	-0.461	0.927	0.23
DRB*04 (present)	-0.027	-0.131	0.313	-0.756	0.471	0.21
DRB*01 (present)	0.043	0.207	0.453	-0.667	1.108	0.21
DRB*03 (present)	-0.047	-0.179	0.23	-0.604	0.297	0.26
MHCI*02 (present)	0.030	0.146	0.335	-0.507	0.807	0.21
MHCI*03 (present)	0.036	0.143	0.206	-0.278	0.529	0.25
(b)						
(Intercept)	**2.520**	**2.520**	**0.385**	**1.948**	**3.121**	
Age (adult)	**-1.034**	**-1.034**	**0.161**	**-1.342**	**-0.714**	1.00*
Season (Summer)[a]	**-0.573**	**-0.573**	**0.164**	**-0.893**	**-0.251**	1.00*
Season (Autumn)[a]	**-0.624**	**-0.624**	**0.221**	**-1.056**	**-0.191**	"
Class I allele diversity	0.046	0.136	0.139	-0.141	0.401	0.34
DRB heterozygosity	-0.044	-0.145	0.182	-0.487	0.219	0.31

[a] Spring was the reference category

* Included in all models

^ Relative importance is the sum of Akaike weights for all models (ΔAICc < 7) containing the parameter

Table S6.6 The association of (a) presence or absence of alleles or (b) MHC genes heterozygosity and prevalence (presence or absence) of *Salmonella* infection in badgers: Parameter estimates with their unconditional standard errors (SE), 95% confidence intervals and relative importance after model averaging (models with ΔAICc < 7). Estimates where the 95% confidence intervals do not overlap zero are in bold.

Parameter	Estimate (with shrinkage)	Estimate	Unconditional SE	95% Confidence interval		Relative importance^
(a)						
(Intercept)	-0.614	-0.614	0.725	-2.035	0.807	
Age (adult)	**-1.501**	**-1.501**	**0.53**	**-2.54**	**-0.463**	1.00*
Season (Summer)[a]	**2.662**	**2.662**	**0.569**	**1.548**	**3.777**	1.00*
Season (Autumn)[a]	**2.733**	**2.733**	**0.577**	**1.603**	**3.863**	"
MHCI*03 (present)	-0.401	-0.779	0.552	-1.861	0.303	0.51
DRB*03 (present)	0.129	0.440	0.609	-0.753	1.633	0.29
MHCI*01 (present)	0.201	0.637	0.758	-0.849	2.124	0.32
MHCI*04 (present)	-0.156	-0.490	0.532	-1.532	0.553	0.32
MHCI*07 (present)	-0.108	-0.443	0.793	-1.997	1.112	0.24
MHCI*02 (present)	-0.109	-0.445	0.869	-2.149	1.258	0.24
DRB*04 (present)	-0.249	-0.736	0.799	-2.302	0.83	0.34
DRB*01 (present)	0.004	0.018	0.708	-1.37	1.406	0.22
(b)						
(Intercept)	-0.107	-0.107	1.160	-2.380	2.166	
Age (adult)	**-1.559**	**-1.559**	**0.522**	**-2.583**	**-0.536**	1.00*
Season (Summer)[a]	**2.665**	**2.665**	**0.564**	**1.560**	**3.771**	1.00*
Season (Autumn)[a]	**2.764**	**2.764**	**0.575**	**1.637**	**3.891**	"
Class I allele diversity	-0.293	-0.541	0.343	-1.214	0.132	0.54
DRB heterozygosity	0.004	0.014	0.413	-0.794	0.823	0.25

[a] Spring was the reference category

* Included in all models

^ Relative importance is the sum of Akaike weights for all models (ΔAICc < 7) containing the parameter

Table S6.7 The association of (a) presence or absence of alleles or (b) MHC genes heterozygosity and infection intensity of *Salmonella* in badgers: Parameter estimates with their unconditional standard errors (SE), 95% confidence intervals and relative importance after model averaging (models with ΔAICc < 7). Estimates where the 95% confidence intervals do not overlap zero are in bold.

Parameter	Estimate (with shrinkage)	Estimate	Unconditional SE	95% Confidence interval		Relative importance^
(a)						
(Intercept)	**2.631**	**2.631**	**0.301**	**2.01**	**3.189**	
Age (adult)	-0.176	-0.176	0.117	-0.422	0.038	1.00*
MHCI*03 (present)	-0.083	-0.182	0.117	-0.393	0.065	0.46
SH[a]	**-0.666**	**-0.666**	**0.302**	**-1.265**	**-0.08**	1.00*
MHCI*04 (present)	-0.059	-0.166	0.136	-0.393	0.142	0.35
DRB*01 (present)	0.032	0.136	0.18	-0.237	0.468	0.24
DRB*03 (present)	-0.026	-0.100	0.129	-0.367	0.137	0.26
MHCI*07 (present)	0.087	0.272	0.226	-0.15	0.735	0.32
DRB*04 (present)	-0.024	-0.108	0.165	-0.392	0.256	0.22
MHCI*01 (present)	0.016	0.076	0.139	-0.179	0.366	0.21
MHCI*02 (present)	-0.003	-0.019	0.223	-0.454	0.421	0.17
(b)						
(Intercept)	**2.628**	**2.628**	**0.331**	**1.986**	**3.179**	
Age (adult)	-0.207	-0.207	0.115	-0.433	0.016	1.00*
SH[a]	**-0.623**	**-0.623**	**0.294**	**-1.200**	**-0.065**	1.00*
DRB heterozygosity	-0.028	-0.094	0.107	-0.303	0.113	0.30
Class I allele diversity	-0.007	-0.029	0.096	-0.220	0.153	0.23

* Included in all models

^ Relative importance is the sum of Akaike weights for all models (ΔAICc < 7) containing the parameter

[a] Standardised microsatellite heterozygosity

Table S6.8 The association of (a) presence or absence of alleles or (b) MHC genes heterozygosity and prevalence (presence or absence) of *Yersinia* infection in badgers: Parameter estimates with their unconditional standard errors (SE), 95% confidence intervals and relative importance after model averaging (models with ΔAICc < 7). Estimates where the 95% confidence intervals do not overlap zero are in bold.

Parameter	Estimate (with shrinkage)	Estimate	Unconditional SE	95% Confidence interval		Relative importance^
(a)						
(Intercept)	-1.141	-1.141	0.681	-2.475	0.193	
Season (Summer)[a]	**1.227**	**1.227**	**0.451**	**0.342**	**2.111**	1.00*
Season (Autumn)[a]	**1.154**	**1.154**	**0.453**	**0.266**	**2.042**	"
MHCI*07 (present)	-0.910	-1.483	0.896	-3.24	0.273	0.61
MHCI*04 (present)	**0.862**	**1.052**	**0.491**	**0.09**	**2.013**	0.82
DRB*04 (present)	0.362	0.828	0.625	-0.398	2.054	0.44
MHCI*02 (present)	0.115	0.448	0.785	-1.091	1.987	0.26
MHCI*03 (present)	-0.040	-0.161	0.464	-1.07	0.748	0.25
DRB*03 (present)	-0.026	-0.109	0.464	-1.018	0.801	0.24
MHCI*01 (present)	0.029	0.114	0.633	-1.128	1.356	0.25
DRB*01 (present)	-0.005	-0.019	0.768	-1.524	1.486	0.26
(b)						
(Intercept)	-0.574	-0.574	0.784	-2.110	0.961	
Season (Summer)[a]	**1.120**	**1.120**	**0.434**	**0.269**	**1.971**	1.00*
Season (Autumn)[a]	**1.145**	**1.145**	**0.442**	**0.278**	**2.012**	"
DRB heterozygosity	0.070	0.242	0.372	-0.487	0.970	0.29
Class I allele diversity	-0.035	-0.131	0.315	-0.748	0.485	0.27

[a] Spring was the reference category

* Included in all models

^ Relative importance is the sum of Akaike weights for all models (ΔAICc < 7) containing the parameter

Table S6.9 The association of (a) presence or absence of alleles or (b) MHC genes heterozygosity and infection intensity of *Yersinia* in badgers: Parameter estimates with their unconditional standard errors (SE), 95% confidence intervals and relative importance after model averaging (models with ΔAICc < 7). Estimates where the 95% confidence intervals do not overlap zero are in bold.

Parameter	Estimate (with shrinkage)	Estimate	Unconditional SE	95% Confidence interval		Relative importance^
(a)						
(Intercept)	**1.270**	**1.270**	**0.243**	**0.719**	**1.671**	
DRB*01 (present)	**0.423**	**0.536**	**0.238**	**0.058**	**0.993**	0.79
DRB*04 (present)	**-0.290**	**-0.397**	**0.188**	**-0.75**	**-0.015**	0.73
MHCI*04 (present)	-0.084	-0.240	0.247	-0.692	0.277	0.35
MHCI*01 (present)	0.063	0.194	0.205	-0.219	0.584	0.33
MHCI*07 (present)	0.169	0.479	0.398	-0.287	1.274	0.35
MHCI*02 (present)	-0.037	-0.144	0.258	-0.692	0.32	0.25
MHCI*03 (present)	-0.014	-0.062	0.156	-0.365	0.245	0.23
DRB*03 (present)	-0.007	-0.033	0.158	-0.346	0.273	0.22
(b)						
(Intercept)	**1.577**	**1.577**	**0.288**	**0.719**	**1.671**	
DRB heterozygosity	-0.048	-0.135	0.132	-0.380	0.125	0.35
Class I allele diversity	0.019	0.069	0.132	-0.195	0.306	0.27

^ Relative importance is the sum of Akaike weights for all models (ΔAICc < 7) containing the parameter

Table S6.10 The association of (a) presence or absence of alleles or (b) MHC genes heterozygosity and prevalence (presence or absence) of *Campylobacter* infection in badgers: Parameter estimates with their unconditional standard errors (SE), 95% confidence intervals and relative importance after model averaging (models with ΔAICc < 7). Estimates where the 95% confidence intervals do not overlap zero are in bold.

Parameter	Estimate (with shrinkage)	Estimate	Unconditional SE	95% Confidence interval		Relative importance^
(a)						
(Intercept)	0.452	0.452	0.438	-0.406	1.311	
DRB*04 (present)	0.174	0.521	0.548	-0.553	1.596	0.33
DRB*03 (present)	0.215	0.541	0.49	-0.42	1.503	0.40
MHCI*03 (present)	-0.183	-0.507	0.503	-1.493	0.478	0.36
MHCI*07 (present)	-0.135	-0.486	0.687	-1.832	0.861	0.28
MHCI*01 (present)	0.060	0.241	0.523	-0.783	1.265	0.25
DRB*01 (present)	0.090	0.342	0.584	-0.802	1.487	0.26
MHCI*04 (present)	-0.017	-0.075	0.447	-0.95	0.8	0.22
MHCI*02 (present)	-0.009	-0.041	0.692	-1.396	1.315	0.22
(b)						
(Intercept)	-0.030	-0.030	1.124	-2.234	2.174	
DRB heterozygosity	0.391	0.628	0.360	-0.078	1.334	0.62
Class I allele diversity	-0.145	-0.354	0.304	-0.949	0.241	0.41

^ Relative importance is the sum of Akaike weights for all models (ΔAICc < 7) containing the parameter

Table S6.11 The association of (a) presence or absence of alleles or (b) MHC genes heterozygosity and infection intensity of *Campylobacter* in badgers: Parameter estimates with their unconditional standard errors (SE), 95% confidence intervals and relative importance after model averaging (models with ΔAICc < 7). Estimates where the 95% confidence intervals do not overlap zero are in bold.

Parameter	Estimate (with shrinkage)	Estimate	Unconditional SE	95% Confidence interval		Relative importance^
(a)						
(Intercept)	**1.773**	**1.773**	**0.34**	**1.049**	**2.381**	
MHCI*03 (present)	-0.103	-0.221	0.127	-0.462	0.035	0.47
Season (Summer)[b]	**-0.367**	**-0.367**	**0.121**	**-0.596**	**-0.123**	1.00
Season (Autumn)[b]	-0.194	-0.194	0.117	-0.413	0.045	"
DRB*03 (present)	-0.070	-0.176	0.13	-0.433	0.077	0.40
MHCI*01 (present)	0.049	0.176	0.171	-0.182	0.487	0.28
MHCI*02 (present)	-0.105	-0.282	0.223	-0.707	0.168	0.37
DRB*01 (present)	-0.030	-0.125	0.187	-0.441	0.291	0.24
MHCI*04 (present)	0.006	0.028	0.131	-0.211	0.304	0.19
SH [a]	-0.547	-0.547	0.317	-1.182	0.061	1.00
DRB*04 (present)	-0.066	-0.202	0.189	-0.516	0.227	0.33
MHCI*07 (present)	-0.006	-0.036	0.268	-0.534	0.517	0.18
(b)						
(Intercept)	**2.055**	**2.055**	**0.459**	**1.005**	**2.734**	
Season (Summer)[b]	**-0.343**	**-0.343**	**0.121**	**-0.582**	**-0.110**	1.00
Season (Autumn)[b]	-0.186	-0.186	0.117	-0.409	0.049	"
SH [a]	-0.565	-0.565	0.31	-1.179	0.037	1.00
DRB heterozygosity	**-0.146**	**-0.216**	**0.109**	**-0.429**	**-0.005**	0.68
Class I allele diversity	-0.013	-0.051	0.107	-0.277	0.140	0.25

* Included in all models

^ Relative importance is the sum of Akaike weights for all models (ΔAICc < 7) containing the parameter

[a] Standardised microsatellite heterozygosity

[b] Spring was the reference category

Table S6.12 The association of (a) presence or absence of alleles or (b) MHC genes heterozygosity and prevalence (presence or absence) of fleas infection in badgers: Parameter estimates with their unconditional standard errors (SE), 95% confidence intervals and relative importance after model averaging (models with ΔAICc < 7). Estimates where the 95% confidence intervals do not overlap zero are in bold.

Parameter	Estimate (with shrinkage)	Estimate	Unconditional SE	95% Confidence interval		Relative importance^
(a)						
(Intercept)	**2.082**	**2.082**	**0.808**	**0.498**	**3.665**	
Age (adult)	**1.867**	**1.867**	**0.462**	**0.961**	**2.773**	1.00*
Season (Summer)[a]	**1.797**	**1.797**	**0.348**	**1.115**	**2.479**	1.00*
Season (Autumn)[a]	**1.727**	**1.727**	**0.412**	**0.92**	**2.534**	“
Season (Winter)[a]	-0.104	-0.104	0.681	-1.438	1.231	“
Weight/length ratio	**-0.263**	**-0.263**	**0.09**	**-0.439**	**-0.086**	1.00*
DRB*01 (present)	-0.177	-0.593	0.677	-1.92	0.734	0.30
MHCI*07 (present)	-0.080	-0.340	0.699	-1.709	1.029	0.23
MHCI*01 (present)	0.027	0.122	0.375	-0.612	0.857	0.22
DRB*03 (present)	-0.022	-0.099	0.327	-0.74	0.542	0.22
MHCI*04 (present)	-0.005	-0.025	0.366	-0.743	0.693	0.22
MHCI*02 (present)	-0.016	-0.073	0.468	-0.989	0.843	0.21
DRB*04 (present)	-0.020	-0.093	0.435	-0.946	0.76	0.22
MHCI*03 (present)	-0.002	-0.011	0.315	-0.627	0.606	0.21
(b)						
(Intercept)	**2.087**	**2.087**	**0.852**	**0.417**	**3.757**	
Age (adult)	**1.882**	**1.882**	**0.454**	**0.993**	**2.772**	1.00*
Season (Summer)[a]	**1.791**	**1.791**	**0.348**	**1.109**	**2.472**	1.00*
Season (Autumn)[a]	**1.724**	**1.724**	**0.410**	**0.920**	**2.529**	“
Season (Winter)[a]	-0.109	-0.109	0.679	-1.440	1.221	“
Weight/length ratio	**-0.264**	**-0.264**	**0.089**	**-0.439**	**-0.089**	1.00*
DRB heterozygosity	-0.058	-0.189	0.284	-0.745	0.367	0.30
Class I allele diversity	-0.016	-0.061	0.257	-0.565	0.443	0.27

[a] Spring was the reference category

* Included in all models

^ Relative importance is the sum of Akaike weights for all models (ΔAICc < 7) containing the parameter

Table S6.13 The association of (a) presence or absence of alleles or (b) MHC genes heterozygosity and infection intensity of fleas in badgers: Parameter estimates with their unconditional standard errors (SE), 95% confidence intervals and relative importance after model averaging (models with ΔAICc < 7). Estimates where the 95% confidence intervals do not overlap zero are in bold.

Parameter	Estimate (with shrinkage)	Estimate	Unconditional SE	95% Confidence interval		Relative importance^
(a)						
(Intercept)	**1.666**	**1.666**	**0.167**	**1.338**	**1.993**	
MHCI*04 (present)	0.100	0.189	0.123	-0.051	0.43	0.53
Season (Summer)[a]	**0.194**	**0.194**	**0.07**	**0.057**	**0.332**	1.00*
Season (Autumn)[a]	0.144	0.144	0.076	-0.004	0.292	"
Season (Winter)[a]	**-0.549**	**-0.549**	**0.272**	**-1.082**	**-0.016**	"
DRB*01 (present)	0.058	0.170	0.184	-0.192	0.531	0.34
DRB*03 (present)	0.032	0.101	0.12	-0.134	0.336	0.32
MHCI*02 (present)	0.058	0.171	0.174	-0.171	0.512	0.34
MHCI*07 (present)	0.039	0.144	0.229	-0.305	0.593	0.27
DRB*04 (present)	-0.020	-0.078	0.151	-0.375	0.218	0.26
MHCI*03 (present)	0.000	-0.001	0.117	-0.231	0.229	0.24
MHCI*01 (present)	-0.004	-0.019	0.126	-0.266	0.229	0.23
(b)						
(Intercept)	**1.643**	**1.643**	**0.234**	**1.184**	**2.102**	
Season (Summer)[a]	**0.195**	**0.195**	**0.070**	**0.058**	**0.332**	1.00*
Season (Autumn)[a]	0.143	0.143	0.076	-0.005	0.292	"
Season (Winter)[a]	**-0.547**	**-0.547**	**0.272**	**-1.079**	**-0.014**	"
Class I allele diversity	0.055	0.115	0.081	-0.044	0.273	0.48
DRB heterozygosity	0.009	0.034	0.102	-0.167	0.234	0.28

***** Included in all models

^ Relative importance is the sum of Akaike weights for all models (ΔAICc < 7) containing the parameter

[a] Spring was the reference category

Table S6.14 The association of (a) presence or absence of alleles or (b) MHC genes heterozygosity and prevalence (presence or absence) of lice infection in badgers: Parameter estimates with their unconditional standard errors (SE), 95% confidence intervals and relative importance after model averaging (models with ΔAICc < 7). Estimates where the 95% confidence intervals do not overlap zero are in bold.

Parameter	Estimate (with shrinkage)	Estimate	Unconditional SE	95% Confidence interval		Relative importance^
(a)						
(Intercept)	**3.130**	**3.130**	**0.68**	**1.797**	**4.463**	
Season (Summer) [a]	**0.655**	**0.655**	**0.269**	**0.128**	**1.183**	1.00*
Season (Autumn) [a]	0.484	0.484	0.305	-0.113	1.081	"
Season (Winter) [a]	0.912	0.912	0.733	-0.526	2.349	"
Sex (female)	-0.515	-0.515	0.242	-0.989	0.04	1.00*
Weight/length ratio	**-0.239**	**-0.239**	**0.054**	**-0.345**	**-0.132**	1.00*
DRB*04 (present)	0.264	0.560	0.418	-0.26	1.38	0.47
MHCI*02 (present)	0.068	0.262	0.449	-0.619	1.142	0.26
MHCI*01 (present)	-0.045	-0.175	0.378	-0.916	0.566	0.26
DRB*03 (present)	0.046	0.175	0.275	-0.363	0.714	0.26
MHCI*04 (present)	-0.020	-0.086	0.309	-0.691	0.52	0.23
DRB*01 (present)	-0.017	-0.077	0.469	-0.995	0.842	0.23
MHCI*03 (present)	0.005	0.023	0.272	-0.511	0.557	0.23
MHCI*07 (present)	0.024	0.104	0.602	-1.076	1.285	0.23
(b)						
(Intercept)	**2.861**	**2.861**	**0.798**	**1.297**	**4.424**	
Season (Summer) [a]	**0.657**	**0.657**	**0.269**	**0.130**	**1.184**	1.00*
Season (Autumn) [a]	0.484	0.484	0.304	-0.111	1.079	"
Season (Winter) [a]	0.902	0.902	0.734	-0.538	2.342	"
Sex (female)	**-0.511**	**-0.511**	**0.241**	**-0.982**	**-0.039**	1.00*
Weight/length ratio	**-0.239**	**-0.239**	**0.054**	**-0.345**	**-0.133**	1.00*
DRB heterozygosity	0.111	0.278	0.244	-0.199	0.755	0.4
Class I allele diversity	0.008	0.028	0.218	-0.400	0.456	0.27

* Included in all models

^ Relative importance is the sum of Akaike weights for all models (ΔAICc < 7) containing the parameter

[a] Spring was the reference category

Table S6.15 The association of (a) presence or absence of alleles or (b) MHC genes heterozygosity and infection intensity of lice in badgers: Parameter estimates with their unconditional standard errors (SE), 95% confidence intervals and relative importance after model averaging (models with ΔAICc < 7). Estimates where the 95% confidence intervals do not overlap zero are in bold.

Parameter	Estimate (with shrinkage)	Estimate	Unconditional SE	95% Confidence interval		Relative importance^
(a)						
(Intercept)	1.028	1.028	0.073	0.910	1.168	
Sex (female)	**-0.086**	**-0.086**	**0.027**	**-0.137**	**-0.033**	1.00*
Weight/length ratio	**-0.033**	**-0.033**	**0.006**	**-0.044**	**-0.022**	1.00*
MHCI*02 (present)	0.026	0.061	0.048	-0.026	0.156	0.42
MHCI*04 (present)	0.014	0.038	0.034	-0.020	0.101	0.37
MHCI*03 (present)	-0.007	-0.025	0.033	-0.079	0.027	0.30
MHCI*01 (present)	-0.007	-0.026	0.039	-0.090	0.041	0.29
DRB*01 (present)	0.010	0.035	0.054	-0.055	0.138	0.28
DRB*04 (present)	0.007	0.027	0.045	-0.059	0.095	0.28
MHCI*07 (present)	-0.008	-0.033	0.067	-0.168	0.087	0.25
DRB*03 (present)	0.001	0.005	0.036	-0.065	0.051	0.24
(b)						
(Intercept)	**1.040**	**1.040**	**0.075**	**0.916**	**1.169**	
Sex (female)	**-0.085**	**-0.085**	**0.027**	**-0.137**	**-0.032**	1.00*
Weight/length ratio	**-0.033**	**-0.033**	**0.006**	**-0.044**	**-0.022**	1.00*
DRB heterozygosity	0.003	0.013	0.027	-0.041	0.063	0.27
Class I allele diversity	-0.002	-0.007	0.024	-0.051	0.039	0.26

* Included in all models

^ Relative importance is the sum of Akaike weights for all models (ΔAICc < 7) containing the parameter

Table S6.16 The association of (a) presence or absence of alleles or (b) MHC genes heterozygosity and co-infection status in badgers: Parameter estimates with their unconditional standard errors (SE), 95% confidence intervals and relative importance after model averaging (models with ΔAICc < 7). Estimates where the 95% confidence intervals do not overlap zero are in bold.

Parameter	Estimate (with shrinkage)	Estimate	Unconditional SE	95% Confidence interval		Relative importance^
(a)						
(Intercept)	1.806	1.806	0.097	1.616	1.995	
MHCI*03 (present)	-0.051	-0.117	0.093	-0.300	0.066	0.43
DRB*04 (present)	0.027	0.095	0.118	-0.136	0.326	0.29
MHCI*01 (present)	0.013	0.052	0.108	-0.160	0.265	0.25
MHCI*07 (present)	-0.031	-0.115	0.17	-0.448	0.218	0.27
DRB*03 (present)	0.000	0.001	0.111	-0.216	0.218	0.24
DRB*01 (present)	0.014	0.059	0.128	-0.192	0.309	0.24
MHCI*02 (present)	0.004	0.018	0.162	-0.299	0.336	0.22
MHCI*04 (present)	0.000	-0.001	0.098	-0.193	0.191	0.22
(b)						
(Intercept)	1.832	1.832	0.171	1.497	2.167	
Class I allele diversity	-0.025	-0.068	0.067	-0.200	0.064	0.36
DRB heterozygosity	0.011	0.040	0.078	-0.112	0.192	0.28

^ Relative importance is the sum of Akaike weights for all models (ΔAICc < 7) containing the parameter

Chapter 7

Discussion

\

7.1 Abstract

This thesis has focused on the evolution of MHC, in particular investigating whether MHC-based mate choice and pathogen resistance occurred in a population of wild mammals, the European badger. I examined the evolutionary history of both major histocompatibility complex (MHC) class I and class II genes. Phylogenetic analysis evidenced *trans*-species polymorphism of class II genes. For class I genes I found evidence for separate evolutionary histories of the α1 and α2/α3 domains, with orthology or *trans*-species polymorphism of the α1 domain and concerted evolution of the α2/α3 domains. There was an intermediate level of diversity for both the MHC class I genes and class II DRB genes, which showed evidence of balancing selection. There was limited variability in these genes and it was possible to determine the number of loci present; consequently they proved useful for investigating questions about mate choice and pathogen resistance. A high number of bacterial species (56 operational taxonomic units) was found in the subcaudal gland secretion, which has been shown to encode individual-specific information in its chemical profile. These bacteria communities differed between cubs and adults. MHC genotype may influence the microbiota and chemical profile of subcaudal gland secretion, modifying the animals scent profile in a way that may cue its MHC constitution to potential mates. MHC genes have the potential to act as the target of mate choice and provide genetic benefits with respect to pathogen resistance. I thus investigated whether females showed a MHC-based preference towards breeding with compatible males (genetic compatibility hypothesis). Neighbouring-group parent-pairs had higher MHC similarity (based on class II DRB genes) compared to simulated random neighbouring-group pairs. MHC heterozygosity had no influence on the co-infection status; however, four MHC alleles associated with specific pathogen resistance and susceptibility for two pathogen species, providing evidence of genetic benefits from MHC genes. This study, using long-term data from a wild population, provides a valuable contribution to our understanding of MHC evolution with respect to mate choice and pathogen-mediated selection as there are few similar studies in wild mammals and unique for the superfamily Musteloidea. The results revealed the evolutionary process that underlie MHC gene diversity and how contemporary selection acts on the MHC genes, which increases our knowledge of how individual genetic constitution may provide indirect genetic benefits to mate choice.

7.2 Mate choice and immunity

The Red Queen hypothesis (vanValen, 1973) proposes that sexual reproduction provides the genetic diversity necessary to counter the rapid adaptations of pathogens during the coevolutionary arms race. Under this hypothesis, individuals are expected to choose mates to maximize the adaptive benefit of sexual reproduction, with respect to the evolutionary interaction between pathogens and their hosts (Hamilton & Zuk, 1982; Hamilton *et al.*, 1990).

It has been proposed that indirect genetic benefits, resulting via offspring genotype, are important drivers of mate choice (Kokko *et al.*, 2003; Consuegra & Garcia-de-Leaniz, 2008). The Compatibility Hypothesis predicts that a female should select a mate whose genes, in combination with her own genes, provide the offspring with the best immune response against pathogens. This requires the following assumptions (Milinski, 2006): (a) there is genetic variation in the population; (b) females can recognize the immune genes of potential mates; and (c) females "know" their immune genes. These assumptions appear to be fulfilled in the case of the major histocompatibility complex (MHC) genes in vertebrates (Penn & Potts, 1999; Janeway *et al.*, 2005; Milinski, 2006).

The MHC is a cluster of genes involved in the immune response, and MHC differences between individuals are associated with differences in pathogen resistance and auto-immune disease (Apanius *et al.*, 1997). The MHC of an individual may be distinguishable through their odour. Individual odour is created through production of soluble proteins or proteins which bind volatile molecules, and/or through influencing bacterial community (Yamazaki *et al.*, 1990; Toivanen *et al.*, 2001; Leinders-Zufall *et al.*, 2004; Lanyon *et al.*, 2007). Therefore, MHC genes may serve as a basis for individual recognition and mate choice. In this thesis, I investigated MHC evolution and how the MHC may affect mate choice and pathogen resistance in a high-density population of group-living badgers; for which we had the very unusual advantage of knowing their genetic relatedness and detailed life histories (Macdonald & Newman, 2002; Dugdale *et al.*, 2007; Annavi et al. in prep; Dugdale *et al.*, 2008; Macdonald *et al.*, 2008).

7.3 MHC diversity and evolution

MHC genes have extremely high levels of diversity at three levels: (a) number of gene loci, (b) number of alleles in a locus, and (c) divergence between sequences. The nucleotide diversity within MHC genes is generally attributed to balancing selection (Bergström & Gyllensten, 1995), which maintains large numbers of alleles in populations. Balancing selection greatly enhances the persistence of allelic lineages over long periods of time (Richman, 2000), even before species diverged, which is known as trans-species polymorphism (Klein, 1987). For example, some mammalian MHC allelic lineages are more than a million years old as they were maintained after speciation events (Figueroa *et al.*, 1988). In contrast to class II genes, class I genes tend to form monophyletic groups within species and therefore it is difficult to establish orthologous relationships among closely related mammalian species (Takahashi *et al.*, 2000). Plausible explanations are that class I loci undergo: (1) a faster rate of birth-and-death evolution than do class II loci (Nei *et al.*, 1997; Vogel *et al.*, 1999; Takahashi *et al.*, 2000; Piontkivska & Nei, 2003), or (2) concerted evolution that yields genes within a species that are more similar to each other than they are to the orthologs in closely related species (Rada *et al.*, 1990; Hess & Edwards, 2002; Nei & Rooney, 2005).

To date, most studies have focused on a small portion of the MHC (e.g. class II genes only), which may not be representative of the MHC as a whole. I characterized both MHC class I and class II genes and combined them in the analyses to provide a comprehensive overview. Based on the differences between mustelids and other carnivores, inferences can be made on MHC evolution in carnivores. The characterization of the badger MHC therefore not only provides insight into the evolutionary processes shaping the badger MHC, but may also contribute to our understanding of the evolution of the mammalian MHC. Phylogenetic reconstructions revealed trans-species polymorphism in the badger class II genes (Chapter 2 - Sin *et al.*, 2012b), while phylogenetic inconsistency between the α1 and α2/α3 domains of the badger class I genes indicates that these domains have different evolutionary histories (Chapter 3 - Sin *et al.*, 2012c). Concerted evolution provides a more plausible hypothesis than birth-and-death evolution in the α2 and α3 domains, on the other hand the α1 domain is likely to be subject to balancing selection (Nei *et al.*, 1997).

MHC sequence characteristics can also be used to infer the selection processes acting on the MHC within a species. The MHC class II genes in mammals are divided into gene clusters such as the *DR* and *DQ* genes, which I had examined in Chapter 2 (Sin *et al.*, 2012b). The *DRB* gene in badgers had the highest variability compared to other class II genes (i.e. *DQB*, *DQA* and *DRA*), and was moderately variable in comparison to closely related species. The high d_N/d_S ratio and positive selection detected among antigen-binding site (ABS) and trans-species polymorphism discovered indicate balancing selection to be the potential force that maintains genetic diversity. The high genetic diversity, high d_N/d_S ratio and positively selected sites of the α1 domain of the badger class I genes were also potentially maintained by balancing selection (Chapter 3 - Sin *et al.*, 2012c), which could be driven by sexual selection and pathogen-mediated selection.

The high degree of sequence diversity among alleles were due principally to the diverse exons that encode the domains forming the ABS, both in badgers and in general (Hughes & Yeager, 1998). The ABS is responsible for peptide binding and plays an important role in immune responses, therefore it is under selective pressure from pathogens (Spurgin & Richardson, 2010). Alternatively, but not mutually exclusive, the peptide-binding property of the ABS may influence body odour and hence individual recognition, which if detectable may be used in mate choice (Yamazaki *et al.*, 1982; Brennan & Kendrick, 2006).

The identified badger MHC sequences were confirmed to be putatively functional (seven for class I genes and four for class II DRB genes) by examining whether they were transcribed or not. The gene-specific primers developed, and the knowledge about badger MHC genes obtained in these studies (e.g. intermediate variability and number of loci), facilitated investigations between MHC and mate choice or pathogen resistance in subsequent chapters. To answer questions regarding the selection mechanisms on MHC genes of badgers, a method to quickly and accurately genotype individuals was required. Chapter 5 describes how I developed a high throughput, high resolution and inexpensive genotyping method, the Reference Strand-mediated Conformation Analysis (RSCA), to successfully screen >1100 individuals for the MHC class I genes and class II DRB genes. The RSCA method

allowed me to screen individuals for all the identified putatively functional sequences for both class I and class II DRB genes, and to identify any previously unidentified sequences.

7.4 Subcaudal gland secretion: Bacterial community and scent profile

MHC genes can influence odour and mating preferences (e.g. in mice; Penn & Potts, 1998b; Penn & Potts, 1999; Penn, 2002). The microflora hypothesis suggests that MHC genes control odour indirectly, by influencing the growth of commensal microbiota that contribute to variation in individual odour (Yamazaki *et al.*, 1990; Toivanen *et al.*, 2001; Lanyon *et al.*, 2007). However, studies examining the odour of mice reared in germ-free environments have produced mixed results (Yamazaki *et al.*, 1990). Whether MHC genes influence odour by controlling MHC-dependent microbiota remains unsolved.

The usage of odour is an important way of communication in mammals, as scent marking is common and can convey information about the marking individual (Brown & Macdonald, 1985). The chemical profile of scent marks is likely to be affected by bacteria metabolizing the primary gland products (Albone *et al.*, 1974; Albone & Perry, 1975), hence bacteria are proposed to play a role in scent marking. The variation in bacterial communities between different individuals may underpin olfactory communication and individual recognition (fermentation hypothesis; Albone *et al.*, 1974; Albone & Perry, 1975; Gorman, 1976). Badgers use their unique subcaudal gland, which produces a paste-like secretion that is stored in the subcaudal pouch (Stubbe, 1971), to scent marking objects (Buesching & Macdonald, 2004) or conspecifics (Buesching *et al.*, 2003). The lower molecular weight components of the odour are thought to be by-products from bacteria metabolizing the primary gland products (Kruuk *et al.*, 1984), and the time-related changes in the chemical composition of the secretion are believed to be the results of chemical and microbial degradation (Buesching *et al.*, 2002b). The microbial communities inhabiting the subcaudal pouch are thus likely to play an important role in generating the actual odour of the subcaudal gland secretion.

I characterised the microbial communities in the subcaudal gland secretion (Chapter 4) and correlated the microbiota with different individual-specific parameters (age, sex, body condition, reproductive

status, and season). I discovered a high number of bacterial species (56 operational taxonomic units from four phyla: Actinobacteria, Firmicutes, Proteobacteria, and Bacteroidetes), dominated by Actinobacteria (76.0%). The bacterial communities of cubs and adults were different; cubs possessed considerably more diverse communities dominated by Firmicutes while adults had less diverse communities dominated by Actinobacteria. Some Actinobacteria play a major role in the transformation of odourless steroids into odourous derivatives (Gower *et al.*, 1986; Kohl *et al.*, 2001), the dominance of Actinobacteria is thus likely contribute to the actual odour of the secretion.

My collaborative work, using gas chromatography–mass spectrometry (GC-MS), investigated whether that the subcaudal gland secretion encodes information about group membership and individuality (Chapter 4 & Appendix 2) as has been previously suggested (Gorman *et al.*, 1984; Buesching *et al.*, 2002c). Although there was a correlation between some bacteria in the subcaudal pouch and the chemical composition of the subcaudal gland secretion (Appendix 2), bacterial communities did not differ between social groups (Appendix 2) or other parameters (except for age group and spring body condition; Chapter 4). This could be because the bacteria communities are not responsible for generating the odour encoding this information, or the method (i.e. terminal restriction fragment length polymorphism; Liu *et al.*, 1997) used for determining bacterial community has limited resolution. Alternatively, other factors such as kinship (Weisfeld *et al.*, 2003) and host genotypes (Kuhn & Natsch, 2009), particularly the MHC (Lanyon *et al.*, 2007; Zomer *et al.*, 2009), may have a greater influence on shaping the community of the symbiotic bacteria. If MHC genes influence the scent profile of the subcaudal gland secretion, this would underpin the mechanism of MHC-based mate choice in this species.

7.5 Mate choice based on the MHC

The genetic diversity at the MHC is maintained by balancing selection (see above), which comprises different mechanisms such as pathogen-mediated selection and MHC-based mate choice (Potts & Wakeland, 1990; Hedrick, 1994; Hughes & Hughes, 1995). In the case of pathogen-mediated selection, balancing selection is caused by the coevolutionary arms race between hosts and pathogens,

which maintains MHC diversity. Mating preference may further enhance the pathogen resistance of an individual's progeny during the same arms race (Hamilton & Zuk, 1982; Hamilton *et al.*, 1990). Different forms of mating preference, such as preference for good genes and compatible genes (Trivers, 1972; Tregenza & Wedell, 2000; Neff & Pitcher, 2005), is predicted to result in balancing selection, as they increase the likelihood that alleles are maintained in the population rather than being lost.

The genetic compatibility hypothesis predicts that females choose compatible mates, by selecting males with particular alleles with reference to the alleles possessed by the female (Penn & Potts, 1998a; Mays & Hill, 2004). If the parents have genes that are more compatible, the offspring are expected to have higher fitness than those that have less compatible parents (Zeh & Zeh, 1996). Genetic compatibility is non-additive genetic variance in fitness (Puurtinen *et al.*, 2005), which is where variation is caused by interactions between genes. The variation in fitness of compatible genes would not respond to directional selection due to this interactive effect, but the mechanism (e.g. mate preference) of acquiring compatible genes would (Neff & Pitcher, 2005).

As MHC genes are codominant, MHC heterozygous individuals are predicted to show higher fitness than homozygotes. Selection could therefore favour MHC-dependent mate choice if offspring MHC heterozygosity is associated with fitness benefits. MHC-disassortative mating could maximize the MHC heterozygosity of offspring, either as a mechanism to maximize the number of MHC alleles in offspring or inbreeding avoidance (Penn & Potts, 1999). In Chapter 5 I investigated whether MHC-based mate choice occurred by comparing the MHC allele sharing and amino acid distance (a measure of functional difference) of the assigned parental pairs to those of simulated pairs that were randomly generated by pairing up the breeding females with potential mates from certain groups in the population. Genetic backgrounds were also examined, by genotyping 35 microsatellite loci, to investigate the possible association between MHC-based mating and inbreeding avoidance (Penn & Potts, 1999).

In contrast to MHC-disassortative mating, which predicts preference for MHC-dissimilar mates, badgers were more likely to breed with neighbouring-group mates that had a more similar MHC-genotype than random neighbouring-group mates (Chapter 5). This overall effect was found in MHC class II DRB genes. This is different to the paradigm of MHC-disassortative mating that has been reported in other species (Landry *et al.*, 2001; Miller *et al.*, 2009; Juola & Dearborn, 2012), though MHC-assortative mating has also been reported in mice (Yamazaki *et al.*, 1978), humans (Roberts *et al.*, 2005) and tiger salamanders *Ambystoma tigrinum* (Bos *et al.*, 2009).

Puurtinen *et al*. (2005) pointed out that compatibility does not equal dissimilarity. The highest level of fitness (compatibility) is not necessarily achieved by possessing the highest level of dissimilarity (Puurtinen *et al.*, 2005). The lack of preference for MHC-dissimilar individuals suggests a possible lack of reproductive advantage for MHC-divergent pairs. Possible disadvantages that arise in offspring whose parents have highly dissimilar MHC, are increased loss of T-cells (Lawlor *et al.*, 1990; Nowak *et al.*, 1992), increased risk of autoimmune diseases and / or disruption of co-adapted gene complex (Kaufman, 1999; Puurtinen *et al.*, 2005).

For group-living mammals the chances of meeting mating partners is highest within the same social group. Based on the assumption that extra-group offspring are generally the result of preferred matings (i.e. females of socially monogamous species are free in their choice of extra-pair partners; Kempenaers *et al.*, 1992; Kempenaers *et al.*, 1997), difference between within- and extra-group mate partners could be used as evidence for mate choice for genetic benefits (Kempenaers *et al.*, 1997; Foerster *et al.*, 2003; Dreiss *et al.*, 2008). In the examined badger population, MHC-assortative mating was identified for neighbouring-group mating, considering all potential neighbouring-group mates (Chapter 5). Neighbouring-group pairs showed an effect of MHC-assortative mate choice, which suggests that females may be able to seek MHC-similar neighbouring-group males. I speculated that, due to their reproductive biology (females may conceive throughout the year, and implant in December, due to delayed implantation, which means they can conceive while being pregnant [superfoetation] – they have oestrus cycles, but may also be induced ovalutors – they may reabsorb

implanted embryos), badgers may be able to select for compatible mates through pre- and/or post-copulatory mate choice, even though they are promiscuous.

By the examination of the genome-wide background variation (of within-group and neighbouring group mates compared with within-group and neighbouring group candidate mates, respectively) I demonstrated that in the years in which MHC-based mate choice occurred, there was no associated mate-choice based on microsatellite heterozygosity (Chapter 4). This indicated that the MHC-assortative mate choice that was identified did not arise as a result of mating between close relatives, therefore MHC genes were the actual target of selection.

The inbreeding avoidance hypothesis predicts that MHC-dependent mating preferences may function to avoid kin matings and deleterious consequences of inbreeding (Penn & Potts, 1999). We found negative correlation between MHC class I genes amino acid distance and relatedness at microsatellite loci, which suggests that MHC class I genes-based mating preferences could be used as a mechanism to avoid inbreeding, by selecting mates that were more distant in class I genes amino acid distance to avoid mating with close relatives.

7.6 MHC and pathogen burden

Indirect genetic benefits, with respect to pathogen resistance, are proposed to be important drivers of MHC-associated mate choice. A female is predicted to select males with MHC genes that would, in combination with her own genes, provide the offspring with the best immune response against pathogens (Mays & Hill, 2004; Puurtinen *et al.*, 2005). In this thesis, I investigated the associations between the MHC and pathogen burden of an organism to detect contemporary selection on MHC genes (Spurgin & Richardson, 2010). The MHC plays a pivotal role in the immune system, and it has long been suggested that pathogen-mediated selection is the force driving and maintaining the diversity of MHC genes (Potts & Wakeland, 1990; Hedrick, 1994; Hughes & Hughes, 1995).

I examined the association between the MHC and pathogen burden, using a wide variety of pathogen taxa (Chapter 6). The co-infection status was also considered, since the relationship between MHC and pathogens can be tested properly only by examining multiple pathogens simultaneously, as the advantage of being able to recognize and bind a wider range of antigens may only be realized when the host is challenged by multiple infections (Hughes & Nei, 1992; McClelland *et al.*, 2003). Associations between specific pathogens and particular MHC alleles were discovered for both class I and class II alleles, with two alleles conferred quantitative resistance and two alleles conferred qualitative or quantitative susceptibility to pathogen(s). However, MHC heterozygotes did not possess a lower pathogen burden with respect to their co-infection status (Chapter 6), even though thirteen pathogenic species were examined. Although there was weak evidence of a heterozygote advantage for individual pathogen infections, it may be due to the presence of specific alleles. This indicates a greater role for a rare-allele advantage and/or fluctuating selection, than a heterozygote advantage in this system.

The Rare-allele advantage and the fluctuating selection hypotheses have the same predictions on associations between the MHC and pathogens, thus this study is not able to distinguish between these two mechanisms (Spurgin & Richardson, 2010). Other evolutionary processes, such as sexual selection, also contribute towards shaping MHC diversity in addition to pathogen-mediated selection (Penn & Potts, 1999). MHC-based mate choice may operate together with pathogen-mediated selection to achieve an optimal MHC constitution (Eizaguirre *et al.*, 2009). If MHC heterozygotes are more resistant to pathogens than homozygotes, then MHC-disassortative mating preferences will subsequently be favoured as a mechanism to produce MHC-heterozygous offspring. If MHC diversity is maintained by rare-allele advantages, then MHC-disassortative mate choice will provide a moving target to pathogens that evade immune recognition by producing progeny that are MHC-dissimilar from their parents rather than heterozygous *per se* (Penn & Potts, 1999). This means that if MHC diversity is maintained by pathogen-mediated selection through a heterozygote and/or rare-allele advantage, then MHC-disassortative mate choice will also be favoured by selection (Penn & Potts, 1999).

However, in contrast to MHC-disassortative mate choice, an MHC-assortative mating preference, which does not maximize heterozygosity, was found in this population (Chapter 5). This is actually in line with the finding that no (or weak) heterozygote advantage had been discovered (Chapter 6). Possible explanations of MHC-assortative rather than MHC-disassortative mate choice are: (1) there were no heterozygote and rare-allele advantages detected, thus MHC-disassortative mate choice was not favoured by selection (hence the MHC allele and pathogen associations were due to fluctuating selection); or (2) there was a rare-allele advantage but selective pressure for MHC-disassortative mate choice was outweighed by potential benefits of MHC-assortative mate choice (see the earlier section "Mate choice based on the MHC"). Although pathogen-mediated selection was likely to be one of the mechanisms maintaining MHC diversity in badgers, this system was different to the paradigm that heterozygote and/or rare-allele advantage lead to the evolution of MHC-disassortative mate choice (Penn & Potts, 1999).

7.7 Conclusions

This is the first published study of MHC and mate choice in wild carnivores, combining information from both MHC class I and class II genes to provide comprehensive results. Characterization of MHC genes in the European badger and development of the RSCA allowed genotyping of a large number of individuals. The knowledge we have on the study species allowed me to investigate how the genetic constitution of an individual may provide indirect genetic benefits. In this thesis I had demonstrated that the MHC genes of badgers were potentially under balancing selection. As the MHC is primarily responsible for immune responses, I investigated its association with pathogen burden and I identified a rare-allele advantage or fluctuating selection, but not a heterozygote advantage, to be the major pathogen-mediated selective force. The role that MHC genes play in the immune system and their potential to act as the target for indirect fitness during mate choice is the focus of this thesis. MHC-assortative mate choice is evidence that females preferred MHC-similar neighbouring-group mates over other neighbouring-group candidate mates. The results from this study differ from MHC-disassortative mating that is usually detected and does not demonstrate MHC heterozygote advantage with respect to co-infection as well. I highlighted that long-term studies of wild populations and detailed knowledge on the animal's mating system are essential to improve understanding of MHC

evolution and how the MHC can affect fitness of an animal. The results presented in this thesis are in line with the growing evidence that reveals a complicated and context-dependent pattern of MHC-based reproduction rather than simple MHC-disassortative mating. Therefore it is important to study these processes in natural systems that experience variations. The findings revealed evolutionary process and contemporary selection on the MHC genes, which increase our knowledge of how the genetic constitution may provide indirect benefit and affect the behaviour of an animal.

7.8 Future research

My work has shed much light on MHC-based mate choice in badgers, which in turn has opened up further potential avenues of research, and here I briefly mention the specific questions on MHC and odour that lead on from this thesis.

The MHC and olfactory cues

MHC genotypes might be recognizable through olfactory cues (Yamazaki *et al.*, 1982; Brennan & Kendrick, 2006) and underpin individual recognition and mate choice. Several mechanisms have been proposed for how MHC genes influence odour: (1) the peptide hypothesis suggests that MHC genes influence odour directly through MHC molecules or by altering the pool of peptides in the body (Leinders-Zufall *et al.*, 2004); (2) the microflora hypothesis suggests that MHC genes influence the bacterial community that contributes to variation in individual odour (Yamazaki *et al.*, 1990; Toivanen *et al.*, 2001; Lanyon *et al.*, 2007); and (3) the peptide-microflora hypothesis suggests that MHC-bound peptides provide the precursors that are metabolised and made volatile by commensal microflora (Zomer *et al.*, 2009). There is still no consensus on how MHC genes influence odour.

Badgers are known to exchange information about individual specific parameters through subcaudal gland scent (Buesching *et al.*, 2002c) (Appendix 2), which may also include information on the MHC genotype of the host. The bacterial community of the subcaudal gland secretion was shown to be highly diverse (Chapter 4) and some bacteria were correlated to chemical composition of the secretion

(Appendix 2). Future study could investigate whether the bacterial community and scent chemical profile are associated with particular MHC genes or combination of genes, to determine whether and how the MHC influence odour, and thus to elucidate the role that MHC genes may play in pre-copulatory MHC-based mate choice in this species.

References

References

Ackerman, A.L. & Cresswell, P. 2004. Cellular mechanisms governing cross-presentation of exogenous antigens. *Nature Immunology* **5:** 678–684.

Akcay, E. & Roughgarden, J. 2007. Extra-pair paternity in birds: review of the genetic benefits. *Evolutionary Ecology Research* **9:** 855-868.

Alberts, S.C. & Ober, C. 1993. Genetic variability of the MHC: a review of non-pathogen mediated selective mechanisms. *Yearbook of physical anthropology* **36:** 71-89.

Albone, E.S., Eglinton, G., Walker, J.M. & Ware, G.C. 1974. The anal sac secretion of red fox (*Vulpes vulpes*); its chemistry and microbiology. A comparison with the anal sac secretion of the lion (*Panthera leo*). *Life Sciences* **14:** 387-400.

Albone, E.S. & Perry, G.C. 1975. Anal sac secretion of the red fox, *Vulpes vulpes*; volatile fatty acids and diamines: Implications for a fermentation hypothesis of chemical recognition. *Journal of Chemical Ecology* **2:** 101–111.

Albone, E.S. & Shirley, S.G. 1984. *Mammalian Semiochemistry: The Investigation of Chemical Signals between Mammals*. Wiley, Chichester.

Aldridge, B.M., Bowen, L., Smith, B.R., Antonelis, G.A., Gulland, F. & Stott, J.L. 2006. Paucity of class I MHC gene heterogeneity between individuals in the endangered Hawaiian monk seal population. *Immunogenetics* **58:** 203–215.

Alexy, K., Gassett, J., Osborn, D., Miller, K. & Russell, S. 2003. Bacterial fauna of the tarsal tufts of white-tailed deer (*Odocoileus virginianus*). *American Midland Naturalist* **149:** 237–240.

Allen, A.R., Minozzi, G., Glass, E.J., Skuce, R.A., McDowell, S.W.J., Woolliams, J.A. & Bishop, S.C. 2010. Bovine tuberculosis: the genetic basis of host susceptibility. *Proceedings of the Royal Society B: Biological Sciences* **277:** 2737–2745.

Alonso, M.C., Navarrete, C., Solana, R., Torres, A., Pena, J. & Festenstein, H. 1985. Differential expression of HLA-DR and HLA-DQ antigens on normal cells of the myelomonocytic lineage. *Tissue Antigens* **26:** 310-317.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215:** 403–410.

Amann, R.I., Ludwig, W. & Schleifer, K.H. 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological Reviews* **59:** 143–169.

Anderson, D.R. 2008. *Model-based inference in the life sciences: a primer on evidence*, 2nd edn. Springer.

Anderson, R.M. & Trewhella, W. 1985. Population dynamics of the badger (*Meles meles*) and the epidemiology of bovine tuberculosis (*Mycobacterium bovis*). *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **310:** 327-381.

Andersson, M. 1994. *Sexual Selection*. Princeton University Press.

Andersson, M.B. & Simmons, L.W. 2006. Sexual selection and mate choice. *Trends in Ecology & Evolution* **21:** 296–302.

Anisimova, M., Bielawski, J.P. & Yang, Z. 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Molecular Biology and Evolution* **18:** 1585–1592.

Anisimova, M., Bielawski, J.P. & Yang, Z. 2002. Accuracy and power of Bayes prediction of amino acid sites under positive selection. *Molecular Biology and Evolution* **19:** 950–958.

Annavi, G., Dawson, D.A., Horsburgh, G.J., Greig, C., Dugdale, H.L., Newman, C., Macdonald, D.W. & Burke, T. 2011. Characterisation of twenty-one European badger (*Meles meles*) microsatellite loci facilitates the discrimination of second-order relatives. *Conservation Genetics Resources* **3:** 515–518.

Annavi, G., Newman, C., Buesching, C.D., Macdonald, D.W., Burke, T. & Dugdale, H.L. subm. Heterozygosity-fitness correlations in a wild mammal population: single locus, paternal and environmental effects.

Anwar, A., Knaggs, J., Service, K.M., McLaren, G.W., Riordan, P., Newman, C., Delahay, R.J., Cheesman, C. & Macdonald, D.W. 2006. Antibodies to *Toxoplasma gondii* in Eurasian badgers. *Journal of Wildlife Diseases* **42:** 179–181.

Anwar, M., Newman, C., Macdonald, D., Woolhouse, M. & Kelly, D. 2000. Coccidiosis in the European badger (*Meles meles*) from England, an epidemiological study. *Parasitology* **120:** 255–260.

Apanius, V., Penn, D., Slev, P.R., Ruff, L.R. & Potts, W.K. 1997. The nature of selection on the major histocompatibility complex. *Critical reviews in immunology* **17:** 179–224.

Archie, E.A. & Theis, K.R. 2011. Animal behaviour meets microbial ecology. *Animal Behaviour* **82:** 425-436.

Babik, W., Pabijan, M., Arntzen, J.W., Cogalniceanu, D., Durka, W. & Radwan, J. 2009. Long-term survival of a urodele amphibian despite depleted major histocompatibility complex variation. *Molecular Ecology* **18:** 769–781.

Barton, K. 2009. MuMIn: multi-model inference. R package, version 0.12.2. http://r-forge.r-project.org/projects/mumin/.

Bates, D. & Maechler, M. 2009. lme4: Linear mixed-effects models using S4 classes. R package version 0.999375–31. http://CRAN.R-project.org/package=lme4.

Beauchamp, G.K., Yamazaki, K., Bard, J. & Boyse, E.A. 1988. Preweaning experience in the control of mating preferences by genes in the major histocompatibility complex of the mouse. *Behavior Genetics* **18:** 537–547.

Beck, S., Geraghty, D., Inoko, H. & Rowen, L. 1999. Complete sequence and gene map of a human major histocompatibility complex. *Nature* **401:** 921–923.

Becker, L., Nieberg, C., Jahreis, K. & Peters, E. 2009. MHC class II variation in the endangered European mink *Mustela lutreola* (L. 1761)—consequences for species conservation. *Immunogenetics* **61:** 281–288.

Bent, S.J., Pierson, J.D. & Forney, L.J. 2007. Measuring species richness based on microbial community fingerprints: the emperor has no clothes. *Applied and Environmental Microbiology* **73:** 2399–2401.

Bergström, T. & Gyllensten, U. 1995. Evolution of Mhc class II polymorphism: the rise and fall of class II gene function in primates. *Immunological Reviews* **143:** 13–31.

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.

Bevan, M.J., Hogquist, K.A. & Jameson, S.C. 1994. Selecting the T-cell repertoire. *Science* **264:** 796–797.

Bilham, K., Sin, Y.W., Newman, C., Buesching, C.D. & Macdonald, D.W. 2013. An example of life history antecedence in the European badger (*Meles meles*): rapid development of juvenile

antioxidant capacity, from plasma vitamin E analogue. *Ethology Ecology & Evolution***:** DOI:10.1080/03949370.03942013.03767861.

Bininda-Emonds, O.R.P., Gittleman, J.L. & Purvis, A. 1999. Building large trees by combining phylogenetic information: a complete phylogeny of the extant Carnivora (Mammalia). *Biological Reviews* **74:** 143–175.

Birkhead, T.R. & Pizzari, T. 2002. Postcopulatory sexual selection. *Nature Reviews Genetics* **3:** 262–273.

Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L. & Wiley, D.C. 1987. The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature* **329:** 512–518.

Bjorkman, P.J. & Parham, P. 1990. Structure, function, and diversity of class I major histocompatibility complex molecules. *Annual Review of Biochemistry* **59:** 253–288.

Bollmer, J.L., Dunn, P.O., Freeman-Gallant, C.R. & Whittingham, L.A. 2012. Social and extra-pair mating in relation to major histocompatibility complex variation in common yellowthroats. *Proceedings of the Royal Society B* **279:** 4778–4785

Bonneaud, C., Chastel, O., Federici, P., Westerdahl, H. & Sorci, G. 2006a. Complex Mhc-based mate choice in a wild passerine. *Proceedings of the Royal Society B* **273:** 1111–1116.

Bonneaud, C., Perez-Tris, J., Federici, P., Chastel, O. & Sorci, G. 2006b. Major histocompatibility alleles associated with local resistance to malaria in a passerine. *Evolution* **60:** 383–389.

Borghans, J.A.M., Noest, A.J. & DeBoer, R.J. 2003. Thymic selection does not limit the individual MHC diversity. *European Journal of Immunology* **33:** 3353–3358.

Bos, D.H. & Waldman, B. 2006. Evolution by recombination and transspecies polymorphism in the MHC class I gene of *Xenopus laevis*. *Molecular Biology and Evolution* **23:** 137–143.

Bos, D.H., Williams, R.N., Gopurenko, D., Bulut, Z. & Dewoody, J.A. 2009. Condition-dependent mate choice and a reproductive disadvantage for MHC-divergent male tiger salamanders. *Molecular Ecology* **18:** 3307–3315.

Bowen, L., Aldridge, B.M., Gulland, F., Woo, J., VanBonn, W., DeLong, R., Stott, J.L. & Johnson, M.L. 2002. Molecular characterization of expressed DQA and DQB genes in the California sea lion (*Zalophus californianus*). *Immunogenetics* **54:** 332–347.

Bowen, L., Aldridge, B.M., Gulland, F., VanBonn, W., DeLong, R., Melin, S., Lowenstine, L.J., Stott, J.L. & Johnson, M.L. 2004. Class II multiformity generated by variable MHC-DRB region configurations in the California sea lion (*Zalophus californianus*). *Immunogenetics* **56:** 12–27.

Bowen, L., Aldridge, B., Beckmen, K., Gelatt, T., Rea, L., Burek, K., Pitcher, K. & Stott, J.L. 2006a. Differential expression of immune response genes in Steller Sea Lions (*Eumetopias jubatus*): An indicator of ecosystem health? *EcoHealth* **3:** 109–113.

Bowen, L., Aldridge, B.M., Miles, A.K. & Stott, J.L. 2006b. Expressed MHC class II genes in sea otters (*Enhydra lutris*) from geographically disparate populations. *Tissue Antigens* **67:** 402–408.

Brennan, P.A. & Kendrick, K.M. 2006. Mammalian social odours: attraction and individual recognition. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **361:** 2061-2078.

Brouwer, L., Barr, I., vandePol, M., BURKE, T., Komdeur, J. & Richardson, D.S. 2010. MHC-dependent survival in a wild population: evidence for hidden genetic benefits gained through extra-pair fertilizations. *Molecular Ecology* **19:** 3444–3455.

Brown, J.H., Jardetzky, T.S., Gorga, J.C., Stern, L.J., Urban, R.G., Strominger, J.L. & Wiley, D.C. 1993. Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* **364:** 33–39.

Brown, J.L. & Eklund, A. 1994. Kin recognition and the major histocompatibility complex: an integrative review. *The American naturalist* **143:** 435-461.

Brown, R.E. & Macdonald, D.W. 1985. *Social odours in mammals*. Clarendon Press, Oxford.

Buesching, C.D. & Macdonald, D.W. 2001. Scent-marking behaviour of the European badger (*Meles meles*): Resource defence or individual advertisement? In: *Chemical Signals in Vertebrates 9* (A. Marchlewska-Koj, J. J. Lepri & D. Müller-Schwarze, eds), pp. 321–327. Plenum Publishers, New York.

Buesching, C.D., Newman, C. & Macdonald, D.W. 2002a. Variations in colour and volume of the subcaudal gland secretion of badgers (*Meles meles*) in relation to sex, season and individual-specific parameters. *Mammalian Biology* **67:** 147–156.

Buesching, C.D., Waterhouse, J.S. & Macdonald, D.W. 2002b. Gas-chromatographic analyses of the subcaudal gland secretion of the European badger (*Meles meles*) Part II: Time-related variation in the individual-specific composition. *Journal of Chemical Ecology* **28:** 57–69.

Buesching, C.D., Waterhouse, J.S. & Macdonald, D.W. 2002c. 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.

Buesching, C.D., Stopka, P. & Macdonald, D.W. 2003. The social function of allo-marking in the European badger (*Meles meles*). *Behaviour* **140:** 965–980.

Buesching, C.D. & Macdonald, D.W. 2004. Variations in scent-marking behaviour of European badgers *Meles meles* in the vicinity of their setts. *Acta Theriologica* **49:** 235–246.

Buesching, C.D., Heistermann, M. & Macdonald, D.W. 2009. Seasonal and inter-individual variation in testosterone levels in badgers *Meles meles*: evidence for the existence of two endocrinological phenotypes. *Journal of Comparative Physiology A* **195:** 865–871.

Burnett, R.C., DeRose, S.A., Wagner, J.L. & Store, R. 1997. Molecular analysis of six dog leukocyte antigen class I sequences including three complete genes, two truncated genes and one full-length processed gene. *Tissue Antigens* **49:** 484–495.

Burnham, K.P. & Anderson, D.R. 2002. *Model selection and multimodel inference: a practical information-theoretic approach*, 2nd edn.

Burnham, K.P., Anderson, D.R. & Huyvaert, K.P. 2011. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behavioral Ecology and Sociobiology* **65:** 23–35.

Burri, R., Hirzel, H.N., Salamin, N., Roulin, A. & Fumagalli, L. 2008. Evolutionary patterns of MHC class II B in owls and their implications for the understanding of avian MHC evolution. *Molecular Biology and Evolution* **25:** 1180–1191.

Burtscher, H. & Url, A. 2007. Evidence of canine distemper and suggestion of preceding parvovirus-myocarditis in a Eurasian badger (*Meles meles*). *Journal of Zoo and Wildlife Medicine* **38:** 139–142.

Buss, D.M. 1989. Sex differences in human mating preferences: evolutionary hypotheses tested in 37 cultures. *Behavioral and Brain Sciences* **12:** 1–49.

Butler, J.M. & Roper, T.J. 1996. Ectoparasites and sett use in European badgers. *Animal Behaviour* **52:** 621–629.

Cadavid, L.F., Hughes, A.L. & Watkins, D.I. 1996. MHC class I-processed pseudogenes in New World primates provide evidence for rapid turnover of MHC class I genes. *Journal of Immunology* **157:** 2403–2409.

Canivenc, R. & Bonnin, M. 1979. Delayed implantation is under environmental control in the badger (*Meles meles* L.). *Nature* **26:** 849–850.

Carpenter, P.J., Pope, L.C., Greig, C., Dawson, D.A., Rogers, L.M., Erven, K., Wilson, G.J., Delahay, R.J., Cheeseman, C.L. & Burke, T. 2005. Mating system of the Eurasian badger, *Meles meles*, in a high density population. *Molecular Ecology* **14:** 273–284.

Carrington, M., Nelson, G.W., Martin, M.P., Kissner, T., Vlahov, D., Goedert, J.J., Kaslow, R., Buchbinder, S., Hoots, K. & O'Brien, S.J. 1999. HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. *Science* **283:** 1748–1752.

Charmantier, A. & Sheldon, B.C. 2006. Testing genetic models of mate choice evolution in the wild. *Trends in Ecology and Evolution* **21:** 417-419.

Cheeseman, C.L., Wilesmith, J.W. & Stuart, F.A. 1989. Tuberculosis: The disease and its epidemiology in the badger, a review. *Epidemiology and Infection* **103:** 113–125.

Chen, J.-M., Cooper, D.N., Chuzhanova, N., Férec, C. & Patrinos, G.P. 2007. Gene conversion: mechanisms, evolution and human disease. *Nature Reviews Genetics* **8:** 762–775.

Chen, S.L., Zhang, Y.X., Xu, M.Y., Ji, X.S., Yu, G.C. & Dong, C.F. 2006. Molecular polymorphism and expression analysis of MHC class II B gene from red sea bream (*Chrysophrys major*). *Developmental and Comparative Immunology* **30:** 407–418.

Cheng, Y., Siddle, H.V., Beck, S., Eldridge, M.D.B. & Belov, K. 2009. High levels of genetic variation at MHC class II DBB loci in the tammar wallaby (*Macropus eugenii*). *Immunogenetics* **61:** 111–118.

Christian, S.F. 1995. Observations of extra-group mating and mate-defence behaviour in badgers, *Meles meles*. *Journal of Zoology* **237:** 668–670.

Clarke, K.R. 1993. Non-parametric multivariate analysis of changes in community structure. *Australian Journal of Ecology* **18:** 117–143.

Clarke, K.R. & Gorley, R.N. 2001. Primer v5: user manual/tutorial. PRIMER-E, Plymouth, UK.

Clifton-Hadley, R.S., Wildesmith, J.W. & Stuart, F.A. 1993. Mycobacterium bovis in the European badger (Meles meles): epidemiological findings in tuberculous badgers from a naturally infected population. *Epidemiology and Infection* **111:** 9–19.

Coltman, D.W., Pilkington, J.G., Smith, J.A. & Pemberton, J.M. 1999. Parasite-mediated selection against inbred Soay sheep in a free-living, island population. *Evolution* **53:** 1259–1267

Consuegra, S. & Garcia-de-Leaniz, C. 2008. MHC-mediated mate choice increases parasite resistance in salmon. *Proceedings of the Royal Society of London. Series B* **275:** 1397–1403.

Corey, L. & Wald, A. 2009. Maternal and neonatal herpes simplex virus infections. *New England Journal of Medicine* **361:** 1376–1385.

Courtenay, O., Reilly, L.A., Sweeney, F.P., Hibberd, V., Bryan, S., Ul-Hassan, A., Newman, C., Macdonald, D.W., Delahay, R.J., Wilson, G.J. & Wellington, E.M.H. 2006. Is *Mycobacterium bovis* in the environment important for the persistence of bovine tuberculosis? *Biology Letters* **2:** 460–462.

Cox, R., Stewart, P.D. & Macdonald, D.W. 1999. The ectoparasites of the European badger, *Meles meles*, and the behavior of the host-specific flea, *Paraceras melis*. *Journal of Insect Behavior* **12:** 245–265.

Cresswell, W.J., Harris, S., Cheeseman, C.L. & Mallinson, P.J. 1992. To breed or not to breed—an analysis of the social and density-dependent constraints on the fecundity of female badgers (*Meles meles*). *Philosophical Transactions of the Royal Society of London B* **338:** 393–407.

Davies, J.M., Lachno, D.R. & Roper, T.J. 1988. The anal gland secretion of the European badger (*Meles meles*) and its role in social communication. *Journal of Zoology* **45:** 455–463.

deBellocq, J.G., Suchentrunk, F., Baird, S.J.E. & Schaschl, H. 2009. Evolutionary history of an MHC gene in two leporid species: characterisation of Mhc–DQA in the European brown hare and comparison with the European rabbit. *Immunogenetics* **61:** 131–144.

deBellocq, J.G. & Leirs, H. 2010. Polymorphism and signatures of selection in the multimammate rat DQB gene. *Immunogenetics* **62:** 59–64.

deGroot, N., Doxiadis, G.G., deGroot, N.G., Otting, N., Heijmans, C., Rouweler, A.J.M. & Bontrop, R.E. 2004. Genetic makeup of the DR region in rhesus macaques: gene content, transcripts, and pseudogenes. *Journal of Immunology* **172:** 6152–6157.

Delahay, R.J., Brown, J.A., Mallinson, P.J., Spyvee, P.D., Handoll, D., Rogers, L.M. & Cheeseman, C.L. 2000. The use of marked bait in studies of the territorial organization of the European Badger (*Meles meles*). *Mammal Review* **30:** 73–87.

Delahay, R.J., Cheeseman, C.L. & Clifton-Hadley, R.S. 2001. Wildlife disease reservoirs: the epidemiology of *Mycobacterium bovis* infection in the European badger (*Meles meles*) and other British mammals. *Tuberculosis* **81:** 43–49.

Deplancke, B., Hristova, K.R., Oakley, H.A., McCracken, V.J., Aminov, R., Mackie, R.I. & Gaskins, H.R. 2000. Molecular ecological analysis of the succession and diversity of sulfate-reducing bacteria in the mouse gastrointestinal tract. *Applied and Environmental Microbiology* **66:** 2166–2174.

Doherty, P. & Zinkernagel, R. 1975. Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature* **256:** 50–52.

Doxiadis, G.G.M., Otting, N., deGroot, N.G. & Bontrop, R.E. 2001. Differential evolutionary MHC class II strategies in humans and rhesus macaques: relevance for biomedical studies. *Immunological Reviews* **183:** 76–85.

Dreiss, A.N., Silva, N., Richard, M., Moyen, F., Théry, M., Møller, A.P. & Danchin, E. 2008. Condition-dependent genetic benefits of extrapair fertilization in female blue tits *Cyanistes caeruleus*. *Journal of Evolutionary Biology* **21:** 1814–1822.

Drury, J.P. 2010. Immunity and mate choice: A new outlook. *Animal Behaviour* **79:** 539–545.

Dugdale, H.L., Macdonald, D.W., Pope, L.C. & Burke, T. 2007. Polygynandry, extra-group paternity and multiple-paternity litters in European badger (*Meles meles*) social groups. *Molecular Ecology* **16:** 5294–5306.

Dugdale, H.L., Macdonald, D.W., Pope, L.C., Johnson, P.J. & Burke, T. 2008. Reproductive skew and relatedness in social groups of European badgers, *Meles meles*. *Molecular Ecology* **17:** 1815–1827.

Dugdale, H.L., Ellwood, S.A. & Macdonald, D.W. 2010. Alloparental behaviour and long-term costs of mothers tolerating other group-members in a plurally breeding mammal. *Animal Behaviour* **80:** 721–735.

Dugdale, H.L., Davison, D., Baker, S.E., Ellwood, S.A., Newman, C., Buesching, C.D. & Macdonald, D.W. 2011a. Female teat size is a reliable indicator of annual breeding success in European badgers: Genetic validation. *Mammalian Biology* **76:** 716–721

Dugdale, H.L., Griffiths, A. & Macdonald, D.W. 2011b. Polygynandrous and repeated mounting behaviour in European badgers, *Meles meles*. *Animal Behaviour* **82:** 1287–1297.

Dugdale, H.L., Pope, L.C., Newman, C., Macdonald, D.W. & Burke, T. 2011c. Age-specific breeding success in a wild mammalian population: selection, constraint, restraint and senescence. *Molecular Ecology* **20:** 3261–3274.

Dunn, P.O., Bollmer, J.L., Freeeman-Gallant, C.R. & Whittingham, L.A. 2012. MHC variation is related to a sexually selected ornament, survival, and parasite resistance in common yellowthroats. *Evolution* **Early view:** DOI: 10.1111/j.1558-5646.2012.01799.x.

DuPasquier, L., Schwager, J. & Flajnik, M.F. 1989. The immune system of *Xenopus*. *Annual review of immunology* **7:** 251–275.

Durairaj, M., Sharma, R., Varghese, J.C. & Kane, K.P. 2003. Requirement for Q226, but not multiple charged residues, in the class I MHC CD loop/D strand for TCR-activated CD8 accessory function. *European Journal of Immunology* **33:** 676–684.

Eberhard, W.G. 1996. *Female control: Sexual selection by cryptic female choice*. Princeton University Press.

Edwards, J.A., Jones, D.B., Evans, P.R. & Smith, J.L. 1985. Differential expression of HLA class II antigens on human fetal and adult lymphocytes and macrophages. *Immunology* **55:** 489–500.

Edwards, S.V. & Hedrick, P.W. 1998. Evolution and ecology of MHC molecules: from genomics to sexual selection. *Trends in Ecology & Evolution* **13:** 305–311.

Egid, K. & Brown, J.L. 1989. The major histocompatibility complex and female mating preferences in mice. *Animal behaviour* **38:** 548–550.

Eizaguirre, C., Yeates, S.E., Lenz, T.L., Kalbe, M. & Milinski, M. 2009. MHC-based mate choice combines good genes and maintenance of MHC polymorphism. *Molecular Ecology* **18:** 3316–3329.

Eklund, A. 1997. The effect of early experience on MHC-based mate preferences in two B10.W strains of mice (*Mus domesticus*). *Behavior Genetics* **27:** 223–229.

Engebretson, J.J. & Moyer, C.L. 2003. Fidelity of select restriction endonucleases in determining microbial diversity by terminal restriction fragment length polymorphism. *Applied and Environmental Microbiology* **69:** 4823–4829.

Evans, P.G.H., Macdonald, D.W. & Cheeseman, C.L. 1989. Social structure of the Eurasian badger (*Meles meles*): genetic evidence. *Journal of Zoology* **218:** 587–595.

Favier, C.F., Vaughan, E.E., Vos, W.M.D. & Akkermans, A.D.L. 2002. Molecular monitoring of succession of bacterial communities in human neonates. *Applied and Environmental Microbiology* **68:** 219–226.

Fell, R.J., Buesching, C.D. & Macdonald, D.W. 2006. The social integration of European badger (*Meles meles*) cubs into their natal group. *Behaviour* **148:** 683–700.

Fernandez-Soria, V.M., Morales, P., Castro, M.J., Suarez, B., Recio, M.J., Moreno, M.A., Paz-Artal, E. & Arnaiz-Villena, A. 1998. Transcription and weak expression of *HLA-DRB6*: a gene with anomalies in exon 1 and other regions. *Immunogenetics* **48:** 16–21.

Figueroa, F., Gúnther, E. & Klein, J. 1988. MHC polymorphism pre-dating speciation. *Nature* **335:** 265–267.

Fisher, R.A. 1935. *The Design of Experiment*. Hafner, New York.

Fitch, W.M. 2000. Homology: a personal view on some of the problems. *Trends in Genetics* **16:** 227–231.

Fitze, P.S., Tschirren, B. & Richner, H. 2004. Life history and fitness consequences of ectoparasites. *Journal of Animal Ecology* **73:** 216–226.

Flynn, J.J., Finarelli, J.A., Zehr, S., Hsu, J. & Nedbal, M.A. 2005. Molecular phylogeny of the Carnivora (Mammalia): assessing the impact of increased sampling on resolving enigmatic relationships. *Systematic Biology* **54:** 317–337.

Foerster, K., Delhey, K., Johnsen, A., Lifjeld, J.T. & Kempenaers, B. 2003. Females increase offspring heterozygosity and fitness through extra-pair matings. *Nature* **425:** 714–717.

Foreyt, W.J. 2001. *Veterinary parasitology reference manual*, Fifth edn. Blackwell Publishing.

Forsberg, L.A., Dannewitz, J., Petersson, E. & M, G. 2007. Influence of genetic dissimilarity in the reproductive success and mate choice of brown trout—females fishing for optimal MHC dissimilarity. *Journal of Evolutionary Biology* **20:** 1859–1869.

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.

Fulton, T.L. & Strobeck, C. 2006. Molecular phylogeny of the Arctoidea (Carnivora): Effect of missing data on supertree and supermatrix analyses of multiple gene data sets. *Molecular Phylogenetics and Evolution* **41:** 165–181.

Furlong, R.F. & Yang, Z. 2008. Diversifying and purifying selection in the peptide binding region of *DRB* in mammals. *Journal of Molecular Evolution* **66:** 384–394.

Garrigan, D. & Hedrick, P.W. 2003. Perspective: detecting adaptive molecular polymorphism: Lessons from the MHC. *Evolution* **57:** 1707–1722.

Gillingham, M.A.F., Richardson, D.S., Løvlie, H., Moynihan, A., Worley, K. & Pizzari, T. 2009. Cryptic preference for MHC-dissimilar females in male red junglefowl, *Gallus gallus*. *Proceedings of the Royal Society B: Biological Sciences* **276:** 1083–1092.

Glaberman, S. & Caccone, A. 2008. Species-specific evolution of class I MHC genes in iguanas (Order: Squamata; Subfamily: Iguaninae). *Immunogenetics* **60:** 371–382.

Goldrath, A.W. & Bevan, M.J. 1999. Selecting and maintaining a diverse T-cell repertoire. *Nature* **402:** 255–262.

Goldsby, R.A., Kindt, T.J., Kuby, J. & Osborne, B.A. 2002. *Immunology*, 5th edition edn. W. H. Freeman.

Gorman, M.L. 1976. A mechanism for individual recognition by odour in *Herpestes auropunctatus* (Carnivora: Viverridae). *Animal Behaviour* **24:** 141-145.

Gorman, M.L., Kruuk, H. & Leitch, A. 1984. Social functions of the sub-caudal scent gland secretion of the European badger (*Meles meles*) (Carnivora: Mustelidae). *Journal of Zoology* **203:** 549–559.

Gower, D.B., Nixon, A., Jackman, P.J.H. & Mallett, A.I. 1986. Transformation of steroids by axillary coryneform bacteria. *International Journal of Cosmetic Science* **8:** 149–158.

Grafen, A. 1990. Do animals really recognize Kin. *Animal behaviour* **39:** 42–54.

Griffith, S.C., Owens, I.P.F. & Thuman, K.A. 2002. Extra pair paternity in birds: A review of interspecific variation and adaptive function. *Molecular Ecology* **11:** 2195–2212.

Hadfield, J.D., Richardson, D.S. & Burke, T. 2006. Towards unbiased parentage assignment: combining genetic, behavioural and spatial data in a Bayesian framework. *Molecular Ecology* **15:** 3715–3730.

Hall, A. 1982. Intestinal helminths of man: The interpretation of egg counts. *Parasitology* **85:** 605–613.

Hamilton, W.D. & Zuk, M. 1982. Heritable true fitness and bright birds: A role for parasites? *Science* **218:** 384–387.

Hamilton, W.D., Axelrod, R. & Tanese, R. 1990. Sexual reproduction as an adaptation to resist parasites (a review). *Proceedings of the National Academy of Sciences of the USA* **87:** 3566–3573.

Hancox, M. 1980. Parasites and infectious diseases of the Eurasian badger (*Meles meles* L.): A review. *Mammal Review* **10:** 151–162.

Handunnetthi, L., Ramagopalan, S.V., Ebers, G.C. & Knight, J.C. 2010. Regulation of major histocompatibility complex class II gene expression, genetic variation and disease. *Genes and Immunity* **11:** 99–112.

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.

Havlicek, J. & Roberts, S.C. 2009. MHC-correlated mate choice in humans: A review. *Psychoneuroendocrinology* **34:** 497—512.

Hawlena, H., Abramsky, Z. & Krasnov, B.R. 2006a. Ectoparasites and age-dependent survival in a desert rodent. *Oecologia* **148:** 30–39.

Hawlena, H., Krasnov, B.R., Abramsky, Z., Khokhlova, I.S., Saltz, D., Kam, M., Tamir, A. & Degen, A.A. 2006b. Flea infestation and energy requirements of rodent hosts: are there general rules? *Functional Ecology* **20:** 1028–1036.

Heath, W.R. & Carbone, F.R. 2001. Cross-presentation in viral immunity and self-tolerance. *Nature Reviews Immunology* **1:** 126–134.

Hedrick, P.W. 1992. Female choice and variation in the major histocompatibility complex. *Genetics* **132:** 575–581.

Hedrick, P.W. 1994. Evolutionary genetics at the major histocompatibility complex. *American Naturalist* **143:** 945–964.

Hedrick, P.W. 1999. Balancing selection and the MHC. *Genetica* **104:** 207–214.

Hedrick, P.W. 2002. Pathogen resistance and genetic variation at MHC loci. *Evolution* **56:** 1902–1908.

Hedrick, P.W. 2004. Comment on "Parasite Selection for Immunogenetic Optimality". *Science* **303:** 957.

Hendry, A.P., Burg, J.K., Bentzen, P., Volk, E.C. & Quinn, T.P. 2000. Rapid evolution of reproductive isolation in the wild: evidence from introduced salmon. *Science* **290:** 516–518.

Hess, C.M. & Edwards, S.V. 2002. The evolution of the major histocompatibility complex in birds. *BioScience* **52:** 423–431.

Hill, A.V.S., Allsopp, E.M., Dwiatkowski, D., Anstey, N.M., Twumasi, P., Rowe, P.A., Bennett, S., Brewster, D., McMichael, A.J. & Greenwood, B.M. 1991. Common West African HLA antigens are associated with protection from severe malaria. *Nature (London)* **352:** 595–600.

Hill, A.V.S. 1998. The immunogenetics of human infectious diseases. *Annual review of immunology* **16:** 593–617.

Hill, R.E. & Hastie, N.D. 1987. Accelerated evolution in the reactive center regions of serine protease inhibitors. *Nature* **326:** 96–99.

Ho, H.N., Yang, Y.S., Hsieh, R.P., Lin, H.R., Chen, S.U., Huang, S.C., Lee, T.Y. & Gill, T.J. 1994. Sharing of human leukocyte antigens in couples with unexplained infertility affects the success of in vitro fertilization and tubal embryo transfer. *American Journal of Obstetrics and Gynecology* **170:** 63–71.

Holmes, N. & Parham, P. 1985. Exon shuffling in vivo can generate novel HLA class I molecules. *EMBO Journal* **4:** 2849–2854.

Houlden, B.A., Greville, W.D. & Sherwin, W.B. 1996. Evolution of MHC class I loci in marsupials: Characterization of sequences from koala (*Phascolarctos cinereus*). *Molecular Biology and Evolution* **13:** 1119–1127.

Huchard, E., Raymond, M., Benavides, J., Marshall, H., Knapp, L.A. & Cowlishaw, G. 2010. A female signal reflects MHC genotype in a social primate. *BMC Evolutionary Biology* **10:** 96.

Hugenholtz, P., Goebel, B.M. & Pace, N.R. 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *Journal of Bacteriology***:** 4765–4774.

Hughes, A. & Hughes, M. 1995. Natural selection on the peptide-binding regions of major histocompatibility complex molecules. *Immunogenetics* **42:** 233–243.

Hughes, A.L. & Nei, M. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335:** 167–170.

Hughes, A.L. & 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, A.L. & Nei, M. 1992. Maintenance of MHC polymorphism. *Nature* **355:** 402–403.

Hughes, A.L. 1995. Origin and evolution of HLA class I pseudogenes. *Molecular Biology and Evolution* **12:** 247–258.

Hughes, A.L. & Yeager, M. 1998. Natural selection at major histocompatibility complex loci of vertebrates. *Annual Review of Genetics* **32:** 415–434.

Hurst, J.L., Thom, M.D., Nevison, C.M., Humphries, R.E. & Beynon, R.J. 2005. MHC odours are not required or sufficient for recognition of individual scent owners. *Proceedings of the Royal Society B* **272:** 715–724.

Huseby, E.S., White, J., Crawford, F., Vass, T., Becker, D., Pinilla, C., Marrack, P. & Kappler, J.W. 2005. How the T cell repertoire becomes peptide and MHC specific. *Cell* **122:** 247–260.

Huson, D.H. & Bryant, D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23:** 254–267.

Jaatinen, K., Noreikiene, K., Merila, J. & Öst, M. 2012. Kin association during brood care in a facultatively social bird: active discrimination or by-product of partner choice and demography? *Molecular Ecology* **21:** 3341–3351.

Jakobsen, I.B., Wilson, S.R. & Easteal, S. 1998. Patterns of reticulate evolution for the classical class I and II HLA loci. *Immunogenetics* **48:** 312–323.

Janeway, C.A. & Travers, P. 1994. *Immunobiology*, Garland, New York.

Janeway, C.A., Travers, P., Walport, M. & Schlomchik, M. 2005. *The immune system in health and disease*, 6th edn. Garland Science Publishing.

Jeffrey, K.J. & Bangham, R.M. 2000. Do infectious diseases drive MHC diversity? *Microbes and infection* **2:** 1335–1341.

Jennions, M.D. & Petrie, M. 2000. Why do females mate multiply? A review of the genetic benefits. *Biological Reviews* **75:** 21–64.

Johnson, D.D.P., Stopka, P. & Macdonald, D. 2004. Ideal flea constraints on group living: Unwanted public goods and the emergence of cooperation. *Behavioral Ecology* **15:** 181–186.

Joly, E. & Rouillon, V. 2006. The orthology of HLA-E and H2-Qa1 is hidden by their concerted evolution with other MHC class I molecules. *Biology Direct* **1:** doi:10.1186/1745-6150-1181-1182.

Jones, C.A. 2001. Maternal transmission of infectious pathogens in breast milk. *Journal of paediatrics and child health* **37:** 576–582.

Jordan, W.C. & Bruford, M.W. 1998. New perspectives on mate choice and the MHC. *Heredity* **81:** 239–245.

Jukes, T.H. & Cantor, C.R. 1969. Evolution of protein molecules. In: *Mammalian protein metabolism* (H. N. Munro, ed., pp. 21–132. Academic Press, New York.

Juola, F.A. & Dearborn, D.C. 2012. Sequence-based evidence for major histocompatibility complex-disassortative mating in a colonial seabird. *Proceedings of the Royal Society of London. Series B* **279:** 153–162.

Kalbe, M., Eizaguirre, C., Dankert, I., Reusch, T.B.H., Sommerfeld, R.D., Wegner, K.M. & Milinski, M. 2009. Lifetime reproductive success is maximized with optimal major histocompatibility complex diversity. *Proceedings of the Royal Society B* **276:** 925–934.

Kaplan, C.W., Astaire, J.C., Sanders, M.E., Reddy, B.S. & Kitts, C.L. 2001. 16S ribosomal DNA terminal restriction fragment pattern analysis of bacterial communities in feces of rats fed *Lactobacillus acidophilus* NCFM. *Applied and Environmental Microbiology* **67:** 1935–1939.

Kaufman, J., Salomonsen, J. & Flajnik, M. 1994. Evolutionary conservation of MHC class I and class II molecules—different yet the same. *Seminars in Immunology* **6:** 411–424.

Kaufman, J. 1999. Co-evolving genes in MHC haplotypes: the 'rule' for nonmammalian vertebrates? *Immunogenetics* **50:** 228–236.

Kelley, J., Walter, L. & Trowsdale, J. 2005. Comparative genomics of major histocompatibility complexes. *Immunogenetics* **56:** 683–695.

Kempenaers, B., Verheyen, G.R., VandenBroeck, M., Burke, T., VanBroeckhoven, C. & Dhondt, A.A. 1992. Extra-pair paternity results from female preference for high-quality males in the blue tit. *Nature* **357:** 494–496.

Kempenaers, B., Verheyen, G.R. & Dhondt, A.A. 1997. Extra-pair paternity in the blue tit (*Parus caeruleus*): female choice, male characteristics, and offspring quality. *Behavioural Ecology* **8:** 481–492.

Kempenaers, B. 2007. Mate choice and genetic quality: a review of the heterozygosity theory. *Advances in the Study of Behavior* **37:** 189–278.

Kennedy, L.J., Altet, L., Angles, J.M., Barnes, A., Carter, S.D., Francino, O., Gerlach, J.A., Happ, G.M., Ollier, W.E.R., Polvi, A., Thomson, W. & Wagner, J.L. 1999. Nomenclature for factors of the dog major histocompatibility system (DLA), 1998. First report of the ISAG DLA Nomenclature Committee. *Tissue Antigens* **54:** 312–321.

Kennedy, L.J., Angles, J.M., Barnes, A., Carter, S.D., Francino, O., Gerlach, J.A., Happ, G.M., Ollier, W.E.R., Thomson, W. & Wagner, J.L. 2001. Nomenclature for factors of the dog major

histocompatibility system (DLA), 2000: second report of the ISAG DLA Nomenclature Committee. *Tissue Antigens* **58:** 55–70.

Kennedy, L.J., Ryvar, R., Gaskell, R.M., Addie, D.D., Willoughby, K., Carter, S.D., Thomson, W., Ollier, W.E.R. & Radford, A.D. 2002. Sequence analysis of MHC DRB alleles in domestic cats from the United Kingdom. *Immunogenetics* **54:** 348–352.

Kennedy, L.J., Quarmby, S., Happ, G.M., Barnes, A., Ramsey, I.K., Dixon, R.M., Catchpole, B., Rusbridge, C., Graham, P.A., Hillbertz, N.S., Roethel, C., Dodds, W.J., Carmichael, N.G. & Ollier, W.E.R. 2006. Association of canine hypothyroidism with a common major histocompatibility complex DLA class II allele. *Tiss Antig* **68:** 82–86.

Keverne, E.B. 2004. Importance of olfactory and vomeronasal systems for male sexual function. *Physiology & Behavior* **83:** 177–187.

Kilshaw, K., Newman, C., Buesching, C., Bunyan, J. & Macdonald, D. 2009. Coordinated latrine use by European badgers, *Meles meles*: Potential consequences for territory defense. *Journal of Mammalogy* **90:** 1188–1198.

King, D.P., Mutukwa, N., Lesellier, S., Cheeseman, C., Chambers, M.A. & Banks, M. 2004. Detection of mustelid herpesvirus-1 infected European badgers (*Meles meles*) in the British Isles. *Journal of Wildlife Diseases* **40:** 99–102.

Klein, J. 1986. *Natural history of the major histocompatibility complex*, 1st edn. Wiley, New York.

Klein, J. 1987. Origin of major histocompatibility complex polymorphism: the trans-species hypothesis. *Human Immunology* **19:** 155–162.

Klein, J., Sato, A., Nagl, S. & O'hUigín, C. 1998. Molecular trans-species polymorphism. *Annual Review of Ecology and Systematics* **29:** 1–21.

Klein, J., Sato, A. & Nikolaidis, N. 2007. MHC, TSP, and the origin of species: from immunogenetics to evolutionary genetics. *Annual Review of Genetics* **41:** 281–304.

Knapp, L.A. 2007. Selection on MHC? A matter of form over function. *Heredity* **99:** 241–242.

Koepfli, K.-P., Deere, K.A., Slater, G.J., Begg, C., Begg, K., Grassman, L., Lucherini, M., Veron, G. & Wayne, R.K. 2008. Multigene phylogeny of the Mustelidae: resolving relationships, tempo and biogeographic history of a mammalian adaptive radiation. *BMC Biology* **6:** doi:10.1186/1741-7007-1186-1110.

Kohl, J.V., Atzmueller, M., Fink, B. & Grammer, K. 2001. Human pheromones: Integrating neuroendocrinology and ethology. *Neuroendocrinology Letters* **22:** 309-321.

Kokko, H., Brooks, R., Jennions, M.D. & Morley, J. 2003. The evolution of mate choice and mating biases. *Proceedings of the Royal Society B: Biological Sciences* **270:** 653–664.

Koller, B.H. & Orr, H.T. 1985. Cloning and complete sequence of an HLA-A2 gene: analysis of two HLA-A alleles at the nucleotide level. *Journal of Immunology* **134:** 2727–2733.

Kostyu, D.D., Dawson, D.V., Elias, S. & Ober, C. 1993. Deficit of HLA homozygotes in a Caucasian isolate. *Human Immunology* **37:** 135–142.

Kruuk, H. 1978a. Foraging and spatial organisation of the European badger, *Meles meles* L. *Behavioral Ecology and Sociobiology* **4:** 75–89.

Kruuk, H. 1978b. Spatial organisation and territorial behaviour of the European badger *Meles meles*. *Journal of Zoology* **184:** 1–19.

Kruuk, H., Gorman, M.L. & Leitch, A. 1984. Scent-marking with the subcaudal gland by the European badger, *Meles meles* L. *Animal Behaviour* **32:** 899–907.

Kuhn, F. & Natsch, A. 2009. Body odour of monozygotic human twins: a common pattern of odorant carboxylic acids released by a bacterial aminoacylase from axilla secretions contributing to an inherited body odour type. *Journal of the Royal Society Interface* **6:** 377–392.

Kurtz, J., Kalbe, M., Aeschlimann, P.B., M. A. Haberli, Wegner, K.M., Reusch, T.B.H. & Milinski, M. 2004. Major histocompatibility complex diversity influences parasite resistance and innate immunity in sticklebacks. *Proceedings of the Royal Society B: Biological Sciences* **271:** 197–204.

Ladel, C.H., Daugelat, S. & Kaufmann, S.H.E. 1995. Immune response to *Mycobacterium bovis* bacille Calmette Guerin infection in major histocompatibility complex class I- and II-deficient knock-out mice: contribution of CD4 and CD8 T cells to acquired resistance. *European Journal of Immunology* **25:** 377–384.

Landry, C., Garant, D., Duchesne, P. & Bernatchez, L. 2001. Good genes as heterozygosity: The major histocompatibility complex and mate choice in Atlantic salmon (*Salmo salar*). *Proceedings of the Royal Society of London. Series B* **268:** 1279–1285.

Langefors, Å., Lohm, J., Grahn, M., Andersen, Ø. & Schantz, T.v. 2001. Association between major histocompatibility complex class IIB alleles and resistance to *Aeromonas salmonicida* in Atlantic salmon. *Proceedings of the Royal Society B* **268:** 479–485.

Lanyon, C.V., Rushton, S.P., O'Donnell, A.G., Goodfellow, M., Ward, A.C., Petrie, M., Jensen, S.P., Gosling, L.M. & Penn, D.J. 2007. Murine scent mark microbial communities are genetically determined. *FEMS Microbiology Ecology* **59:** 576–583.

Lawlor, D.A., Zemmour, J., Ennis, P.D. & Parham, P. 1990. Evolution of class-I MHC genes and proteins: from natural selection to thymic selection. *Annual Review of Immunology* **8:** 23–63.

LeibundGut-Landmann, S., Waldburger, J., Krawczyk, M., Otten, L.A., Suter, T., Fontana, A., Acha-Orbea, H. & Reith, W. 2004. Specificity and expression of CIITA, the master regulator of MHC class II genes. *European Journal of Immunology* **34:** 1513–1525.

Leinders-Zufall, T., Brennan, P., Widmayer, P., Chandramani, P., Maul-Pavicic, A., Jager, M., Li, X.H., Breer, H., Zufall, F. & Boehm, T. 2004. MHC class I peptides as chemosensory signals in the vomeronasal organ. *Science* **306:** 1033–1037.

Liu, W.T., Marsh, T.L., Cheng, H. & Forney, L.J. 1997. Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Applied and Environmental Microbiology* **63:** 4516–4522.

Lively, C.M. & Dybdahl, M.F. 2000. Parasite adaptation to locally common host genotypes. *Nature* **405:** 679–681.

Lizundia, R., Newman, C., Buesching, C.D., Ngugi, D., Blake, D., Sin, Y.W., Macdonald, D.W., Wilson, A. & McKeever, D. 2011. Evidence for a role of the host-specific flea (*Paraceras melis*) in the transmission of *Trypanosoma (Megatrypanum) pestanai* to the European badger. *PLoS ONE* **6:** e16977.

Lloyd, S. & Smith, J. 1997. Pattern of Cryptosporidium parvum oocyst excretion by experimentally infected dogs. *International Journal for Parasitology* **27:** 799–801.

Loiseau, C., Zoorob, R., Robert, A., Chastel, O., Julliard, R. & Sorci, G. 2011. Plasmodium relictum infection and MHC diversity in the house sparrow (*Passer domesticus*). **278:** 1264–1272.

Lu, J., Idris, U., Harmon, B., Hofacre, C., Maurer, J.J. & Lee, M.D. 2003. Diversity and Succession of the Intestinal Bacterial Community of the Maturing Broiler Chicken. *Applied and Environmental Microbiology* **69:** 6816–6824.

Lu, Y. 2001. Roles of lineage sorting and phylogenetic relationship in the genetic diversity at the self-incompatibility locus of Solanaceae. *Heredity* **86:** 195–205.

Macdonald, D.W., Anwar, M., Newman, C., Woodroffe, R. & Johnson, P.J. 1999. Inter-annual differences in the age-related prevalences of *Babesia* and *Trypanosoma* parasites of European badgers (*Meles meles*). *Journal of Zoology* **247:** 65–70.

Macdonald, D.W. & Newman, C. 2002. Population dynamics of badgers (*Meles meles*) in Oxfordshire, UK: numbers, density and cohort life histories, and a possible role of climate change in population growth. *Journal of Zoology* **256:** 121–138.

Macdonald, D.W., Newman, C., Stewart, P.D., Domingo-Roura, X. & Johnson, P.J. 2002. Density-dependent regulation of body mass and condition in badgers (*Meles meles*) from Wytham Woods. *Ecology* **83:** 2056–2061.

Macdonald, D.W., Newman, C., Dean, J., Buesching, C.D. & Johnson, P.J. 2004. The distribution of Eurasian badger, *Meles meles*, setts in a high density area: Field observations contradict the sett dispersion hypothesis. *Oikos* **106:** 295–307.

Macdonald, D.W., Riordan, P. & Mathews, F. 2006. Biological hurdles to the control of TB in cattle: A test of two hypotheses concerning wildlife to explain the failure of control. *Biological Conservation* **131:** 268–286.

Macdonald, D.W., Newman, C., Buesching, C.D. & Johnson, P.J. 2008. Male-biased movement in a high-density population of the Eurasian badger (*Meles meles*). *Journal of Mammalogy* **89:** 1077–1086.

Macdonald, D.W., Newman, C., Nouvellet, P.M. & Buesching, C.D. 2009. An analysis of Eurasian badger (*Meles meles*) population dynamics: implications for regulatory mechanisms. *Journal of Mammalogy* **90:** 1392–1403.

Macdonald, D.W., Newman, C., Buesching, C.D. & Nouvellet, P.M. 2010. Are badgers 'Under The Weather'? Direct and indirect impacts of climate variation on European badger (*Meles meles*) population dynamics. *Global Change Biology***:** doi: 10.1111/j.1365-2486.2010.02208.x.

Maddison, W.P. & Knowle, L.L. 2006. Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology* **55:** 21–30.

MAFF 1986. *Manual of veterinary parasitological techniques*, London, HMCO. .

Manning, C.J., Wakeland, E.K. & Potts, W.K. 1992. Communal nesting patterns in mice implicate MHC genes in kin recognition. *Nature* **360:** 581–583.

Marmi, J., Lopez-Giraldez, J.F. & Domingo-Roura, X. 2004. Phylogeny, evolutionary history and taxonomy of the Mustelidae based on sequences of the cytochrome b gene and a complex repetitive flanking region. *Zoologica Scripta* **33:** 481–499.

Martin, D. & Rybicki, E. 2000. RDP: detection of recombination amongst aligned sequences. *Bioinformatics* **16:** 562–563.

Martin, D.P., Posada, D., Crandall, K.A. & Williamson, C. 2005. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Research and Human Retroviruses* **21:** 98–102.

Martin, D.P., Lemey, P., Lott, M., Moulton, V., Posada, D. & Lefeuvre, P. 2010. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* **26:** 2462–2463.

Mathews, F., Macdonald, D.W., Taylor, G.M., Gelling, M., Norman, R.A., Honess, P.E., Foster, R., Gower, C.M., Varley, S., Harris, A., Palmer, S., Hewinson, G. & Webster, J.P. 2006a. Bovine tuberculosis (*Mycobacterium bovis*) in British farmland wildlife: the importance to agriculture. *Proceedings of the Royal Society of London. Series B* **273:** 357–365.

Mathews, F., Macdonald, D.W., Taylor, G.M., Gelling, M., Norman, R.A., Honess, P.E., Foster, R., Gower, C.M., Varley, S., Harris, A., Palmer, S., Hewinson, G. & Webster, J.P. 2006b. Bovine tuberculosis (*Mycobacterium bovis*) in British farmland wildlife: the importance to agriculture. *Proceedings of the Royal Society B: Biological Sciences* **273:** 357–365.

Mayer, W.E., O'huigin, C. & Klein, J. 1993. Resolution of the *HLA-DRB6* puzzle: a case of grafting a *de novo* generated exon on an existing gene. *Proceedings of the National Academy of Sciences of the United States of America* **90:** 10720–10724.

Mays, H.L.J. & Hill, G.E. 2004. Choosing mates: good genes versus genes that are a good fit. *Trends in Ecology and Evolution* **19:** 554–559.

Mazerolle, M.J. 2011. AICcmodavg: model selection and multi-model inference based on (Q)AIC(c). R package, version 1.15 http://CRAN.R-project.org/package=AICcmodavg.

McClelland, E.E., Penn, D.J. & Potts, W.K. 2003. Major histocompatibility complex heterozygote superiority during coinfection. *Infection and immunity* **71:** 2079–2086.

McLaren, G.W., Thornton, P.D., Newman, C., Buesching, C.D., Baker, S.E., Mathews, F. & Macdonald, D.W. 2005. The use and assessment of ketamine–medetomidine–butorphanol combinations for field anaesthesia in wild European badgers (*Meles meles*). *Veterinary Anaesthesia and Analgesia* **32:** 367–372.

Mesa, C.M., Thulien, K.J., Moon, D.A., Veniamin, S.M. & Magor, K.E. 2004. The dominant MHC class I gene is adjacent to the polymorphic *TAP2* gene in the duck, *Anas platyrhynchos*. *Immunogenetics* **56:** 192–203.

Messing, J. 1983. New M13 vectors for cloning. *Methods in Enzymology* **101:** 20–78.

Meyer-lucht, Y. & Sommer, S. 2005. MHC diversity and the association to nematode parasitism in the yellow-necked mouse (*Apodemus flavicollis*). *Molecular Ecology* **14:** 2233–2243.

Michael, E. & Bundy, D.A.P. 1989. Density dependence in establishment, growth and worm fecundity in intestinal helminthiasis: The population biology of *Trichuris muris* (Nematoda) infection in CBA/Ca mice. *Parasitology* **53:** 63–84.

Milinski, M., Griffiths, S., Wegner, K.M., Reusch, T.B.H., Haas-Assenbaum, A. & Boehm, T. 2005. Mate choice decisions of stickleback females predictably modified by MHC peptide ligands. *Proceedings of the National Academy of Sciences of the United States of America* **102:** 4414–4418.

Milinski, M. 2006. The major histocompatibility complex, sexual selection, and mate choice. *Annual Review of Ecology, Evolution, and Systematics* **37:** 159–186.

Miller, H.C., Moore, J.A., Nelson, N.J. & Daugherty, C.H. 2009. Influence of major histocompatibility complex genotype on mating success in a free-ranging reptile population. *Proceedings of the Royal Society of London. Series B* **276:** 1695–1704.

Mombaerts, P. 2004. Genes and ligands for odorant, vomeronasal and taste receptors. *Nature Reviews Neuroscience* **5:** 263–278.

Montes, I., Newman, C., Mian, R. & Macdonald, D.W. 2011. Radical health: ecological corollaries of body condition, transport stress and season on plasma antioxidant capacity in the European badger. *Journal of Zoology* **284:** 114–123.

Muyzer, G., Waal, E.C.d. & Uitterlinden, A.G. 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Applied and Environmental Microbiology* **59:** 695–700.

Muyzer, G., Teske, A. & Wirsen, C.O. 1995. Phylogenetic relationships of Thiomicrospira species and their identification in deep-sea hydrothermal vent samples by denaturing gradient gel electrophoresis of 16S rDNA fragments. *Archives of Microbiology* **164:** 165–172.

Myers, S., Freeman, C., Auton, A., Donnelly, P. & McVean, G. 2008. A common sequence motif associated with recombination hot spots and genome instability in humans. *Nature Genetics* **40:** 1124–1129.

Myers, S., Bowden, R., Tumian, A., Bontrop, R.E., Freeman, C., MacFie, T.S., McVean, G. & Donnelly, P. 2010. Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. *Science* **327:** 876–879

Natsch, A., Derrer, S., Flachsmann, F. & Schmid, J. 2006. A broad diversity of volatile carboxylic acids, released by a bacterial aminoacylase from axilla secretions, as candidate molecules for the determination of human-body odor type. *Chemistry & Biodiversity* **3:** 1-20.

Neal, E. & Cheeseman, C.L. 1996. *Badgers*. T. & A.D. Poyser, London.

Neff, B.D. 2004. Stabilizing selection on genomic divergence in a wild fish population. *Proceedings of the National Academy of Sciences USA* **101:** 2381–2385.

Neff, B.D. & Pitcher, T.E. 2005. Genetic quality and sexual selection: An integrated framework for good genes and compatible genes. *Molecular Ecology* **14:** 19–38.

Neff, B.D., Garner, S.R., Heath, J.W. & Heath, D.D. 2008. The MHC and non-random mating in a captive population of Chinook salmon. *Heredity* **101:** 175–185.

Nei, M. & Gojobori, T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution* **3:** 418–426.

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 USA* **94:** 7799–7806.

Nei, M. & Rooney, A.P. 2005. Concerted and birth-and-death evolution of multigene families. *Annual Review of Genetics* **39:** 121–152.

Newman, C., Macdonald, D.W. & Anwar, M.A. 2001. Coccidiosis in the European badger, *Meles meles* in Wytham Woods: infection and consequences for growth and survival. *Parasitology* **123:** 133–142.

Newman, C., Buesching, C.D. & MacDonald, D.W. 2004. First report of *Cheyletiella parasitovorax* infestation in the Eurasian badger (*Meles meles*). *Veterinary Record* **155:** 180–181.

Newman, C., Zhou, Y.-B., Buesching, C.D., Kaneko, Y. & Macdonald, D.W. 2011. Contrasting sociality in two widespread, generalist, mustelid genera, *Meles* and *Martes*. *Mammal Study* **36:** 169-188.

Nielsen, R. 2005. Molecular signatures of natural selection. *Annual Review of Genetics* **39:** 197–218.

Nikolich-Žugich, J., Fremont, D.H., Miley, M.J. & Messaoudi, I. 2004. Review The role of mhc polymorphism in anti-microbial resistance. *Microbes and Infection* **6:** 501–512.

Nordstrom, N.K.M. & Noble, W.C. 1984. Colonization of the axilla by *Propionibacterium avidum* in relation to age. *Applied and Environmental Microbiology* **47:** 1360–1362.

Nouvellet, P., Buesching, C.D., Dugdale, H.L., Newman, C. & Macdonald, D.W. 2010. Mouthing off about developmental stress: Individuality of palate marking in the European badger and its relationship with juvenile parasitoses. *Journal of Zoology* **283:** 52–62.

Nowak, M.A., Tarczy-Hornoch, K. & Austyn, J.M. 1992. The optimal number of major histocompatibility complex molecules in an individual. *Proceedings of the National Academy of Sciences of the United States of America* **89:** 10896–10899.

O'Brien, S.J. & Yuhki, N. 2006. Comparative genome organization of the major histocompatibility complex: Lessons from the Felidae. *Immunological Reviews* **167:** 133–144.

Ober, C. 1992. The maternal-fetal relationship in human pregnancy: an immunological perspective. *Experimental and Clinical Immunogenetics* **9:** 1–14.

Ober, C. 1995. HLA and reproduction: lessons from studies in the Hutterites. *Placenta* **16:** 569–577.

Ober, C., Weitkamp, L.R., Cox, N., Dytch, H., Kostyu, D. & Elias, S. 1997. HLA and mate choice in humans. *American Journal of Human Genetics* **61:** 497–504.

Ogden, N.H., Cripps, P., Davison, C.C., Owen, G., Parry, J.M., Timms, B.J. & Forbes, A.B. 2000. The ixodid tick species attaching to domestic dogs and cats in Great Britain and Ireland. *Medical and Veterinary Entomology* **14:** 332–338.

Ohta, T. 1991. Role of diversifying selection and gene conversion in evolution of major histocompatibility complex loci. *Proceedings of the National Academy of Sciences of the United States of America* **88:** 6716–6720.

Oliver, M.K., Telfer, S. & Piertney, S.B. 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.k. 2003. Major histocompatibility complex and mate choice in sand lizards. *Proceedings of the Royal Society B* **270:** S254–S256.

Orcutt, B., Bailey, B., Staudigel, H., Tebo, B.M. & Edwards, K.J. 2009. An interlaboratory comparison of 16S rRNA gene-based terminal restriction fragment length polymorphism and sequencing methods for assessing microbial diversity of seafloor basalts. *Environmental Microbiology* **11:** 1728–1735.

Owczarzy, R., Tataurov, A.V., Wu, Y., Manthey, J.A., McQuisten, K.A., Almabrazi, H.G., Pedersen, K.F., Lin, Y., Garretson, J., McEntaggart, N.O., Sailor, C.A., Dawson, R.B. & Peek, A.S. 2008. IDT SciTools: a suite for analysis and design of nucleic acid oligomers. *Nucleic Acids Research* **36:** W163-W169.

Pace, N.R. 1997. A molecular view of microbial diversity and the biosphere. *Science* **276:** 734 – 740.

Padidam, M., Sawyer, S. & Fauquet, C.M. 1999. Possible emergence of new geminiviruses by frequent recombination. *Virology* **265:** 218–225.

Pan, H.-J., Wan, Q.-H. & Fang, S.-G. 2008. Molecular characterization of major histocompatibility complex class I genes from the giant panda (*Ailuropoda melanoleuca*). *Immunogenetics* **60:** 185–193.

Parham, P. & Ohta, T. 1996. Population biology of antigen presentation by MHC class I molecules. *Science* **272:** 67–74.

Parker, B.B. & Duszynski, D.W. 1986. Coccidiosis of sandhill crane (*Grus canadensis*) wintering in New Mexico. *Journal of Wildlife Diseases* **22:** 25–35.

Paterson, S. & Pemberton, J.M. 1997. No evidence for major histocompatibility complex-dependent mating patterns in a free-living ruminant population. *Proceedings of the Royal Society, Series B* **264:** 1813–1819.

Paterson, S., Wilson, K. & Pemberton, J.M. 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 USA* **95:** 3714–3719.

Peace-Brewer, A.L., Tussey, L.G., Matsui, M., Li, G., Quinn, D.G. & Frelinger, J.A. 1996. A point mutation in HLA-A*0201 results in failure to bind the TAP complex and to present virus-derived peptides to CTL. *Immunity* **4:** 505–514.

Penn, D. & Potts, W. 1998a. MHC-disassortative mating preferences reversed by cross-fostering. *Proceedings of the Royal Society of London B, Biological Sciences* **265:** 1299–1306.

Penn, D. & Potts, W. 1998b. How do major histocompatibility complex genes influence odor and mating preferences? *Advances in Immunology* **69:** 411–436.

Penn, D.J. & Potts, W.K. 1999. The evolution of mating preferences and major histocompatibility complex genes. *American Naturalist* **153:** 145–164.

Penn, D.J. 2002. The scent of genetic compatibility: sexual selection and the major histocompatibility complex. *Ethology* **108:** 1–21.

Penn, D.J., Damjanovich, K. & Potts, W.K. 2002. MHC heterozygosity confers a selective advantage against multiple-strain infections. *Proceedings of the National Academy of Sciences USA* **99:** 11260–11264.

Piertney, S.B. & Oliver, M.K. 2006. The evolutionary ecology of the major histocompatibility complex. *Heredity* **96:** 7–21.

Piontkivska, H. & Nei, M. 2003. Birth-and-death evolution in primate MHC class I genes: divergence time estimates. *Molecular Biology and Evolution* **20:** 601–609.

Potts, W.K. & Wakeland, E.K. 1990. Evolution of diversity at the major histocompatibility complex. *Trends in Ecology & Evolution* **5:** 181–187.

Potts, W.K., Manning, C.J. & Wakeland, E.K. 1991. Mating patterns in seminatural populations of mice influenced by MHC genotype. *Nature* **352:** 619–621.

Potts, W.K., Manning, C.J., Wakeland, E.K. & Hughes, A.L. 1994. The role of infectious disease, inbreeding and mating preferences in maintaining MHC genetic diversity: an experimental test. *Philosophical Transactions of the Royal Society B* **346:** 369-378

Potts, W.K. 2002. Wisdom through immunogenetics. *Nature Genetics* **30:** 130–131.

Puurtinen, M., Ketola, T. & Katiaho, J.S. 2005. Genetic compatibility and sexual selection. *Trends in Ecology and Evolution* **20:** 157–158.

Queller, D.C. & Goodnight, K.F. 1989. Estimating relatedness using genetic-markers. *Evolution* **43:** 258–275.

Råberg, L., Sim, D. & Read, A.F. 2007. Disentangling genetic variation for resistance and tolerance to infectious diseases in animals. *Science* **318:** 812–814.

Rada, C., Lorenzi, R., Powis, S.J., Bogaerde, J.v.d., Parham, P. & Howard, J.C. 1990. Concerted evolution of class I genes in the major histocompatibility complex of murine rodents. . *Proceedings of the National Academy of Sciences of the United States of America* **87:** 2167–2171.

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.

Reche, P.A. & Reinherz, E.L. 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.

Rees, L.E., Ayoub, O., Haverson, K., Birchall, M.A. & Bailey, M. 2003. Differential major histocompatibility complex class II locus expression on human laryngeal epithelium. *Clinical & Experimental Immunology* **134:** 497–502.

Reusch, T.B.H., Haberli, M.A., Aeschlimann, P.B. & Milinski, M. 2001. Female sticklebacks count alleles in a strategy of sexual selection explaining MHC polymorphism. *Nature* **414:** 300-302.

Richardson, D.S., Komdeur, J., Burke, T. & vonSchantz, 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.

Richman, A. 2000. Evolution of balanced genetic polymorphism. *Molecular Ecology* **9:** 1953–1963.

Richman, A., Herrera, L., Nash, D. & Schierup, M. 2003. Relative roles of mutation and recombination in generating allelic polymorphism at an MHC class II locus in *Peromyscus maniculatus*. *Genetical Research* **82:** 89–99.

Riordan, P., Delahay, R.J., Cheeseman, C., Johnson, P.J. & Macdonald, D.W. 2011. Culling-induced changes in badger (*Meles meles*) behaviour, social organisation and the epidemiology of bovine tuberculosis. *PLoS ONE* **6:** e28904. doi:28910.21371/journal.pone.0028904.

Robbins, F., Hurley, C.K., Tang, T., Yao, H., Lin, Y.S., Wade, J., Goeken, N. & Hartzman, R.J. 1997. Diversity associated with the second expressed *HLA-DRB* locus in the human population. *Immunogenetics* **46:** 104–110.

Roberts, S.C., Little, A.C., Gosling, L.M., Jones, B.C., Perrett, D.I., Carter, V. & Petrie, M. 2005. MHC-assortative facial preferences in humans. *Biology Letters* **1:** 400–403.

Robinson, J., Waller, M., Parham, P., Bodmer, J.G. & Marsh, S.G.E. 2001. IMGT/HLA Database — a sequence database for the human major histocompatibility complex. *Nucleic Acids Research* **29:** 210– 213.

Rodgers, J.R. & Cook, R.G. 2005. MHC class Ib molecules bridge innate and acquired immunity. *Nature Reviews Immunology* **5:** 459–471.

Rogers, L.M., Cheeseman, C.L. & Mallinson, P.J. 1997. The demography of a high-density badger (*Meles meles*) population in the west of England. *Journal of Zoology* **242:** 705–728.

Ronquist, F. & Huelsenbeck, J.P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19:** 1572–1574.

Ronquist, F., Huelsenbeck, J.P. & van-der-Mark, P. 2005. MrBayes v. 3.1 Manual. Available from: http://mrbayes.sourceforge.net/manual.php.

Roper, T.J., Conradt, L., Butler, J., Christian, S.E., Ostler, J. & Schmid, T.K. 1993. Territorial marking with faeces in badgers (Meles meles): a comparison of boundary and hinterland latrine use. *Behaviour* **123:** 3-4.

Roper, T.J. 2010. *Badger*. Collins, London.

Rosalino, L.M., Macdonald, D.W. & Santos-Reis, M. 2004. Spatial structure and land-cover use in a low-density Mediterranean population of Eurasian badgers. *Canadian Journal of Zoology* **82:** 1493–1502.

Rosalino, L.M., Torres, J. & Santos-Reis, M. 2006. A survey of helminth infection in Eurasian badgers (*Meles meles*) in relation to their foraging behaviour in a Mediterranean environment in southwest Portugal. *European Journal of Wildlife Research* **52:** 202–206.

Rozas, J., Sánchez-DelBarrio, J., Messenguer, X. & Rozas, R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19:** 2469–2497.

Roze, U., Leung, K.T., Nix, E., Burton, G. & Chapman, D.M. 2010. Microanatomy and bacterial flora of the perineal glands of the North American porcupine. *Canadian Journal of Zoology* **88:** 59–68.

Rülicke, T., Chapuisat, M., Homberger, F.R., Macas, E. & Wedekind, C. 1998. MHC-genotype of progeny influenced by parental infection. *Proceedings of the Royal Society B: Biological Sciences* **265:** 711–716.

San, E.D.L. 2007. Ectoparasite categories of the Eurasian Badger (*Meles meles*) in Western Switzerland. *Wildlife Biology in Practice* **3:** 52–59.

Santelli, C.M., Orcutt, B.N., Banning, E., Bach, W., Moyer, C.L., Sogin, M.L., Staudigel, H. & Edwards, K.J. 2008. Abundance and diversity of microbial life in ocean crust. *Nature* **453:** 653–656.

Sardell, R.J., Arcese, P., Keller, L.F. & Reid, J.M. 2011. Sex-specific differential survival of extra-pair and within-pair offspring in song sparrows, Melospiza melodia. *Proc. R. Soc. B*.

Schad, J., Ganzhorn, J.U. & Sommer, S. 2005. Parasite burden and constitution of major histocompatibility complex in the Malagasy mouse lemur, *Microcebus murinus*. *Evolution* **59:** 439–450.

Schloss, P.D. & Handelsman, J. 2005. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Applied and Environmental Microbiology* **71:** 1501–1506.

Schmoll, T., Schurr, F., Winkel, W., Epplen, J. & Lubjuhn, T. 2011. Lifespan, lifetime reproductive performance and paternity loss of within-pair and extra-pair offspring in the coal tit Periparusater. *Proceedings of the Royal Society B: Biological Sciences* **276:** 337-345.

Schwensow, N., Eberle, M. & Sommer, S. 2008a. Compatibility counts: MHC-associated mate choice in a wild promiscuous primate. *Proceedings of the Royal Society, Series B* **275:** 555–564.

Schwensow, N., Fietz, J., Dausmann, K. & Sommer, S. 2008b. MHC-associated mating strategies and the importance of overall genetic diversity in an obligate pair-living primate. *Evolutionary Ecology* **22:** 617–636.

Scott, M.E. & Dobson, A. 1989. The role of parasites in regulating host abundance. *Parasitology Today* **5:** 176–183.

Seivwright, L.J., Redpath, S.M., Mougeot, F., Watt, L. & Hudson, P.J. 2004. Faecal egg counts provide a reliable measure of Trichostrongylus tenuis intensities in free-living red grouse Lagopus lagopus scoticus. *Journal of Helminthology* **78:** 69–76.

Sepil, I., Lachish, S., Hinks, A.E. & Sheldon, B.C. 2013. Mhc supertypes confer both qualitative and quantitative resistance to avian malaria infections in a wild bird population. *Proceedings of the Royal Society B* **280:** 20130134.

Setchell, J.M., Charpentier, M.J.E., Abbott, K.M., Wickings, E.J. & Knapp, L.A. 2010. Opposites attract: MHC-associated mate choice in a polygynous primate. *Journal of Evolutionary Biology* **23:** 136–148.

Shum, B.P., Rajalingam, R., Magor, K.E., Azumi, K., Carr, W.H., Dixon, B., Stet, R.J.M., Adkison, M.A., Hedrick, R.P. & Parham, P. 1999. A divergent non-classical class I gene conserved in salmonids. *Immunogenetics* **49:** 479–490.

Shum, B.P., Guethlein, L., Flodin, L.R., Adkison, M.A., Hedrick, R.P., Nehring, R.B., Stet, R.J.M., Secombes, C. & Parham, P. 2001. Modes of salmonid MHC class I and II evolution differ from the primate paradigm. *Journal of Immunology* **166:** 3297–3308.

Sin, Y.W., Buesching, C.D., Burke, T. & Macdonald, D.W. 2012a. Molecular characterization of the microbial communities in the subcaudal gland secretion of the European badger (*Meles meles*). *FEMS Microbiology Ecology* **81:** 648–659.

Sin, Y.W., Dugdale, H.L., Newman, C., Macdonald, D.W. & Burke, T. 2012b. MHC class II genes in the European badger (*Meles meles*): characterization, patterns of variation, and transcription analysis. *Immunogenetics* **64:** 313–327. doi: 310.1007/s00251-00011-00578-00259.

Sin, Y.W., Dugdale, H.L., Newman, C., Macdonald, D.W. & Burke, T. 2012c. Evolution of MHC class I genes in the European badger (*Meles meles*). *Ecology and Evolution* **3:** 285. doi: 210.1002/ece1003.1285.

Sin, Y.W. 2013. The major histocompatibility complex, mate choice and pathogen resistance in the European badger *Meles meles*, University of Oxford. Oxford. **Doctor of Philosophy:** 301.

Singh, P.B., Herbert, J., Roser, B., Arnott, L., Tucker, D. & Brown, R. 1990. Rearing rats in a germ-free environment eliminates their odors of individuality. *Journal of Chemical Ecology* **16:** 1667–1682.

Skarstein, F., Folstad, I., Liljedal, S. & Grahn, M. 2005. MHC and fertilization success in the Arctic charr (*Salvelinus alpinus*). *Behavioral Ecology and Sociobiology* **57:** 374–380.

Slade, R.W. & McCallum, H.I. 1992. Overdominant vs frequency-dependent selection at MHC loci. *Genetics* **132:** 861-862.

Smith, C.J., Danilowicz, B.S., Clear, A.K., Costello, F.J., Wilson, B. & Meijer, W.G. 2005. T-Align, a web-based tool for comparison of multiple terminal restriction fragment length polymorphism profiles. *FEMS Microbiology Ecology* **54:** 375–380.

Smith, G.C. & Wilkinson, D. 2002. Modelling disease spread in a novel host: rabies in the European badger *Meles meles*. *Journal of Applied Ecology* **39:** 865–874.

Smith, J.M. 1992. Analyzing the mosaic structure of genes. *Journal of Molecular Evolution* **34:** 126–129.

Sokal, R.R. & Rohlf, F.J. 1994. *Biometry: principles and practice of statistics in biological research*, 3rd edn. W.H. Freeman, New York.

Sommer, S. 2005a. Major histocompatibility complex and mate choice in a monogamous rodent. *Behavioral Ecology and Sociobiology* **58:** 181–189.

Sommer, S. 2005b. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Frontiers in Zoology* **2**.

Spehr, M., Kelliher, K.R., Li, X.H., Boehm, T., Leinders-Zufall, T. & Zufall, F. 2006. Essential role of the main olfactory system in social recognition of major histocompatibility complex peptide ligands. *Journal of Neuroscience* **26:** 1961-1970.

Spurgin, L., VanOosterhout, C., Illera, J., Bridgett, S., Gharbi, K., Emerson, B. & Richardson, D. 2011. Gene conversion rapidly generates major histocompatibility complex diversity in recently founded bird populations. *Molecular Ecology* **20:** 5213–5225.

Spurgin, L.G. & Richardson, D.S. 2010. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings of the Royal Society B: Biological Sciences* **277:** 979–988.

Stewart, P.D., Macdonald, D.W., Newman, C. & Cheeseman, C.L. 2001. Boundary faeces and matched advertisement in the European badger (Meles meles): a potential role in range exclusion. *Journal of Zoology* **225:** 191–198.

Stone, J.L. 2004. Sheltered load associated with S-alleles in *Solanum carolinense*. *Heredity* **92:** 335–342.

Stubbe, M. 1971. Die analen Markierungsorgane des Dachses (*Meles meles*). *Zoologischen Garten* **40:** 125-135.

Swain, S.L. 1983. T cell subsets and the recognition of MHC class. *Immunological Reviews* **74:** 129–142.

Swanson, W.J. & Vacquier, V.D. 2002. The rapid evolution of reproductive proteins. *Nature Reviews Genetics* **3:** 137–144.

Symonds, M.R.E. & Moussalli, A. 2011. A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behavioral Ecology and Sociobiology* **65:** 13–21.

Takahashi, K., Rooney, A.P. & Nei, M. 2000. Origins and divergence times of mammalian class II MHC gene clusters. *Journal of Heredity* **91:** 198–204.

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., Dudley, J., Nei, M. & Kumar, S. 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24:** 1596–1599.

Thody, A.J. & Shuster, S. 1989. Control and function of sebaceous glands. *Physiological Reviews* **69:** 383–408.

Thom, M.D., Johnson, D.D.P. & Macdonald, D.W. 2004. The evolution and maintenance of delayed implantation in the Mustelidae (Mammalia: Carnivora). *Evolution* **58:** 175–183.

Thornton, P.S. 1988. Density and distribution of Badgers in south-west England–a predictive model. *Mammal Review* **18:** 11-23.

Thursz, M.R., Thomas, H.C., Greenwood, B.M. & Hill, A.V. 1997. Heterozygote advantage for HLA class-II type in hepatitis B virus infection. *Nature Genetics* **17:** 11–12.

Toivanen, P., Vaahtovuo, J. & Eerola, E. 2001. Influence of major histocompatibility complex on bacterial composition of fecal flora. *Infection and Immunity* **69:** 2372–2377

Tollenaere, C., Bryja, J., Galan, M., Cadet, P., Deter, J., Chaval, Y., Berthier, K., Salvador, A.R., Voutilainen, L., Laakkonene, J., Henttonen, H., Cosson, J.-F. & Charbonnel, N. 2008. Multiple parasites mediate balancing selection at two MHC class II genes in the fossorial water vole: insights from multivariate analyses and population genetics. *21***:** 1307–1320.

Trachtenberg, E., Korber, B., Sollars, C., Kepler, T.B., Hraber, P.T., Hayes, E., Funkhouser, R., Fugate, M., Theiler, J., Hsu, Y.S., Kunstman, K., Wu, S., Phair, J., Erlich, H. & Wolinsky, S. 2003. Advantage of rare HLA supertype in HIV disease progression. *Nature Medicine* **9:** 928–935.

Tregenza, T. & Wedell, N. 2000. Genetic compatibility, mate choice and patterns of parentage: Invited Review. *Molecular Ecology* **9:** 1013–1027.

Trivers, R.L. 1972. *Parental investment and sexual selection*, Aldine, Chicago.

van de Crommenacker, J., Richardson, D.S., Koltz, A.M., Hutchings, K. & Komdeur, J. 2012. Parasitic infection and oxidative status are associated and vary with breeding activity in the Seychelles warbler. *Proceedings of the Royal Society B* **279:** 1466–1476.

van Oosterhout, C. 2009. A new theory of MHC evolution: beyond selection on the immune genes. *Proceedings of the Royal Society of London. Series B* **276:** 657-665.

van Valen, L. 1973. A new evolutionary law. *Evolutionary Theory* **1:** 1–30.

Villard, J., Lisowska-Grospierre, B., vandenElsen, P., Fischer, A., Reith, W. & Mach, B. 1997. Mutation of RFXAP, a regulator of MHC class II genes, in primary MHC class II deficiency. *The New England Journal of Medicine* **337:** 748–753.

Vogel, T.U., Evans, D.T., Urvater, J.A., O'Connor, D.H., Hughes, A.L. & Watkins, D.I. 1999. Major histocompatibility complex class I genes in primates: co-evolution with pathogens. *Immunological Reviews* **167:** 327–337.

von Schantz, T., Wittzell, H., Goransson, G., Grahn, M. & Persson, K. 1996. MHC genotype and male ornamentation: genetic evidence for the Hamilton-Zuk model. *Proceedings of the Royal Society of London. Series B* **263:** 265-271.

von Schantz, T., Wittzell, H., Goransson, G. & Grahn, M. 1997. Mate choice, male condition-dependent ornamentation and MHC in the pheasant. *Hereditas* **127:** 133-140.

Wagner, J.L., Burnett, R.C. & Storb, R. 1996. Molecular analysis of the DLA-DR region. *Tissue Antigens* **48:** 549–553.

Wagner, J.L., Burnett, R.C. & Storb, R. 1999. Organization of the canine major histocompatibility complex: current perspectives. *Journal of Heredity* **90:** 35–38.

Wagner, J.L. 2003. Molecular organization of the canine major histocompatibility complex. *Journal of Heredity* **94:** 23–26.

Walsh, P.S., Metzger, D.A. & Higuchi, R. 1991. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *BioTechniques* **10:** 506–513.

Wang, J. & Santure, A.W. 2009. Parentage and sibship inference from multilocus genotype data under polygamy. *Genetics* **181:** 1579–1594.

Ware, G.C. & Gosden, P.E. 1980. Anaerobic microflora of the anal sac of the red fox (*Vulpes vulpes*). *Journal of Chemical Ecology* **6:** 97–102.

Weber, D.S., Stewart, B.S., Schienman, J. & Lehman, N. 2004. Major histocompatibility complex variation at three class II loci in the northern elephant seal. *Molecular Ecology* **13:** 711-718.

Wedekind, C., Seebeck, T., Bettens, F. & Paepke, A. 1995. MHC-dependent mate preferences in humans. *Proceedings of the Royal Society of London B, Biological Sciences* **260:** 245–249.

Wedekind, C., Chapuisat, M., Macas, E. & Rülicke, T. 1996. Non-random fertilization in mice correlates with the MHC and something else. *Heredity* **77:** 400–409.

Wedekind, C. & Füri, S. 1997. Body odour preferences in men and women: do they aim for specific MHC combinations or simply heterozygosity? *Proceedings of the Royal Society of London B, Biological Sciences* **264:** 1471–1479.

Wegner, K.M., Kalbe, M., Kurtz, J., Reusch, T.B.H. & Milinski, M. 2004. Response to Comment on "Parasite Selection for Immunogenetic Optimality". *Science* **303:** 957b.

Wegner, K.M.e.a., Wegner, K.M., Kalbe, M., Kurtz, J., Reusch, T.B.H. & Milinski, M. 2003. Parasite selection for immunogenetic optimality. *Science* **301:** 1343.

Weidt, G., Deppert, W., Utermohlen, O., Heukeshoven, J. & Lehmann-Grube, F. 1995. Emergence of virus escape mutants after immunization with epitope vaccine. *Journal of Virology* **69:** 7147–7151.

Weisfeld, G.E., Czilli, T., Phillips, K.A., Gall, J.A. & Lichtman, C.M. 2003. Possible olfaction-based mechanisms in human kin recognition and inbreeding avoidance. *Journal of Experimental Child Psychology* **85:** 279–295.

Westerdahl, H. 2004. No evidence of an MHC-based female mating preference in great reed warblers. *Molecular Ecology* **13:** 2465–2470.

Westerdahl, H., Waldenström, J., Hansson, B., Hasselquist, D., Schantz, T.r.v. & Bensch, S. 2005. Associations between malaria and MHC genes in a migratory songbird. *Proceedings of the Royal Society B* **272:** 1511–1518.

Westerdahl, H., Asghar, M., Hasselquist, D. & Bensch, S. 2012. Quantitative disease resistance: to better understand parasite-mediated selection on major histocompatibility complex. *Proceedings of the Royal Society B* **279:** 577–584.

Wilson, J.S., Hazel, S.M., Williams, N.J., Phiri, A., French, N.P. & Hart, C.A. 2003. Nontyphoidal Salmonellae in United Kingdom Badgers: Prevalence and Spatial Distribution. *Applied and Environmental Microbiology* **67:** 4312–4315.

Woelfing, B., Traulsen, A., Milinski, M. & Boehm, T. 2009. Does intra-individual major histocompatibility complex diversity keep a golden mean? *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **364:** 117-128.

Woodroffe, R. 1995. Body condition affects implantation date in the European badger, *Meles meles*. *Journal of Zoology* **236:** 183–188.

Worley, K., Collet, J., Spurgin, L.G., Cornwallis, C., Pizzari, T. & Richardson, D.S. 2010. MHC heterozygosity and survival in red junglefowl. *Molecular Ecology* **19:** 3064–3075.

Xu, Y., Brereton, R.G., Trebesius, K., Bergmaier, I., Oberzaucher, E., Grammerc, K. & Penn, D.J. 2007. A fuzzy distance metric for measuring the dissimilarity of planar chromatographic profiles with application to denaturing gradient gel electrophoresis data from human skin microbes: demonstration of an individual and gender-based fingerprint. *Analyst* **132:** 638–646.

Yamaguchi, N., Dugdale, H.L. & Macdonald, D.W. 2006. Female receptivity, embryonic diapause, and superfetation in the European badger (*Meles meles*): implications for the reproductive tactics of males and females. *Quarterly Review of Biology* **81:** 33–48.

Yamazaki, K., Boyse, E.A., Mike, V., Thaler, H.T., Mathieson, B.J., Abbott, J., Boyse, J., Zayas, Z.A. & Thomas, L. 1976. Control of mating preference in mice by genes in the major histocompatibility complex. *The Journal of Experimental Medicine* **144:** 1324–1335.

Yamazaki, K., Yamaguchi, M., Andrews, P., Peake, B. & Boyse, E.A. 1978. Mating preferences of F2 segregants of crosses between MHC-congenic mouse strains. *Immunogenetics* **6:** 253–259.

Yamazaki, K., Beauchamp, G.K., Bard, J., Thomas, L. & Boyse, E.A. 1982. Chemosensory recognition of phenotypes determined by the Tla and H-2K regions of chromosome 17 of the mouse. *Proceedings of the National Academy of Sciences of the United States of America* **79:** 7828-7831.

Yamazaki, K., Beauchamp, G.K., Imai, Y., Bard, J., Phelan, S.P., Thomas, L. & Boyse, E.A. 1990. Odor types determined by the major histocompatibility complex in germfree mice. *Proceedings of the National Academy of Sciences USA* **87:** 8413–8416.

Yang, Z., Nielsen, R., Goldman, N. & Pedersen, A.-M.K. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155:** 431–449.

Yang, Z. 2002. Inference of selection from multiple species alignments. *Current Opinion in Genetics and Development* **12:** 688–694.

Yang, Z., Wong, W.S.W. & Nielsen, R. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution* **22:** 1107–1118.

Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24:** 1586–1591.

Yonezawa, T., Nikaido, M., Kohno, N., Fukumoto, Y., Okada, N. & Hasegaw, M. 2007. Molecular phylogenetic study on the origin and evolution of Mustelidae. *Gene* **396:** 1–12.

Yuhki, N., Beck, T., Stephens, R.M., Nishigaki, Y., Newmann, K. & O'Brien, S.J. 2003. Comparative genome organization of human, murine, and feline MHC class II region. *Genome Research* **13:** 1169–1179.

Yuhki, N., Beck, T., Stephens, R., Neelam, B. & O'Brien, S.J. 2007. Comparative genomic structure of human, dog, and cat MHC: HLA, DLA, and FLA. *Journal of Heredity* **98:** 390–399.

Zeh, J.A. & Zeh, D.W. 1996. The Evolution of Polyandry I: Intragenomic Conflict and Genetic Incompatibility. *Proceedings: Biological Sciences* **263:** 1711-1717.

Zelano, B. & Edwards, S.V. 2002. An MHC component to kin recognition and mate choice in birds: predictions, progress, and prospects. *American Naturalist* **160:** S225–237.

Zhong, J.F., Harvey, J.T. & Boothby, J.T. 1998. Characterization of a harbor seal class I major histocompatability complex cDNA clone. *Immunogenetics* **48:** 422–424.

Ziegler, A., Dohr, G. & Uchanska-Ziegler, B. 2002. Possible roles for products of polymorphic MHC and linked olfactory receptor genes during selection processes in reproduction. *American Journal of Reproductive Immunology* **48:** 34–42.

Ziegler, A., Kentenich, H. & Uchanska-Ziegler, B. 2005. Female choice and the MHC. *Trends in Immunology* **26:** 496-502.

Zomer, S., Dixon, S.J., Xu, Y., Jensen, S.P., Wang, H., Lanyon, C.V., O'Donnell, A.G., Clare, A.S., Gosling, L.M., Penne, D.J. & Brereton, R.G. 2009. Consensus multivariate methods in gas chromatography mass spectrometry and denaturing gradient gel electrophoresis: MHC-congenic and other strains of mice can be classified according to the profiles of volatiles and microflora in their scent-marks. *Analyst* **134:** 114–123.

Zuur, A.F., Ieno, E.N., N J Walker, Saveliev, A.A. & Smith, G. 2009. *Mixed Effects Models and Extensions in Ecology with R*, 1st edn. Springer, New York.

Appendix 1

Evidence for a Role of the Host-Specific Flea (*Paraceras melis*) in the Transmission of *Trypanosoma* (*Megatrypanum*) *pestanai* to the European Badger[4]

[4] This chapter has been published as: Lizundia, R., Newman, C., Buesching, C.D., Ngugi, D., Blake, D., Sin, Y.W., Macdonald, D.W., Wilson, A. & McKeever, D. 2011. Evidence for a role of the host-specific flea (*Paraceras melis*) in the transmission of *Trypanosoma (Megatrypanum) pestanai* to the European badger. *PLoS ONE* **6:** e16977.

OPEN ACCESS Freely available online

PLoS one

Evidence for a Role of the Host-Specific Flea (*Paraceras melis*) in the Transmission of *Trypanosoma (Megatrypanum) pestanai* to the European Badger

Regina Lizundia[1], Chris Newman[2], Christina D. Buesching[2], Daniel Ngugi[1], Damer Blake[1], Yung Wa Sin[2], David W. Macdonald[2], Alan Wilson[1], Declan McKeever[1]*

1 Royal Veterinary College, University of London, Hatfield, United Kingdom, **2** Wildlife Conservation Research Unit, Department of Zoology, University of Oxford, The Recanati-Kaplan Centre, Abingdon, United Kingdom

Abstract

We investigated the epidemiology of *Trypanosoma pestanai* infection in European badgers (*Meles meles*) from Wytham Woods (Oxfordshire, UK) to determine prevalence rates and to identify the arthropod vector responsible for transmission. A total of 245 badger blood samples was collected during September and November 2009 and examined by PCR using primers derived from the 18S rRNA of *T. pestanai*. The parasite was detected in blood from 31% of individuals tested. *T. pestanai* was isolated from primary cultures of Wytham badger peripheral blood mononuclear cells and propagated continually *in vitro*. This population was compared with cultures of two geographically distinct isolates of the parasite by amplified fragment length polymorphism (AFLP) and PCR analysis of 18S rDNA and ITS1 sequences. High levels of genotypic polymorphism were observed between the isolates. PCR analysis of badger fleas (*Paraceras melis*) collected from infected individuals at Wytham indicated the presence of *T. pestanai* and this was confirmed by examination of dissected specimens. Wet smears and Giemsa-stained preparations from dissected fleas revealed large numbers of trypanosome-like forms in the hindgut, some of which were undergoing binary fission. We conclude that *P. melis* is the primary vector of *T. pestanai* in European badgers.

Citation: Lizundia R, Newman C, Buesching CD, Ngugi D, Blake D, et al. (2011) Evidence for a Role of the Host-Specific Flea (*Paraceras melis*) in the Transmission of *Trypanosoma (Megatrypanum) pestanai* to the European Badger. PLoS ONE 6(2): e16977. doi:10.1371/journal.pone.0016977

Editor: David Ojcius, University of California Merced, United States of America

Received November 9, 2010; **Accepted** January 14, 2011; **Published** February 14, 2011

Copyright: © 2011 Lizundia et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The fieldwork was supported by a grant to DWM from the Peoples' Trust for Endangered Species. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: dmckeever@rvc.ac.uk

Introduction

A variety of trypanosome species are found in domesticated and free-living British mammalian fauna (Table 1). As stercorarian trypanosomes, all of these parasites undergo development and differentiation within the gut of the arthropod vector and are transmitted to their mammalian hosts contaminatively, through ingestion of the vector or its faeces. Trypanosomes of British cattle (*T. theileri*) and sheep (*T. melophagium*) are classified in the subgenus *Megatrypanum* and are transmitted respectively by tabanid flies [1] and the sheep ked [2]. In contrast, the majority of trypanosomes found in British wild mammals are grouped within the subgenus *Herpetosoma*, of which the rat parasite *T. lewisi* is the type species, and are transmitted by fleas [3]. An exception to this is *T. pestanai*, a parasite of European badgers (*Meles meles*), which is classified in the sub-genus *Megatrypanum*. The arthropod vector of *T. pestanai* is currently unknown. The convention for classification of the Stercoraria has been questioned because it is based largely on morphological parameters and host species [4]. Indeed, a number of analyses at the molecular level have indicated that both the *Herpetosoma* and *Megatrypanum* are polyphyletic [5,6]. A recent study of the evolutionary relationships of *T. rangeli*, a parasite generally accepted as belonging to the subgenus *Herpetosoma*, concluded that the use of these classifications should be discontinued [5]. Since its first description in Portugal in 1905 [7], *T. pestanai* has been reported in badgers from France [8], England [9] and Ireland [10]. The prevalence of the parasite in a badger population resident in Wytham Woods, Oxfordshire, has been investigated previously through microscopic analysis of blood smears [11] where seasonal and age-related differences were observed. However, interpretation of these observations has been confounded by the lack of information on the transmission vector. A number of blood-feeding ectoparasites are found on badgers, including the flea *Paraceras melis* and tick species such as *Ixodes hexagonus*, *I. ricinus* [12] and *I. canisuga*, although ticks are rarely present on animals trapped at Wytham (C. Newman, unpublished observations). In contrast, *P. melis* is highly prevalent among Wytham badgers - and badgers generally [13] - with some animals experiencing substantial infestations [14]. Given the prominent role of flea species in transmission of *Herpetosoma* trypanosomes of other British wild fauna, these observations present *P. melis* as a compelling candidate vector for *T. pestanai*. We have therefore examined the role of *P. melis* in transmission of *T. pestanai* between badgers, using a PCR-based parasite detection system in association with morphological analysis of fleas collected from PCR^{+ve} badgers. We investigated whether the flea supports development of the insect stages of the parasite which would indicate that it represents the principal transmission vector. The

Table 1.

Mammalian host	Trypanosome	Sub-genus	Vector	Reference
Wood mouse *(Apodemus sylvaticus)*	*T. grosi*	*Herpetosoma*	Flea	[34]
Bank vole *(Clethrionomys glareolus)*	*T. evotomys*	*Herpetosoma*	Flea	[34,35]
Rat (*Rattus* sp)	*T. lewisi*	*Herpetosoma*	Flea	[18,36]
Field vole *(Microtus agrestis)*	*T. microti*	*Herpetosoma*	Flea	[37]
Rabbit (*Oryctolagus cuniculus*)	*T. nabiasi*	*Herpetosoma*	Flea	[38]
European Badger *(Meles meles)*	*T. pestanai*	*Megatrypanum*	Unknown	[7]
Cow *(Bos Taurus)*	*T. theileri*	*Megatrypanum*	Tabanid flies	[1]
Sheep *(Ovis aries)*	*T. melophagium*	*Megatrypanum*	Sheep ked	[2]
Mole (*Talpa europaea*)	*T. talpae*	*Megatrypanum*	Mite	[39]

doi:10.1371/journal.pone.0016977.t001

use of PCR techniques also allowed us to extend our previous observations of *T. pestanai* prevalence in Wytham badgers, by achieving higher levels of sensitivity. In addition, we also investigated whether genetic diversity exists between geographically distinct isolates of *T. pestanai*.

Results

Prevalence of *Trypanosoma pestanai* and dynamics of infection and transmission

In total, 245 blood samples were collected from 207 badgers during trapping sessions in September and November 2009. DNA extracted from each blood sample was analysed by PCR using primers (TPEF1, TPEB1) derived from the 18S rRNA of *T. pestanai*. A product of the expected size (~513 bp) was amplified in the infected samples. Of the 245 blood samples screened, 78 samples tested positive (31% prevalence) for *T. pestanai*. Of 207 individual badgers analysed, 170 were trapped only once, 36 were trapped twice and 1 was trapped 3 times. The PCR prevalence of *T. pestanai* infection in individual badgers (no repeats) from the first trapping was 29.3%. To study the dynamics of infection and transmission of *T. pestanai*, several parameters were evaluated: trapping session, sex, age, body condition, location (sett) and number of fleas. No significant differences were observed between prevalence rates for the September and November trapping sessions (36% and 28% prevalence respectively, $p = 0.289$, Pearson Chi-Square test). However, the prevalence of *T. pestanai* was significantly higher in males (42%) than in females (27%) ($p = 0.025$, Chi-Square test). Cubs (less than one year old) had significantly higher rates of infection (40% prevalence) than both young adults (1–5 years old, 35% prevalence) and mature adults (more than 5 years old, 16% prevalence) ($p = 0.04$, linear by linear association test). After adjusting for age, evidence for an association between sex and the presence of *T. pestanai* in blood was apparent in a multivariable logistic regression analysis ($p = 0.033$). Males were twice as likely to be infected as females (OR = 1.90; 95% CI: 1.05–3.43). After adjusting for sex in a multivariable logistic regression test, there was evidence for a difference in infection rate between cubs and adults ($p = 0.04$), with adults being more resistant to infection (OR: 0.29; 95% CI: 0.093–0.916). No evidence of association between body condition and prevalence of *T. pestanai* in blood was observed ($p = 0.563$, Linear by linear association test). Similarly, no evidence was found for an association between flea burden and presence of *T. pestanai* in blood ($p = 0.122$, Linear by linear association test). To examine the dynamics of *T. pestanai* infection over time, blood samples from 36 badgers that were caught in both trapping sessions were examined by PCR. Of these, 18 (48%) were negative on both occasions, and 9 (24%) showed persistent infection (or concurrent recrudescence of infection) across trappings. Four badgers observed to be infected in September tested negative in November (10%), while 5 animals that were negative in September had become infected by November (13%). These data are consistent with a cyclical pattern of *T. pestanai* prevalence.

Isolation of *T. pestanai* and morphological characteristics of axenic cultures

Live motile *T. pestanai* parasites were invariably observed in cultures of peripheral blood mononuclear cells established from PCR^{+ve} blood samples. Moreover, these parasites continued to multiply under the culture conditions used, often giving rise to rosette-like aggregates (Figure 1). Giemsa-stained smears (Figure 2) illustrate characteristic trypanosome features (e.g. kinetoplast and flagellum) observed in cultured *T. pestanai* parasites. A variety of parasite morphologies were observed, including slender (Fig. 2A), broad and intermediate forms (Fig. 2B), and parasites undergoing division/binary fission (Fig. 2C) and degeneration as manifested by transformation to a spherical form with granular degeneration of the protoplasm (Fig. 2D). All three *T. pestanai* isolates (East Anglia, Oxford and France) showed similar morphologies in culture.

Genetic characterisation of three geographically distinct cultured *T. pestanai* isolates

Total DNA extracted from cultures of three geographically distinct *T. pestanai* isolates was analysed by PCR (Fig. 3). PCR analysis using primers derived from the 18S rRNA of *T. pestanai* resulted in amplification of an identical band from all three isolates when using TPEF1/ TPEB1 (Fig. 3A) and TPEF2/TPEB2 (Fig. 3B) primers. However, PCR analysis using primers specific for the ITS1 sequence (KIN1, KIN2) revealed different size bands (Fig. 3C). More detailed genetic characterization using AFLP revealed clear genetic polymorphism between all three isolates (Figure S1). Four selective AFLP primer combinations were used, yielding 56 markers, 41 of which were polymorphic for one or more isolates. The Jaccard index of similarity ranged from 34 to 64%, indicating elevated levels of genetic heterogeneity among the isolates. The highest coefficient of similarity was found between the France and East Anglia isolates (64%), followed by the Oxford and East Anglia isolates (48%) and by the Oxford and France isolates (34%).

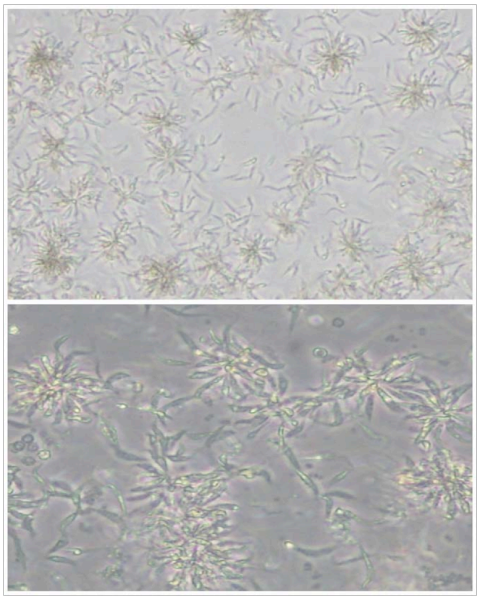


Figure 1. *T. pestanai* (Oxford isolate) in axenic culture. Formation of rosettes as a result of incomplete separation of daughter cells observed by inverted phase contrast microscopy at different magnifications (upper panel, ×20; lower panel, ×40).
doi:10.1371/journal.pone.0016977.g001

Detection of robust IgG responses of badgers against the parasite

Western blot analysis of sera from infected (PCR^{+ve} in blood) and uninfected (PCR^{-ve} in blood) badgers showed a broadly specific IgG response against *T. pestanai* lysates without exception, with male and female animals showing similar breadth of response (Fig. 4A). Antigenic differences were evident between the Oxford and East Anglia isolates when probed with individual badger sera (Fig. 4B). No seronegative badgers were observed over the trapping period.

Detection of *T. pestanai* in badger fleas *(P. melis)*

PCR amplification was carried out on DNA extracted from 26 individual fleas collected from infected badgers (n = 18) and 12 fleas from uninfected individuals (n = 8) using the *T. pestanai*-specific primers TPEF1 and TPEB1. Of those collected from infected badgers, 16 fleas (61%) tested positive for *T. pestanai*, while 2 (16%) of those from uninfected animals yielded a PCR product. The difference in *T. pestanai* PCR prevalence between fleas collected from infected versus uninfected badgers was significant ($p = 0.015$), as revealed by a Fisher's exact test. Similarly, numbers of infected fleas obtained from infected badgers were significantly higher ($p = 0.026$) than those found on uninfected individuals. To confirm that *T. pestanai*-specific PCR products observed in a proportion of fleas reflects development of the parasite within the insect rather than simple contamination of a blood meal, a number of fleas collected from infected badgers were dissected and their gut tissues examined by phase contrast microscopy. In a proportion of these fleas, large numbers of trypanosomes were observed in the hind gut. These were motile and in some cases exhibited the same rosette formation observed in *in vitro* cultures (Fig. 5A and 5B).

Discussion

Available literature on *T. pestanai* is sparse and, like many of the *Megatrypanum* trypanosomes, the parasite has been classified on the basis of its morphological appearance in blood smears [9]. Criteria applied in this regard include large size, a small kinetoplast located close to the nucleus, and a pointed posterior end. Such parameters are of questionable value where parasitaemias are low and the biology of the parasite is incompletely understood. For example, our measurements of cultured *T. pestanai* indicate an overall length of 25.33–33.33 μm ($n = 82$) whereas Pierce & Neal (1974) recorded values of 25.6–41.4 μm in blood smears. In addition, the extent to which morphological classification reflects true evolutionary relationships is in some doubt. Hence, published information on trypanosome 18S rRNA phylogenies generally places *T. pestanai* apart from *T. theileri*, the type species of subgenus *Megatrypanum* [5,15], with its closest relative being a species isolated from a wombat. Indeed, *Megatrypanum* trypanosomes fail to form discrete clusters under this type of analysis, distributing instead in apparently unrelated phylogenetic groupings [16]. We now provide compelling evidence that the badger flea *P. melis* is the definitive invertebrate host of *T. pestanai*. This ectoparasite is highly prevalent among badgers from this focal Wytham Woods badger population [11] and its host specificity is consistent with an evolutionary relationship with *T. pestanai*. In particular, badgers exhibit extensive grooming habits to control fleas [17], providing continuous opportunity for ingestion of infected fleas and their faeces. Badgers that tested positive for *T. pestanai* by PCR were consistently found to harbour fleas that also yielded PCR products. This is unlikely to reflect simple carry-over of parasites in the blood meal, given the low parasitaemias observed in badgers [11] and the sensitivity of the PCR system used. Trypanosomes of mammals will be considered under two major groups, Salivarian or Stercorarian, based on whether they undergo anterior station (foregut) development or posterior station (hindgut) development in the insect vector, respectively. Indeed, where trypanosomes were found in dissected fleas from badgers, these were located in the hindgut rather than the foregut and, in many instances, formed rosette-like structures characteristic of replicating *T. pestanai* cultures. These parasites were distinct from crithidial forms and gregarine protists that were also present in a proportion of dissected fleas (data not shown). Precise details of the life cycle of *T. pestanai* in the insect vector remain to be determined. However, by analogy with *T. lewisi*, it is likely that it undergoes a period of cyclic development in the flea to produce metacyclic forms infective for the mammalian host [3][18]. After ingestion by the flea, trypomastigote forms of *T. lewisi* penetrate the epithelial cells of the stomach of the flea and replicate. Upon rupture of the infected cell, daughter trypomastigotes enter the lumen of the flea stomach and migrate to the hindgut and rectum, where they transform into epimastigotes. These undergo further multiplication, producing large numbers of infective metacyclic forms in the rectum, which are discharged in the faeces. The mammalian host becomes infected contaminatively by ingestion of flea faeces or intact fleas. *Megatrypanum* parasites are transmitted by a diverse range of vectors, including tabanid flies (*T. theileri*), keds (*T. melophagium*) and mites (*T. talpae*) (see Table 1). Our evidence that *T. pestanai* is transmitted by *P. melis* fleas makes it unique among the *Megatrypanum* trypanosomes and is perhaps indicative of a

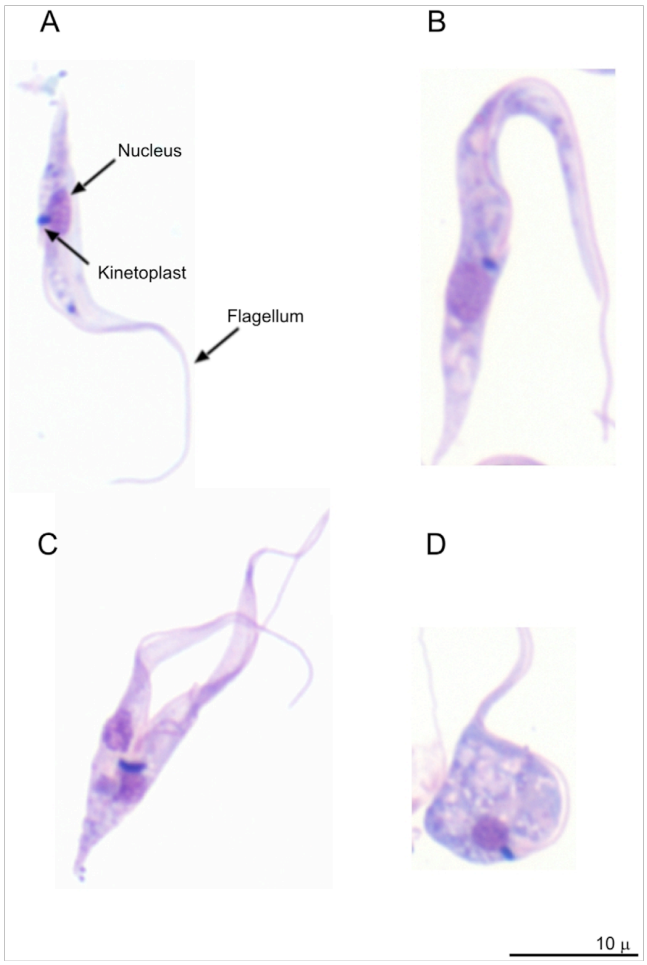


Figure 2. Giemsa-stained smears showing different *T. pestanai* forms in axenic culture. (**A**) Epimastigote-like long slender form. (**B**) Epimastigote-like swollen form. (**C**) Dividing epimastigote. (**D**) Large pear-shaped form (''degenerative'' form).
doi:10.1371/journal.pone.0016977.g002

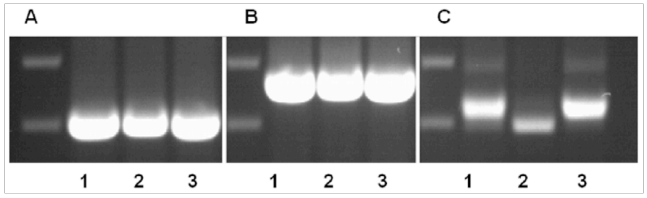


Figure 3. Comparison by PCR of 3 geographically distinct isolates of *T. pestanai*. DNA extracted from different *T. pestanai* isolates in axenic culture (1. East Anglia isolate; 2. Oxford isolate; 3. France isolate) was analysed by PCR using primers derived from the 18S rRNA (A. TPEF1/B1 primers; B. TPEF2/B2 primers) and ITS1 sequences of *T. pestanai*. (C. KIN1/KIN2 primers).
doi:10.1371/journal.pone.0016977.g003

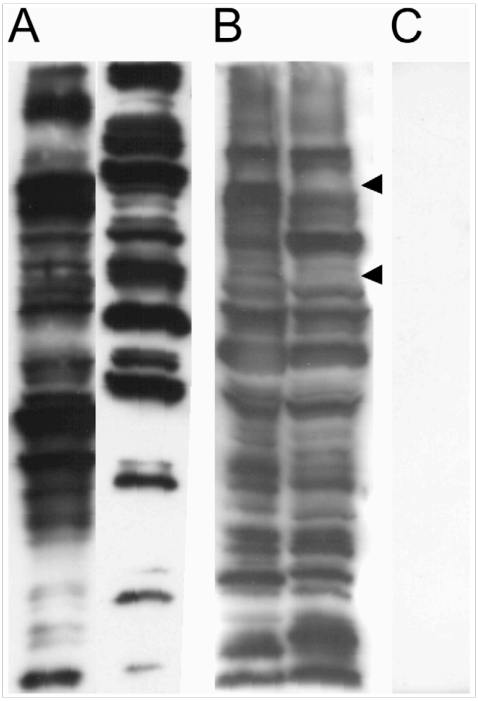


Figure 4. Western blot showing a robust IgG response of badgers against *T. pestanai* lysates. (A, left) Female badger serum, PCR negative in blood. (A, right) male badger serum, PCR positive in blood. (B, left) badger serum response against East Anglia isolate. (B, right) same badger serum response against Oxford isolate. (C) Western blot in the absence of badger serum (HRP-conjugated anti-badger IgG antibody only).
doi:10.1371/journal.pone.0016977.g004

closer relationship with those parasites currently classified within the subgenus *Herpetosoma*. Although recent phylogenetic analysis based on 18S rRNA and the glyceraldehyde phosphate dehydrogenase (GAPDH) gene [15,19] are consistent with this possibility, the precise relationships remain unclear. Macdonald *et al.* [11] observed trypanosomes in only 33 (4.6%) of 718 blood smears collected from the Wytham badgers between 1989–1991. Of the 263 badgers examined during that period, only 20 (7.7%) yielded at least one positive test. This was not considered to be an absolute reflection of prevalence, as detection of *T. pestanai* in blood smears is difficult, especially where parasitaemia is low. Indeed, *T. pestanai* was isolated in culture from four blood samples that were negative by blood smear analysis in that study. The PCR-based methodologies used in the present study clearly provide a more reliable and precise diagnostic tool and reveal a considerably higher infection prevalence in the study population. Significantly higher prevalence of *T. pestanai* infection was observed in male badgers than in females. Several host-parasite systems, including those of humans, exhibit sex-related differences [20]. Although either sex can show a higher rate of infection, many studies suggest that females mount a more effective immune response [21,22,23]. Testosterone has been observed to enhance susceptibility of rats to *T. cruzi*, as evidenced by higher parasitaemia associated with reduced T cell responsiveness [24]. It is therefore possible that the gender bias observed in *T. pestanai* prevalence among the Wytham badgers relates to a hormonal influence on immunity. However, we observed broad specificity of serum IgG responses in both male and female badgers in immunoblots of *T. pestanai* lysates. Furthermore, studies of the coccidian parasites *Eimeria melis* and *Isospora melis* in the Wytham badger population revealed no evidence that prevalence, and hence immune-susceptibility or risk of exposure to either parasite species varied with gender at any stage of maturity [25,26]. Given the communal sleeping habits of badgers, and the frequency with which fleas move between hosts [14], we consider it unlikely that the differences in *T. pestanai* prevalence between males and females arises from exposure to the vector. However, Macdonald *et al.* (2008) [27] observed within a high density population of badgers that males move more between groups than do females. Dispersing males tended to move to larger groups and to groups with a preponderance of females. This bias was influenced by season, occurring more in autumn and spring. Our data are derived from material collected from the Wytham badgers during the autumn trapping, when higher rates of

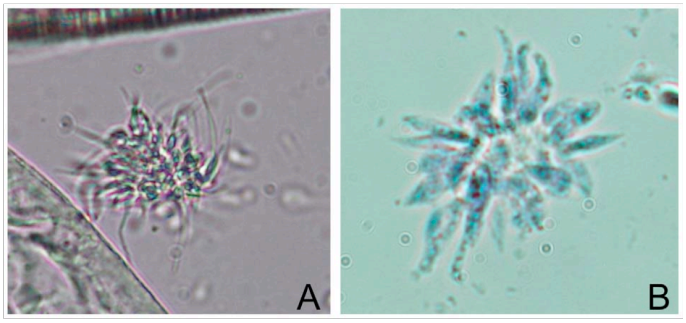


Figure 5. Wet-smear of *T. pestanai* detected in the flea of infected badger. (A) Presence of live motile *T. pestanai* parasites in the hindgut of an infected flea observed by phase contrast microscopy (×40). (B) Detection of rosette in the hindgut of an infected flea observed by phase contrast microscopy (×100).
doi:10.1371/journal.pone.0016977.g005

movement would be expected among males. It is therefore possible that the male bias in *T. pestanai* infection prevalence relates to badger dispersal and flea exchange. Under these circumstances, males might be expected to pick up fleas from other badger groups, enhancing their potential for infection with *T. pestanai*. We also observed a strong association between *T. pestanai* infection and age, with cubs showing substantially higher prevalence rates than adults. This is likely to arise from exposure to fleas. Cubs spend more time in the sett, where they are exposed to contaminated bedding [12] and as a result of close contact with their mother during suckling and grooming. Our data also indicate that Wytham badgers undergo multiple infections with *T. pestanai* over time and that they can remain infected for prolonged periods. Some individuals remained PCR-positive over two trapping exercises, while others appeared to lose the parasite during the interval between trappings. In addition, the complexity of the patterns observed with badger sera on immunoblots of *T. pestanai* lysate appeared to increase with age (data not shown). Indeed, no seronegative badgers were observed over the trapping. These observations are consistent with a model based on repeated infections that perhaps evolve to a prolonged or intermittent carrier state. Precise resolution of this situation will await further evaluation in the Wytham badgers and identification of appropriate molecular markers. Our preliminary observations with AFLP and the 18S rRNA and ITS1 sequences are consistent with genetic polymorphism between isolates of *T.pestanai*. Indeed, the two UK populations, although moderately disparate in origin (Oxfordshire and East Anglia), are sufficiently diverse to classify as distinct subspecies. Acknowledging that these data are derived from small sample numbers, they do suggest that *T. pestanai* is genotypically diverse. The origins of this diversity are unclear, in the absence of information on sexual recombination in the flea vector, but it may relate to maintenance of the parasite in the face of what appears to be a robust IgG response in infected badgers. Such diversity confounds interpretations of prevalence data based on relatively conserved 18S rRNA sequences. In particular, elucidation of whether persistence of the parasite as revealed by 18S rRNA PCR relates to repeated infection or prolonged carrier status will require a more refined set of genotypic markers.

In conclusion, we have determined the prevalence of *T. pestanai* infection in Wytham Woods badgers using PCR of 18S rRNA sequences. We observe considerably higher prevalence than that previously reported on the basis of blood smear examination. In addition we report higher prevalence in male badgers and in those <1 year old. We provide compelling evidence that *T. pestanai* in this population is transmitted by the badger flea *P. melis* and further show that badgers mount a vigorous antibody response against the parasite. Finally, we reveal that *T. pestanai* parasites are genotypically diverse with substantial variation being evident between isolates derived from relatively adjacent locations. We propose that this diversity is driven in part by the badger immune response.

Materials and Methods

2.1. Ethics Statement

This study was approved by Natural England and carried out under Natural England Licenses, currently 20001537 and Home Office License PPL 30/2318. All trapping and handling procedures were in accord with the UK Animals (Scientific Procedures) Act, 1992, and approved by the institutional ethical review committee.

2.2. Study population

All samples were collected between September and November 2009 from a badger population at Wytham Woods; a 424ha mixed semi-natural woodland site, in Oxfordshire, UK (GPS ref: 51:46:26N; 1:19:19W). This population has been studied continuously since the 1970s [28,29] and currently comprises ca. 220 adults with ca. 50 cubs per year. As part of ongoing long-term monitoring studies, the population is trapped 1–4 times annually using cage-traps baited with peanuts set in the vicinity of each sett. Upon capture, badgers are transported to a central handling facility and sedated by intramuscular injection of 0.2ml ketamine hydrochloride (100 mg ml^{-1}) per kg body weight [30]. At first capture (generally as cubs) every individual is marked with a permanent tattoo number in the left inguinal region. For each badger, the location of capture ('sett'), sex and age is recorded. Morphometric measures recorded include weight (kg), length (tip of snout to base of sacrum - mm) and body-condition, using a subcutaneous fat score originally developed for sheep on a scale of 1 = emaciated to 5 = very good condition. For the present study, the number of fleas was also counted at each capture during a cursory examination over ~20 sec interval, by parting the fur and examining the badger's back and underside [14]. Blood samples were taken by jugular venipuncture using a potassium-EDTA vacutainer (BD Vacutainer Systems, Plymouth, UK) and stored at 4°C before being processed in the laboratory within 72h of collection. After sampling, badgers were allowed to recover fully before being released at the site of capture later in the same day.

2.3. Parasite isolation from blood

Badger peripheral blood mononuclear cells (PBMC) were isolated from uncoagulated blood by Ficoll density gradient separation. Briefly, 3 ml of Ficoll were underlaid with approximately 3 ml of uncoagulated blood. After 30 min centrifugation at 700×g the white blood cell layer was collected carefully and washed twice in Alsever's solution (200×g for 10 min). The PBMC cell pellet was re-suspended in Schneider's medium containing 20% FCS and incubated in 24-well plates at 28°C in a humidified atmosphere of 5% CO_2 in air.

The *T. pestanai* France-isolate (LEM 110) was kindly provided by Dr. Wendy Gibson (University of Bristol, UK) while *T. pestanai* East Anglia isolate was isolated from a blood sample kindly provided by Dr. Mark A. Chambers (Veterinary Laboratories Agency, Addlestone, Surrey, UK).

2.4. DNA extraction

Total DNA was extracted from blood samples (300 µl of whole blood per badger) using the Wizard Genomic DNA Purification kit (Promega) following the manufacturer's recommendations. Total DNA from individual fleas was extracted using a DNeasy blood and tissue kit (Qiagen). Each flea was soaked for 5 min in 200 µl of phosphate buffered saline in a petri-dish and crushed with the plunger of a disposable plastic syringe. After a brief centrifugation (16000×g for 1 min) the supernatant was transferred into a new sterile eppendorf tube containing 20 µl of Proteinase K (Fermentas) and 200 µl of buffer AL (Qiagen). DNA extraction was then carried out in line with the kit instructions.

2.5. PCR

Primers designed to target the *T. pestanai* 18S rRNA coding sequence (Accession no: AJ009159; TPE primers) and kinetoplastid ITS1 sequences (KIN) were used for PCR amplification [31].

TPEF1:5′-CCATGCATGCCTCAGAATCACTGC-3
TPEB1: 5′-GGCACTGCCGGCTCTATTTC-3′
TPEF2: 5′-GCAGCGAAAAGAAATAGAGCCGG-3′
TPEB2: 5′-GTTCGTCCTGGTGCGGTCTAA-3′
KIN1: 5′-GCGTTCAAAGATTGGGCAAT-3′
KIN2: 5′-CGCCCGAAAGTTCACC-3′

PCR amplifications were performed in a total volume of 20 µl containing 2 µl of 10× NH_4 PCR buffer (Bioline), 1 µl of 50 mM $MgCl_2$ (Bioline), 0.5 µl of 10 mM dNTP (Bioline), 4 µl of primer mix (Forward and Reverse at 10 pmol each), 0.2 µl (1 unit) of *Taq*DNA polymerase (Bioline), 2 µl of DNA template and 10.3 µl of distilled water. The reaction profile included an initial denaturation step at 94°C for 10 min, followed by 40 cycles of 45 seconds at 94°C, 1 min at 63°C, 1 min at 72°C and a final step of 7 min at 72°C, using a G-STORM thermal cycler. PCR products were electrophoresed for 75 min at 100V in a 1% Tris-acetic acid-EDTA (TAE) agarose gel containing 1× Safeview Nucleic Acid Stain (NBS Biologicals) for visualization. O'GeneRuler 1kb DNA ladder (Fermentas) was used for sizing the DNA fragments in agarose gels.

2.6. Parasite protein lysate preparation

T. pestanai parasites (5×10^7) from axenic culture were washed twice in PBS by centrifugation at 700×g for 10 min. The parasite pellet was resuspended in 80 µl of M-PER lysis buffer (Pierce) containing 1× Halt protease/phosphatase inhibitor cocktail (Pierce). The mixture was vortexed and incubated for 20 min at RT. After centrifugation at 16000×g for 15 min, the supernatant was transferred into a new microcentrifuge tube containing 20 µl of 5× Laemmli buffer and boiled at 90°C for 3 min.

2.7. Western blotting

15 µl of protein lysate were loaded in wells of a precast 10% polyacrylamide Tris HCl gel (Biorad) and electrophoresed at 30 mA for 1h. Proteins were then transferred to a 0.2-µm nitrocellulose membrane (Amersham Biosciences) at 100V for 60 min. The membrane was incubated in blocking solution (PBS containing 0.05% Tween 20 and 5% milk powder) for 1h at RT followed by overnight incubation with badger serum (1:100 in blocking solution). Excess antibody was removed by extensive washing in PBS containing 0.05% Tween 20. The membrane was then incubated with HRP-conjugated anti-badger IgG (clone CF2, kindly provided by Dr. Mark Chambers, VLA) diluted 1:500 in blocking buffer, for 90 min at room temperature. The membranes were washed extensively in PBS 0.05% Tween 20, and bands were visualized using the ECL system (Amersham Biosciences) and a Curix 60 processor (Agfa-Gevaert N.V., Mortsel, Belgium).

2.8. Amplified fragment length polymorphism (AFLP)

Approximately 200 ng total *T. pestanai* genomic DNA was digested using the enzymes *Eco* RI and *Mse* I (Fermentas) prior to ligation to adapters as described previously [32,33]. Primer pairs (MWG Biotech (UK) Ltd.) were based on the adaptor sequences. Pre-amplification was performed using the *Eco* RI associated primer with no selective base and the *Mse* I associated primer with a single selective base (A). Selective amplifications were performed with primers including two selective bases (*Eco* RI primer −AC+*Mse* I primer −AC; *Eco* RI primer −AC+*Mse* I primer −AT; *Eco* RI primer −CG+*Mse* I primer −AC; *Eco* RI primer −CG+*Mse* I primer −AT). The selective *Eco* RI primers were labelled with $^{33}P\gamma ATP$ using T4 polynucleotide kinase (Invitrogen). AFLP products were resolved by denaturing polyacrylamide gel electrophoresis (6% acrylamide, UreaGel 6, National Diagnostics) and visualised by autoradiography. Electrophoretic patterns were converted into binary matrices (1 for presence, 0 for absence of a band) and used for calculation of the Jaccard index for each pair-wise comparison [33] (calculated as the number of common bands/the total number of bands ×100, to quantify sampled genetic similarity as a percentage).

2.9. Flea dissection and staining

Individual fleas collected from infected and uninfected badgers were dissected under a dissecting-microscope. Each flea was placed on a glass slide in a drop of phosphate buffered saline and transected at the thoraco-abdominal junction with a scalpel blade. The head and thorax were then discarded and the abdomen was immobilised with a fine needle and opened by cutting along the midline of the dorsal tergites with iris scissors. The abdominal contents were then carefully removed with the aid of fine forceps, and the salivary glands, fat body and (in the case of females) uterine tissues were dissected away and discarded. The residual tissues comprising proventriculus, foregut, intestine, malphigian tubules and hindgut were arranged on the slide before covering the preparation with a coverslip. Slides were examined at 100× magnification under phase contrast to identify trypanosome forms and images were obtained using an Olympus CX41 microscope mounted camera (Olympus DP20).

2.9. Statistical analysis

Pearson Chi-square test and Fisher's exact test were used to assess associations between categorical variables and prevalence of infection in blood. Linear by linear association Chi-square test was used to assess association between ordinal variables and prevalence of infection in blood. Logistic regression models were employed to estimate the effects of sex, age groups, body condition scores, number of fleas, trapping session and risk of blood infection. Odds ratio (OR) and 95% confidence interval (CI) are presented. All analyses were carried out using Stata 9 software package (StataCorp, Texas, USA).

Supporting Information

Figure S1 AFLP fingerprints generated from DNA samples of 3 geographically distinct isolates of *T. pestanai* (1: France isolate; 2: Oxford isolate; 3: East Anglia isolate) with 4 different primer combinations. The selective *Eco*RI (E) primers and *Mse*I (M) primers included two added bases (either −AC, −AT or −CG).
(TIF)

Acknowledgments

We are grateful to Geetha Annavi and Stephen Ellwood for assistance with badger trapping for this study.

Author Contributions

Conceived and designed the experiments: DM AW DB. Performed the experiments: RL DN DM DB. Analyzed the data: RL DM DB. Contributed reagents/materials/analysis tools: DM CN CB DB DWM YWS. Wrote the manuscript: DM RL CN CB.

References

1. Bose R, Heister NC (1993) Development of *Trypanosoma* (*M.*) *theileri* in tabanids. J Eukaryot Microbiol 40: 788–792.
2. Hoare CA (1923) An experimental study of the sheep-trypanosome (*T. melophagium Flu*, 1908), and its transmission by the sheep-ked (*Melophagus ovinus L.*). Parasitology 15: 365–424.
3. Molyneux DH (1976) Biology of trypanosomes of the subgenus *Herpetosoma*. In: Lumsden WHR, Evans DA, eds. Biology of the *Kinetoplastida*. London: Academic Press. pp 285–325.
4. Hoare CA (1972) The trypanosomes of mammals. Oxford: Blackwell Scientific Publications. pp 3–29.

5. Stevens JR, Teixeira MM, Bingle LE, Gibson WC (1999a) The taxonomic position and evolutionary relationships of *Trypanosoma rangeli*. Int J Parasitol 29: 749–757.
6. Stevens JR, Noyes HA, Schofield CJ, Gibson W (2001) The molecular evolution of *Trypanosomatidae*. Adv Parasitol 48: 1–56.
7. Bettencourt A, Franca C (1905) Sur un trypanosome du blaireau (*Meles taxus* Schr.). Arch Inst R Bacteriol 1: 73–78.
8. Rioux JA, Albaret JL, Bres A, Dumas A (1966) Presence of *Trypanosoma pestanai* Bettencourt and Franca, 1905, in badgers from the south of France. Ann Parasitol Hum Comp 41: 281–288.
9. Pierce MA, Neal C (1974) *Trypanosoma* (*Megatrypanum*) *pestanai* in British badgers (*Meles meles*). Int J Parasitol 4: 439–440.
10. McCarthy G, Shiel R, O'Rourke L, Murphy D, Corner L, et al. (2009) Bronchoalveolar lavage cytology from captive badgers. Vet Clin Pathol 38: 381–387.
11. Macdonald DW, Anwar M, Newman C, Woodroffe R, Johnson PJ (1999) Inter-annual differences in the age-related prevalences of *Babesia* and *Trypanosoma* parasites of European badgers (*Meles meles*). Journal of Zoology 247: 65–70.
12. Butler JM, Roper TJ (1996) Ectoparasites and sett use in European badgers. Animal Behaviour 52: 621–629.
13. Neal EG, Cheeseman C (1996) Badgers. London: T & AD Poyser.
14. Cox R, Stewart PD, Macdonald DW (1999) The ectoparasites of the European badger, *Meles meles*, and the behavior of the host-specific flea, *Paraceras melis*. J Insect Behav 12: 245–265.
15. Hamilton PB, Gibson WC, Stevens JR (2007) Patterns of co-evolution between trypanosomes and their hosts deduced from ribosomal RNA and protein-coding gene phylogenies. Molecular Phylogenetics and Evolution 44: 15–25.
16. Stevens JR, Noyes HA, Dover GA, Gibson WC (1999b) The ancient and divergent origins of the human pathogenic trypanosomes, *Trypanosoma brucei* and *T. cruzi*. Parasitology 118(Pt 1): 107–116.
17. Stewart PD, Bonesi L, Macdonald DW (1999) Individual differences in den maintenance effort in a communally dwelling mammal: the Eurasian badger. Anim Behav 57: 153–161.
18. Minchin EA, Thomson JD (1911) The transmission of *Trypanosoma lewisi* by the rat-flea (*Ceratophyllus Fasciatus*). Br Med J 1: 1309–1310.
19. Hamilton PB, Stevens JR, Gidley J, Holz P, Gibson WC (2005) A new lineage of trypanosomes from Australian vertebrates and terrestrial bloodsucking leeches (*Haemadipsidae*). International Journal for Parasitology 35: 431–443.
20. Brabin L, Brabin BJ (1992) Parasitic infections in women and their consequences. Adv Parasitol 31: 1–81.
21. Schuurs AH, Verheul HA (1990) Effects of gender and sex steroids on the immune response. J Steroid Biochem 35: 157–172.
22. Olsson M, Wapstra E, Madsen T, Silverin B (2000) Testosterone, ticks and travels: a test of the immunocompetence-handicap hypothesis in free-ranging male sand lizards. Proc Biol Sci 267: 2339–2343.
23. Weinstein Y, Ran S, Segal S (1984) Sex-associated differences in the regulation of immune responses controlled by the MHC of the mouse. J Immunol 132: 656–661.
24. Filipin Mdel V, Caetano LC, Brazao V, Santello FH, Toldo MP, et al. (2010) DHEA and testosterone therapies in *Trypanosoma cruzi*-infected rats are associated with thymic changes. Res Vet Sci 89: 98–103.
25. Newman C, Macdonald DW, Anwar MA (2001) Coccidiosis in the European badger, *Meles meles* in Wytham Woods: infection and consequences for growth and survival. Parasitology 123: 133–142.
26. Anwar MA, Newman C, MacDonald DW, Woolhouse ME, Kelly DW (2000) Coccidiosis in the European badger (*Meles meles*) from England, an epidemiological study. Parasitology 120(Pt 3): 255–260.
27. Macdonald DW, Newman C, Buesching CD, Johnson PJ (2008) Male-biased Movement in a High-density Population of the Eurasian Badger (*Meles meles*). Journal of Mammalogy 89: 1077–1086.
28. Kruuk H (1978a) Spatial organisation and territorial behaviour of the European Badger *Meles meles*. Journal of Zoology 184: 1–19.
29. Kruuk H (1978b) Foraging and spatial organisation of the European badger *Meles meles* L Behavioural. Ecology and Sociobiology 4: 14.
30. Thornton PD, Newman C, Johnson PJ, Buesching CD, Baker SE, et al. (2005) Preliminary comparison of four anaesthetic techniques in badgers (*Meles meles*). Vet Anaesth Analg 32: 40–47.
31. Thekisoe OM, Honda T, Fujita H, Battsetseg B, Hatta T, et al. (2007) A trypanosome species isolated from naturally infected *Haemaphysalis hystricis* ticks in Kagoshima Prefecture, Japan. Parasitology 134: 967–974.
32. Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, et al. (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23: 4407–4414.
33. Blake DP, Smith AL, Shirley MW (2003) Amplified fragment length polymorphism analyses of *Eimeria spp.*: an improved process for genetic studies of recombinant parasites. Parasitol Res 90: 473–475.
34. Noyes HA, Ambrose P, Barker F, Begon M, Bennet M, et al. (2002) Host specificity of *Trypanosoma* (*Herpetosoma*) species: evidence that bank voles (*Clethrionomys glareolus*) carry only one *T.* (*H.*) *evotomys* 18S rRNA genotype but wood mice (*Apodemus sylvaticus*) carry at least two polyphyletic parasites. Parasitology 124: 185–190.
35. Molyneux DH (1969) The morphology and biology of *Trypanosoma* (*Herpetosoma*) evotomys of the bank-vole, *Clethrionomys glareolus*. Parasitology 59: 843–857.
36. Strickland C (1911) The mechanism of transmission of *Trypanosoma lewisi* from rat to rat by the rat flea. Br Med J 1: 1049.
37. Molyneux DH (1969) The morphology and life-history of *Trypanosoma* (*Herpetosoma*) *microti* of the field-vole, *Microtus agrestis*. Ann Trop Med Parasitol 63: 229–244.
38. Hamilton PB, Stevens JR, Holz P, Boag B, Cooke B, et al. (2005) The inadvertent introduction into Australia of *Trypanosoma nabiasi*, the trypanosome of the European rabbit (*Oryctolagus cuniculus*), and its potential for biocontrol. Mol Ecol 14: 3167–3175.
39. Mohamed HA, Molyneux DH, Wallbanks KR (1987) On *Trypanosoma* (*Megatrypanum*) *talpe* from *Talpa europaea*: method of division and evidence of *haemogamasinae* as vectors. J Parasitol 73: 1050–1052.

Appendix 2

The effects of pouch bacteria on the chemical composition of subcaudal gland secretions in the European badger *Meles meles*: Maintaining the balance between shared group-odour and individual-specific information in scent-marks

The effects of pouch bacteria on the chemical composition of subcaudal gland secretions in the European badger *Meles meles*: Maintaining the balance between shared group-odour and individual-specific information in scent-marks

C.D. Buesching[1], H.V. Tinnesand[2], Y.W. Sin[1,3], F. Rosell[2], T. Burke[3] & D.W. Macdonald[1]

1 ***Wildlife Conservation Research Unit, Department of Zoology, University of Oxford, Recanati-Kaplan Centre, Tubney House, Tubney, Abingdon Road, OX13 5QL, UK***

2 *Faculty of Arts and Sciences, Department of Environmental and Health Studies, Telemark University College, N-3800 Bø, Telemark, Norway.*

3 NERC Biomolecular Analysis Facility – Sheffield, Department of Animal & Plant Sciences, Alfred Denny Building, University of Sheffield, Western Bank, Sheffield, S10 2TN, UK

INTRODUCTION

Social mammals often face an olfactory dilemma: maintaining a common group-odour without losing the means of self-advertisement (Albone 1984). In many species, a shared group-odour has been implicated in social acceptance (e.g. Hurst *et al*. 1993) and group-cohesion (e.g. peccaries: Byers 1985; badgers: Buesching *et al*. 2003), although few semio-chemical studies have investigated its chemical characteristics. The existence of individual-specific odours, in contrast, has been established in a wide variety of animals (see Wyatt 2010). In many mammals, these individual-specific scent-characteristics are in part dependent on the individual's genetic make-up (Lanyon *et al*. 2007), the endocrinological (Theis *et al*. 2008) and physiological (Buesching *et al*. 2002a) state of the animal, as well as the time of year (Buesching *et al*. 2002b) and other environmental parameters (Walkden-Brown *et al*. 1994). Thus, individual-specific odours are generally assumed to encode fitness-related information in addition to individuality (Wyatt 2010), which is especially important in the context of mate choice and oestrus-advertisement (e.g. Buesching & Macdonald 2001).

To convey discrete information to their conspecifics, many mammals have developed specialised scent-glands as well as highly adaptive scent-marking behaviours (e.g. Brown & Macdonald, 1985). Mammalian olfactory signals are generally complex chemical mixtures (Wyatt 2000), and information is often encoded in relative ratios (analogue coding) rather than presence/ absence of chemical components (digital coding: Wyatt 2010). In many cases, the primary gland products excreted by the animal are metabolised by bacteria resulting in secondary metabolites generating the characteristic odour of the scent-mark (Fermentation hypothesis: e.g. Archie & Theis 2011). The bacteria can live in the gland itself, in the lumen where the secretion is stored prior to scent-marking, or on the substrate where the scent-mark is deposited (Archie & Theis 2011). Therefore, the odour-profile depends partly on the primary gland products excreted by the mammal, partly on the bacterial flora metabolising these primary gland products, and partly on the age of the scent-mark (e.g. Buesching *et al*. 2002b). As in most social species offspring of at least one sex disperses at least once in their lives, and may change group-affiliation more than once (e.g. to avoid inbreeding: Greenwood 1980), neither group-membership nor most fitness-related parameters (such as age, body condition, reproductive status etc.)

are constant throughout their lifespan. Odour-characteristics relating to these parameters must therefore be flexible yet reliable.

In this study, we use the European badger (*Meles meles*) as a model to investigate the suitability of bacterial assimilation in glandular secretions amongst group-members as a mechanism to solve the olfactory dilemma in mammals generating and maintaining a shared group-odour without overriding individual-specific information in their scent-marks. The importance of scent-marking in the intra-specific communication of mustelids has long been recognised, with most species using their faeces and anal gland secretions as primary olfactory signals (e.g. reviewed by Macdonald 1985). In addition, badgers have evolved a subcaudal gland which is unique amongst the Carnivora (Macdonald 1985) and consists of several layers of apocrine and sebaceous cells, which secrete a margarine-like paste into a common lumen, the subcaudal pouch (Stubbe 1971), where it is stored until scent-marking (Buesching *et al*. 2002c). The pouch opens at a 20-80mm wide slit situated between the anus and the base of the tail. Both sexes possess this gland and use the secretion for scent-marking by pressing the slit onto the substrate (Östborn 1976; Buesching *et al*. 2003; Buesching & Macdonald 2004). Cubs start to produce traces of subcaudal gland secretion when they are approximately 4 months old (Buesching *et al*. 2002c). GCMS analyses have shown that subcaudal gland secretions are of high chemical complexity and contain mainly long-chained carboxylic acids, water and protein (Gorman *et al.* 1984). Secretions are highly individual-specific, and encode information about sex, age, body condition, and reproductive status of the donor (Östborn 1976; Gorman *et al.* 1984; Buesching *et al.* 2002a), but change with season and age of the scent-mark (Buesching *et al*. 2002b). In addition, odour profiles are more similar amongst the members of a social group than they are between groups, indicating the existence of a shared group-odour (Buesching *et al*. 2002a)

As long-chained carboxylic acids are usually of bacterial origin rather than being primary gland products (Albone 1984) and the subcaudal pouches of badgers have been shown to support a rich bacterial flora (Albone *et al.* 1978; Sin *et al*., 2012), it has been hypothesised that individuals in the same social group have a more similar bacterial flora (Albone *et al.* 1978; Gorman *et al.* 1984). Mutual allo-marking, a behaviour unique to badgers, during which two individuals (always from the same social group) press the openings of their subcaudal pouches against each other simultaneously (Kruuk 1989), has been suggested as the mechanism to facilitate the exchange of pouch bacteria

between group members (Kruuk *et al.* 1984; Buesching *et al*. 2003), thus resulting in a group-specific composition of pouch bacteria, and hence leading to a shared group-specific smell. However, a direct empirical link between pouch bacteria and scent-profiles remains to be established.

Here, we analysed the bacterial pouch flora in relation to the chemical scent-profile of adults belonging to six different social groups. The aim of this study was three-fold: 1) To investigate how group-membership is encoded in subcaudal gland secretions, 2) to investigate the overall relation between the bacterial communities and the chemical composition of the subcaudal gland secretion, and 3) to investigate potential group-differences in the bacterial communities of badger subcaudal gland secretions.

Material and Methods

Study Animals and Collection of Samples

Subcaudal gland secretions were collected from a badger population in Wytham Woods, Oxfordshire, England (Ordnance survey reference: SP462080; for detailed description of the study site see Kruuk 1989; for details of study population see Macdonald and Newman 2002) during two different trapping events: in spring (27.5.-9.6., N=31), and in summer (6.9.-18.9. 2010, N=35). All samples were used to investigate the correlation between microbiota and chemical composition. To investigate differences between social groups, we used secretions collected from a total of 6 different social groups (N_{CH}=5, N_{GAH}=5, N_{LS}=6, N_{MT}=6, N_{PO}=6, N_{RC}=5) as established by bi-annual bait-marking surveys (following the methodology described by Delahay *et al*. 2000), where we caught a minimum of 5 adult animals. To avoid pseudoreplication, only one sample from each individual was included (N=33, N_{Males}=14, $N_{Females}$=19).

As part of an ongoing population study, and following the methodology described by Macdonald and Newman (2002), badgers were trapped over night in cage traps baited with peanuts. At first capture (usually as cubs), all animals were tattooed with an individual number on the left inguinal region. Thus, after sedation with 0.2 ml ketamine hydrochloride/kg body weight (Thornton *et al*. 2005), all badgers could be identified individually, sexed, and aged.

Dispersal (i.e. permanent change of social group) is low in the study population (<20% of males) and is also dependent on social group membership (Macdonald *et al*. 2008). However, as less than

40% of badgers are never trapped outside their natal group (Macdonald *et al*., 2008), we recorded if individuals were caught in any other territories within twelve months prior to sampling to investigate possible relationships between chemical and bacteriological subcaudal gland characteristics and group fidelity.

Subcaudal gland secretion was scooped out of the subcaudal pouch using a rounded stainless steel spatula. Secretions were sub-divided into two samples. Both samples were frozen immediately and stored at -20'C until further analyses. Between sampling each individual, the spatula was wiped clean and sterilized three times with 90% Ethanol. Following sterilization, the ethanol was flamed off to avoid contamination of the GC-MS sample. To avoid contact with plasticizing agents, the GC-MS sample was stored first in a glass vial with Teflon lid, whilst the sample for the bacteriological analyses was stored afterwards in a microcentrifuge tube. Contact of the spatula with body parts other than the inside of the subcaudal pouch was strictly avoided to exclude bacteriological cross-contamination.

Chemical Analyses

All chemical analyses were performed in February 2011. For extraction, 1.0ml of dichloromethane (Pestanal® Grade, Sigma-Aldrich®, Oslo, Norway) was added to 0.1g of subcaudal gland secretion. Solutions were left at room temperature for 1 hour and centrifuged at 3000rpm through a 0.45µm filter for 5 minutes. The resulting particle-free solution was transferred to a GC-vial.

Analyses were carried out using a Hewlett-Packard (HP, Oslo, Norway) 6890 Series II gas chromatograph equipped with a non-polar HP-5 MS 5% phenyl-methyl-siloxane column (30.0m long × 0.25mm ID × 0.25µm film thickness), connected to a HP 5973 Series mass spectrometer detector with a split/splitless inlet in splitless mode. Helium was used as the carrier gas at a constant flow of 0.9ml/min. Injection port temperature was set at 300°C. The purge flow to split vent was 49.8m/min at 1.00 min. An auto-injection system (Agilent 7683 Series Injector, Oslo, Norway) was used to inject 1.0µl of the SGS solution into the GC-MS. The samples were cold-trapped at 40°C on the column tip for 2 minutes and separated using a temperature program of 8°C/min from 40º to 150°C, then 6°C/min from 150°C to 200°C, and finally 4°C/min from 200º to 240°C (holding for 15min). A solvent delay of 5 minutes was set for every run to prevent solvent damage to the detector. Samples were analyzed

in random order, and after every 5 samples, a blank was run to ensure that there was no contamination left in the column. A mixture of unbranched alkanes between C_8 and C_{40} (Sigma-Aldrich, Oslo, Norway) was also run after every 5 samples to calculate Kováts Retention Indices (KRI), allowing standardization of retention times. The instrument was calibrated every morning to detect possible changes in sensitivity.

Compounds were matched between profiles by their retention times and mass spectra, and given an individual peak number. A tentative identification of the analytes was provided by cross-checking the best suggested matches from the Wiley 275 spectral library (Scientific Instrument Services Inc., Ringoes, USA) with the calculated KRIof the analytes. Identification of 12 compounds was confirmed through injection of commercial standards (Sigma-Aldrich, Oslo, Norway). Where the Wiley library did not provide a good match, analytes were given a name based on its retention time and added to a new library.

Bacteriological Analyses

DNA extraction

Secretions were transferred to microcentrifuge tubes and resuspended in 180µl enzymatic lysis buffer [20mMTris·Cl (pH 8.0), 2mM EDTA, 1.2% Triton X-100] containing 20mg/ml lysozyme, and incubated at 37°C for at least 1 hr in a shaking incubator to lyse cell walls of Gram-positive bacteria before DNA purification. Samples were then incubated with proteinase K at 56°C for at least 1hr, followed by the addition of 200µl ethanol. DNA isolation was continued by pipetting the mixture into the spin column according to the manufacturer's instructions for animal tissue samples using the DNeasy Blood & Tissue kit (Qiagen, Hilden, Germany).

T-RFLP analysis

Internal fragments of 16S rRNA genes were amplified from the isolated DNA using universal bacterial primers 341f (5'-CCTACGGGAGGCAGCAG-3'; Muyzer *et al.*, 1993) and 926r (5'-CCGTCAATTCMTTTRAGTTT-3'; Muyzer *et al.*, 1995). Primer 341f was labeled at its 5' end with the dye 6-carboxyfluorescein (6-FAM) and 926r labeled with hexachloro-6-carboxyfluorescein (HEX). PCR mixtures comprised 1X PCR buffer, 1mM $MgCl_2$, each dNTP at a concentration of 0.2mM, each primer at a concentration of 0.6µM, as well as 0.5U Taq polymerase (Bioline).

Amplification was carried out in two separate 25µL reaction mixtures under the following conditions: 95°C for 3min, amplified for 35 cycles of 95°C/ 30s; 55°C/ 60s; 72°C/ 75s, and a final extension at 72°C for 10min. For each sample, duplicate fluorescently labeled PCR products were pooled after PCR. PCR product sizes were confirmed by agarose gel electrophoresis using 100bp DNA ladders as size markers and by staining with ethidium bromide. The amplified products (around 586bp) were purified using the QIA quick gel extraction kit (Qiagen).

Fluorescently labeled PCR products were digested with MspI restriction enzymes (Applied Biosystems ABI) for 6h at 37°C, followed by 20min at 80°C for enzyme inactivation. Digested fragments were separated on the ABI 3730 Genetic Analyzer (ABI). Sizes of the fluorescently labeled terminal restriction fragments (TRFs) were determined by comparison with the GeneScan 500 ROX size standard (ABI). Before injection, 0.5µL of the DNA sample was denatured in the presence of 9.5µL Hi-Di formamide and 0.05µL GS 500 ROX size standard (ABI) at 95°C for 5 min. Injection was performed electrokinetically at 2kV for 10s, and electrophoresis was run at 15kV for 30min. After electrophoresis, terminal restriction fragment length polymorphism (T-RFLP) electropherograms were imaged using the GENEMAPPER software (v. 3.7; ABI). The lengths of TRFs were determined by comparison with the internal standard, within the lower threshold at 50bp and upper threshold at 500bp. Only peaks with heights exceeding 50 fluorescence units were evaluated.

T-RFLP profiles were aligned using T-Align (Smith *et al.*, 2005), which identified all fragments unequivocally with ±0.5bp in all profiles generated, and determined the presence and absence of TRFs and their relative fluorescence in all samples. Log (n+1) transformation was performed in order to normalize the data.

The bacterial groups that contributed to the TRFs were identified by cloning and sequencing of the 16S rRNA genes, detailed in Sin *et al.* (2012).

Statistical analyses

All TRFs and chemical compounds that occurred in only one sample were excluded from all analyses.

Analyses of ScentProfiles

To account for variation in the total quantity of injected secretion, peak areas of each compound were standardized using the peak area of hexadecanoic acid, which was present as one of the biggest peaks in all sample-profiles. All data were square-root transformed to reduce the influence of the most abundant variables (Clarke & Warwick 2001), and a Bray–Curtis similarity matrix calculated between each possible sample pairing.

Analyses of correlation between TRFs & Chemical Composition

To investigate any linkage between the TRFs and the chemical composition of the secretion, we performed the BEST procedure (Clarke *et al*. 2008). The rationale of BEST is to find the "best match" between the multivariate among-sample patterns of two data matrices, one being a fixed resemblance matrix (often biotic variables, or in our case the chemical compounds) and one an explanatory matrix (often environmental variables, in our case the TRF data; Clarke & Gorley 2006). The BVSTEP procedure in BEST selects the combination of explanatory variables that best explains the structure in the fixed matrix, and the global BEST test is an overall significance test for the final subset of variables (Clarke *et al*. 2008).

To investigate correlations between specific TRFs and chemical compounds, we constructed a correlation matrix between all variables that occurred in > 10 % of the samples. All TRF-compound combinations with a correlation > 0.4 were tested for statistical significance using two-tailed tests. Correlation analyses were performed with the software packages JMP 10 (SAS Institute Inc., Cary, NC, USA) and SPSS Statistics 19 (SPSS Inc., New York, USA).

Analyses of Group-differences in Chemical Composition and TRFs

To investigate differences in microbiota and chemical profiles between social groups, we first used principal coordinate (PCO) analysis based on the Bray–Curtis similarity matrix in order to visualize the patterns of differences among samples (Gower 1966). If the same animal was caught during both trapping events, only one sample, selected at random, was included in the analysis. Differences between social groups were then compared with a single factor PERMANOVA (Anderson 2001, McArdle & Anderson 2001) using 9999 permutations, applying both main and pairwise tests. PERMANOVA allows distance-based tests of significance for comparing a priori groupings.

Significant ($P < 0.05$) and marginally significant ($P < 0.1$) differences between groups were investigated further using canonical analysis of principal coordinates (CAP; Anderson & Willis 2003), which obtains predictive models that search the multivariate data for the best discrimination between a priori groups. The number of PCO axes, *m,* to use in the model and the predictive ability of the model to discriminate between the social groups was assessed by leave-one-out cross-validation (Anderson & Robinson 2003, Mardon *et al*. 2010). The software PRIMER V6.1.13 (Clarke & Gorley 2006) with the PERMANOVA+ V1.0.3 add-on package (Anderson *et al*. 2008) was used in all analyses.

Results

In total, 50 TRFs and 125 different chemical compounds were found in the GCMS-profiles of subcaudal gland secretions.

Group-Differences in Chemical Composition

When comparing chemical group differences including all compounds, an unconstrained 2D PCO explained 49.2 % of the variation in the data, and the third axis explained further 8.7 %, but there was no separation between groups (Fig. 1).

The main PERMANOVA comparing the 6 groups with ≥ 5 adults sampled was significant (pseudo $F_{=}1.6692$, $P = 0.0141$), and pairwise tests revealed significant differences between four pairs of groups (Table 1). The CAP analysis classified 33.3 % of the chemical profiles into the correct group using leave-one-out cross-validation and $m = 16$ axes ($\delta_1^2 = 0.89814$, $P = 0.028$, Fig. 2).

However, if only compounds shared by all members in each group were included, the unconstrained 2D PCO explained 54.8 % of the variation in the data (Fig. 3), and the third axis explained further 18.1 %, with separations between social groups.

The PERMANOVA comparing the 6 groups was highly significant (pseudo $F = 12.39$, $P = 0.0001$), as were all pairwise comparisons (Table 1).

The CAP analysis classified 100 % of the chemical profiles into the correct group using $m = 5$ axes ($\delta_1^2 = 0.98934$, $P = 0.0001$, Fig. 4).

Correlation of Microbiota and Chemical Composition

Using 9999 permutations, the BEST procedure showed an overall significant correlation between microbiota and chemical composition (Rho = 0.228, P = 0.041), with a subset of four TRFs (99.2, 138.9, 194.2 and 212.0) best explaining the structure in the chemical matrix (Table 2).

In the investigation of correlations between specific TRFs and chemical compounds, 5 TRFs and 23 chemical compounds showed a correlation > 0.4. All correlations were statistically significant (P< 0.01, Table 2).

Group-Differences in Microbiota

Comparing the TRFs to investigate differences in microbiota between groups, an unconstrained 2D PCO explained 82.3% of the variation in the data (Fig. 5), and the third axis explained further 12.2 %, but there was no separation between social groups, and no significant differences in the TRFs of the 6 social groups was found (PERMANOVA: pseudo F =1.007, P = 0.4317, Table 1).

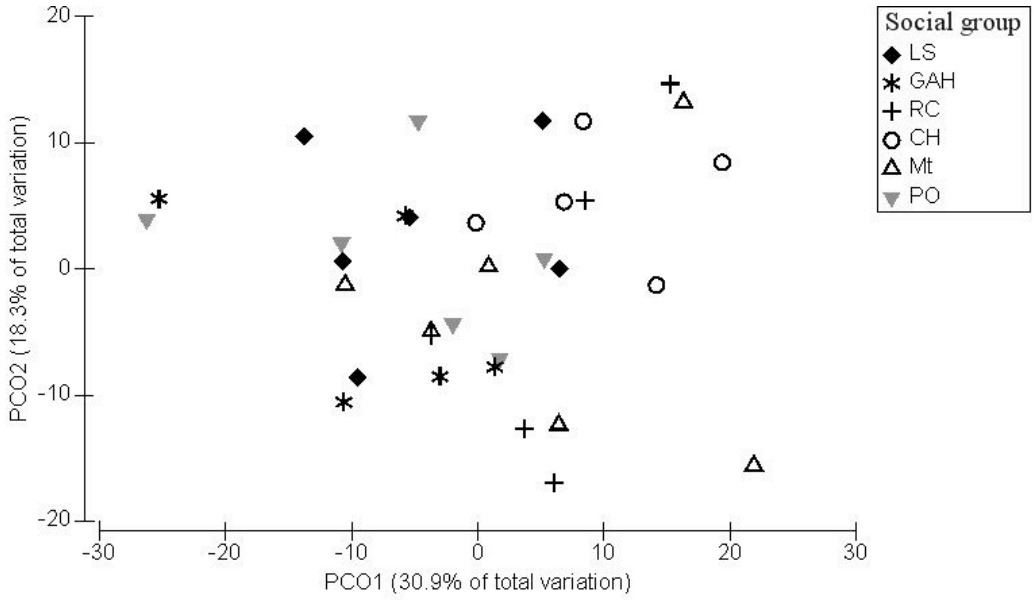


Fig. 1: Unconstrained 2D PCO including all chemical compounds from adults of six social groups with ≥5 SCG samples

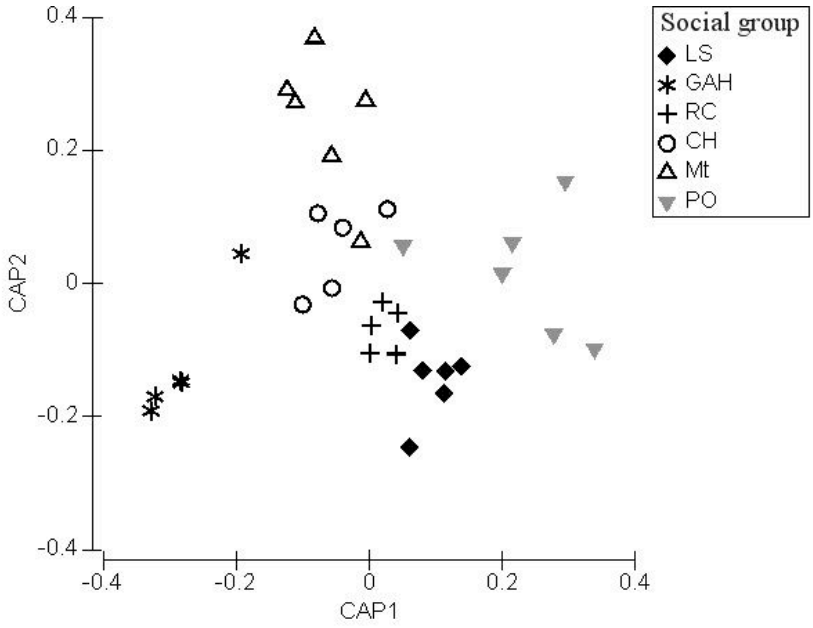


Fig.2: CAP analysis including all chemical compounds from adults of six social groups with ≥5 SCG samples

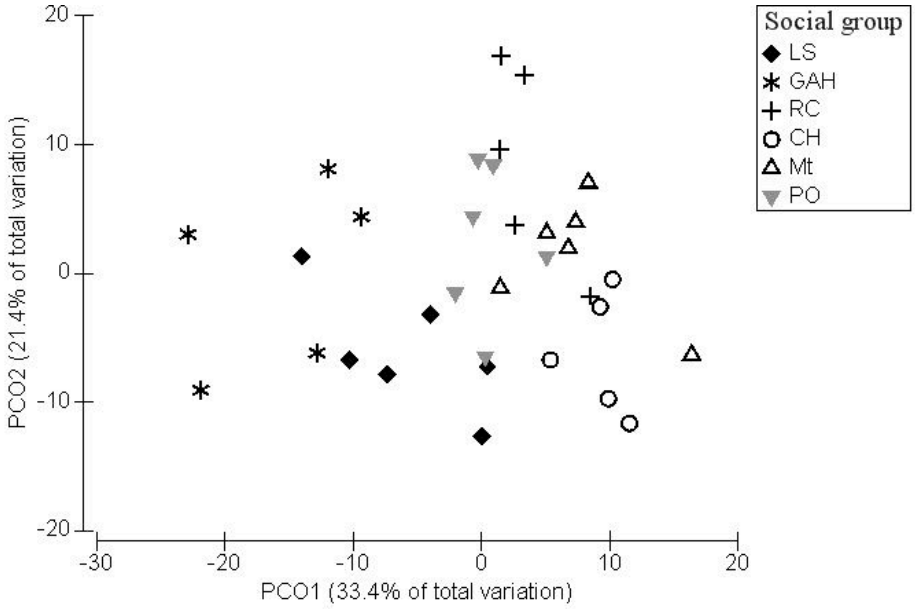


Fig. 3: Unconstrained 2D PCO includingonlychemical compounds which were shared by all adults belonging to the same social group (six social groups with ≥5 SCG samples)

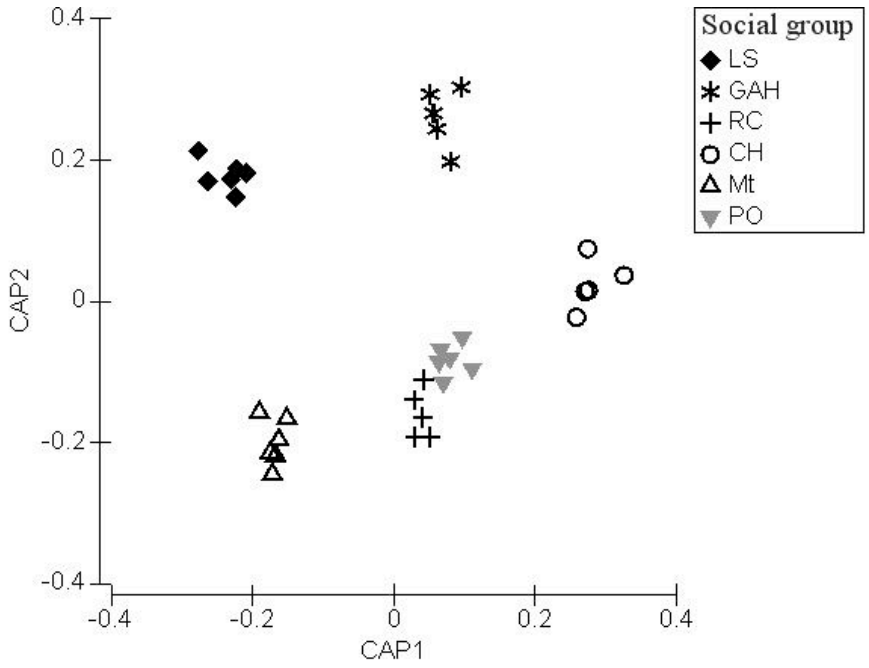


Fig. 4: CAP analysis including only chemical compounds which were shared by all adults belonging to the same social group (six social groups with ≥5 SCG samples)

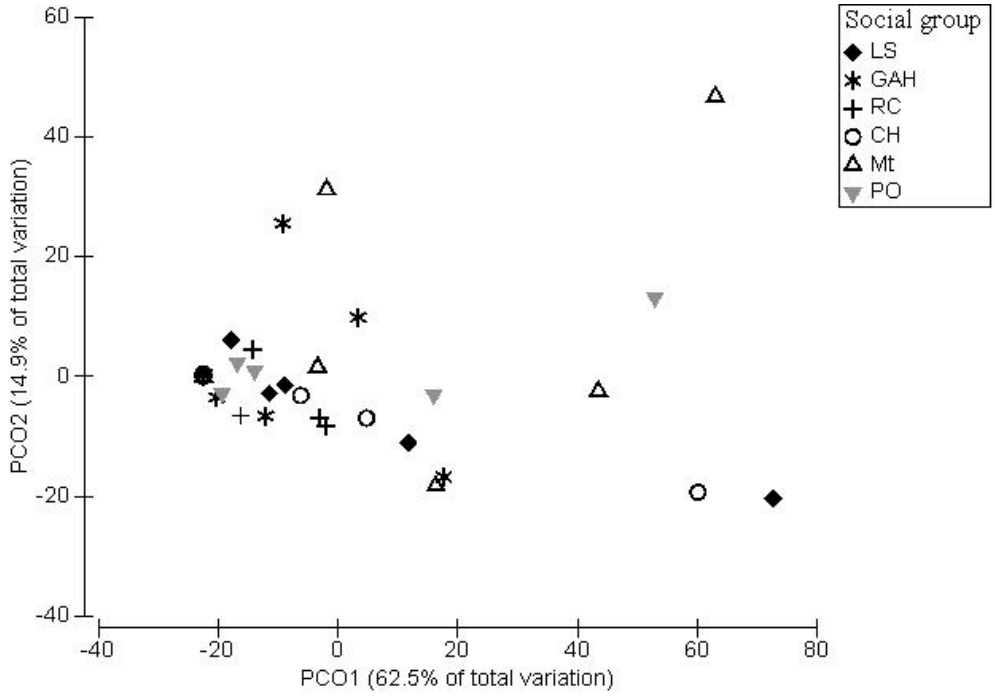


Fig. 5: Unconstrained 2D PCO comparing the TRFs from 6 social groups with ≥5 adults sampled

Table 1: PERMANOVA results for pairwise comparisons between the chemical profiles and the microbiota of members of 6 different social groups (significant results highlighted in bold)

	Chemistry				**Microbiota**	
	All compounds included		Only compounds shared by all group members included		TRFs	
Group	t	P(perm)	t	P(perm)	t	P(perm)
LS vs. GAH	0.9084	0.5719	3.0404	**0.0013**	1.0487	0.3468
LS vs. RC	1.2738	0.1096	3.6784	**0.0035**	0.94561	0.4736
LS vs. CH	1.6147	**0.0281**	4.4595	**0.0025**	0.59353	0.6728
LS vs. Mt	1.3406	0.1041	3.5608	**0.0026**	1.052	0.313
LS vs. PO	0.9924	0.4065	3.4059	**0.0022**	0.64851	0.6137
GAH vs. RC	1.1914	0.1737	3.3558	**0.0018**	0.88187	0.6028
GAH vs. CH	1.5818	**0.0175**	4.2611	**0.0016**	0.7999	0.6341
GAH vs. Mt	1.1415	0.2353	4.5744	**0.0004**	1.6775	**0.0478**
GAH vs. PO	0.99478	0.4149	3.1995	**0.0008**	0.46955	0.8486
RC vs. CH	1.1533	0.1965	3.4795	**0.0094**	0.91271	0.5374
RC vs. Mt	0.91766	0.555	3.2032	**0.002**	1.5689	0.076
RC vs. PO	1.1016	0.2727	2.3542	**0.0036**	0.99276	0.4195
CH vs. Mt	1.1669	0.2369	3.9912	**0.0026**	0.96919	0.4436
CH vs. PO	1.441	**0.0535**	3.3926	**0.0021**	0.75341	0.5956
Mt vs. PO	0.93082	0.4903	3.0985	**0.0029**	1.239	0.1921

Discussion

Most mammals communicate a wide variety of information important in intra- and inter-specific behaviours using olfactory signals (Brown & Macdonald 1985). Some species, such as the European badger, encode a wide variety of odour cues ranging from individual-specific information to group-membership, seasonality, and age of the scent-mark (Buesching & Macdonald 2001) in a single scent-mark. Thus, the reliable maintenance of olfactory characteristics encoding important information becomes a fine balance to avoid the over-writing of e.g. individual-specific information in favour of group-membership.

Our results confirm the presence of a shared group-odour in badgers. However, while Buesching *et al*. (2002a) found clear group-differences in samples collected from the same population 15 years prior to the present study also when all components detected by the GCMS were included in the analyses, in the present study group-differences were obvious only if analyses were restricted to those components present in all members of the same social group. While pairwise comparisons revealed significant

Table 2: All combinations of TRF and chemical compounds with a Pearson's correlation coefficint > 0.4. * Significant at the 0.01 level (2-tailed). ** Identified with authentic chemical standard. TRF 99.2 = Alphaproteobacteria / Betaproteobacteria; TRF 132.4 = Unknown; TRF 138.9 = Actinobacteria; TRF 160.6 = Gammaproteobacteria/ Betaproteobacteria/ Actinobacteria; TRF 212.0 = Firmicutes.

Retention time	Compound	TRF 99.2	TRF 132.4	TRF 138.9	TRF 160.6	TRF 212
7.93	Benzaldehyde**			0.60*		
9.30	Limonene			0.56*		
15.33	Badger_SGS_15.325					0.54*
15.76	Badger_SGS_15.756	0.46*				0.46*
17.44	Badger_SGS_17.443	0.49*				
18.74	Dodecanoic acid **	0.41*				
19.67	Badger_SGS_19.672					0.95*
20.55	Badger_SGS_20.548	0.45*				
20.77	Badger_SGS_20.765	0.55*				
20.98	Badger_SGS_20.981	0.56*				
21.32	Tetradecanoic acid, methyl ester**					0.43*
21.56	Badger_SGS_21.556					0.78*
21.82	Badger_SGS_21.821		0.41*			
22.11	Tetradecanoic acid**	0.46*				
22.57	Badger_SGS_22.566					0.46*
23.00	Badger_SGS_23.000					0.53*
23.05	Badger_SGS_23.054	0.43*				
23.06	Badger_SGS_23.060					0.41*
23.69	Badger_SGS_23.690			0.50*		
28.57	Linoleic acid**					0.58*
28.96	Oleic acid, ethyl ester				0.40*	
29.03	Badger_SGS_29.025			0.56*		
29.40	Octadecanoic acid, ethyl ester				0.65*	

differences between the groups, the high number of axes necessary to separate all groups in the statistical analyses indicates that the coding of a specific group scent is complex and multidimensional. Although dispersal rates remain low in this population (Macdonald *et al*. 2008), and overall population density remained comparatively stable (Macdonald *et al.* 2009), individuals now move more frequently between social groups (i.e. excursions and visits to neighbours: Macdonald *et al*. 2008) and extra-territorial matings have increased (Dugdale *et al*. 2007), and these changes are associated with less pronounced group differences in odour profiles.

Nevertheless, by restricting group-specific information to the relative amounts and ratios of a subset of components, other chemical compounds can be utilized effectively to encode other information e.g. relating to individuality. Our results mirror those from other studies in that no two individuals shared the same scent profile, indicating the coding of individual-specific parameters (Gorman *et al*. 1984; Buesching *et al*. 2002a/ b).

The fermentation hypothesis predicts that mammalian scent profiles are heavily affected by bacteria metabolizing the primary gland products excreted by the individual (Albone *et al.*, 1974; Albone & Perry, 1975; Gorman, 1976). Here, we found a significant correlation between the microbiota in the subcaudal pouch and the chemical composition of the subcaudal gland secretion. A subset of four TRFs (97.6, 138.9, 194.2 and 212.0) explained most of the effect. Sin *et al*. (2012) found TRFs 138.94 to belong to phylum Actinobacteria, and 97.6 and 211.97 to phylum Firmicutes (no phylum could be assigned from cloning and sequencing TRF 194.2). Actinobacteria and Furmicutes include several well-known odour producers, and are also abundant in the paste of spotted hyaenas (*Crocuta crocuta*: Theis *et al*. 2012). Several genera in these phyla produce a diverse array of short- and medium-chain fatty acids, which are prominent in the subcaudal gland secretion of badgers (Buesching *et al*. 2002a). Some Actinobacteria have been found to play a major role in the transformation of odourless steroids into odourous derivatives (e.g. Gower *et al*. 1986, Kohl *et al*. 2001). It is thus highly likely that bacteria belonging to these phyla will affect the chemical composition in the secretion.

It has been suggested that in badgers shared group-odours are generated by assimilating the bacterial flora in the subcaudal pouches through mutual allo-marking (Kruuk *et al*. 1984; Gorman *et al.* 1984*;* Buesching *et al*. 2003). This behaviour occurs only between members of the same social group (Buesching *et al*. 2003), and serves to exchange small amounts of subcaudal gland secretion (containing large numbers of bacteria) through pressing the openings of the subcaudal pouches against each other simultaneously. In this case, badgers have to maintain a fragile balance between assimilating their pouch flora enough to maintain a shared group-odour, without over-riding individual-specific information.

TRF profiles were indeed highly individual-specific (Sin *et al*. 2012), indicating that many individual-advertisement signals might in fact be affected by pouch bacteria (Buesching *et al*. 2003). Interestingly, however, we found no group-specific differences in the microbiota, neither in the composition nor between any pair of social groups as determined by TRFs. While this could indicate that other factors are more important in creating the group odour in badgers, the sensitivity of the method applied in this study restricted the sensitivity of our analyses as the number of OTUs exceeded the numbers of TRFs, and a single TRF can comprise several OTUs, thus limiting resolution (Sin *et al*. 2012). In contrast, Theis *et al*. (2012) used a high-throughput sequencing approach to analyse the bacterial communities inhabiting the scent organs of spotted hyaenas, which afforded a far more detailed analysis and allowed for the most comprehensive view of the bacterial communities inhabiting any specialized mammalian scent gland to date. As it is likely that the variation in metabolic activity is found at the species-, subspecies-, or even strain-level, in badgers high-throughput sequencing can be expected to reveal more subtle differences in the microbial communities between social groups.

References

Albone E. 1984. Mammalian Semiochemistry: The investigation of chemical signals between vertebrates. Wiley and Sons, Chichester.

Albone ES & Perry GC 1975 Anal sac secretion of the red fox, *Vulpes vulpes*; volatile fatty acids and diamines: Implications for a fermentation hypothesis of chemical recognition. *J Chem Ecol* 2: 101-111.

Albone ES, Eglinton G, Walker JM & Ware GC 1974 The anal sac secretion of red fox (*Vulpes vulpes*); its chemistry and microbiology. A comparison with the anal sac secretion of the lion (*Panthera leo*). *Life Sci* 14: 387-400.

Albone, E.S., Gosden, P.E., Ware, G.C., Macdonald, D.W. & Hough, N.G. 1978. Bacterial action and chemical signalling in the red fox (*Vulpes vulpes*) and other mammals. *In:* R.W. Bullard (Ed). Flavour chemistry of animal foods, ACS Symposium Series 67: 78-91. American Chemical Society, Washington, D.C.

Anderson MJ. 2001. A new method for non-parametric multivariate analysis of variance. *Austral Ecol* 26: 32-46.

Anderson MJ, Willis TJ. 2003. Canonical analysis of principal coordinates: a useful method of constrained ordination for ecology. *Ecology* 84: 511-525.

Anderson MJ, Gorley RN, Clarke KR. 2008. PERMANOVA+ for PRIMER: guide to software and statistical methods. PRIMER-E Ltd, Plymouth UK.

Archie, E. A. & Theis, K. R. 2011. Animal behaviour meets microbial ecology. *Anim. Behav*. 82, 425–436.

Brown RE & Macdonald DW. (eds). 1985. Social odours in mammals. Volume I &Volume II, Clarendon Press, Oxford.

Buesching, C.D. & Macdonald, D.W. 2001. Scent-marking behaviour of the European badger (*Meles meles*): resource defence or individual advertisement. In: *Chemical signals in vertebrates 9* (Ed. by Anna Marchlewska-Koj, John L. Lepri & Dietland Müller-Schwarze) Kluwer Academic/Plenum Press, New York, pp 321-327.

Buesching, C.D., Waterhouse, J.P & Macdonald, D.W. 2002a. Gas chromatographic analysis of the subcaudal gland secretion of the European badger (*Meles meles*) Part I: Chemical differences related to individual-specific parameters. *J. of Chem. Ecol.*. 28:41-56.

Buesching, C.D., Waterhouse, J.P & Macdonald, D.W. 2002b. Gas chromatographic analysis of the subcaudal gland secretion of the European badger (*Meles meles*) Part II: Time-related variation in the individual-specific composition. *J. Chem. Ecol.* 28: 57-69.

Buesching, C.D., Newman, C. & Macdonald, D.W. 2002c. Variations in colour and volume of the subcaudal gland secretion of badgers (*Meles meles*) in relation to sex, season and individual-specific parameters. *Zeitschrift für Säugetierkunde.* 67: 1-10.

Buesching, C.D., Stopka, P. & Macdonald, D.W. 2003. The Social Function of Allo-marking Behaviour in the European Badger (*Meles meles*). *Behaviour.* 140: 965-980.

Buesching, C.D. & Macdonald, D.W. (004. Variations in Object-Marking Activity and Over-Marking Behaviour of European Badgers (*Meles meles*) in the Vicinity of their Setts. *Acta Theriologica.* 49: 235-246.

Buyers, JA (1985). Olfaction related behaviour in peccaries. *Z. Tierpsychol.* 70:201-210.

Clarke KR, Gorley RN. 2006. PRIMER v6: user manual/tutorial. PRIMER-E Ltd, Plymouth UK.

Clarke KR, Warwick RM. 2001. Change in marine communities, 2nd edition. PRIMER-E Ltd, Plymouth UK.

Clarke KR, Somerfield PJ, Gorley RN. 2008. Testing of null hypotheses in exploratory community analyses: similarity profiles and biota-environment linkage. *J Exp Mar BiolEcol* 366: 56-69.

Delahay, R. J., Brown, J. A., Mallinson, P. J., Spyvee, P. D., Handoll, D., Rogers, L. M. and Cheeseman, C. L. 2000, The use of marked bait in studies of the territorial organization of the European Badger (*Meles meles*) . *Mammal Review*, 30: 73–87

Dugdale HL, Macdonald DW, Pope LC, Burke T. 2007 Polygynandry, extra-group paternity and multiple-paternity litters in European badger (*Meles meles*) social groups. *Mol. Ecol.* 16:5294-5306.

Gorman ML 1976 A mechanism for individual recognition by odour in *Herpestes auropunctatus* (Carnivora: Viverridae). *Anim. Behav.* 24: 141-145.

Gorman, ML, Kruuk, H & Leitch, A 1984. Social functions of the sub-caudal scent gland secretion of the European badger (Meles meles) (Carnivora: Mustelidae). *J. Zool. Lond,* 203:549-559

Gower JC. 1966. Some distance properties of latent root and vector methods used in multivariate analyses. *Biometrika* 53: 325-338.

Gower DB, Nixon A, Jackman PJH & Mallett AI 1986 Transformation of steroids by axillary coryneform bacteria. *Int. J. Cosmetic Sci.* 8: 149–158.

Greenwood, PJ 1980. Mating systems, philopatry and dispersal in birds and mammals. *Anim. Behav*. 28:1140-1162.

Hurst JL, Fang J & Barnard CJ 1993. The role of substrate odours in maintaining social tolerance between male house mice, *Mus musculus domesticus. Anim. Behav*.45:997-1006.

Kohl JV, Atzmueller M, Fink B & Grammer K 2001 Human pheromones: Integrating neuroendocrinology and ethology. *Neuroendocrinol Lett* 22: 309-321.

Kruuk, H. 1989. The social badger: ecology and behaviour of a group-living carnivore (*Meles meles*). Oxford University Press, Oxford.

Kruuk, H., Gorman, M.L. & Leitch, A. 1984. Scent-marking with the subcaudal gland by the European badger, *Meles meles* L. *Anim. Behav*. 32:899-907.

Lanyon CV, Rushton SP, O'Donnell AG, Goodfellow M, Ward AC, Petrie M, Jensen SP, Gosling LM & Penn DJ. 2007 Murine scent mark microbial communities are genetically determined. *FEMS Microbiol. Ecol.* 59: 576–583

Macdonald, D.W. & Newman, C. 2002 Badger (*Meles meles*) population dynamics in Oxfordshire, UK, Numbers, Density and Cohort life histories, and a possible role of climate change in population growth. *J. Zool.* 256: 121-138.

Macdonald, D.W., Newman, C., Dean, J., Buesching, C.D. & Johnson, P.J. 2004. The distribution of European badger *Meles meles* setts in a high density area: Field observations contradict the sett dispersion hypothesis. *Oikos*.106: 295-307.

Macdonald, D.W., C. Newman, C.D. Buesching and P.J. Johnson. 2008. Male-biased movement in a high-density population of the Eurasian badger (*Meles meles*). *J. Mammal*. 89:1077-1086.

Macdonald, D.W., Newman, C., Nouvellet, P.M. & Buesching, C.D. 2009. An analysis of Eurasian badger (*Meles meles*) population dynamics: Implications for regulatory mechanisms. *J. Mammal* 90: 1392-1403.

McArdle BH & Anderson MJ. 2001. Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology* 82: 290-297

Őstborn, H. (1976). Doftmarkering hos graveling. Zool. Revy. 38:103-112

Sin, Y.W., Buesching C.D., Burke, T., Macdonald, D.W. 2012. Molecular characterization of the microbial communities in the subcaudal gland secretion of the European badger (*Meles meles*). *FEMS Microbiol. Ecol.* 81(3):648-59. doi: 10.1111/j.1574-6941.2012.01396.x.

Smith CJ, Danilowicz BS, Clear AK, Costello FJ, Wilson B & Meijer WG (2005) T-Align, a web-based tool for comparison of multiple terminal restriction fragment length polymorphism profiles. *FEMS Microbiol. Ecol.* 54: 375–380.

Stubbe, M. 1971. Die analen Markierungsorgane des Dachses (*Meles meles). Zool. Garten N.F.* 40:125-135.

Thornton, P.D., Newman, C., Johnson, P.J., Buesching, C.D., Baker, S.E., Slater, D., Johnson D.D.P. & Macdonald, D.W. 2005. Preliminary comparison of four anaesthetic regimes in badgers (*Meles meles*). *Veterinary Anaesthesia and Analgesia.* 32: 40-47.

Walkden-Brown, SW, Restall, BJ, Norton, BW, Scaramuzzi, RJ & Martin, GB 1994. Effect of nutrition on seasonal patterns of LH, FSH, and testosterone concentration, testicular mass, sebaceous gland volume and odour in Australian cashmere goats. *J. Reprod. Fert.* 102:351-360.

Appendix 3

Socio-ecological correlates of extra-group paternity in the European badger (*Meles meles*)

Socio-ecological correlates of extra-group paternity in the European badger (*Meles meles*)

Abstract

Extra-group paternity (EGP) occurs commonly among group-living mammals and plays an important role in the mating systems and the dynamics of sexual selection. The extents to which socio-ecological conditions, such as the number of breeders, influence the rate of EGP within-species are currently under explored. We used genetic paternity and demographic data from 1988 to 2010, to study the pattern of EGP, and the factors driving EGP, in a high-density European badger (*Meles meles*) population. We found the numbers of within-group and neighbouring-group candidate fathers (but not the ratio of within-group candidate fathers to within-group mother/neighbouring-group candidate fathers) had opposing influences on EGP rates in litters - the former had a negative effect while the latter had a positive effect on EGP rates. EGPs were related positively to the average pair-wise relatedness between mothers and within-group candidate fathers, but not to the average, or maximum, within-group heterozygosity of candidate fathers. Although the number of within-group candidate fathers was correlated negatively with resident females' extra-group paternity rates, inbreeding avoidance appears to drive females to attain a high frequency of EGP.

Introduction

Extra-group paternity (EGP) refers to the fathering of offspring (number/proportion) by males from outside of the female's social group (or social pair in socially monogamous species). EGP is a widespread phenomenon, occurring in more than two-third of those mammal species so far investigated (Isvaran & Clutton-Brock, 2007; Soulsbury, 2010) and 90% of birds (Griffith et al., 2002). There is, however, considerable variation in the rates of extra-group paternity across mammalian taxa, where some species within a family exhibit exclusive within-group paternity whereas others do not (e.g., EGP rate in Rodentia: 0–61%, Artiodactyla: 0–38%, Carnivora: 0–80% and Primates: 0–44%; Isvaran & Clutton-Brock, 2007). EGP can play an important role in the mating system and the socio-genetic structuring of populations (e.g., Young et al., 2007; Schubert et al., 2011), yet there is limited understanding of the effects of socio-ecological factors on the frequency of EGP, especially in mammals (see Cohas et al., 2006; Isvaran & Clutton-Brock, 2007) and with regard to within-species variation between populations (Cameron et al., 2011).

Intra-specific variation in the rate of EGP between different individuals within a population may be linked to ecological factors such as (i) the number of breeding males within groups (Isvaran & Clutton-Brock, 2007; Spong et al., 2008), (ii) the number of potential mates outside females social groups (Westneat et al., 1990), (iii) the ratio of within-group males to within-group females (Kokko & Rankin, 2006), and (iv) the ratio of within-group males to neighbouring-group males. There is also the possibility for a null hypothesis strategy, where despite living in (ecologically determined) social groups, these groups may not represent coherent, exclusive, breeding groups – and thus the female may be free to encounter extra-group males while out in extra-territorial foraging grounds, and to mate with them, i.e., random, unconstrained mating. EGPs are expected to decrease as the number or proportion of males that may monopolise, defend or mate-guard the within-group females increase (van Noordwijk & van Schaik, 2004; Kokko & Rankin, 2006; Isvaran & Clutton-Brock, 2007) and are predicted to increase as the number of neighbouring-group males increase through unrestrained random contact, as well as deliberate attempts to procure matings. The literature, however, is still lacking concerning how these ecological parameters may shape individual mating strategies,

especially in mammals (e.g., mammals, Cohas et al., 2006; Isvaran & Clutton-Broack, 2007; birds, Westneat et al., 1990; Griffith et al., 2002; Westneat & Stewart, 2003; reptiles, Uller &Olsson, 2008).

EGP may be an adaptive tactic through which females derive either direct (e.g., fertility assurance (Sheldon, 1994; Vedder, et al., 2011), avoidance of infanticide (Wolff, 1993; Borries et al., 2011) or indirect benefits (e.g., 'Good-genes-as-heterozygosity Hypothesis', Brown, 1997; 'Compatible-genes Hypothesis', Neff & Pitcher 2005; Kempenaers, 2007). The 'Good-genes-as-heterozygosity Hypothesis' (Brown, 1997), posits that 'specific' alleles, or 'general' allelic diversity increase fitness, independent of genomic architecture. This predicts that females should value heterozygosity in their offspring and therefore, under certain conditions, in their mates. Any advantage to heterozygous offspring will depend on the relationship between fitness and heterozygosity, and the reported effect sizes of this relationship vary widely (see Britten, 1996; David, 1998; Hansson & Westerberg, 2002; Coltman & Slate, 2003). In studies that show positive effects, heterozygosity has been associated with increased survival rates (Daniels & Walthers, 2000; Cohas et al., 2009; Mainguy et al., 2009), reproductive success (Kruuk et al., 2002; Slate et al., 2004), disease resistance (Coltman et al., 1999; Whiteman et al., 2006; Reid et al., 2007) and developmental stability (reviewed in Kempenaers, 2007). The 'Compatible-gene Hypothesis', proposes that alleles increase fitness in a specific genetic context or via gene–gene interactions (e.g., epistasis, Neff & Pitcher, 2005; Kempenaers, 2007). EGP has been hypothesised to have evolved to avoid inbreeding in animals where inbreeding can be demonstrably costly (Moore & Ali, 1984; Pusey & Wolf, 1996) and this has been demonstrated in some bird studies (e.g., Szulkin et al., 2009; Brouwer et al., 2011). In circumstances where heterozygosity confers fitness benefits to offspring, and where inbreeding can be avoided through extra-group mating, EGP rates would be predicted to correlate with group genetic compositions, such as the degree of relatedness within-group males (Cohas et al., 2006) or the level of heterozygosity of within-group males.

Extra-group (or extra-pair) paternity does, however, occur in many species without advantages being apparent. Indeed, there are risks associated with (extra-group) promiscuity, such as sexually transmitted disease, injury or the possible rejection of extra-group young (Westneat & Stewart, 2003) if extra-group paternity can be detected. Alternatively, EGP could be considered non-adaptive if females mated randomly with whomever they encountered and by chance they mated with extra-group

males (Kokko & Rankin, 2006). A non-adaptionist scenario has been explored recently by Forstmeier et al. (2011), to address this puzzling observation. Forstmeier et al. (2011), propose the 'Within-sex genetic correlation Hypothesis' where the genetic basis of a female's response to courtship by her social mate (within-group males) and extra-group males (extra-pair males) is the same. Alleles for resistance to infidelity may, however, also convey resistance to fidelitous copulations, leading to infertility (Arnquist & Kilpatrick, 2005).

An alternative plausible resolution for circumstances where promiscuity (pair, or extra-group) appears to be a maladaptive behaviour is that the propensity for males and females to engage in promiscuous mating may be affected by the same set of alleles (Halliday & Arnold, 1987). There is likely to be strong positive selection for alleles that enhance promiscuous behaviour in males (Albrecht, 2007). If promiscuous behaviour is heritable (Reid et al., 2011), pleiotropic effects on these sexually selected alleles could also cause promiscuous behaviour to evolve in females, despite antagonistic selection (Rice et al., 1992), known as the 'Between-sex genetic correlation Hypothesis' (Halliday & Arnold, 1987; Forstmeier et al., 2011).

In this study we examine the effects of ecological factors and group genetic compositions on extra-group paternity rates, as evidenced by offspring paternity, in a high-density population of European badgers (*Meles meles*), benefiting from 23 years of genetic pedigree data. Badgers provide an informative model species because they live in social groups that vary from 2–7 individuals in low density population (Revilla & Palomares, 2002; Rosalino et al., 2004) and range between 2–10, up to 29 individuals in high-density population (da Silva et al., 1994; Delahay et al., 2000). In this study population up to seven male and female breeders have been recorded per social-group-year, with a mean (± 95% confidence interval) of 1.9 ± 0.1 (range = 2–7, Dugdale et al., 2007). Crucially, in high-density populations, around half of all cubs are fathered by extra-group males, primarily from neighbouring groups (Carpenter et al., 2005; Dugdale et al., 2007; Annavi et al., in prep) and multiple-paternity do occur (Carpenter et al., 2005; Dugdale et al., 2007).

Badgers exhibit a high degree of natal philopatry (Woodroffe et al., 1995); only 19% of badgers disperse successfully, mainly to neighbouring groups (Macdonald et al., 2008). An individual's breeding social groups and neighbouring-groups are therefore containing close relatives (Dugdale et al., 2008). Under benign environmental conditions (optimal summer rainfall, equating

with sufficient food availability; Nouvellet et al., *subm*), cubs that have fathers with high levels of heterozygosity (i.e., the proportion of markers with heterozygous alleles) had a higher survival probability in their first year than cubs fathered by males with higher homozygosity (Annavi et al., in prep). In combination, these factors lead to hypothesise that EGP may confer a selective advantage when within-group males are genetically more similar to resident females and are less heterozygous than extra-group males.

Even without inferring active male defence of females (anti-kleptogamy, Revilla & Palomares, 1999), simply by proportion, a larger number of males in a group may have the potential to decrease EGP rates (see also Spong et al., 2008; meerkats). However since female badgers forage solitarily and are known to transgress and make incursions into neighbouring group ranges (Bohm et al., 2009) and frequently get caught making temporary visits to other groups (given the sampling rate of 4/365 our trapping permits; Macdonald et al., 2008), the capacity for within-group males to actively control female access to extra-group males will be limited, although if females mate indiscriminately (polygynandry, Dugdale et al., 2007) then simple extra-group male encounter rate could be influential. Like most mustelids (e.g., Sundqvist et al., 1988), badgers are thought to be induced ovulators (a coital stimulus is required to trigger the ovulation, Mead, 1989), that could restrict their reproduction especially in females living in low densities where the number of males and the opportunity for being choosey are very limited. Thus, female may mate randomly with any males they encounter.

Here we test: [1] the effect of ecological factors on the rate of EGP at individual level (per litter, see Carpenter et al., 2005, Dugdale et al., 2007). In particular we ask: [1a] Does the rate of EGP decrease with the number of males or the ratio of within-group males to within-group breeding females in a social group (Kleptogamy defence)? [1b] Does the rate of EGP increase with the number of neighbouring-group males (as 86% EGP [paternity assignment was at ≥ 95% confidence] has been attributed to neighbouring-group males; Dugdale et al., 2007)? [1c] Does the rate of EGP decrease with the ratio of within-group to neighbouring-group males? we further test for: [2] The effect of group genetic compositions on the rates of EGP: [2a] Does EGP rate increase when the within-group average pair-wise relatedness between within-group breeding females and males is high, indicating inbreeding avoidance, [2b] Does EGP rate correlate negatively with within-group males mean or maximum heterozygosity? - This would imply offspring heterozygosity benefits. [3] we then consider

whether badgers of either sex exhibit a lifetime bias in mating strategy towards exclusive within-group or extra-group fidelity and relate our observations to the null hypothesis of random mating.

Materials and methods

Study site and field methods

A high-density population of badgers inhabiting Wytham Woods; a 424-ha site situated 5km north west of Oxford, UK (51: 46: 26N; 1: 19: 19W), has been studied intensively since the 1970s (Kruuk, 1978a; Kruuk, 1978b). A detailed description of the study site (e.g., soil, microclimates and vegetation) is provided elsewhere (Morecroft et al., 1998; Savill, 2010).

Since 1987, a systematic capture-mark-recapture regime has attempted to mark all individuals in the population (Macdonald & Newman, 2002; Macdonald et al., 2009). Live-trapping was conducted three to four times per annum; generally over two weeks in June, September and November, with one week of trapping in January of some years (Macdonald et al., 2009). Badgers were caught in mesh-traps, baited with peanuts placed near the entrances of active setts (Macdonald & Newman, 2002; Macdonald et al., 2009; Tuyttens et al., 1999). Captured badgers were then transferred to holding cages and transported to a central handling facility. Badgers were sedated by an intra-muscular injection using ketamine hydrochloride at 0.2 ml/kg body weight (McLaren et al., 2005). Upon their first capture all badgers were tattooed with a unique individual number on the left inguinal region for permanent individual identification. The sex, age (cub or adult based on body size) and location (sett name) of each badger were recorded. For genetic analysis, hair samples and/or blood from the jugular vein (c. 3 ml) were collected from all individuals upon first capture.

Social group ranges were established using a 'bait-marking' technique approximately every two years (see Kilshaw et al., 2009). The number of social groups occupying this study site has increased steadily (Macdonald et al., 2004b) with a mean of 19 ± 2 (ranges = 14–26) social groups per year (social-group-year) between 1987 and 2005 (Dugdale et al., 2008).

Macdonald et al. (2008) report that only 19% of badgers with at least 4 captures in this population dispersed (if the 2 most recent captures and at least one or two captures before were within a non-natal group), mainly to neighbouring social groups. This indicates that badgers in this

population exhibit high sett-fidelity (often through natal philopatry; see Woodroffe et al., 1995). Temporary group-movements do however occur at quite a high rate (16.4%, Macdonald et al., 2008; based on a maximum of 4 sampling opportunities, from the trapping protocol, per year). By taking into account both permanent dispersal and temporary movements, we define here the social group of residence of each individual per year based on their trapping history according to the following rules:

(a) Badgers first caught as cubs (and thus of known natal group, N = 975) were considered to be resident in their natal group unless, they satisfied the dispersal rules provided by Macdonald et al. (2008) above.

(b) Badgers first caught as adults (N = 421) were assigned to a social group based on the site that trapping revealed them to be affiliated to most frequently, unless clear dispersal events were recorded.

(c) Badgers (unknown natal group affiliation) that exhibited equal affiliation to two or more social groups (N = 14) within their lifetime were assigned to the social group where they were captured initially. Breeding females, however, were allocated to the social group where they were known to have bred most recently.

(d) Badger social groups can undergo fission, i.e., the splitting of a social group into two (or more) new social units (da Silva et al., 1993; Macdonald et al., 2004b), dividing the physical range occupied by the former group (as defined by bait-marking). Clearly, where a subsidiary sett within the former group range splits off and becomes independent from the main sett range, individuals occupying this subsidiary sett range do not move spatially in the process of becoming identified as a new social group in the population records. Badgers gaining a new group affiliation in this way (N = 41), were assigned to the new social group after the group split (with condition they also trapped at least once in the new group before split), unless they were subsequently re-trapped in the former group.

Genotyping and parentage analysis

Details of the DNA extraction, microsatellite characterisation and genotyping methods are presented in Dugdale et al., (2007b) and Annavi et al. (2011; in prep). In summary, we genotyped 1172

individuals at 35 microsatellite loci, of which 813 were cubs born between 1988 and 2010 inclusive (see Annavi et al., in prep). Mean observed heterozygosity (± standard error) was estimated at 0.45 ± 0.03, with 4.46 ± 0.34 alleles per locus. Exact tests, performed using GENEPOP 4.0.10 (Raymond & Rousset, 1995), revealed that no locus, or pair of loci, departed consistently from Hardy-Weinberg equilibrium or linkage equilibrium (Annavi et al., in prep). Two types of potential genotyping error; allelic drop-out rate (ε_1=0.005) and stochastic error rate (ε_2=0.005), were estimated using PEDANT 1.0 (Johnson & Haydon, 2007).

Parentage was determined sequentially using MasterBayes 2.47 (Hadfield et al., 2006) implemented in the R statistics package 2.12.2 (R Development Core Team 2010) and Colony 2.0 (Wang & Santure, 2009). For detailed description about the selection of candidate parents, and their social group assignment, see Dugdale et al., (2007b); for the MasterBayes and Colony analyses see Annavi et al., in prep. we was unable to assign a mother and father at 80% confidence to 16% (N = 130) and 19% (N = 158) of the cubs, respectively.

Offspring that were fathered by males from their natal social group in the year of conception were categorised as within-group offspring (WGO, N = 340), whereas offspring fathered by males that did not reside in the natal social group were categorised as extra-group offspring (EGO, N = 315). Cubs where paternity was unknown (N = 158) were excluded from this analysis.

Heterozygosity and pair-wise relatedness

Individual standardised heterozygosity was estimated as the proportion of loci that were heterozygous, divided by the population mean heterozygosity for these loci (Coltman et al., 1999) using GENHET 2.2 in R 2.12.2 (Coulon, 2010). The average and maximum standardised heterozygosity of within-group adult males (candidate fathers) was then derived per social-group-year. Pair-wise relatedness values between females that were assigned maternity, and their potential within-group mates, were estimated using Coancestry 1.0.0.1 (Wang, 2011) and averaged per social-group-year. Two marker-based pair-wise relatedness estimates were computed and compared: The Queller and Goodnight's estimator (QG; Queller & Goodnight, 1989) and Lynch and Ritland's estimator (LR; Lynch & Ritland,

1999). These estimators were chosen because QG is the method most commonly used and simulations show that LR performs well for most population compositions (Csillery et al., 2006).

Rates of group infidelity

Of the 502 candidate mothers and 612 candidate fathers trapped between 1987 and 2010, 228 females (45%) and 201 males (32%), were assigned parentage. we calculated the number of males and females in the population that were engaged in group infidelity, where infidelity was defined as being assigned at least one EGO in their lifetime. we limited our data to individuals that were born between 1987 and 2008, which were assumed to be dead (i.e., to have completed their lifetime reproductive output) by 2010 (see Dugdale et al., 2008). we estimated the proportion of individuals engaged in group infidelity by including: (i) all, and (ii) restricting analyses to individuals that bred in more than one year. Males and females that were assigned offspring, but no EGO, in their lifetime were categorised as exhibiting within-group fidelity. We examined whether the group infidelity of parents was correlated with that of their offspring. For this, we estimated the proportion of EGO that was subsequently assigned by extra-group offspring of their own.

Statistical analyses

We used R 2.13.2 (R Development Core Team 2007) to fit generalised linear mixed models (GLMMs, using the *lmer* function with Laplace approximation in the lme4 package, Bates & Sarkar, 2007). The rate of EGP was investigated at the individual level (i.e., litter level). Initially, We tested whether the distribution of EGO across litter corresponded to an expected distribution generated through binomial processes. We examined the effects of the number of males in proximity to the breeding female, using (i) the number of within-group candidate fathers, (ii) the number of neighbouring-group candidate fathers, (iii) the ratio of within-group candidate fathers to within-group mothers and (iv) the ratio of within-group candidate fathers to neighbouring-group candidate fathers. We considered the effects of the within-group genetic parameters of candidate parents, using (i) the average and maximum heterozygosity of within-group candidate fathers, and (ii) the average pair-wise relatedness of within-group mothers and candidate fathers. We also examined the quadratic effect of pair-wise relatedness between mothers and candidate fathers on EGP. We derived two indices of EGP: the absolute number

of individuals fathered by extra-group males and the relative proportion of EGP at litter (N_{EGO}/ [N_{EGO} + N_{WGO}]). To examine the absolute number of EGPs within a litter in relation to the explanatory terms, we ran a GLMM, controlling for the total number of cubs per litter, with Poisson error structure and log link function. To investigate the proportion of EGPs per litter, we used a binomial error distribution and a logit link function. Social group and year were included as random effects in all models. We also included female identity as a random effect, because litters produced by the same mother across years are not statistically independent. We reran the models above with the rate of EGP calculated as the number of mate-pairs (either within-group mate-pairs, WGM or extra-group mate-pairs, EGM) rather than the number of offspring (WGO or EGO), because a single copulation could lead to several offspring within a litter and therefore we tested whether this alters the results.

We fitted a maximal model that included all explanatory terms. We eliminated non-significant terms using a backwards selection procedure to derive a minimal adequate model. The significance of including each explanatory term in the model was assessed using a likelihood ratio test (chi-square distribution) by comparing the two nested models (with- and without- the term under consideration). Effect sizes (and their standard error) are based on the final minimal models. Means are given ± standard error.

In order to test EGP per litter (measured as the number of offspring and mate-pairs), We restricted the analysis to social-group-years in which all individuals were genotyped and all offspring were assigned both parents (Data RD_{LTR}: SGY = 81, no of litters = 137, representing 97 mothers (30 mothers had litters for 2-4 years), 194 cubs: 100 WGO and 94 EGO; mothers had only WGM = 76 [17 had one and 5 had two WGMs], only EGM = 58 [50 had one and 8 had two EGMs] and both WGM + EGM =3 [all had one WGM and one EGM]). We then re-ran the test with an unrestricted dataset that included social-group-years in which all trapped within-group candidate parents were genotyped (Data URD_{LTR}: SGY = 124, no of litters = 220, representing 138 mothers (52 mothers had litters for 2–4 years), 310 cubs: 174 WGO and 136 EGO; mothers had only WGM = 122 [115 had one and 7 had two WGMs], only EGM = 91 [81 had one and 10 had two EGMs] and both WGM + EGM =7 [all had one WGM and one EGM]). We used a restricted dataset that had an accurate measure of EGP. This dataset however, was limited by the issue of low statistical power and excluding groups where not all cubs

were assigned paternity could bias EGP rate. Thus, we also re-ran an open analysis with unrestricted dataset to check whether both datasets give similar results.

Results

Patterns of EGP

Nearly half of assigned paternities (48%: 315 of 655 cubs) were extra-group, where the majority were attributable to neighbouring-group males (85%: 268 of 315 cubs). EGP was detected in almost half (47%: 178 of 378, Table 1) of litters where all offspring were assigned fathers at 80% confidence (81%: 378 of 467). Mean litter size was 1.46 (CI: 1.43, 1.49; range = 1–5) and the mean number of EGO at litters that contained at least one EGO was 1.40 (CI: 1.32, 1.49; 1–3, mode = 1), 64%, 32% and 4% of litters included one, two and three EGO respectively. No litters were fathered by more than two males. Among the 178 litters that included EGO, 26 were fathered by two males (16 had two extra-group fathers and 10 had one extra-group and one within-group father). The remaining 152 litters were fathered exclusively by one extra-group male. The distribution of EGP among litters differs from what is expected under the binomial distribution (Table 1, $\chi^2 = 59.90$, df = 3, $P < 0.001$). This result showed a non-random distribution of EGO over litter in badger, with more females than expected had either many or no EGO (Figure 1).

Table 1 Distribution of EGP among litters that include only cubs with both parents assigned. The numbers of litters expected from binomial probabilities are shown in parentheses.

	No. of EGO per litter				
Litter Size	**0**	**1**	**2**	**3**	**Total litters**
1	132 (124.3)	107 (114.7)	-	-	239
2	57 (31.6)	6 (58.4)	54 (27.0)	-	117
3	10 (2.8)	1 (7.8)	2 (7.2)	7 (2.2)	20
4	1 (0.1)	-	1 (0.7)	-	2
Total	200 (158.8)	114 (180.9)	57 (34.9)	7 (2.2)	378

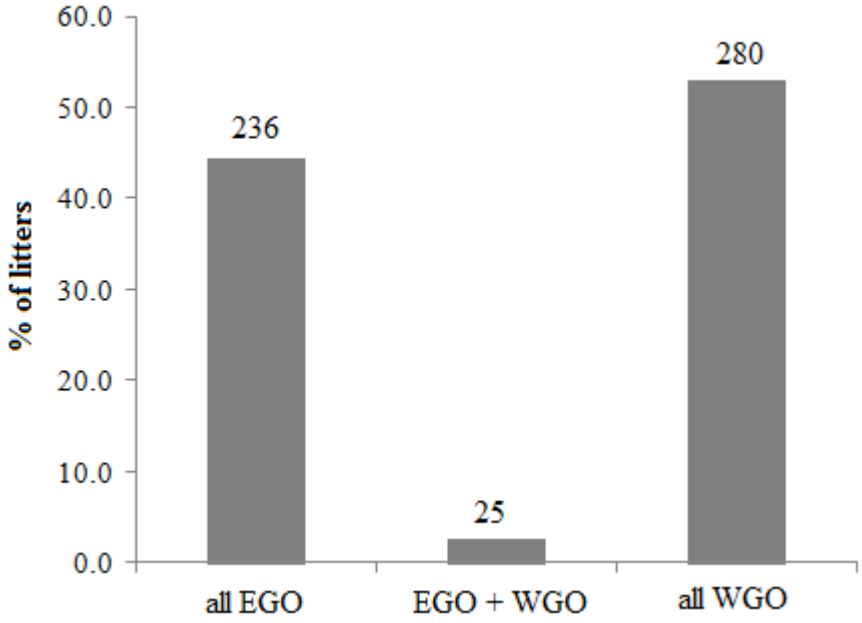


Figure 1 Percentage of litters with only within-group offspring (WGO), only extra-group offspring (EGO) and mixed with both WGO and EGO. Data restricted to litters that include only cubs with both parents assigned (N = 378).

Rate of EGP

One third of male and female breeders (28% and 30% respectively, when full datasets were used) only mated within their social group, and never extra-group, throughout their lifetimes (Table 2). We also found that 64% of extra-group offspring were subsequently assigned at least one EGO, when they themselves bred. When we tested descendents of male and female parents separately, the results were similar; 67% of the offspring produced by either sex were subsequently assigned at least one extra-group offspring. This indicates that tendencies towards extra-group mating strategies are an inherited trait in the badger.

Table 2 Percentage of individuals engaging in infidelity and fidelity to within-group mates in their lifetime

	N	Fidelity (only WGO)	Infidelity (at least one EGO)	Only EGO in their lifetime
(i) Full dataset				
Male	137	28%	72%	38%
Female	111	30%	70%	43%
All	248	29%	71%	40%
(ii) Restricted dataset				
Male	64	17%	83%	20%
Female	60	32%	68%	18%
All	124	24%	76%	20%

Number of breeders, group genetic compositions and the rate of EGP

EGO occurred in approximately half of the litters (Data URD_{LTR} vs. RD_{LTR}: 45% vs. 50%, Figure 2). Forty-four percent of females (in Data URD_{LTR}; 45% in RD_{LTR}) paired with an extra-group male (EGM) with a maximum of two and a mode of 1 mate-pairs. We present results derived from unrestricted datasets (Data URD_{LTR}) because restricted models for Data RD_{LTR} did not converge, probably due to small sample size. Since QG and LR pair-wise relatedness estimators were highly correlated (Spearman's Rank correlation coefficient = 0.91, $P < 0.001$) and yielded similar results we used QG estimator for the main analysis (see Supplementary Table S1 and S2).

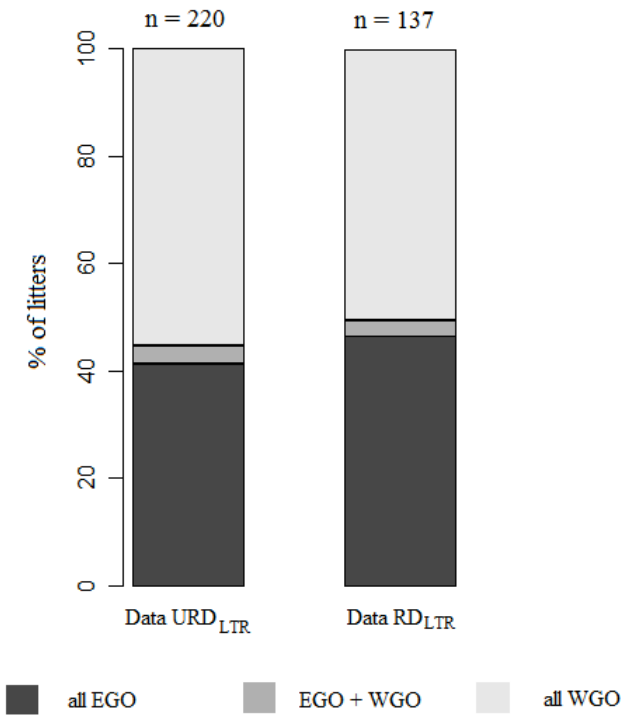


Figure 2 Percentage of litters with cubs that were assigned to extra-group males only (EGO), within-group males only (WGO), and mixed paternity (EGO + WGO) in restricted and unrestricted data.

Relative and absolute numbers of within-group and extra-group males present

The number of EGO within a litter (Data URD_{LTR}) ranged from 1 to 4 with a mean of 1.4 ± 0.1. EGP rates varied among litters, ranging from 33% to 100%. The absolute and relative rates of EGP, measured as the number of EGO and EGM at individual level (per litter) produced similar results (Table 3 and 4). These results showed that bias in over estimating the number of EGO, such that mating with one EGM may result in two EGOs, did not alter our findings. Higher numbers of within-group candidate fathers correlated with lower numbers of EGO (and EGM) per litter in the relative model (Table 4), but not in the absolute model (Table 3). The absolute and relative rates of EGP in litters were associated positively with the number of neighbouring-group candidate fathers (Table 3 and 4). EGP rate was associated neither with the ratio of within-group candidate fathers to within-group mothers, nor with the ratio of within-group candidate fathers to neighbouring-group candidate fathers (Table 3 and 4).

Group compositions

The average pair-wise relatedness between mothers and candidate fathers within-groups had the greatest explanatory power for both the relative and absolute rates of EGP across litters (Table 3 and 4). Where the average within-group candidate parents' relatedness was greater, the number of EGO (absolute) and the proportion of EGO (relative) were higher (Figure 3). Relative and absolute rates of EGP did not correlate with the mean and maximum heterozygosity of within-group candidate fathers (Table 3 and 4).

Table 3 Parameter estimates from GLMMs analysing *the absolute number of extra-group paternity (EGP) or extra-mate pair* in a litter in relation to group density and composition. Mean SH = Individual standardised heterozygosity and QG = average Queller and Goodnight's pair-wise estimator between mothers and within-group candidate fathers. Unrestricted dataset, UR_{LTR}: SGY = 124, no of litters = 220 representing 310 cubs: 174 WGO and 136 EGO.

No	Explanatory variable	Extra-group paternity			Extra-mate pairs		
		Estimate ± SE	**$\chi^2_{[df=1]}$**	**P**	**Estimate ± SE**	**$\chi^2_{[df=1]}$**	**P**
	Final model						
1	Intercept	-2.75 ± 0.40	-	-	-1.71 ± 0.33	-	-
2	Average pair-wise relatedness (QG)	2.30 ± 0.57	15.07	**<0.001**	1.99 ± 0.60	10.16	**0.001**
3	No of neighbouring-group candidate fathers	0.02 ± 0.01	8.39	**0.004**	0.02 ± 0.01	5.74	**0.017**
4	Litter size	0.71 ± 0.12	30.46	**<0.001**	-	-	**-**
	Rejected terms						
1	Quadratic effect of QG	-3.90 ± 2.44	2.81	0.093	-3.38 ± 2.71	1.75	0.186
2	Average SH within-group candidate fathers or	-0.50 ± 0.89	0.32	0.574	-0.04 ± 0.91	0.00	0.964
	Maximum SH within-group candidate fathers	-0.20 ± 0.68	0.09	0.770	0.00 ± 0.72	0.00	0.995
3	Ratio of within-group candidate fathers to neighbouring-group candidate fathers	2.31 ± 0.55	2.72	0.099	-0.32 ± 0.40	0.67	0.414
4	No of within-group candidate fathers	-0.05 ± 0.03	0.20	0.656	-0.05 ± 0.03	2.70	0.101
5	Ratio of within-group candidate fathers to within-group mothers	-0.02 ± 0.03	0.74	0.391	0.01 ± 0.05	0.01	0.927

In the absolute model of extra-group paternity, the random effect for social group, year and mother ID were zero. For extra-mate pairs model, it was zero for year and mother ID but 0.05 ± 0.22 for social group. Significant effects are in bold.

Table 4 Parameter estimates from GLMMs analysing *the relative number of extra-group paternity (EGP) or extra-mate pair* in a litter in relation to group density and composition. Mean SH = Individual standardised heterozygosity and QG = average Queller and Goodnight's pair-wise estimator between mothers and within-group candidate fathers. Unrestricted dataset, UR_{LTR}: SGY = 124, no of litters = 220 representing 310 cubs: 174 WGO and 136 EGO.

No	Explanatory variable	Extra-group paternity			Extra-mate pairs		
		Estimate ± SE	$\chi^2_{[df = 1]}$	P	Estimate ± SE	$\chi^2_{[df = 1]}$	P
	Final model						
1	Intercept	-3.31 ± 1.38			-1.98 ± 0.74		
2	Average pair-wise relatedness (QG)	10.89 ± 2.78	18.66	**<0.001**	5.82 ± 1.43	17.07	**<0.001**
3	No of neighbouring-group candidate fathers	0.13 ± 0.04	13.00	**<0.001**	0.07 ± 0.02	11.00	**<0.001**
4	No of within-group candidate fathers	-0.37 ± 0.13	6.95	**0.008**	-0.19 ± 0.08	4.77	**0.029**
	Rejected terms						
1	Quadratic effect of QG	-10.86 ± 9.15	1.49	0.222	-6.06 ± 5.33	1.23	0.268
2	Average SH within-group candidate fathers or	-0.86 ± 3.49	0.06	0.812	-0.11 ± 2.06	0.00	0.959
	Maximum SH within-group candidate fathers	-1.96 ± 2.81	0.44	0.507	-0.70 ± 1.65	0.16	0.689
3	Ratio of within-group candidate fathers to neighbouring-group candidate fathers	-1.44 ± 1.49	1.02	0.312	-0.55 ± 0.80	0.45	0.504
4	Ratio of within-group candidate fathers to within-group mothers	0.08 ± 0.14	0.31	0.575	0.05 ± 0.10	0.21	0.646
5	Litter size	0.85 ± 0.49	2.89	0.089	5.92 ± 0.30	3.34	0.068

In the relative model of extra-group paternity, the random effect for social group, year and mother ID were 4.48 ± 2.12, 1.26 ± 1.12 and 10.15 ± 3.19, respectively. For extra-mate pairs model, they were 1.07 ± 1.03, 0.33 ± 0.58 and 1.23 ± 1.11, respectively.

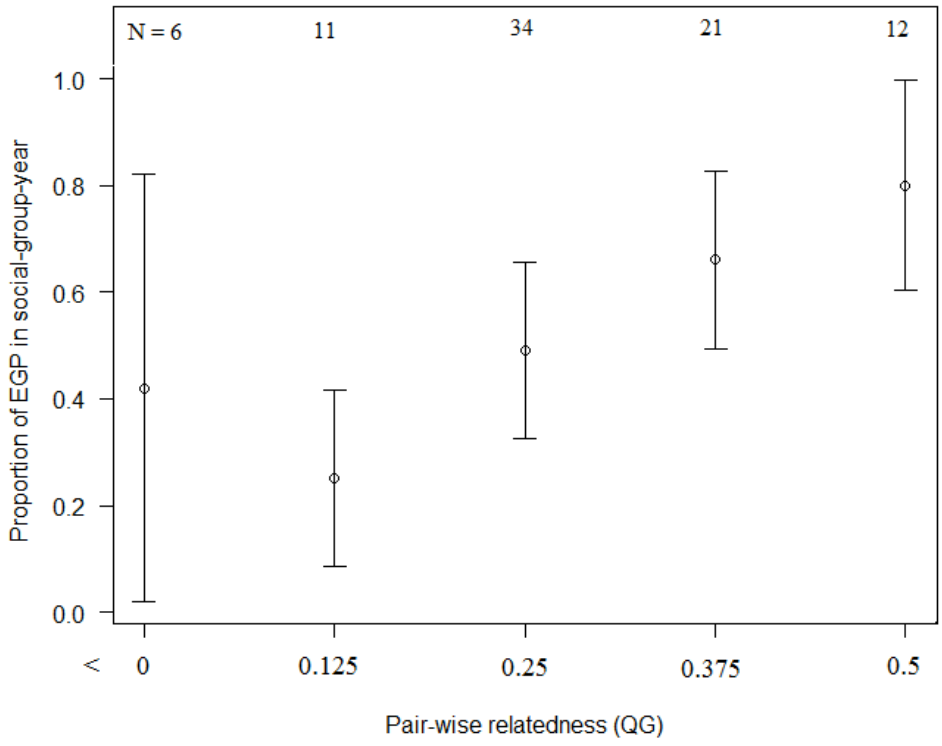


Figure 3 The relationship between the average pair-wise relatedness (Queller and Goodnight's estimator) of within social-group-year candidate parents and the proportion of offspring that are extra-group young at the social group level. For ease of visualisation, pair-wise values were grouped on the x-axis with error bars show ± SE. Numbers along the top of the figure indicate the number of social-group-years in each category.

Discussion

Our findings contribute to the contentious debate in evolutionary biology concerning the role of promiscuity and extra-pair (or group) mating, with evidence here of extra social group infidelity. We show that, at the litter level, EGP correlated negatively with the number of within-group candidate fathers, and positively with the number of neighbouring-group candidate fathers. The ratios of within-group fathers to within-group mothers, or to neighbouring-group candidate fathers, did not correlate with the rate of EGP. While we observed high rates of extra-group paternity, relating to the number of intra- and inter-group candidate fathers and their genetic composition (pair-wise relatedness), we found no evidence that rates of EGP corresponded to the heterozygosity of within-group males (Arnqvist & Kilpatrick, 2005; Griffith, 2007; Eliassen & Kokko, 2008; Forstmeier et al., 2011). Neither mean nor the maximum level of heterozygosity of within-group candidate fathers correlated with EGP rate in litters.

Within-group vs extra-group fidelity

Almost half of offspring that were assigned parentage involved infidelity from the parents' social group affiliation. This indicates that the badger social group unit does not correspond to a breeding unit (Carpenter et al., 2005; Dugdale et al., 2007). In meerkats, EGP rates decrease with higher numbers of defending males in a group (Spong et al., 2008). We observed that the number of candidate within-group males had a negative influence on the proportion of the females' offspring sired by extra-group candidate fathers. Similarly, EGP rate correlated positively with the number of neighbouring-group candidate fathers present. While a higher number of within-group males may deter intrusions by extra-group males it is also equally plausible that we are observing a simple density-dependent effect, where females mate *ad libitum* but that lower within-group male presence militates for more matings with extra-group males.

A greater proportion of EGO may result at the litter level if females actively solicit extra-group copulations (EGCs). Female solicitation of extra-group copulations is well documented in birds (e.g., Schwartz et al., 1999; Pryke et al., 2010) and mammals (Wolff & Macdonald, 2004). Female badgers may solicit EGCs to reduce the risk of infanticide. Circumstantial infanticide by males has been reported in badgers (Kruuk, 1989; Lups & Roper, 1990). By mating with potentially infanticidal males, female may increase the chances of survival of their offspring. Female snub-nosed langurs (*Rhinopithecus* spp) for example, reduce infanticide by mating with the invading males (Ren et al., 2011). It is not known to what extent the solicitation of extra-group matings is led by males or females, or equally, in badgers.

The majority of EGP matings were attributed to candidate fathers from neighbouring-groups. Badger movements between social groups are predominantly temporary and of short distance, i.e., to adjacent groups (Macdonald et al., 2008). Female badgers forage solitarily (Carr & Macdonald, 1986; da Silva et al., 1993), and are known to make transient extra-territorial excursions (Huck et al., 2008; Macdonald et al., 2008; see also Stewart et al., 1997) where they transgress into neighbouring groups (Macdonald et al., 2008) and appear compliant to EGCs, possibly in order to mitigate aggression (Macdonald et al., 2004a; Dugdale et al., 2011) or due to a fundamental evolved mating strategy, from

low(er) density conditions, where the opportunity to copulate is restricted and therefore all opportunities are taken (e.g., Erlinge & Sandell, 1986). Mate guarding (Brotherton & Komers, 2003) has not been observed in badgers – rather non-competitive serial copulations with different partners underscore a freely promiscuous system (Dugdale et al., 2007). Thus we cannot reject the null hypothesis / the simplest possible scenario; however we proceed to include further plausible socio-biological interpretation.

Inbreeding avoidance and heterozygosity

Extra-group mating has been proposed as an alternative strategy to avoid inbreeding in a population where dispersal is restricted or does not occur (e.g., Brouwer et al., 2011). Dispersal and inbreeding avoidance are generally related (see Moore & Ali, 1984). Durant and Hughes (2005) found that in the Australian magpie (*Gymnorhina tibicen*), EGP was inversely related to dispersal rates between two contrasting populations such that high philopatry corresponded with high EGP. Similarly, in some socially monogamous fairy-wrens, *Malurus splendens*, promiscuity provided a potential inbreeding avoidance mechanism (Brooker et al., 1990).

The high incidence of EGP and a negative correlation between EGP and average pair-wise relatedness between mothers and candidate fathers that have been shown in the current study (Figure. 3) indicate that female badgers may have evolved extra-group mating strategy to avoid inbreeding. At high density, badger social groups contain close relatives (Dugdale et al., 2008), because dispersal is limited (Pope et al., 2006), and both sexes are recruited into their natal groups. The lifetime reproductive success of inbred badgers in this population is lower than for out-bred (Dugdale, 2007a). Thus, female badgers may engage in extra-group copulations preferably, or whenever matings with close relatives is unavoidable, to reduce the probability of producing inbred offspring.

Females may copulate with extra-group males to obtain fitness benefits e.g., heterozygous advantage, good and compatible genes (Jennions & Petrie, 2000). The 'Good-gene-as-heterozygosity Hypothesis' predicts that females may seek extra-group copulation if within-group males have low level of heterozygosity (Brown, 1997). A paternal heterozygosity advantage on first-year survival probability (under most suitable environment conditions) has been observed in this badger population

(Annavi et al., in prep). This suggests that mating with males with higher levels of heterozygosity brings survival benefits. We would thus predict that when there are heterozygous males within-groups, females will mate with them, but when only males with low heterozygosity are available, females may benefit from mating outside of their social group. In this study, EGP rates were not related to the average or maximum heterozygosity of within-group candidate fathers. Small sample size for females with mixed paternity litters, however (within-group and extra-group candidate fathers assigned paternity to offspring in the same litter, N = 10), did not allow us to test for the biological significance of differences between within-group and extra-group fathers in terms of heterozygosity.

Pleiotropy

Female badgers clearly engage in promiscuous matings (Dugdale et al., 2007 & 2011). Two-third of males and females were exclusive extra-group breeders over their lifetimes, 31–56% exhibited a mixed strategy, while 24–29 % were observed to produce offspring only from within-group matings (Table 2). Importantly, females engaging in group infidelity tended to produce sons that were assigned more EGO, consequently any direct maladaptive costs to the breeding female may be remediated by selection for traits favouring a greater number of grandchildren (i.e., a "promiscuous-son benefit," similar to the "sexy-son benefit" in Fisher's runaway sexual selection; Fisher 1930; Weatherhead & Robertson, 1979).

Three mechanisms can produce positive genetic correlations between measures of male and female extra-pair mating behaviour (see Forstmeier et al., 2011): (*i*) Pleiotropic effects, where alleles for male infidelity also enhance infidelity in females carrying those alleles (Halliday & Arnold, 1987); (*ii*) Alleles that enhance infidelity, where may be in strong linkage disequilibrium between the sexes; (*iii*) assortative mating, where may also contribute to the genetic correlation between the sexes, due to linkage disequilibrium between physically unlinked alleles.

Conclusions

Despite almost half of offspring-yielding matings taking place between individuals from different, proximate, social groups, we found no evidence for mate selection for genetic quality (i.e., heterozygosity). A positive correlation between the rates of EGP and the average pair-wise relatedness between mothers and within-group candidate fathers, suggesting EGPs may play an important role in reducing inbreeding depression. Most interesting is that within-group versus extra-group parentages do not appear to be opposing mating strategies in badgers; rather these exist within a continuum of polygynandrous mating events. Failure to realise fitness advantages through superior extra-group offspring survival, recruitment or life-time reproductive success would confirm that infidelity in badgers is either socially adaptive, or inevitable, through sex-linked genetic correlation; a 'self-fulfilling prophecy' (Merton, 1968), rather than a selection for superior genetic quality.

References

Albrecht, T., Schnitzer, J., Kreisinger, J., Exnerova, A., Bryja, J. & Munclinger, P. 2007. Extrapair paternity and the opportunity for sexual selection in long-distant migratory passerines. *Behavioral Ecology*, **18,** 477–486.

Annavi, G., Dawson, D., Horsburgh, G., Greig, C., Dugdale, H., Newman, C., Macdonald, D. & Burke, T. 2011. Characterisation of twenty-one European badger (Meles meles) microsatellite loci facilitates the discrimination of second-order relatives *Conservation Genetic Resources,* **3,** 515–518.

Arnqvist, G. & Rowe, L. 2005. Sexual conflict. Princeton, NJ: Princeton University Press.

Bates, D. & Sarkar, D. 2007. lme4: Linear Mixed-Effects Models Using S4 Classes. R package version 0.9975-12, URL http://CRAN.R-project.org/.

Bohm, M., Hutchings, M.R. & White, P. 2009. Contact networks in a wildlife-livestock host community: Identifying high-risk individuals in the transmission of Bovine TB among badgers and cattle. *PLoS ONE*, **4,** e5016.

Borries, C., Savini, T. & Koenig, A. 2011. Social monogamy and the threat of infanticide in larger mammals. *Behavioral Ecololy & Sociobiology*, **65,** 685–693.

Britten, H. 1996. Meta-analyses of the association between multilocus heteroyzygosity and fitness. *Evolution*, **50**, 2158–2164.

Brooker, M.G., Rowley, I., Adams, M., Baverstock, P.R. 1990. Promiscuity: an inbreeding avoidance mechanism in a socially monogamous species? *Behavioral Ecology & Sociobiology*, **26,** 191–199.

Brotherton, P.N. & Komers, P.E. 2003. Mate guarding and the evolution of monogamy in mammals: Cambridge: Cambridge University Press.

Brouwer, L., Van de Pol, M., Atema, E. & Cockburn, A. 2011. Strategic promiscuity helps avoid inbreeding at multiple levels in a cooperative breeder where both sexes are philopatric. *Molecular Ecology*, **20,** 4796–4807.

Brown, J.L. 1997. A theory of mate choice based on heterozygosity. *Behavioral Ecology*, **8,** 60–65.

Cameron, C., Berteaux, D. & Dufresne, F. 2011. Spatial variation in food availability predicts extrapair paternity in the arctic fox. *Behavioral Ecology*, **22,** 1364–1373.

Carpenter, P.J., Pope, L.C., Greig, C., Dawson, D.A., Rogers, L.M., Erven, K., Wilson, G.J., Delahay, R.J., Cheeseman, C.L. & Burke, T. 2005. Mating system of the European badger, *Meles meles*, in a high density population. *Molecular Ecology*, **14,** 273–284.

Carr, G. & Macdonald, D.W. 1986. The sociality of solitary foragers: a model based on resourse dispersion. *Animal Behaviour*, **34**, 1540–1549.

Cohas, A., Bonenfant, C., Kempenaers, B. & Allaine, D. 2009. Age-specific effect of heterozygosity on survival in alpine marmots, *Marmota marmota*. *Molecular Ecology*, **18,** 1491–1503.

Cohas, A., Yoccoz, N.G., Da Silva, A., Goossens, B. & Allaine, D. 2006. Extra-pair paternity in the monogamous alpine marmot (*Marmota marmota*): the roles of social setting and female mate choice. *Behavioral Ecology & Sociobiology*, **59,** 597–605.

Coltman, D.W., Pilkington, J.G., Smith, J.A. & Pemberton, J.M. 1999. Parasite-mediated selection against inbred soay sheep in a free-living, Island population. *Evolution*, **53,** 1259–1267.

Coltman, D.W. & 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.

Csillery, K., Johnson, T., Beraldi, D., Clutton-Brock, T.H., Coltman, D., Hansson, B., Spong, G. & Pemberton, J.M. 2006. Performance of marker-based relatedness estimators in natural populations of outbred vertebrates. *Genetics*, **173,** 2091–2101.

da Silva, J., Macdonald, D.W. & Evans, P.G.H. 1994. Net costs of group living in a solitary forager, the Eurasian badger (*Meles meles*). *Behavioral Ecology*, **5,** 151–158.

da Silva, J., Woodroffe, R. & Macdonald, D.W. 1993. Habitat, food availability and group territoriality in the European badger, *Meles meles*. *Oecologia*, **95,** 558–564.

Daniels, S. & Walters, J. 2000. Inbreeding depression and its effects on natal dispersal in red-cockaded woodpeckers. *Condor*, **102,** 482–491.

Delahay, R.J., Brown, J.A., Mallinson, P. J., Spyvee, P. D., Handoll, D., Rogers, L. M. & Cheeseman, C. L. 2000. The use of marked bait in studies of the territorial organization of the European Badger *(Meles meles)*. *Mammal Review,* **30,** 73–87.

Dugdale, H.L., Griffiths, A. & Macdonald, D.W. 2011. Polygynandrous and repeated mounting behaviour in European badgers, *Meles meles*. *Animal Behaviour*, **82,** 1287–1297.

Dugdale, H.L., Macdonald, D.W., Pope, L.C. & Burke, T. 2007. Polygynandry, extra-group paternity and multiple-paternity litters in European badger (*Meles meles*) social groups. *Molecular Ecology*, **16,** 5294–5306.

Dugdale, H.L., Macdonald, D.W., Pope, L.C., Johnson, P.J. & Burke, T. 2008. Reproductive skew and relatedness in social groups of European badgers, *Meles meles*. *Molecular Ecology*, **17,** 1815–1827.

Durrant, K.L. & Hughes, J.M. 2005. Differing rates of extra-group paternity between two populations of the Australian magpie (*Gymnorhina tibicen*). *Behavioral Ecology & Sociobiology*, **57,** 536–545.

Eliassen, S. & Kokko, H. 2008. Current analyses do not resolve whether extra-pair paternity is male or female driven. *Behavioral Ecology & Sociobiology*, **62,** 1795–1804.

Erlinge, S. & Sandell, M. 1986. Seasonal changes in the social organisation of stoats, *Mustela erminea*: An effect on shifts between two decisive resources. *Oikos,* **47,** 57–62.

Fisher, R.A. 1930. The genetical theory of natural selection. Oxford: Oxford University Press.

Forstmeier, W., Martin, K., Bolund, E., Schielzeth, H. & Kempenaers, B. 2011. Female extrapair mating behavior can evolve via indirect selection on males. *Proceedings of the National Academy of Sciences*, **108,** 10608–10613.

Griffith, S.C. 2007. The evolution of infidelity in socially monogamous passerines: Neglected components of direct and indirect selection. *The American Naturalist*, **169,** 274–281.

Griffith, S.C., Owens, I.P.F. & Thuman, K.A. 2002. Extra pair paternity in birds: a review of interspecific variation and adaptive function. *Molecular Ecology*, **11,** 2195–2212.

Hadfield, J.D., Richardson, D.S. & Burke, T. 2006. Towards unbiased parentage assignment: combining genetic, behavioural and spatial data in a Bayesian framework. *Molecular Ecology*, **15,** 3715–3730.

Halliday, T. & Arnold, S.J. 1987. Multiple mating by females - A perspective from quantitative genetics. *Animal Behaviour*, **35,** 939–941.

Hansson, B. & Westerberg, L. 2002. On the correlation between heterozygosity and fitness in natural populations. *Molecular Ecology*, **11,** 2467–2474.

Huck, M., Frantz, A.C., Dawson, D.A., Burke, T.J. & Roper, T. 2008. Low genetic variability, female-biased dispersal and high movement rates in an urban population of Eurasian badgers *Meles meles*. *Journal of Animal Ecology*, **77,** 905–915.

Isvaran, K. & Clutton-Brock, T.H. 2007. Ecological correlates of extra-group paternity in mammals. *Proceedings of the Royal Society of London B: Biological Science*, **274,** 219–224.

Jennions, M.D. & Petrie, M. 2000. Why do females mate multiply? A review of the genetic benefits. *Biological Reviews*, **75,** 21–64.

Johnson, P.C.D. & Haydon, D.T. 2007. Maximum-likelihood estimation of allelic dropout and false allele error rates from microsatellite genotypes in the absence of reference data. *Genetics*, **175,** 827–842.

Kempenaers, B. 2007. Mate choice and genetic quality: A review of the heterozygosity theory. *Advances in the Study of Behavior*, **37,** 189–278.

Kilshaw, K., Newman, C., Buesching, C.D, Bunyan, J. & Macdonald D.W. 2009. Coordinated latrine use by European badgers, *Meles meles*: Potential consequences for territory defense. *Journal of Mammalogy*, **90,** 1188–1198.

Kokko, H. & Rankin, D.J. 2006. Lonely hearts or sex in the city? Density-dependent effects in mating systems. *Philosophical Transactions of the Royal Society B: Biological Science*, **361,** 319–334.

Kruuk, H. 1978a. Foraging and spatial organisation of the European Badger, *Meles meles* L. *Behavioral Ecology & Sociobiology,* **4,** 75–89.

Kruuk, H. 1978b. Spatial organisation and territorial behaviour of the European badger, *Meles meles*. *Journal of Zoology (London),* **184,** 1–19.

Kruuk, H. 1989. The Social Badger: Ecology and behaviour of a group living carnivore (*Meles meles*). Oxford: Oxford University Press.

Kruuk, L.E.B., Sheldon, B.C. & Merila, J. 2002. Severe inbreeding depression in collared flycatchers (*Ficedula albicolis*). *Proceedings of the Royal Society B: Biological Sciences*, **269,** 1581–1589.

Lups, P. & Roper, T. 1990. Cannibalism in a female badger (*Meles meles*): infanticide or predation? *Journal of Zoology*, **221,** 314–315.

Lynch, M. & Ritland, K. 1999. Estimation of pairwise relatedness with molecular markers. *Genetics*, **152,** 1753–1766.

Macdonald, D.W., Harmsen, B.J., Johnson, P.J. & Newman, C. 2004a. Increasing frequency of bite wounds with increasing population density in Eurasian badgers, *Meles meles*. *Animal Behaviour*, **67,** 745–751.

Macdonald, D.W. & Newman, C. 2002. Population dynamics of badgers (*Meles meles*) in Oxfordshire, U.K.: numbers, density and cohort life histories, and a possible role of climate change in population growth. *Journal of Zoology (London)*, **256,** 121–138.

Macdonald, D.W., Newman, C., Buesching, C.D. & Johnson, P.J. 2008. Male-biased movement in a high-density population of the European badger (*Meles meles)*. *Journal of Mammalogy*, **89,** 1077–1086.

Macdonald, D.W., Newman, C., Dearr, J., Buesching, C.D. & Johnson, P.J. 2004b. The distribution of Eurasian badger, *Meles meles,* setts in a high density area: field observations contradict the sett dispersion hypothesis. *Oikos*, **106,** 295–307.

Macdonald, D.W., Newman, C., Nouvellet, P.M. & Buesching, C.D. 2009. An Analysis of Eurasian badger (*Meles meles*) population dynamics: Implication for regulatory mechanisms. *Journal of Mammalogy*, **90,** 1392–1403.

Mainguy, J., Cote, S. & Coltman, D. 2009. Multilocus heterozygosity, parental relatedness and individual fitness components in a wild mountain goat, *Oreamnos americanus* population. *Molecular Ecology*, **18,** 2297–2306.

McLaren, G.W., Thornton, P.D., Newman, C., Buesching, C.D, Baker, S.E., Mathews, F. & Macdonald, D.W. 2005. The use and assessment of ketamine–medetomidine– butorphanol combinations for field anaesthesia in wild European badgers (*Meles meles*). *Veterinary Anaesthesia & Analgesia*, **32,** 367–372.

Mead, R. 1989. *Conservation Biology and Black-Footed Ferret*. Dexter, MI: Thomson-Shore.

Merton, R. 1968. Social Theory and Social Structure. New York: Free Press.

Moore, J. & Ali, R. 1984. Are dispersal and inbreeding avoidance related? *Animal Behaviour*, **32,** 94–112.

Morecroft, M., Taylor, M. & Oliver, H. 1998. Air and soil microclimates of deciduous woodland compared to an open site. *Agricultural & Forest Meteorology*, **90,** 141–156.

Neff, B. & Pitcher, T. 2005.Genetic quality and sexual selection: an integrated framework for good genes and compatible genes. *Molecular Ecology*, **14,** 19–38.

Pope, L.C., Domingo-Roura, X., Erven, K. & Burke, T. 2006. Isolation by distance and gene flow in the Eurasian badger (*Meles meles*) at both a local and broad scale. *Molecular Ecology*, **15,** 371–386.

Pryke, S.R., Rollins, L.A. & Griffith, S.C. 2010. Females use multiple mating and genetically loaded sperm competition to target compatible genes. *Science*, **329,** 964–967.

Pusey, A. & Wolf, M. 1996. Inbreeding avodance in animals. *Trends in Ecology & Evolution*, **11,** 201–206.

Queller, D.C. & Goodnight, K.F. 1989. Estimating relatedness using genetic-markers. *Evolution*, **43,** 258–275.

R Development Core Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Raymond, M. & Rousset, F. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86,** 248–249.

Reid, J.M., Arcese, P., Keller, L.F., Elliot, K.H., Sampson, L. & Hasselquist, D. 2007. Inbreeding effects on immune response in free-living song sparrow (*Melospiza melodia*). *Proceedings of the Royal Society B: Biological Sciences*, **274,** 697–706.

Reid, J.M., Arcese, P., Sardell, R.J. & Keller, L.F. 2011. Heritability of female extra-pair paternity rate in song sparrows (*Melospiza melodia*). *Proceedings of the Royal Society B: Biological Sciences*, **278,** 1114–1120.

Ren, B., Li, D., He, X., Qiu, J. & Li, M. 2011. Female resistance to invading males increases infanticide in langurs. *PLoS ONE*, **6,** 1–4.

Revilla, E. & Palomares, F. 1999. Changes in the behaviour of a male European badger: evidence in favour of the anti-kleptogamy hypothesis? *Acta Theriologica*, **44,** 471–476.

Rice, W.R. 1992. Sexually antagonistic genes: Experimental evidence. *Science*, **256,** 1436–1439.

Rosalino, L., Macdonald, D.W. & Santos-Reis, M. 2004. Spatial structure and land-cover use in a low-density Mediterranean population of Eurasian badgers. *Canadian Journal of Zoology,* **82,** 1493–1502.

Savill, P.S., Perrins, C.M., Kirby, K.J., Fisher, N. 2010. *Wytham Woods: Oxford's ecological laboratory*. Oxford, UK: Oxford University Press

Schubert, G., Stoneking, C.J., Aranjelovic, M., Boesch, C., Eckhardt, N., Hohmann, G., Langergarber, K., Lukas, D. & Vigilant, L. 2011. Male-mediated gene flow in patrilocal primates. *PLoS ONE*, **6,** e21514.

Schwartz, M.K., Boness, D.J., Schaeff, C.M., Majluf, P., Perry, E.A. & Fleischer, R.C. 1999. Female-solicited extrapair matings in Humboldt penguins fail to produce extrapair fertilizations. *Behavioral Ecology*, **10,** 242–250.

Sheldon, B. 1994. Male phenotype, fertility, and the pursuit of extra-pair copulations by female birds. *Proceedings of the Royal Society B: Biological Sciences,* **257,** 25–30.

Slate, J., David, P., Dodds, K.G., Veenvilet, B.A., Glass, B.C., Broad, T.E. & McEwan, J.C. 2004. Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data. *Heredity*, **93,** 255–265.

Soulsbury, C.D. 2010. Genetic patterns of paternity and testes size in mammals. *PLoS ONE*, **5,** e9581.

Spong, G.F., Hodge, S.J., Young, A.J. & Clutton-Brock, T.H. 2008. Factors affecting the reproductive success of dominant male meerkats. *Molecular Ecology*, **17,** 2287–2299.

Stewart, P. D., Ellwood, S. A. & Macdonald, D. W. 1997. Remote video-surveillance of wildlife – an introduction from experience with the European badger *Meles meles*. *Mammal Review*, **27,** 185–204

Sundqvist, C., Ellis, L. & Bartke, A. 1988. Reproductive endocrinology of the mink (*Mustela vison*). *Endocrine Reviews,* **9,** 247–266.

Szulkin, M., Zelazowski, P., Nicholson, G. & Sheldon, B.C. 2009. Inbreeding avoidance under different null models of random mating in the great tit. *Journal of Animal Ecology,* **78,** 778–788.

Tuyttens, F.A.M., Macdonald, D.W., Delahay, R., Rogers, L.M., Mallinson, F.J., DOnnelly, C.A. & Newman, C. 1999. Differences in trappability of European badgers *Meles meles* in three populations in England. *Journal of Applied Ecology*, **36,** 1051–1062.

Uller, T. & Olsson, M. 2008. Multiple paternity in reptiles: patterns and processes. *Molecular Ecology,* **17,** 2566–2580.

van Noordwijk, M.A. & van Schaik, C.P. 2004. Sexual selection and the careers of primate males: paternity concentration, dominance-acquisition tactics and transfer decisions. In: *Sexual selection in primates: new and comparative perspectives* (Ed. by Kappeler, P.M. & van Schaik, C.M.). Cambridge, UK: Cambridge University Press.

Vedder, O., Komdeur, J., Van der Velde, M., Schut, E. & Magrath, M.J.L. (2011). Polygyny and extra-pair paternity enhance the opportunity for sexual selection in blue tits. *Behavioral Ecology & Sociobiology*, **65,** 741–752.

Wang, J. 2011. COANCESTRY: a program for simulating, estimating and analysing relatedness and inbreeding coefficients. *Molecular Ecology Resources*, **11,** 141–145.

Wang, J. & Santure, A.W. 2009. Parentage and sibship inference from multilocus genotype data under polygamy. *Genetics*, **181,** 1579–1594.

Weatherhead, P.J. & Robertson, R.J. 1979. Offspring quality and the polygyny threshold –Sexy son hypothesis. *The American Naturalist*, **113,** 201–208.

Westneat, D.F, Sherman, P.W. & Morton, M.L. 1990. The ecology and evolution of extra-pair copulations in birds. *Current Ornithology*, 7, 331–369.

Westneat, D.F. & Stewart, I.R.K. 2003. Extra-pair paternity in birds: Causes, correlates, and conflict. *Annual Review of Ecology, Evolution, & Systematics*, **34,** 365–396.

Whiteman, N., Matson, K., Bollmer, J. & Parker, P. 2006. Disease ecology in the Galápagos Hawk (*Buteo galapagoensis*): host genetic diversity, parasites, and natural antibodies. *Proceedings of the Royal Society B: Biological Sciences*, **273,** 797–804.

Wolff, J.O. 1993. Why are female small mammals territorial? *Oikos,* **68,** 364–370.

Wolff, J.O. & Macdonald, D.W. 2004. Promiscuous females protect their offspring. *Trends in Ecology & Evolution*, **19,** 127–134.

Woodroffe, R., Macdonald, D.W. & da Silva, J. 1995. Dispersal and philopatry in the European badger, *Meles meles*. *Journal of Zoology*, **237,** 227–239.

Young, A.J., Spong, G. & Clutton-Brock, T. 2007. Subordinate male meerkats prospect for extra-group paternity: alternative reproductive tactics in a cooperative mammal. *Proceedings of the Royal Society B: Biological Sciences*, **274,** 1603–1609.

Supplementary materials

Supplementary Table S1 Parameter estimates from GLMMs analysing *the absolute number of extra-group paternity (EGP) or extra-mate pairs* in a litter in relation to group density and composition. Mean SH = Individual standardised heterozygosity and LR = average Lynch and Ritland's pair-wise estimator between mothers and within-group candidate fathers. Unrestricted dataset, UR_{LTR}: SGY = 124, no of litters = 220 representing 310 cubs: 174 WGO and 136 EGO.

No	Explanatory variable	Extra-group paternity			Extra-mate pairs		
		Estimate ± SE	$\chi^2_{[df=1]}$	P	Estimate ± SE	$\chi^2_{[df=1]}$	P
	Final model						
1	Intercept	-2.74 ± 0.39	-	-	-1.69 ± 0.31	-	-
2	Average pair-wise relatedness (LR)	2.45 ± 0.56	17.07	**<0.001**	2.18 ± 0.57	12.44	**<0.001**
3	No of neighbouring-group candidate fathers	0.02 ± 0.01	1.10	**0.008**	0.02 ± 0.01	5.01	**0.025**
4	Litter size	0.73 ± 0.12	32.04	**<0.001**	-	-	-
	Rejected terms						
1	Quadratic effect of LR	-4.48 ± 2.44	3.89	0.050	-4.80 ± 2.82	3.42	0.064
2	Average SH within-group candidate fathers or	-1.13 ± 0.90	1.46	0.227	-0.65 ± 0.95	0.47	0.492
	Maximum SH within-group candidate fathers	-0.43 ± 0.69	0.39	0.534	-0.27 ± 0.75	0.02	0.900
3	Ratio of within-group candidate fathers to neighbouring-group candidate fathers	-0.47 ± 0.37	1.78	0.182	-0.29 ± 0.41	0.55	0.460
4	No of within-group candidate fathers	-0.04 ± 0.03	1.91	0.167	-0.05 ± 0.03	3.30	0.069
5	Ratio of within-group candidate fathers to within-group mothers	0.05 ± 0.05	0.90	0.342	0.01 ± 0.06	0.01	0.921

In the absolute model of extra-group paternity, the random effect for social group, year and mother ID were 0.08 ± 0.29, 0.01 ± 0.10 and 0, respectively. For extra-mate pairs model, they were 0.96 ± 0.98, 0.35 ± 0.59 and 1.25 ± 1.12, respectively.

Supplementary Table S2 Parameter estimates from GLMMs analysing *the relative number of extra-group paternity (EGP) or extra-mate pairs* in a litter in relation to group density and composition. Mean SH = Individual standardised heterozygosity and LR = average Lynch and Ritland's pair-wise estimator between mothers and within-group candidate fathers. Unrestricted dataset, UR_{LTR}: SGY = 124, no of litters = 220 representing 310 cubs: 174 WGO and 136 EGO.

No	Explanatory variable	Extra-group paternity			Extra-mate pairs		
		Estimate ± SE	$\chi^2_{[df=1]}$	P	Estimate ± SE	$\chi^2_{[df=1]}$	P
	Final model						
1	Intercept	-3.65 ± 1.41			-2.09 ± 0.73		
2	Average pair-wise relatedness (LR)	12.83 ± 2.91	23.59	**<0.001**	6.82 ± 1.47	21.93	**<0.001**
3	No of neighbouring-group candidate fathers	0.13 ± 0.04	12.08	**<0.001**	0.07 ± 0.02	9.87	**0.002**
4	No of within-group candidate fathers	-0.37 ± 0.14	6.71	**0.010**	-0.18 ± 0.08	4.51	**0.034**
	Rejected terms						
1	Quadratic effect of LR	-19.90 ± 10.38	3.84	0.050	-10.36 ± 5.59	3.07	0.080
2	Average SH within-group candidate fathers or	-2.04 ± 3.34	0.09	0.765	-1.05 ± 2.09	0.23	0.630
	Maximum SH within-group candidate fathers	-1.08 ± 2.19	0.00	0.999	-0.27 ± 0.75	0.14	0.713
3	Ratio of within-group candidate fathers to neighbouring-group candidate fathers	-0.79 ± 1.45	0.42	0.517	-0.45 ± 0.82	0.29	0.592
4	Ratio of within-group candidate fathers to within-group mothers	0.08 ± 0.13	0.28	0.594	0.05 ± 0.10	0.23	0.628
5	Litter size	0.68 ± 0.46	2.14	0.143	-	-	-

In the relative model of extra-group paternity, the random effect for social group, year and mother ID were 4.41 ± 2.10, 1.32 ± 1.15 and 11.41 ± 3.38, respectively. For extra-mate pairs model, they were 0.96 ± 0.98, 0.35 ± 0.59 and 1.25 ± 1.12, respectively

Appendix 4

Choosing mates for offspring genetic quality: Extra-group paternity does not enhance fitness in the European badger (*Meles meles*)

Choosing mates for offspring genetic quality: Extra-group paternity does not enhance fitness in the European badger (*Meles meles*)

Abstract

In many group-living species, females often produce offspring fathered by males from outside of their social group (i.e., extra-group paternity, EGP). Why this reproductive strategy is adopted in the mating system of certain species, or under certain socio-ecological conditions remains unclear. The hypothesis that females obtain indirect genetic benefits from EGP has received considerable attention, especially in birds, but the evidence has been equivocal. Here, we test whether extra-group offspring (EGO; offspring sired by extra-group males) exhibit greater fitness (survival and lifetime reproductive success, LRS) than within-group offspring (WGO). EGO had a lower first-year survival probability than WGO, and survived on average for 1.3 years less. Female WGO produced more litters and offspring over their lifetime compared with female EGO. The opposite was observed in males: male EGO produced more offspring in their lifetime than did male WGO. These findings indicate that, for females, the EGP reproductive strategy has a cost, where WGO confer greater lifetime fitness. Sex-specific differences in the LRS advantages in producing EGO or WGO may reflect sexually antagonistic genetic effects, which may determine constraints that maintain the EGP in this population.

Introduction

Over the last decade, the understanding of genetic benefits increasingly informed insights into mate choice evolution (see reviews by Mays & Hill, 2004; Neff & Pitcher, 2005; Roberts et al., 2006; Kempenaers, 2007). In general, individuals are predicted to select for genetic quality in their mating partners when this increases offspring viability and future reproductive success (Ryder et al., 2010), where often females are the more selective sex (e.g., Double & Cockburn, 2000) due to greater investment in offspring (Bateman, 1948).

Two competing, often not mutually exclusive, models have been proposed to explain the effect of mate choice on the genetic quality of offspring, where quality is defined as the contribution of an allele or a genotype to offspring fitness (Kempenaers, 2007);

1) The good allele model: an allele that enhances offspring fitness independent of the genome, and
2) The compatible allele model: a specific male allele, when combined with a specific female allele, enhances offspring fitness in a specific genetic context either via heterozygote advantage or gene–gene interactions (e.g., epistasis, Neff & Pitcher, 2005; Kempenaers, 2007).

Both models predicted mate choice informed by the phenotypic traits of the mating partner, which signal either specific alleles (e.g., advantageous alleles in the traditional 'Good-genes Hypothesis') or general allelic diversity (i.e., heterozygosity).

In the good allele model, one sex selects their mating partner irrespective of their own genotype. In the compatible allele model, however, mate suitability depends largely on the interaction of male and female genotypes; in some cases (e.g., major histocompatibility complex, see Sin et al., 2012), this hypothesis also predicts that genetically dissimilar mates may produce superior offspring / experience greater fitness (reviewed in Mays & Hill, 2004; Kempenaers, 2007).

Extra-group paternity (henceforth EGP), is a common feature of mating systems, found in approximately 90% of bird species (reviewed by Griffits et al., 2002) and in group-living mammals (e.g., Red foxes, *Vulpes vulpes*, Baker et al., 2004; Alpine marmots, *Marmota marmota*, Goossens et al., 1998; Ethiopian wolves, *Canis simensis*, Sillero-Zubiri et al., 1996; Meerkat, *Suricata suricatta*, Young et al., 2007, see also Isvaran and Clutton-Brock, 2007; Soulsbury, 2010). EGP results in females giving birth to offspring that are fathered by males that live outside of their mating-pair or

social group. EGP has major implications for the evolution of mating systems and the dynamics of sexual selection (Isvaran & Clutton-Brock, 2007). Although EGP occurs extensively, the genetic benefits of EGP have been analysed primarily in birds (e.g., (Akcay & Roughgarden, 2007); Ryder et al., 2010; (Sardell *et al.*, 2011; Schmoll *et al.*, 2011), and such benefits in mammalian mating systems have not been investigated as thoroughly (but see Cohas et al., 2007a & b).

Here we make a quantitative assessment of the within-group and extra-group paternity benefits conferred on offspring fitness, in the European badger (*Meles meles*). The European badger exhibits different social strategies, according to resources and population density over its extensive range (see Newman et al., 2011). In our high-density UK study population, badgers live in groups of up to 29 individuals (da Silva et al., 1994), and exhibit relatively well synchronised parturition, with one litter per year – although there is evidence that the badger is an induced ovultator (Canivenc, 1966), with a post-partum mating peak (though also noted to breed later in the year; Creswell et al., 1992) – delayed implantation leads to synchronised parturition (Yamaguchi et al., 2006).

The large social groups, in which Wytham badgers typically live, usually consist of close kin (Dugdale et al., 2008). Dugdale et al. (2007b) report a mean of 5.60 breeding females per social-group-year (95% confidence interval, CI [4.82, 6.38]) and 5.80 males (CI [5.02, 6.58]), with up to seven breeding members of both sexes. Badgers have one litter per year with a mean post-emergence litter size of 1.46 (this study, CI [1.43, 1.49]; range = 1–5). Multiple-paternities occurred in 27% of the multi-parous litters, with a maximum of two fathers per litter (this study).

Badgers provide an informative species with which to confirm or refute selection for genetic quality in terms of good, and compatible, alleles, due to their highly polygynadrous mating system, where females mate with several males and males with several females (Dugdale et al., 2007b; 2011a), Furthermore, badgers transgress into one another's territories (Macdonald et al., 2008) and cubs may be sired either by males from their mother's social group, or from other groups (Carpenter et al., 2005; Dugdale et al., 2007b; Annavi et al., in prep). Consequently, rates of EGP reach almost 50% in some high-density badger populations (Carpenter et al., 2005; Dugdale et al., 2007b).

We assume mates have either the ability to assess genetic quality in their partner using cues and/or the ability to assess a mate's genotype with respect to their own genotype (e.g., morphological, behavioural, olfactory, see Buesching et al., 2002a & b; Ryder et al., 2010). For example, badger sub-

caudal scent contains information on individual specific parameters and group membership (Buesching et al., 2002a & b), and this trait may impact on mate choice.

Here, we use a long-term dataset on the demographics and reproductive patterns (using a genetic pedigree) of a high-density badger population (see Macdonald & Newman, 2002; Macdonald et al., 2009) to investigate the fitness consequences of EGP. We examine whether extra-group offspring (EGO) and within-group offspring (WGO) vary in terms of first-year survival probability or estimated longevity. We consider whether offspring-paternity type is correlated with their lifetime reproductive success (LRS), in terms of the numbers of litters, or offspring, they produce. We test whether estimated longevity and LRS of EGO and WGO depend on sex (there was no sex-specific differences in relative offspring-paternity type first-year survival probability, Annavi et al., in prep). Evidence of sex-specific variation in fitness-related traits between EGO and WGO, with any additive genetic basis, would suggest a sex-specific antagonistic effect (i.e., opposing selection pressure between the sexes) that constrain and maintain extra-group reproduction in the population (Foerster et al., 2007; Brommer et al., 2007). We predict that if extra-group mating increases offspring fitness, EGO will exhibit:

(i) greater first-year survival probability;

(ii) greater lifetime reproductive success; and

(iii) survive to an older age, than WGO.

This informs our evaluation of the selective pressures favouring EGP in badgers. Benefits to EGP are required in order to reject the developing proposition (Annavi et al., in prep) that extra-group promiscuity does not confer fitness advantages to parents, from which we would infer either pleiotropic genetic linkage (of genes for extra-pair mating) between the sexes, or mechanisms connected to an adaptionist social function (Forstmeier et al., 2011).

Materials and methods

Study species, population and field methods

This study was conducted in Wytham Woods, a 424-ha site, 6 km northwest of Oxford in southern England (51: 46: 26N; 1: 19: 19W). The Wytham badger population inhabits a partly discrete

geographical area, which limits, but does not eliminate, the potential for migration into or out of the study area (Macdonald & Newman, 2002). Dispersal is restricted, although temporary movements occur (male-biased), most often to neighbouring groups (Macdonald et al., 2008).

Since 1987 a systematic capture-mark-recapture regime has attempted to trap and mark all badgers in the population (Macdonald & Newman, 2002; Macdonald et al., 2009). Trapping was conducted three to four times per annum; typically over two weeks in June, September and November, with one week of trapping in the January of some years (Macdonald et al., 2009). Badgers were caught in mesh-traps baited with peanuts, placed near the entrances of active setts (Tuyttens et al., 1999; Macdonald & Newman 2002; Macdonald et al., 2009). Captured badgers were transferred to holding cages and transported to a central handling facility, where they were sedated by an intra-muscular injection using ketamine hydrochloride at 0.2 ml/kg body weight (McLaren et al., 2005; Thornton et al., 2005). Upon their first capture all badgers were tattooed with a unique individual number on the left inguinal region (Cheeseman & Harris, 1982) for permanent individual identification. Sex, age (cub or adult) and location of capture (sett and social group affiliation were determined through bait-marking surveys, Kilshaw et al., 2009) were recorded for each badger. For genetic analysis, hair samples and/or blood from the jugular vein (c. 3 ml) were collected from all individuals upon first capture.

Genotyping and parentage analyses

Blood or hair samples, collected from 1988 until 2009, were genotyped using 35 microsatellite markers. Details of the microsatellite characterisations and genotyping procedures used for parentage analysis are described in detail elsewhere (Dugdale et al., 2007b; Annavi et al., 2011 & in prep). Briefly, DNA was extracted using a slightly modified Chelex protocol (Chelex® 100, Bio-Rad, Hercules CA, USA: Walsh et al., 1991), then sequenced and analysed using GENEMAPPER® 3.5 software. Of the 1247 badgers that were first caught as cubs (N = 899) and adults (N = 348, unknown age) between 1988 and 2009, 764 (85%) and 256 (74%) were genotyped, respectively.

Parentage analysis was performed using the software MasterBayes 2.47 (Hadfield et al., 2006), implemented in the R statistics package 2.12.2 (R Development Core Team 2010), and Colony

2.0 (Wang & Santure 2009). Parentage was allocated based on those cubs that were able to catch and sample after weaning / independence (at 12-14 weeks; an ethical constraint to ensure animal welfare standards) to which parentage was assigned with at least 80% confidence. Maternity and paternity were assigned to 648 (85%) and 618 (81%) cubs respectively, born between 1988 and 2009. Both parents were assigned to 534 (70%) cubs.

The social group in which an individual was first trapped as a cub was assumed to be its natal social group. Offspring that were fathered by males in their natal social group were categorised as within-group offspring (WGO), whereas offspring fathered by males that did not reside in their offspring's natal social group in the year of conception (cubs are born one year after conception, due to delayed implantation; see Thom et al., 2004) were categorised as extra-group offspring (EGO). The social group of residence of each individual per year was defined based on their trapping history according to the rules described in detail in Annavi et al., in prep. Litters comprised of only WGO were termed within-group litters (WGL); only EGO were termed extra-group litters (EGL); and, litters including both WGO and EGO were denoted as mixed-litters (ML).

Statistical analyses

First-year survival probability

Cormack-Jolly-Seber models were performed in MARK 6.1 (White & Burnham, 1999), to estimate and compare the first-year survival probability of EGO and WGO born between 1988 and 2009 that were trapped in their first year 1989–2010. The input data consisted of capture-recapture histories for all offspring (e.g., badgers first caught as cubs that were assigned a father, so that the offspring-paternity type [EGO or WGO] could be assigned), over 23 years (1988–2010; Dataset A, the complete dataset, N = 618 offspring [295 EGO and 323 WGO]); data from each trapping session within a year were pooled into one annual capture record. We adopted the basic model infrastructure from Annavi et

al., in prep (see also Macdonald et al., 2009), using: two age-classes (cub < 1 years old; adult > 1 years old) with year-dependent survival applied to the cub age-class and consistent, but sex-specific, survival probabilities applied to the adult age-class, with year-dependent recapture probabilities. Estimates of recapture probabilities were established using the primary model, which were then fixed in all models. Dividing the model deviance by the bootstrapped deviance showed that our basic model was slightly over dispersed ($\hat{c}$ = 1.06; N = 100 replicates); thus, we adjusted the Akaike information criterion (corrected for small sample size [AICc] – Burnham & Anderson, 2002) value accordingly, using quasi-likelihood (QAICc).

We included all effects, shown by previous studies to be associated with first-year survival probability in this population, in our basic model:

(a) standardised total summer rainfall (May–October) (Macdonald et al., 2010; Annavi et al., in prep);

(b) standardised mean winter temperature (November–February) (Macdonald et al., 2010; Annavi et al., in prep);

(c) paternal multilocus heterozygosity (Annavi et al., in prep);

(d) the interaction between paternal multilocus heterozygosity and standardised total summer rainfall (Annavi et al., in prep); and

(e) $\log_e(x+1)$ transformed endo-parasitic (coccidial) load (Newman et al., 2001; Nouvellet et al., 2010).

These five effects (a-e) were retained in all models even if they were not statistically significant; however, very few badgers had coccidial load data, limiting this sample size. We therefore ran each model twice (with and without coccidial load e.g., Dataset A [complete dataset] and Dataset A_{coccid} [complete dataset, restricted coccidial load]). By integrating these potentially confounding effects in our models we were able to reduce the unexplained variance in first-year survival of offspring, thereby increasing the power to detect the real effects of the three main tests that we ran.

We first tested the effect of maternal age (Test 1: The effect of maternal age, a continuous effect), which influences first-year offspring survival rates in other mammalian species (e.g., Descamps et al., 2008; Mech & McRoberts 1990). These data for this analysis (Dataset B, restricted maternal age) were limited to 278 offspring (130 EGO and 148 WGO; from Dataset A) for which the

age of their mother was known. Test 1 was also run on badgers from Dataset B that had coccidial load data (Dataset B_{coccid}, restricted maternal age and coccidial load: 59 offspring [27 EGO and 32 WGO]).

We then investigated the influence of litter-type; that is, whether WGO or EGO born in different litter types ([I] WGL vs. ML; [II] EGL vs. ML) differed in their first-year survival probability (Test 2: The effect of litter type). Our analysis was restricted to litters where all the offspring within a litter had an assigned father (Dataset C, restricted complete litters: [I] N = 281 offspring - 269 WGO in 191 WGL and 12 WGO in 10 ML; [II] N = 233 offspring - 220 EGO in 158 EGL and 13 EGO in 10 ML). Test 2 was also run just for badgers in Dataset C with coccidial load data (Dataset C_{coccid}: [I] N = 57 offspring - 55 WGO in 47 WGL and 2 WGO in 2 ML, [II] N = 48 offspring - 43 EGO in 34 EGL and 5 EGO in 4 ML).

We also tested for difference in first-year survival probability according to offspring-paternity type (Test 3: The effect of offspring-paternity type; Dataset A, complete dataset). Test 3 was again run using the same dataset A, but restricted to badgers for which historic parasite data were availability (Dataset A_{coccid}).

Estimated longevity

Estimated longevity (hereafter longevity), defined in the CMR framework as the age at which an individual's was last captured, was measured for badgers that at least reached sexual maturity (i.e., one year old, Cresswell et al., 1992). Based on a modified 'minimum number alive' technique (Krebs, 1966; see also Macdonald & Newman, 2002), We assessed longevity of individuals first caught as cubs from cohort 1988 to 1996, allowing sufficient time (1997–2010) for each to achieve a maximal lifespan of up to ca. 15 years (see Dugdale et al., 2011b) i.e., no extant representatives of these cohorts remained after 2008, confirmed as no subsequent recapture in either 2009 or 2010 (Supplementary Figure S1; an approach used by Macdonald & Newman, 2002), and 95% of the inter-trap intervals are within 2-years (Dugdale et al., 2007b). In order to investigate the differences in longevity between EGO (N = 77) and WGO (N = 125) [Dataset D, complete longevity], we limited the data to individuals first caught as cubs that were assigned both parents. Because this protocol restricted the minimum possible longevity of individuals to 1 year, the number of years that survived after breeding

age was transformed, by subtracting one, in order to meet Poisson assumptions. Mean adult survival probability is affected by sex, but not by year, cohort or population size (Annavi et al., in prep), thus we included sex as a fixed categorical effect in our model. Offspring-paternity type (EGO or WGO; categorical), and its interaction with sex, were included as fixed effects. Birth year, mother ID and natal group ID were included as random effects, to control for maternal and environmental effects.

Lifetime reproductive success

We compared the lifetime reproductive success (LRS) of EGO and WGO (F_1 offspring) that were caught as cubs that survived to breeding age and were assigned both parents. We restricted the data to F_1 offspring that were assigned parentage themselves in at least one year, or that were assigned at least one F_2 grand-offspring (grand-offspring of F_0 female). The LRS of these F_1 offspring was measured as:

(i) the number of litters (i.e., number of years in which female/male F_1 offspring were assigned maternity/paternity) and,

(ii) the number of trapped and genotyped F_2 grand-offspring (i.e., offspring of the F_1 offspring that survived to independence, at 12–15 weeks of age) they were assigned.

LRS analyses were conducted on two data sub-sets. The first dataset (Dataset E, cohorts 1988–1996) comprised individuals born between 1988 and 1996, as all individuals in these cohorts were dead by 2010 (N= 83; EGO: female = 14, male = 17 and WGO: female = 20, male = 32). The second dataset (Dataset E, cohorts 1988-2008) included the complete 1988–1996 cohorts and incomplete 1997–2008 cohorts, i.e., it also included badgers, from cohorts 1997–2008, that were considered to be dead by the end of 2010 (i.e., adults that were not trapped in the subsequent 2 years, and cubs not trapped after 1 year) but excluding longer-lived badgers from these cohorts, that were not dead by 2010 (N = 173; EGO: female = 32, male = 44 and WGO: female = 44, male = 53).

In addition to the fixed and random effects that were used in longevity analysis, we also included the number of other breeding members (males and females) residing at the social-group/year, which is known to affect LRS of cubs born into that group that year as a fixed effect (Dugdale et al., 2010). These were retained in all models (even if they were not significant across the restricted

dataset). Dependent LRS measures were (-1)-transformed to meet Poisson assumptions. We analysed LRS for each sex separately.

All analyses for longevity, and reproductive success, were performed in R 2.13.1. We fitted a generalised linear mixed model (GLMM, using the *lmer* function with Laplace approximation of the maximum likelihood in the lme4 package, Bates & Sarkar, 2007) with a Poisson error structure and log-link function for the dependent variable. The significance of including each explanatory term in the model was assessed using a likelihood ratio test (chi-square distribution) by comparing the two nested models (with and without the term under consideration). All estimates are given as mean ± SE.

Results

First-year survival probability

Maternal age had no effect on their offspring's first-year survival probability (Test 1, Table 1; Dataset B, restricted maternal age). Maternal age was therefore excluded from further analyses.

Models with the effect of litter-type ([I] WGL vs. ML; [II] EGL vs. ML) had no greater explanatory power than the basic model (QAICc: basic model vs. litter-type model = [I] 1868.41 vs. 1869.93; [II] 1340.39 vs. 1341.79) after including variables (a) to (e) (Test 2, Table 1; Dataset C, restricted complete litters). Thus, WGO born in WGL, or EGO born in EGL, had a similar probability of first-year survival as WGO born in ML, or EGO born in ML, respectively.

First-year survival probability was greater for WGO than for EGO (mean survival probability: WGO = 0.74 ± 0.14, EGO = 0.55 ± 0.18; Test 3, Table 1; Dataset A, complete dataset). This difference was not apparent when the smaller coccidial dataset was analysed (Table 1; Dataset A_{coccid}). EGO and WGO had similar coccidial loads (EGO = 1.09 ± 0.21, WGO = 0.94 ± 0.24; N = 60 EGO and 62 WGO; Mann-Whitney U test: $P = 0.306$).

Estimated longevity

Badger longevity ranged from 1 to 14 years with a median of 5 years (restricted to badgers born in cohorts 1988–1996). Sex was a significant predictor of badger longevity; frequency distributions are

given in Figure 1. Females, on average, survived almost one year longer than did males (female = 5.56 ± 0.39, male = 4.61 ± 0.30, Table 2). After controlling for sex, WGO survived around 1.3 years longer than did EGO (WGO = 6.14 ± 0.32, EGO = 4.84 ± 0.40; Table 2, Dataset D, complete longevity). There was no interactive effect between offspring-paternity type and sex on longevity.

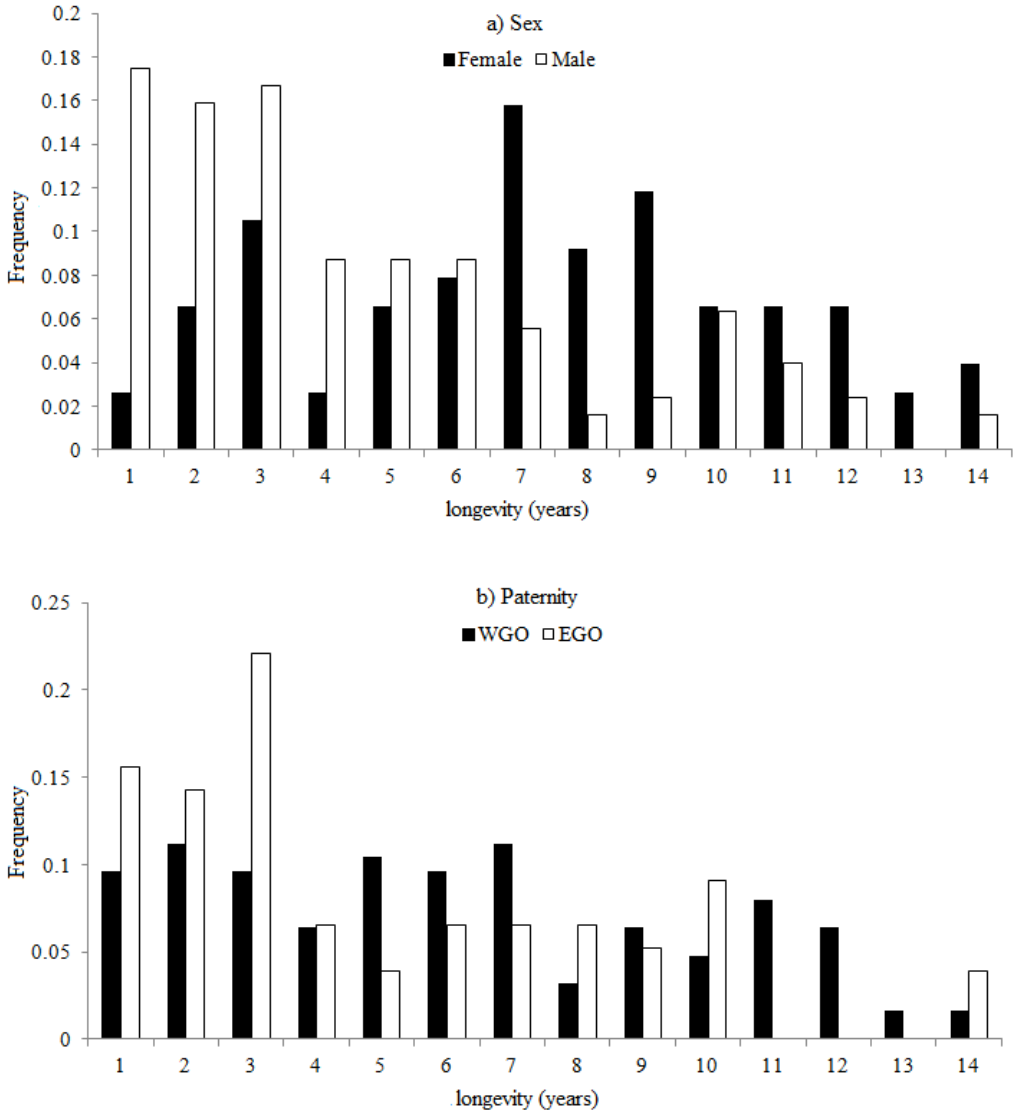


Figure 1 Frequency distributions of longevity in relation to (a) sex (Female = 76, Male = 126) and (b) offspring-paternity type (WGO = 125, EGO = 77) [Dataset D, complete longevity].

Table 1 Tests of the effects of maternal age (Test 1; Dataset B), litter type (Test 2; Dataset C) and offspring-paternity type (Test 3; Dataset A) on first-year survival probability with the inclusion of (a) the standardised total summer rainfall (May–October), (b) standardised mean winter temperature (November–February), (c) paternal multilocus heterozygosity (d) the interaction between paternal multilocus heterozygosity and standardised total summer rainfall and (e) coccidial load. Effect sizes were obtained from the full model.

	Effects	**Without coccidial load**			**With coccidial load**		
		β (SE)	**$\chi^2_{[df = 1]}$**	***P***	**β (SE)**	**$\chi^2_{[df = 1]}$**	***P***
Test 1	Maternal age	0.01 (0.08)	0.02	0.903	0.13 (0.30)	0.20	0.659
	(a) Summer rainfall	-0.47 (1.22)			-2.90 (1.88)		
	(b) Winter temperature	0.10 (0.20)			10.68 (7.63)		
	(c) Paternal heterozygosity	1.40 (1.01)			-8.40 (9.50)		
	(d) Interaction: Paternal heterozygosity*summer rainfall	0.82 (1.14)			-3.69 (6.21)		
	(e) Coccidial load	-			-0.27 (0.68)		
Test 2	Litter type [I]	1.07 (2.07)	0.51	0.475	18.05 (6920.66)	-0.43	0.511
	(a) Summer rainfall	-1.40 (1.10)			0.45 (5.10)		
	(b) Winter temperature	0.17 (0.23)			-0.47 (0.54)		
	(c) Paternal heterozygosity	0.81 (1.14)			-4.19 (4.21)		
	(d) Interaction: Paternal heterozygosity*summer rainfall	1.78 (1.03)			0.44 (4.16)		
	(e) Coccidial load	-			0.44 (0.65)		
	Litter type [II]	0.75 (1.05)	0.69	0.407	-6.02 (5.39)	3.22	0.073
	(a) Summer rainfall	-0.52 (1.29)			5.32 (6.77)		
	(b) Winter temperature	0.36 (0.21)			-2.36 (1.51)		
	(c) Paternal heterozygosity	0.69 (1.07)			-2.72 (7.72)		
	(d) Interaction: Paternal heterozygosity*summer rainfall	0.61 (1.23)			0.41 (5.80)		

	(e) Coccidial load	-			0.06 (0.29)		
Test 3	Offspring-paternity type (WGO vs. EGO)	**-0.86 (0.25)**	**13.08**	**< *0.001***	-0.76 (0.76)	1.05	0.306
	(a) Summer rainfall	0.33 (0.14)			-2.19 (0.94)		
	(b) Winter temperature	-1.11 (0.75)			6.67 (5.25)		
	(c) Paternal heterozygosity	0.75 (0.69)			-6.70 (6.24)		
	(d) Interaction: Paternal heterozygosity*summer rainfall	1.29 (0.71)			-1.31 (4.58)		
	(e) Coccidial load	-			-0.03 (0.28)		

Bold values are significant at $P < 0.05$

Table 3 Generalised Linear Mixed Model Results for the offspring-paternity type effect on LRS measured as (i) number of litters and (II) number of F_2 offspring using individuals from Dataset E-cohorts 1988–1996 and Dataset F-cohorts 1988–2008: Female Only

LRS	Effect	Cohorts 1988–1996 EGO = 14 and WGO = 20				Cohort 1988–2008 EGO = 32 and WGO = 44			
		Random effects	Fixed effects			Random effects	Fixed effects		
		Variance ± SD	β ± SE	$\chi^2_{[df=1]}$	*P*	Variance ± SD	β ± SE	$\chi^2_{[df=1]}$	*P*
Number of litters	Year	0 ± 00				0 ± 00			
	Mother ID	0 ± 00				0 ± 00			
	Natal group ID	0 ± 00				0 ± 00			
	Intercept		0.88 ± 0.43				0.03 ± 0.03		
	Number of other members		-0.06 ± 0.04	2.56	0.110		-0.01 ± 0.02	0.08	0.773
	Offspring-paternity type (EGO)		**-1.08 ± 0.44**	**6.32**	**0.012**		-0.25 ± 0.29	0.81	0.367
Number of F_2 offspring	Year	0 ± 00				0 ± 00			
	Mother ID	0 ± 00				0 ± 00			
	Natal group ID	0 ± 00				0 ± 00			
	Intercept		1.47 ± 0.40				0.70 ± 0.28		
	Number of other members		**-0.07 ± 0.04**	**4.26**	**0.039**		-0.02 ± 0.02	1.08	0.300
	Offspring-paternity type (EGO)		**-1.02 ± 0.37**	**8.09**	**0.004**		-0.35 ± 0.26	1.88	0.170

Significant effects at $P < 0.05$ are in bold. For fixed effect slopes, WPO = 0 and EPO = 1

Table 4 Generalised Linear Mixed Model Results for the offspring-paternity type effect on LRS measured as (i) number of litters and (II) number of F_2 offspring using individuals from Cohorts 1988–1996 (Dataset E) and Cohorts 1988–2008 (Dataset F): Male Only

LRS	Effect	Cohorts 1988–1996 EGO = 17 and WGO = 32				Cohort 1988–2008 EGO = 44 and WGO = 53			
		Random effects	Fixed effects			Random effects	Fixed effects		
		Variance ± SD	β ± SE	$\chi^2_{[df=1]}$	*P*	Variance ± SD	β ± SE	$\chi^2_{[df=1]}$	*P*
Number of litters	Year	0.32 ± 0.57				0.11 ± 0.33			
	Mother ID	0.46 ± 0.68				0.57 ± 0.75			
	Natal group ID	0.35 ± 0.59				0.05 ± 0.23			
	Intercept		-0.86 ± 0.60				-0.79 ± 0.37		
	Number of other members		-0.05 ± 0.05	0.90	0.342		-0.02 ± 0.03	0.70	0.404
	Offspring –paternity type (EGO)		1.00 ± 0.50	3.26	0.071		0.46 ± 0.33	1.97	0.161
Number of F_2 offspring	Year	0.61 ± 0.78				0.19 ± 0.43			
	Mother ID	0.11 ± 0.34				0.70 ± 0.84			
	Natal group ID	0.27 ± 0.52				0.07 ± 0.27			
	Intercept		0.09 ± 0.46				-0.03 ± 0.31		
	Number of other members		-0.04 ± 0.03	1.62	0.203		-0.01 ± 0.02	0.47	0.491
	Offspring-paternity type (EGO)		**1.22 ± 0.52**	**11.16**	**< 0.001**		0.49 ± 0.26	3.30	0.069

Significant effects at $P < 0.05$ are in bold. For fixed effect slopes, WPO = 0 and EPO = 1

Table 2 Results from a generalised linear mixed model of longevity in relation to sex, offspring-paternity type and their interaction (N = 202; dataset D, complete longevity).

Fixed effects	**β (SE)**	$\chi^2_{[df = 1]}$	***P***
Sex (male)	**-0.47 (0.07)**	**50.41**	**< 0.001**
Offspring-paternity type (EGO)	**-0.19 (0.08)**	**5.10**	**0.024**
Sex × Offspring-paternity type	-0.01 (0.14)	0.01	0.932

Bold values are significant at $P < 0.05$. Fixed effects: sex (female = 0 and male = 1); offspring-paternity type (WGO = 0, EGO = 1). The random effects for birth year, natal group ID and mother ID were: 0.01 ± 0.09, 0.08 ± 0.28 and 0.03 ± 0.17, respectively.

Lifetime reproductive success

The proportion of F_1 offspring that went on to breed successfully themselves (i.e., with parentage assigned to at least one genotyped F_2 grand-offspring) was independent of whether these were EGO or WGO (females: 45.2% of EGO, 45.1% of WGO, $\chi^2_{[df=1]} = 0.02$, $P = 0.89$; males: 34.5% of EGO, 39.0% of WGO, $\chi^2_{[1]} = 0.44$, $P = 0.51$; Dataset E, cohorts 1988–1996). This was not affected by increasing the sample size (females: 40.5% of ego, 42.7% of wgo, $\chi^2_{[df=1]} = 0.02$, $p = 0.88$; males: 35.2% of ego, 39.3% of wgo, $\chi^2_{[1]} = 0.30$, $p = 0.58$; dataset f, cohorts 1988–2008).

Offspring-paternity type had sex dependent effects on LRS when analysed with a smaller, more complete dataset (Dataset E, cohorts 1988–1996; Table 3–4). When considering a large dataset that included individuals with incomplete life histories (Dataset F, cohorts 1988–2008), there was no significant relationship between LRS and offspring-paternity type, although the direction of the relationship was in the same direction (Table 3–4). It may be the larger, but incomplete, dataset includes more short-lived individuals that do not show sex effects; consequently this benefit differential effect is only seen among longer-lived individuals.

For females, lifetime number of litters ranged from 1 to 4, with a mean of 2.06 ± 0.18 and for males this ranged from 1 to 6, with a mean of 1.71 ± 0.17 (Dataset E, cohorts 1988-1996, Supplementary Figure S2; for Dataset F, cohorts 1988-2008, female: range = 1-6, mean = 2.05 ± 0.14, male: range = 1-6, mean = 1.69 ± 0.11). F_1 WGO females were assigned significantly more litters over

their lifetime than did F_1 EGO (Dataset E, cohorts 1988-1996; Figure 2, Table 3). The greater lifetime iteroparity of F_1 WGO than EGO females was not the result of greater longevity, as these metrics did not correlate (Dataset E, cohorts 1988-1996, $r_s = 0.28$, $P = 0.107$; sexes combined). Offspring type did not, however, predict the lifetime number of litters for males (i.e., number of years they bred; see Table 4).

The lifetime number of F_2 offspring produced by females ranged from 1 to 8, with a mean of 2.85 ± 0.32, and for males from 1 to 13, with a mean of 2.73 ± 0.33 (Dataset E, cohorts 1988-1996, Supplementary Figure S3; for Dataset F, cohorts 1988-2008, female: range = 1–9, mean = 2.82 ± 0.23, male: range = 1–15, mean = 2.88 ± 0.25). The number of F_2 offspring produced by F_1 offspring correlated positively with the longevity (Dataset E, cohorts 1988-1996, $r_s = 0.35$, $P < 0.001$; sexes combined). F_1 WGO females were assigned significantly more F_2 offspring over their lifetime than did F_1 EGO (Figure 3; Table 3). F_1 EGO males were assigned more F_2 offspring than were F_1 male WGO (Figure 3; Table 4).

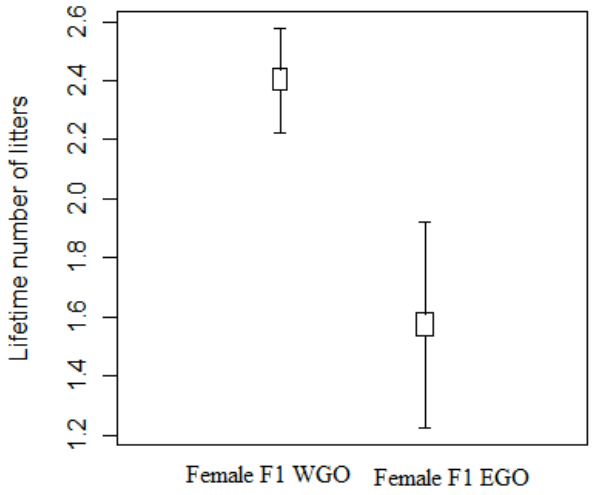


Figure 2 The lifetime number of litters (uncorrected means and SE) for female WGO (N = 20) and EGO (N = 14). (Dataset E).

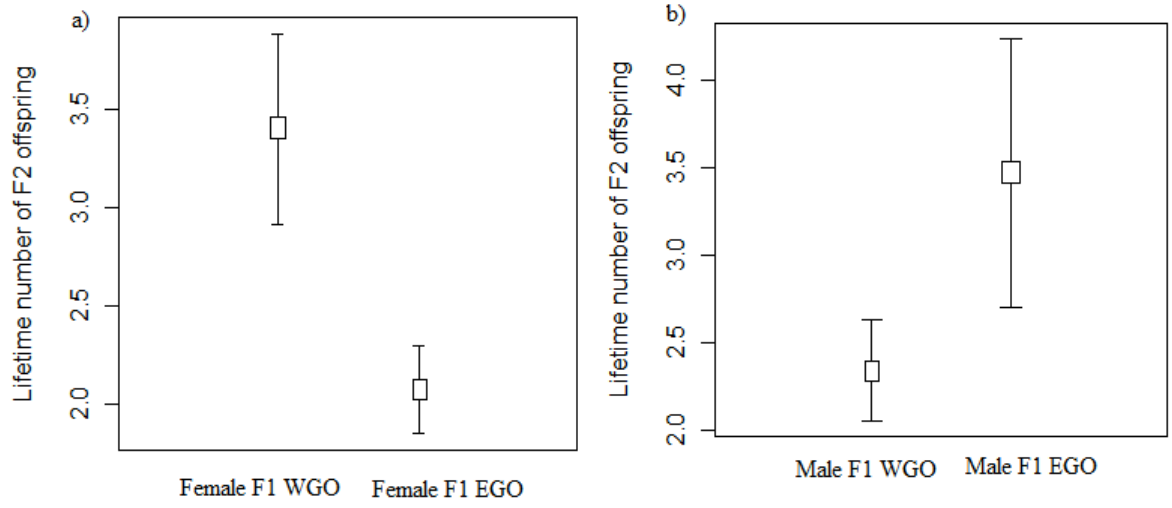


Figure 3 The lifetime number of F_2 offspring (uncorrected means and SE) (a) produced by F_1 WGO females (N = 20) and F_1 EGO females (N = 14) and (b) sired by F_1 WGO males (N = 32) and F_1 EGO males (N = 17). (Dataset E)

Considering the percentage of each type of F_2 grand-offspring (whether F_2 EGO or F_2 WGO) sired by males, F_1 EGO sired 55.3 ± 1.03% of F_2 EGO and 44.7 ± 1.03% of F_2 WGO (1.96xSE, the confidence intervals did not overlap), whereas F_1 WGO sired 50.6 ± 6.45% of F_2 EGO and 49.4 ± 6.45% of F_2 WGO (Figure 4). This indicates that male EGOs may be able to increase their reproductive success through siring extra-group offspring, though these percentages of difference are marginal, given the limitations in our ability to sample the population.

For females, both F_1 EGO and F_1 WGO produced very slightly more F_2 EGO than F_2 WGO (F_1 EGO: 52.6 ± 7.28% F_2 EGO and 47.4 ± 7. 28 F_2 WGO; F_1 WGO: 51.8 ± 8.21% F_2 EGO and 48.2 ± 8.21% F_2 WGO) but their confidence intervals overlapped zero

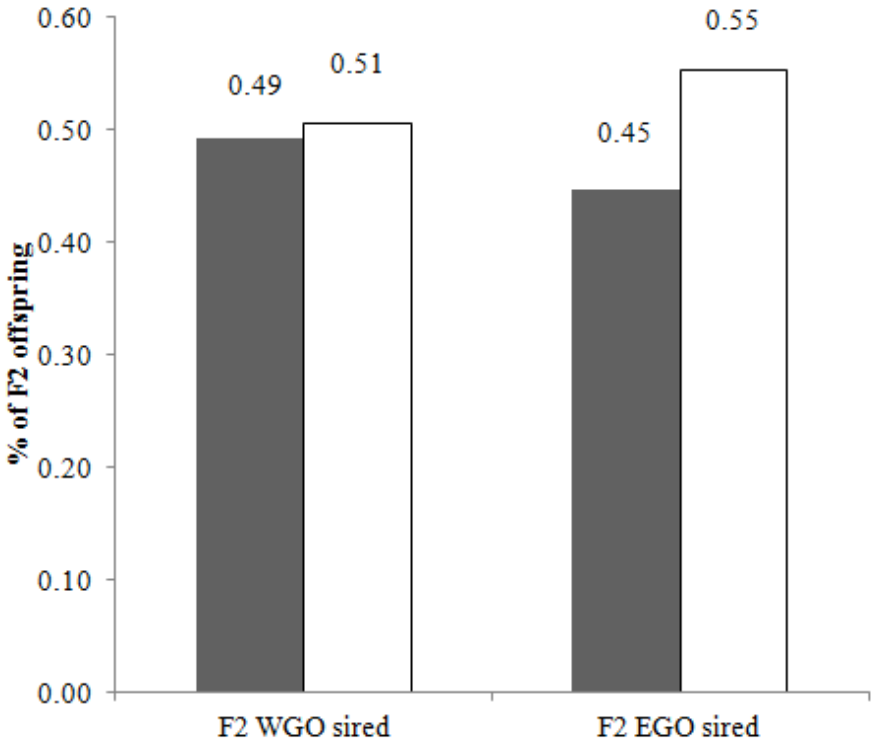


Figure 4 Male extra-group and within-group siring successes. Black bars and white bars represent adult male F_1 WGO and F_1 EGO respectively.

Discussion

EGP is often associated with offspring fitness advantages and therefore in group-living species should be favoured under circumstances where this strategy produces offspring that exhibit higher fitness, in terms of survival or reproductive success than within-group offspring (e.g., Griffith et al., 2002; Gerlach et al., 2011; Sardell et al., 2011; but see Sardell et al., 2012). Here, there were mixed results: while there was no survival advantage for EGO, EGO exhibited ca. 34% lower survival probability through their first year than did WGO; these EGO on average lived 1.3 years shorter than EGO. There was some evidence for a sex-specific reproductive advantage of EGP in badgers. Female EGO exhibited lower lifetime reproductive success than WGO; however, male EGO exhibited higher lifetime reproductive success than male WGO.

To date, many studies have documented the importance of genetic quality on mate choice (Aparicio et al., 2001; Foerster et al., 2003; Masters et al., 2003; Seddon et al., 2004; Suter et al., 2007; Garcia-Navas et al., 2009); yet, few have exposed the type of mating we see here, where high rates of EGP occur, but with little evidence in support of offspring fitness benefits that would drive EGP (e.g., Sardell et al., 2011 & 2012).

Survival advantage

For EGO neither first-year survival probability nor longevity was higher than for WGO. First-year mortality is known to occur in badger cubs largely as a consequence of parasitic infections (Newman et al., 2001) and due to stressful weather conditions e.g., low summer rainfall and temperature (i.e., thus low food availability; Macdonald et al., 2010, Table 1). In Annavi et al., in prep established that the advantages of paternal heterozygosity in badgers are apparent only under benign climatic conditions, whereas stressful weather patterns limit access to food (see also Macdonald et al., 2010) and cause indiscriminate mortality. We speculate that these effects could be so strong as to operate independently and indiscriminately of offspring-paternity type, where offspring-paternity type may have a discriminate influence on more subtle parameters than live/die, e.g., health / fitness parameters. A meta-analysis of 121 studies of birds similarly showed no difference in survival to breeding age between extra-pair offspring and within-pair offspring (Akcay & Roughgarden, 2007). The few studies in birds that compared the relative longevity of offspring in relation to paternal type also found no differences (Sardell et al., 2011; Schmoll et al., 2011). These results combined with our current findings suggest that offspring-paternal type does not correlate with both pre-reproductive survival and/or longevity.

Sex-specific benefits to EGP mating strategies

Male F_1 EGO badgers had a higher lifetime reproductive success than male F_1 WGO (Figure 3). In contrast, female F_1 EGO had a lower LRS than female F_1 WGO. Sex-specific differences in LRS with respect to offspring-paternity type in badgers, support that females may gain an indirect fitness benefit through EGP derived through the greater reproductive success of EGO sons. This study presents the

first evidence of sex-specific EGP benefit shown in a mammal species, where this benefit has been reported for birds (e.g., Gerlach et al., 2011).

Forstmeier et al. (2011) point out that within-sex genetic constraint may drive selection for the same alleles in females, as in males, to promote EGP, where female resistance to extra-group males would also confer resistance to within-group males, leading to reduced reproductive success. Therefore, EGP, which can potentially increase the reproductive output of a male per year (i.e., females mated with), may simultaneously result in females pursuing (or being amenable to) a mating strategy that is disadvantageous to their female EGO. Our data do support this level of EGP benefit: male F1 EGO did sire more F2 offspring than did male WGO, in comparison; female F1 WGO produced more F2 offspring than did female F1 EGO (Figure 3).

Sex-specific, differential fitness-related attributes of EGO versus WGO could potentially reflect a sexually antagonistic effect, such that genes (e.g., good genes or compatible genes) obtained through EGP may be beneficial (i.e., increased reproductive success) when expressed in males, but detrimental when expressed in females. The opposing LRS effects for EGO and WGO shown in males and females in this population, suggest that sexually antagonistic selection may potentially play an important role in the maintenance of EGP in badgers, if the fitness benefits of producing extra-group sons outweigh the costs of producing extra-group daughters (Kokko, 2001).

F_1 EGO male badgers exhibited a greater reproductive success by siring F_2 EGO than by siring F2 WGO (Figure 4). The tendency for F_1 EGO male to then sire F2 EGO has several possible explanations that are not mutually exclusive. It has been reported previously that a female's propensity to produce extra-group offspring can be heritable, i.e., the Baldwin effect (Baldwin & Mark, 1896; e.g., Reid et al., 2011). Alternatively, male F_1 EGO badger could simply be more attractive as mating partners, i.e., have phenotypic attributes that females can detect as cues favouring a greater production of descendents (i.e., 'Sexy-son Hypothesis', Fisher 1930) or produce more competitive sperm than F1 WGO. In that extra-group matings require females to be receptive and complicit, females may increase their indirect reproductive success by choosing sexually attractive mates, if her sons can inherit their father's attractiveness. Moreover, between-sex genetic constraints (Forstmeier et al., 2011) would lead females that inherited alleles coding for extra-group mating from their fathers to pass these to their offspring, where there would plausibly be a grandson advantage (a 'Promiscuous-

grandson Hypothesis', similar to the 'Sexy-son Hypothesis' in Fisher's runaway sexual selection; Fisher, 1930; Weatherhead and Robertson, 1979).

Do WGO perform better than EGO?

While we find little evidence to suggest that EGP is adaptive, female badgers in this study population derive greater lifetime reproductive success if they produce more WGO than EGO, especially female WGO, because F_1 WGO had better survival (first-year survival probability and longevity) than F_1 EGO; also F_1 WGO females produced a greater number of litters, and hence offspring, than F_1 EGO females.

Understanding mate choice, where individuals appear to express strong preferences despite detecting no/little direct benefits from mate choice, remains a challenge to behavioural ecology (Ryder et al., 2010). Indeed, studies such as Forstmeier et al., (2011) have rationalised a hypothetical framework through which within- and between- sex genetic correlations can drive female promiscuity. This can arise due to plieotropic effects (Forstmeier et al., 2011), despite antagonistic selection whereby males may be selected to be promiscuous whereas females may not be selected for this trait directly, if promiscuity does not increase their reproductive output. Our previous research (Annavi et al., in prep) has found no clear advantages to EGP in terms of mate selection, aside from weak evidence of mating to avoid pair-wise genetic similarity. Against this background, support for a genetic basis to the high rates of EGP we observed in this badger population could come from benefits in terms of offspring fitness. There is a paradox, that while EGP rates of 48% are evident (Annavi et al., in prep; see also Dugdale et al., 2007b), F_1 WGO performed better in survival and reproduction than did F_1 EGO, with the exception that F_1 EGO male outperformed F_1 WGO male in the number of F_2 grand-offspring sired.

This leads to 4, not mutually exclusive, possibilities, where, EGP matings would serve a different or additional function in badger society;

i) 'Superior natal territory fit Hypothesis'

There are circumstances where within-group matings would be advantageous: WGO may carry genes that make them a good, or better, fit in their natal territory, with a better chance of breeding successfully (female) and surviving to an older age compared to EGO. WGO may also be more likely to inherit territory / group status from their mothers (Lindstrom, 1986), giving them an advantage over EGO in terms of accessing resources for survival and reproduction. Rather than facilitating genetic quality, EGP mating may serve to alleviate conflict through inter-social group appeasement and inter-group genetic relationships (see Macdonald et al., 2004a; Delahay et al., 2006).

ii) 'Uncontrolled / unconstrained mating Hypothesis'

Isvaran & Clutton-Brock, (2007) report that rates of EGP decline with increasing length of the mating season, which suggests that it is increasingly difficult for males to monopolise their social mates during short mating seasons and when there is high overlap of female oestrus periods. They also found that males are less able to monopolise individual females when there is a higher number of breeding females in a group. Interestingly, they found that the rate of extra-group fertilisations and the payoffs involved did not vary consistently with social mating system.

EGP may be accidental, as the unavoidable consequence in badgers of WG-males being unable to monopolise group females through their oestrus period, in a system where inter-group transgression occurs frequently (Macdonald et al., 2008) and foraging grounds may be common between groups (Macdonald et al., 2004a). Female cub numbers are generally equivalent to male cub numbers at the population level in our study population (Dugdale et al., 2003). These females sometimes require repeated intromission in order to ovulate (Ahnlund, 1980), but they do also experience continuing oestrous cycles (Service et al., 2002), that continue during embryonic diapause, i.e., female badgers are capable of superfoetation (conception during pregnancy) (Yamaguchi et al., 2006).

iii) 'EGO sociological disadvantage Hypothesis'

A further possibility is that, even if EGO were to exhibit genetic quality advantages, these may be countermanded by the sociological disadvantages of having an extra-group father. Infanticide is known to occur in badger society (Lups & Roper 1990; Cresswell et al., 1992; see also Wolf &

Macdonald, 2004). Polygynandrous mating is believed to have evolved in high-density badger populations in an attempt to confuse paternity (Carpenter et al., 2005; Dugdale et al., 2007b; Dugdale et al., 2011a). If males are unable to identify their own young, females may mate with males to confuse paternity while applying cryptic female choice to select against certain males after mating. Alternatively, should males be able to recognise that offspring are not their own, or either sex recognise that offspring were not fathered from within their group (see Packer & Pusey, 1984; Lion & van Baalen, 2007), for example through MHC / scent (Sin et al., 2012; Buesching et al., 2003), they may differentially kill EGO over WGO (see Dunn et al., 2012). EGO exhibited 11% poorer first-year survival than did WGO and on average live 1.3 year less; however there is no evidence that group members can detect that they are EGO.

iv) 'Inbreeding avoidance Hypothesis'

EGP may reduce inbreeding. When the average within-group candidate parents' relatedness was greater, the number of EGO (absolute) and the proportion of EGO (relative) were higher (Annavi et al., in prep). To show that EGP reduces inbreeding requires randomisations evidencing that a reduction in f occurs, rather than that neighbouring individuals are less related than same-group individuals. In high-density populations, as per this study, permanent dispersals are infrequent (Macdonald et al., 2008) and thus EGP (deliberate or accidental) may function to facilitate gene flow between groups (Pope et al., 2006).

Conclusions

The advantages and ecological role of promiscuity depend upon perspective in badgers. It may not inform mating choices based on heterozygosity, i.e., genetic quality (Annavi et al., in prep), nor does it confer advantages to female extra-group offspring, although it does to male EGO. This leads us to posit that rather than being under positive selection, this mating strategy is under opposing selection (Foerster et al., 2007), depending on the sex of the resulting offspring. EGP may also occur as an inevitable consequence of the badger's social system at high density, whereas this species

predominantly exhibits much smaller group sizes, or mating pairs over the majority of its European continental range (see Newman et al. 2011).

Investigating the evolutionary biology of extra-group mating strategies thus requires a comprehensive theoretical framework, more extensive than simply testing for heterozygosity-fitness advantages, where genetic advantages must be weighed against the implied concept of freedom of choice and sociological corollaries of producing young with expatric fathers.

References

Ahnlund, H. 1980. Sexual maturity and breeding season of the badger, *Meles meles* in Sweden. *Journal of Zoology (London)*, **190,** 77–95.

Akcay, E. & Roughgarden, J. 2007. Extra-pair paternity in birds: review of the genetic benefits. *Evolutionary Ecology Research*, **9,** 855–868.

Annavi, G., Dawson, D., Horsburgh, G., Greig, C., Dugdale, H., Newman, C., Macdonald, D. & Burke, T. 2011. Characterisation of twenty-one European badger (Meles meles) microsatellite loci facilitates the discrimination of second-order relatives *Conservation Genetic Resources,* **3,** 515–518.

Aparicio, J., Cordero, P. & Veiga, J. 2001. A test of the hypothesis of mate choice based on heterozygosity in the spotless starling. *Animal Behaviour*, **62,** 1001–1006.

Baldwin, J. & Mark, A. 1896. New factor in evolution. *The American Naturalist,* **30,** 441–451.

Baker, P., Fink, S., Bruford, M. & Harris, S. 2004. Polygynandry in a red fox population: implications for the evolution of group living in canids? *Behavioral Ecology*, **15,** 766–778.

Bateman, A.J. 1948. Intra-sexual selection in Drosophila. *Heredity*, **2,** 349–821.

Bates, D. & Sarkar, D. 2007. lme4: Linear Mixed-Effects Models Using S4 Classes. R package version 0.9975-12, URL http://CRAN.R-project.org/.

Brommer, J., Kirkpatrick, M., Qvarnstrom, A. & Gustafsson, L. 2007. The intersexual genetic correlation for lifetime fitness in the wild and its implications for sexual selection. *PLoS ONE,* **2,** e744.

Buesching, C.B., Waterhouse, J. & Macdonald, D. 2002a. 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.

Buesching, C.B., Waterhouse, J. & Macdonald, DW. 2002b. Gas-chromatographic analyses of the subcaudal gland secretion of the European badger (*Meles meles*) part II: time-related variation in the individual-specific composition. *Journal of Chemical Ecology*, **28,** 57–69.

Buesching, C.D., Stopka, P. & Macdonald, D.W. 2003. The social function of allo-marking in the European Badger (*Meles meles*). *Behaviour*, **140,** 965–980.

Burnham, K.P. & Anderson, D.R. 2002. Model selection and multimodel inference. New York: Springer Verlag.

Canivenc R. 1966. A study of progestation in the European badger (*Meles meles* L.). In: *Symposia of the Zoological Society of London Number 15: Comparative Biology of Reproduction in Mammals* (Ed. by Rowlands, I.W.). London: Academic Press.

Carpenter, P.J., Pope, L.C., Greig, C., Dawson, D.A., Rogers, L.M., Erven, K., Wilson, G.J., Delahay, R.J., Cheeseman, C.L. & Burke, T. 2005. Mating system of the European badger, *Meles meles*, in a high density population. *Molecular Ecology*, **14,** 273–284.

Cheeseman, C.L. & Harris, S. 1982. Methods of marking badgers (*Meles meles*). *Journal of Zoology*, **197,** 289–292.

Cohas, A., Bonenfant, C., Gaillard, J.M. & Allaine, D. 2007a. 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, N. G. & Allaine, D. 2007b. Extra-pair paternity in alpine marmots, *Marmota marmota*: genetic quality and genetic diversity effects. *Behavioral Ecology & Sociobiology*, **61,** 1081–1092.

Cresswell, W.J., Harris, S., Cheeseman, C.L. & Mallinson, P.J. 1992. To breed or not to breed: An analysis of the social and density-dependent constraints on the fecundity of female badgers (*Meles meles*). *Philosophical Transactions of the Royal Society B: Biological Science*, **338,** 393–407.

da Silva, J., Macdonald, D.W. & Evans, P.G.H. 1994. Net costs of group living in a solitary forager, the Eurasian badger (*Meles meles*). *Behavioral Ecology*, **5,** 151–158.

Delahay, R.J., Carter, S.P., Forrester, G.J., Mitchell, A. & Cheeseman, C.L. 2006. Habitat correlates of group size, bodyweight and reproductive performance in a high-density Eurasian badger (*Meles meles*) population.. *Journal of Zoology*, **270,** 437–447.

Descamps, S., Boutin, S., Berteaux, D. & Gaillard, J.M. 2008. Age-specific variation in survival, reproductive success and offspring quality in red squirrels: evidence of senescence. *Oikos*, **117,** 1406–1416.

Double, M. & Cockburn, A. 2000. Pre-dawn infertility: females control extra-pair mating in superb fairy-wrens. *Proceedings of the Royal Society B: Biological Sciences,* **267,** 465–470.

Dugdale, H.L., Griffiths, A. & Macdonald, D.W. 2011a. Polygynandrous and repeated mounting behaviour in European badgers, *Meles meles*. *Animal Behaviour*, **82,** 1287–1297.

Dugdale, H.L., Macdonald, D.W. & Newman, C. 2003. Offspring sex ratio variation in the European Badger, *Meles meles*. *Ecology*, **84,** 40–45.

Dugdale, H.L., Macdonald, D.W., Pope, L.C. & Burke, T. 2007b. Polygynandry, extra-group paternity and multiple-paternity litters in European badger (*Meles meles*) social groups. *Molecular Ecology*, **16,** 5294–5306.

Dugdale, H.L., Macdonald, D.W., Pope, L.C., Johnson, P.J. & Burke, T. 2008. Reproductive skew and relatedness in social groups of European badgers, *Meles meles*. *Molecular Ecology*, **17,** 1815–1827.

Dugdale, H.L., Pope, L.C., Newman, C., Macdonald, D.W. & Burke, T. 2011b. Age-specific breeding success in a wild mammalian population: selection, constraint, restraint, and senescence. *Molecular Ecology*, **20,** 3261–3274.

Dunn, P. O., Bollmer, J., Feeeman-Gallant, R. & Whittingham, L. A. 2012. MHC variation is related to a sexually selected ornament, survival, and parasite resistance in common yellow throats. *DOI: 10.1111/j.1558-5646.2012.01799.x*.

Fisher, R.A. 1930. The genetical theory of natural selection. Oxford: Oxford University Press.

Foerster, K., Coulson, T., Sheldon, B. C., Pemberton, J. M., Clutton-Brock, T. H. & Kruuk, L. E. B. 2007. Sexually antagonistic genetic variation for fitness in red deer. *Nature,* **447,** 1107–1110.

Foerster, K., Delhey, K., Johnsen, A., Lifjeld, J.T. & Kempeneaers, B. 2003. Females increase offspring heterozygosity and fitness through extra-pair matings. *Nature*, **425,** 714–717.

Forstmeier, W., Martin, K., Bolund, E., Schielzeth, H. & Kempenaers, B. 2011. Female extrapair mating behavior can evolve via indirect selection on males. *Proceedings of the National Academy of Sciences*, **108,** 10608–10613.

Garcia-Navas, V., Ortego, J. & Sanz, J. 2009. Heterozygosity-based assortative mating in blue tits (*Cyanistes caeruleus*): implications for the evolution of mate choice. *Proceedings of the Royal Society B: Biological Sciences*, **276**, 2931–2940.

Gerlach, N.M., McGlothlin, J.W., Parker, P.G. & Ketterson, E.D. 2012. Promiscuous mating produces offspring with higher lifetime fitness. *Proceedings of the Royal Society B: Biological Sciences*, **279,** 860–866.

Goossens, B., Graziani, L., Waits, L.P., Farand, E., Magnolon, S., Coulon, J., Bel, M., Taberlet, P. & Allaine, D.1998. Extra-pair paternity in the monogamous alpine marmot revealed by nuclear DNA microsatellite analysis. *Behavioral Ecology & Sociobiology*, **43,** 281–288.

Griffith, S.C., Owens, I.P.F. & Thuman, K.A. 2002. Extra pair paternity in birds: a review of interspecific variation and adaptive function. *Molecular Ecology*, **11,** 2195–2212.

Hadfield, J.D., Richardson, D.S. & Burke, T. 2006. Towards unbiased parentage assignment: combining genetic, behavioural and spatial data in a Bayesian framework. *Molecular Ecology*, **15,** 3715–3730.

Isvaran, K. & Clutton-Brock, T.H. 2007. Ecological correlates of extra-group paternity in mammals. *Proceedings of the Royal Society of London B: Biological Science*, **274,** 219–224.

Kempenaers, B. 2007. Mate choice and genetic quality: A review of the heterozygosity theory. *Advances in the Study of Behavior*, **37,** 189–278.

Kilshaw, K., Newman, C., Buesching, C.D, Bunyan, J. & Macdonald D.W. 2009. Coordinated latrine use by European badgers, *Meles meles*: Potential consequences for territory defense. *Journal of Mammalogy*, **90,** 1188–1198.

Kokko, H. 2001. Fisherian and 'good genes' benefits of mate choice: how (not) to distinguish between them. *Ecology. Letters,* **4,** 322–326.

Krebs, C.J. 1966. Demographic changes in fluctuating populations of *Microtus californicus*. *Ecological Monographs*, **36,** 239–273.

Lion, S. & van Ballen, M. 2007. From infanticide to parental care: Why spatial structure can help adults be good parents. *The American Naturalist*, **170,** E26-E46.

Lups, P. & Roper, T. 1990. Cannibalism in a female badger (*Meles meles*): infanticide or predation? *Journal of Zoology*, **221,** 314–315.

Macdonald, D.W., Harmsen, B.J., Johnson, P.J. & Newman, C. 2004a. Increasing frequency of bite wounds with increasing population density in Eurasian badgers, *Meles meles*. *Animal Behaviour*, **67,** 745–751.

Macdonald, D.W. & Newman, C. 2002. Population dynamics of badgers (*Meles meles*) in Oxfordshire, U.K.: numbers, density and cohort life histories, and a possible role of climate change in population growth. *Journal of Zoology (London)*, **256,** 121–138.

Macdonald, D.W., Newman, C., Buesching, C.D. & Johnson, P.J. 2008. Male-biased movement in a high-density population of the European badger (*Meles meles)*. *Journal of Mammalogy*, **89,** 1077–1086.

Macdonald, D.W., Newman, C., Buesching, C.D. & Nouvellet, P. 2010. Are badgers 'Under The Weather'? Direct and indirect impacts of climate variation on European badger (*Meles meles*) population dynamics. *Global Change Biology*, **16,** 2913–2922.

Macdonald, D.W., Newman, C., Nouvellet, P.M. & Buesching, C.D. 2009. An Analysis of Eurasian badger (*Meles meles*) population dynamics: Implication for regulatory mechanisms. *Journal of Mammalogy*, **90,** 1392–1403.

Masters, B., Hicks, B., Johnson, S. & Erb, L. 2003. Genotype and extra-pair paternity in the house wren: a rare-male effect? *Proceedings of the Royal Society B: Biological Sciences*, **270,** 1393-1397.

Mays,H. & Hill, G. 2004. Choosing mates: good genes versus genes that are a good fit. *Trends in Ecology & Evolution*, **19,** 554–559.

McLaren, G.W., Thornton, P.D., Newman, C., Buesching, C.D, Baker, S.E., Mathews, F. & Macdonald, D.W. 2005. The use and assessment of ketamine–medetomidine– butorphanol combinations for field anaesthesia in wild European badgers (*Meles meles*). *Veterinary Anaesthesia & Analgesia*, **32,** 367–372.

Mech, L.D. & McRoberts. 1990. Survival of white-tailed deer fawns in relation to maternal age. *Journal of Mammalogy*, **71**, 465–467.

Neff, B. & Pitcher, T. 2005.Genetic quality and sexual selection: an integrated framework for good genes and compatible genes. *Molecular Ecology*, **14,** 19–38.

Newman, C., Macdonald, D.W. & Anwar, A. 2001. Coccidiosis in the European badger, *Meles meles* in Wytham Woods: infection and consequences for growth and survival. *Parasitology*, **123,** 133–142.

Newman, C., Zhou, Y.B., Buesching, C.D., Kaneko, Y. & Macdonald, D.W. 2011. Contrasting sociality in two widespread, generalist, Mustelid genera, *Meles* and *Martes*. *Mammal Study*, **36,** 169–188.

Nouvellet, P.M., Buesching, C.D., Dugdale, H.L., Newman, C. & Macdonald, D.W. 2010. Mouthing off about developmental stress: Individuality of palate marking in the European badger and its relationship with juvenile parasitoses. *Journal of Zoology*, **743,** 1–11.

Packer, C. & Pusey, A. 1984. Infanticide in carnivores. *Comparative & Evolutionary Perspectives*, **37,** 6–10.

Pope, L.C., Domingo-Roura, X., Erven, K. & Burke, T. 2006. Isolation by distance and gene flow in the Eurasian badger (*Meles meles*) at both a local and broad scale. *Molecular Ecology*, **15,** 371–386.

R Development Core Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Reid, J.M., Arcese, P., Sardell, R.J. & Keller, L.F. 2011. Heritability of female extra-pair paternity rate in song sparrows (*Melospiza melodia*). *Proceedings of the Royal Society B: Biological Sciences*, **278,** 1114–1120.

Roberts, S., Hale, M. & Petrie, M. 2006. Correlations between heterozygosity and measures of genetic similarity: implications for understanding mate choice. *Journal of Evolutionary Biology*, **19,** 558–569.

Ryder, T., Tori, W., Blake, J., Loiselle, B., Parker, P. 2010. Mate choice for genetic quality: a test of the heterozygosity and compatibility hypotheses in a lek-breeding bird. *Behavioral Ecology*, **21,** 203–210.

Sardell, R.J., Arcese, P., Keller, L.F. & Reid, J.M. 2011. Sex-specific differential survival of extra-pair and within-pair offspring in song sparrows, *Melospiza melodia*. *Proceedings of the Royal Society B: Biological Sciences*, **278,** 3251–3259.

Sardell, R.J., Arcese, P., Keller, L.F. & Reid, J.M. 2012. Are there indirect fitness benefits of female extrapair reproduction? lifetime reproductive success of within-pair and extra-pair offspring. *The American Naturalist*, **179,** 779–793.

Schmoll, T., Schurr, F.M., Winkel, W., Epplen, J.T. & Lubjuhn, T. 2011. Lifespan, lifetime reproductive performance and paternity loss of within-pair and extra-pair offspring in the coal tit *Periparus ater*. *Proceedings of the Royal Society B: Biological Sciences*, **276,** 337–345.

Seddon, N., Amos, W., Mulder, R. & Tobias, J. 2004. Male heterozygosity predicts territory size, song structure and reproductive success in a cooperatively breeding bird. *Proceedings of the Royal Society B: Biological Sciences,* **271,** 1823–1829.

Service, K., Harlow, C. & Harris, S. 2002. Occurrence of monthly oestrus periods in badgers (*Meles meles*): a mechanism for observed variation in breeding patterns? . *Advances in Ethology,* **37,** 149.

Sillero-Zubiri, C., Gottelli, D., Macdonald, D.W. 1996. Male philopatry, extra-pack copulations and inbreeding avoidance in Ethiopian wolves (*Canis simensis*). *Behavioral Ecology & Sociobiology*, **38,** 331–340.

Sin, Y.W., Dugdale, H.L., Newman, C., Macdonald, D.W. & Burke, T. Evolution of MHC class I genes in the European badger (*Meles meles*). Ecology and Evolution, **2,** 1644–1662

Soulsbury, C.D. 2010. Genetic patterns of paternity and testes size in mammals. *PLoS ONE*, **5,** e9581.

Suter, S., Keiser, M., Feignoux, R. & Meyer, D. 2007. Reed bunting females increase fitness through extra-pair mating with genetically dissimilar males. *Proceedings of the Royal Society B: Biological Sciences*, **274,** 2865–2871.

Thom, M.D., Johnson, D.D.P. & Macdonald, D.W. 2004. The evolution and maintenance of delayed implantation in the Mustelidae (Mammalia: Carnivora). Evolution , **58,** 175–183.

Thornton, P.D., Newman, C., Johnson, P.J., Buesching, C.D., Baker, S.E., Slater, D., Johnson, D.D.P. & Macdonald, D.W. 2005. Preliminary comparison of four anaesthetic techniques in badgers (*Meles meles*). *Veterinary Anaesthesia and Analgesia*, **32,** 40–47.

Tuyttens, F.A.M., Macdonald, D.W., Delahay, R., Rogers, L.M., Mallinson, F.J., DOnnelly, C.A. & Newman, C. 1999. Differences in trappability of European badgers *Meles meles* in three populations in England. *Journal of Applied Ecology*, **36,** 1051–1062.

Walsh, P.S., Metzger, D.A. & Higuchi, R. 1991. Chelex® 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques*, **10,** 506–513.

Wang, J. & Santure, A.W. 2009. Parentage and sibship inference from multilocus genotype data under polygamy. *Genetics*, **181,** 1579–1594.

Weatherhead, P.J. & Robertson, R.J. 1979. Offspring quality and the polygyny threshold – Sexy son hypothesis. *The American Naturalist*, **113,** 201–208.

White, G.C. & Burnham, K.P. 1999. Program MARK: survival estimation from populations of marked animals. *Bird Study,* **46,** 120–128.

Wolff, J.O. & Macdonald, D.W. 2004. Promiscuous females protect their offspring. *Trends in Ecology & Evolution*, **19,** 127–134.

Yamaguchi, N., Dugdale, H.L., Macdonald, D.W. 2006. Female receptivity, embryonic diapause, and superfetation in the European badger (*Meles meles*): implications for the reproductive tactics of males and females. *The Quarterly Review of Biology*, **81,** 33–48.

Young, A.J., Spong, G. & Clutton-Brock, T. 2007. Subordinate male meerkats prospect for extra-group paternity: alternative reproductive tactics in a cooperative mammal. *Proceedings of the Royal Society B: B*

Supplementary materials

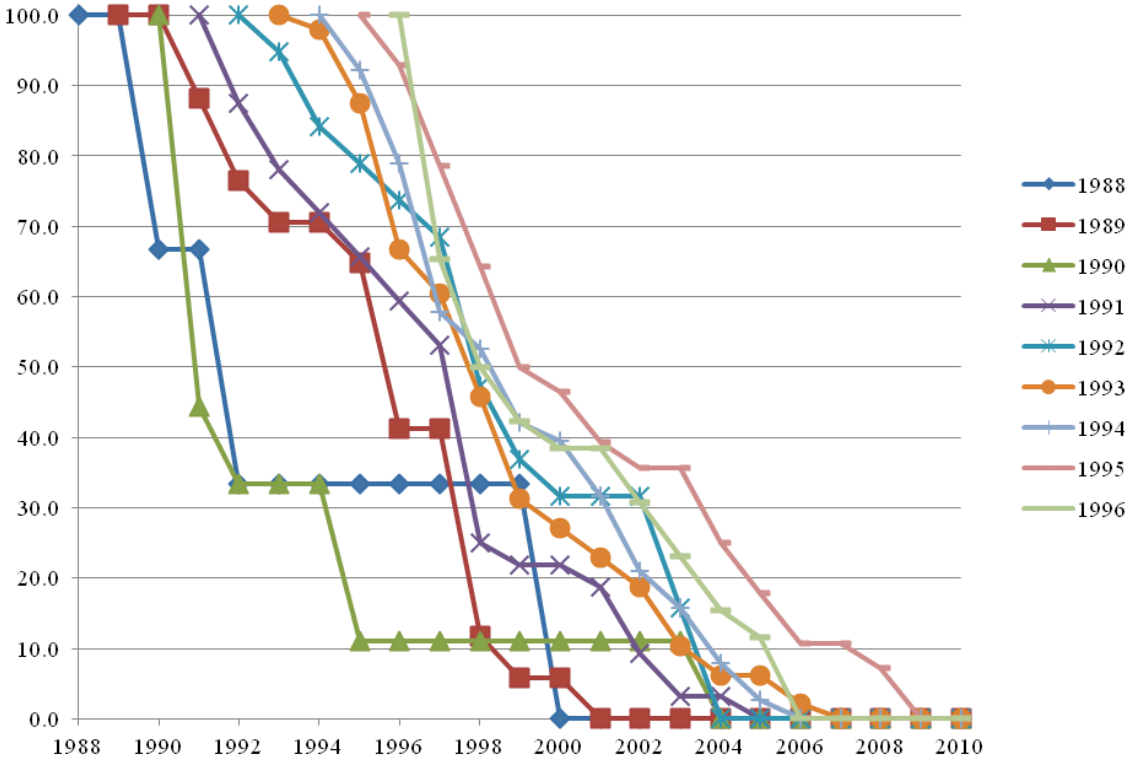


Figure S1 The percentage of badgers recaptured over the period of 1988-2010 for those individuals that survived to breeding age from cohorts 1988 to 1996 (N = 202).

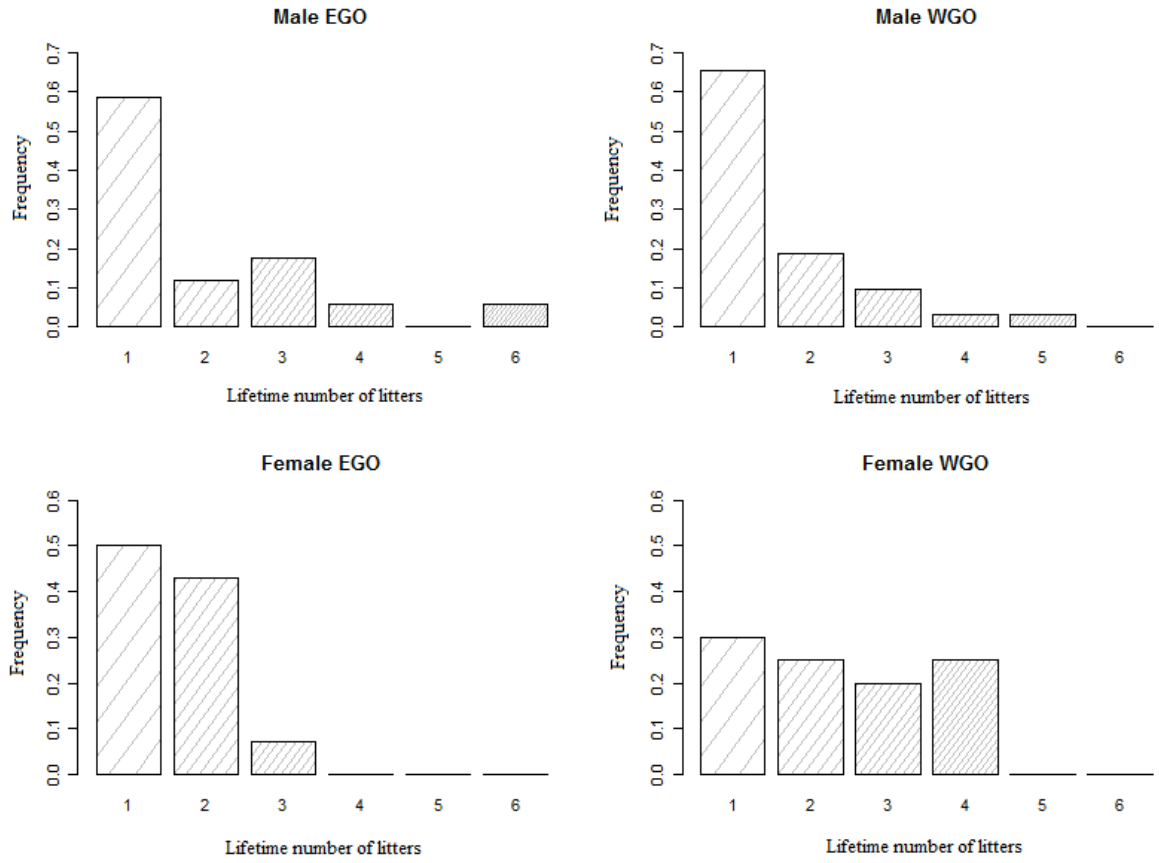


Figure S2 Frequency of the lifetime number of litters in relation to offspring-paternity type for male only and female only. The raw data are based on F_1 offspring born between 1988 and 1996 (EGO: female = 14, male = 17 and WGO: female = 20, male = 32).

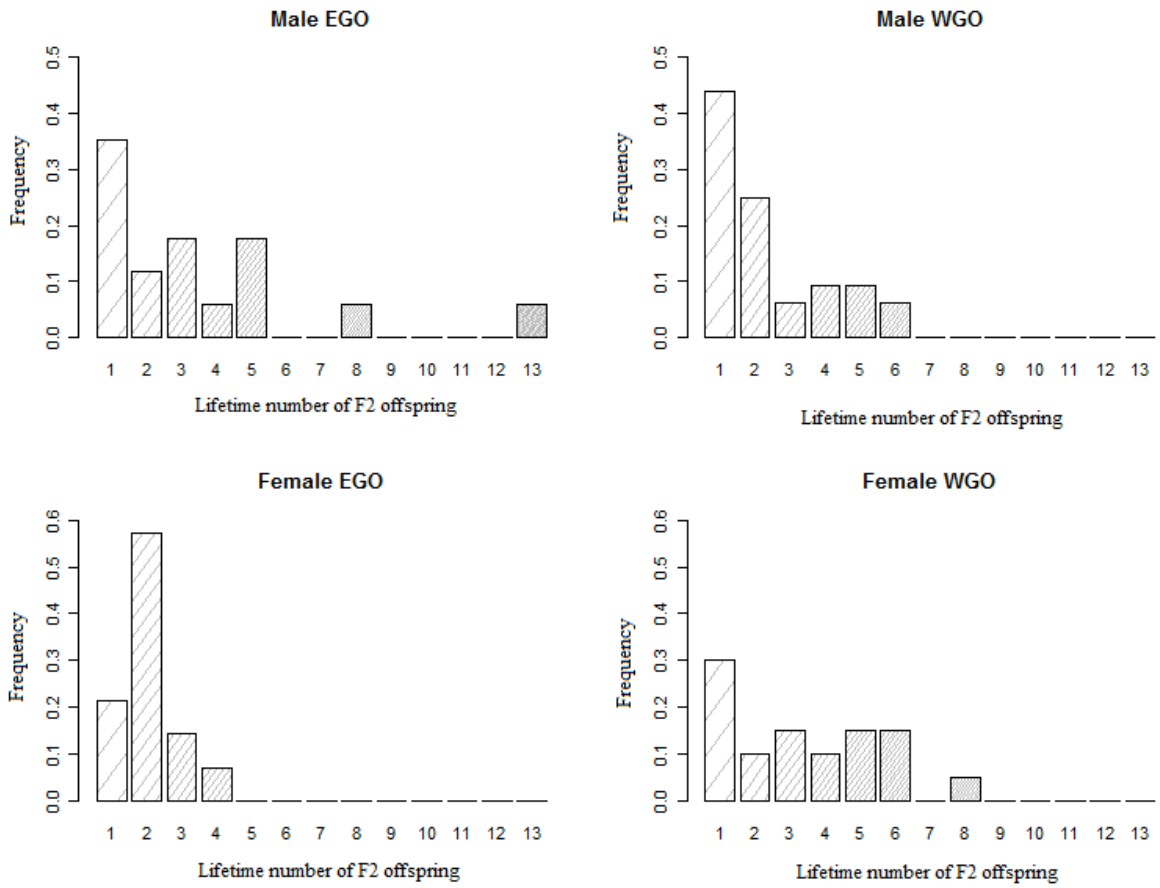


Figure S3 Frequency of the lifetime number of F_2 offspring in relation to offspring-paternity type for both sexes combined together, male only and female only. Data are based on F_1 offspring born between 1988 and 1996 (EGO: female = 14, male = 17 and WGO: female =20, male = 32