the **genetics**society

ARTICLE

Lack of evidence for selection favouring MHC haplotypes that combine high functional diversity

Arnaud Gaigher[1] · Alexandre Roulin[1] · Walid H. Gharib[2] · Pierre Taberlet[3,4] · Reto Burri[5] · Luca Fumagalli[1]

Abstract
High rates of gene duplication and the highest levels of functional allelic diversity in vertebrate genomes are the main hallmarks of the major histocompatibility complex (MHC), a multigene family with a primordial role in pathogen recognition. The usual tight linkage among MHC gene duplicates may provide an opportunity for the evolution of haplotypes that associate functionally divergent alleles and thus grant the transmission of optimal levels of diversity to coming generations. Even though such associations may be a crucial component of disease resistance, this hypothesis has been given little attention in wild populations. Here, we leveraged pedigree data from a barn owl (*Tyto alba*) population to characterize MHC haplotype structure across two MHC class I (MHC-I) and two MHC class IIB (MHC-IIB) duplicates, in order to test the hypothesis that haplotypes' genetic diversity is higher than expected from randomly associated alleles. After showing that MHC loci are tightly linked within classes, we found limited evidence for shifts towards MHC haplotypes combining high diversity. Neither amino acid nor functional within-haplotype diversity were significantly higher than in random sets of haplotypes, regardless of MHC class. Our results therefore provide no evidence for selection towards high-diversity MHC haplotypes in barn owls. Rather, high rates of concerted evolution may constrain the evolution of high-diversity haplotypes at MHC-I, while, in contrast, for MHC-IIB, fixed differences among loci may provide barn owls with already optimized functional diversity. This suggests that at the MHC-I and MHC-IIB respectively, different evolutionary dynamics may govern the evolution of within-haplotype diversity.

Reto Burri and Luca Fumagalli contributed equally to this work.

✉ Arnaud Gaigher
arnaud.gaigher@gmail.com

1 Laboratory for Conservation Biology, Department of Ecology and Evolution, University of Lausanne, Biophore, 1015 Lausanne Switzerland

2 Interfaculty Bioinformatics Unit, University of Bern, 3012 Bren Switzerland

3 CNRS, Laboratoire d'Ecologie Alpine (LECA), Grenoble 38000, France

4 Université Grenoble Alpes, Laboratoire d'Ecologie Alpine (LECA), Grenoble 38000, France

5 Department of Population Ecology, Institute of Ecology and Evolution, Friedrich Schiller University Jena, Dornburger Strasse 159, Jena 07743, Germany

Introduction

Gene duplication is a major mechanism in the evolution of phenotypic complexity (Lynch and Conery 2000; Conant and Wolfe 2008), and has led to one of the most remarkable adaptations in vertebrates, the major histocompatibility complex (MHC). The MHC multigene family has a primordial role in pathogen resistance. Classical MHC class I (MHC-I) and class II (MHC-II) genes encode cell-surface proteins that present antigen-peptides derived from pathogens to T-lymphocytes, in order to trigger an adaptive immune response (Klein and Sato 2000). As a result of the host-pathogen arms race, MHC-I and MHC-II genes have evolved the highest genetic diversity known from any vertebrate genome region to date (Gaudieri et al. 2000; Bernatchez and Landry 2003; Piertney and Oliver 2006). This diversity entails not only the number of different alleles and the high degree of genetic divergence between them, but also the number of duplicated genes. MHC-I and -II diversity is typically distributed across multiple functional

gene copies that are usually situated in tandem (Trowsdale and Parham 2004; Kelley et al. 2005).

Despite the growing amount of data on the characterization of MHC diversity and duplication history, the link between them, i.e., the combination of alleles of each duplicated MHC gene into haplotypes, has received little attention. Yet, diversity within haplotypes may deliver raw material that is selected at the ecological level. Until now, most of our knowledge about MHC haplotype structure is limited to human and poultry. In chicken, MHC-I variants appear to segregate together with co-adapted variants at strongly linked TAP genes that are fine-tuned with respect to their function of loading peptides on the MHC-I molecules (Walker et al. 2011). This coevolution is involved in MHC haplotype-related disease resistance, as for instance in economically important diseases such as Rous sarcoma virus or Marek's disease (Kaufman et al. 1999; Kaufman 2000; Wallny et al. 2006; Koch et al. 2007).

As MHC molecules are directly involved in the presentation of pathogen-peptides, MHC diversity should be optimized for a large number of different MHC molecules in individuals in order to fight a broader range of pathogens and thereby confers them with higher fitness (Doherty and Zinkernagel 1975; Bernatchez and Landry 2003; Sommer 2005; Spurgin and Richardson 2010). Individuals with highly divergent MHC alleles can interact with a wider range of pathogen-peptides than individuals with low allelic divergence (divergent allele advantage, Wakeland et al. 1990; Lenz 2011). Ample evidence has shown that high MHC diversity confers better pathogen resistance via heterozygote advantage or divergent allele advantage (for instance, Penn et al. 2002; Lenz et al. 2009; Oliver et al. 2009; Savage and Zamudio 2011), even if the optimum can be achieved by an intermediate level of MHC diversity due to the negative T-cell selection process (Nowak et al. 1992; Wegner et al. 2003). To optimize an individual's MHC diversity, mate choice for MHC-dissimilar partners may operate to increase the diversity in offspring. Alternatively, high-diversity haplotypes encompassing tightly linked MHC-I and/or MHC-II genes may ensure the transmission of a high amount of individual MHC diversity to progeny, even under random mating (Dearborn et al. 2016).

Under the latter hypothesis, high-diversity MHC haplotypes should be favored by selection. In natural populations, selection for such haplotypes may be expressed in one of two ways. In the most extreme case, the diversity of observed haplotypes (i.e., the ones found in the population) exceeds the diversity levels expected for random subsets of the possible haplotypes (i.e., of all possible combinations of variants across duplicated genes, including haplotypes absent from the population); low-diversity haplotypes are purged from the population. More likely, however, high-diversity MHC haplotypes are found at higher frequencies

than low-diversity haplotypes, and the average observed within-haplotype diversity should exceed the one expected under equal haplotype frequencies. These predictions should especially hold true for functional MHC diversity, i.e., the diversity observed at the residues of the peptide-binding region (PBR) involved in the detection of pathogen-derived peptides.

In most species, determining whether MHC haplotypes lock up higher than randomly expected diversity has been limited by the ability to reconstruct MHC haplotypes. In addition, establishing haplotypes is notoriously difficult in species exhibiting high number of duplicated MHC loci, such as observed in many bird species (for instance Promerová et al. 2009; Zagalska-Neubauer et al. 2010; Strandh et al. 2011; Sepil et al. 2012; Buehler et al. 2013). Here, we took advantage of extensive pedigree data to reconstruct MHC-I and MHC-IIB haplotypes, in order to investigate whether haplotypes combining a high MHC diversity were favored in a natural population of barn owl (*Tyto alba*), a species with only two MHC-I and MHC-IIB duplicates (Burri et al. 2008; Gaigher et al. 2016). To this end, our main objectives in the present study were to: (i) characterize the evolutionary mechanisms that shape MHC diversity; (ii) estimate the degree of linkage between MHC loci; and (iii) test whether the haplotypes' genetic diversity is higher than expected under random allelic combinations.

Material and methods

Sampling and DNA extraction

We focused our study on a single population of barn owls breeding in nest boxes in western Switzerland. We collected blood and feather samples from adults and their offspring between 1997 and 2003 resulting in a total of 937 barn owls. These samples included 823 individuals from 140 families. Each family was formed of two parents and on average 4.5 (range 1–17) offspring.

DNA was extracted using the DNeasy blood and tissue kit following the manufacturer's instructions (Qiagen, Hilden, Germany). All individuals were genotyped at 10 microsatellite markers (multiplex sets 3 and 4 in Burri et al. 2016) to verify parent-offspring relationships using CERVUS (Kalinowski et al. 2007).

MHC sequencing and genotyping

We investigated exon 3 of MHC class Iα (MHC-I) genes and exon 2 of MHC class IIβ (MHC-IIB) genes, which encode for polymorphic sequences encoding the respective genes' PBR. MHC-I primers were developed to specifically co-amplify the exon 3 of the two genes (see details in

Gaigher et al. 2016). For specific amplification of both MHC-IIB genes (DAB1 and DAB2), we used forward primers Tyal-int1F and Tyal-DAB2-int1F together with the single reverse primer Tyal-int2R (Burri et al. 2008; Supplementary Methods).

Because each MHC class was sequenced at a different time period, and since the most updated technologies available at that time were used, libraries of the MHC-I and MHC-IIB genes were sequenced with the Illumina MiSeq technology and the 454 Titanium pyrosequencing protocol, respectively. All molecular protocols are described in Gaigher et al. (2016) and Burri et al. (2008), for MHC-I and MHC-IIB, respectively, and in the Supplementary Methods. In brief, all individuals were amplified for both MHC classes with individual barcoded primers. PCR products were quantified (either visually on agarose gels or using the QIAxcel screening system (Qiagen)), purified by pooling eight PCR products of similar amplification intensity per column, and finally pooled according to equimolar concentrations of purified PCR products. Library preparation and high-throughput sequencing were performed at Fasteris (Plan-les-Ouates, Switzerland).

The MHC-I data used in the current study were previously published, and details about the genotyping procedure can be found in Gaigher et al. (2016). Briefly, the Illumina approach used to sequence MHC-I yielded a very high coverage per individual ($\sim$3000$\times$). To identify and estimate the number of MHC-I alleles per individual, we used the degree of change (DOC) (Lighten et al. 2014), that uses sequencing depth to distinguish true alleles from artifacts. Based on the pattern of allelic segregation within families, we have demonstrated that the DOC method provides accurate MHC genotyping (Gaigher et al. 2016). In addition, allelic segregation patterns, together with high per-individual sequencing coverage, revealed allele sharing among loci, as well as the presence of copy number variation (CNV) in the barn owl MHC-I (Gaigher et al. 2016).

The MHC-IIB data were generated for this study. The 454 technology used to sequence MHC-IIB loci resulted in an average coverage of 78 reads per individual. However, from these data a high proportion of artifacts was detected (mainly attributed to indels, but also including substitutions or chimera errors generated during PCR or sequencing). Consequently, in order to increase the coverage of true alleles to facilitate their identification we deployed a sequence similarity-based clustering approach to gather true alleles with all their potential artifacts, an approach in the same line of reasoning as Stutz and Bolnick (2014) and Sebastian et al. (2016). Our procedure relied on the three assumptions that: (i) in the whole dataset true alleles should be found at higher frequency than artifacts; (ii) artifacts should be highly similar to true alleles, differing only by 1 or 2 indels (especially in homopolymer regions) and/or

substitutions; and (iii) artifacts have to co-occur with their true alleles within an individual. Generated clusters (i.e., the true allele plus its artifacts) were used to define MHC-IIB genotypes. Due to the independent amplification of both MHC-IIB loci, a maximum of two clusters per loci and per individual was expected. For details of the procedure see the Supplementary Methods. The MHC-IIB genotyping were judged reliable due to the correct matches in the pattern of allelic segregation within families. Furthermore, a subset of around 100 individuals were genotyped using the cloning/Sanger method, and showed congruent genotype results with the 454 sequencing.

Characterization of MHC-I and MHC-IIB

All identified alleles were designated according to standard nomenclature (Klein et al. 1990) and deposited in GenBank. Alignments of MHC-I and MHC-IIB alleles were performed separately using ClustalW (Thompson et al. 1994) implemented in MEGA5 (Tamura et al. 2011). For each MHC class, the average number of pairwise differences per base pair (π) was estimated in DnaSP (Librado and Rozas 2009), and Poisson corrected amino acid distances were obtained in MEGA5. These analyzes were run on three data partitions: (i) the entire exon; (ii) codons of the PBR exclusively; and (iii) codons of the non-PBR exclusively. PBR codons were defined from Human HLA and Chicken BF for MHC-I (Bjorkman et al. 1987; Wallny et al. 2006) and from Human HLA for MHC-IIB (Brown et al. 1993).

In order to investigate the phylogenetic relationships among MHC alleles, we built a molecular phylogeny for each MHC class separately, using MrBayes v3.2.3 (Ronquist and Huelsenbeck 2003) based on the GTR $+$ Γ model, which was considered the best-fitting nucleotide substitution model by jModelTest (Darriba et al. 2012). Bayesian inference analyzes were performed with two independent MCMC runs of 2×10^7 generations (three heated chains with a temperature of 0.15). Parameter values and tree topologies were sampled every 2000 generations. Posterior probabilities were calculated after removing the first 25% of the topologies as burn-in. Convergence was estimated using the average standard deviation of split frequencies between runs, the estimated sample size and the potential scale reduction factor (PSRF) using MrBayes and Tracer v1.6 (Rambaut et al. 2014).

Recombination events were inferred using multiple methods implemented in RDP4, including RDP (Martin and Rybicki 2000), MaxChi (Smith 1992), and Chimerae (Posada and Crandall 2001). All default parameters were applied with a highest acceptable P-value of 0.05 and Bonferroni correction for multiple comparisons. In addition, we performed the Φw test (Bruen et al. 2006) in SplitsTree 4 (Huson and Bryant 2006), and estimated the minimal

number of historical recombination events (Hudson and Kaplan 1985) using the four-gamete test in DnaSP. Finally, gene conversion events were tracked using Geneconv 1.81 (Sawyer 1999) with 10,000 permutations.

In order to investigate footprints of positive selection, we estimated maximum likelihood site-models using CodeML implemented in PAML v4.7 (Yang 2007). These analyzes were performed independently for each MHC gene using the identified alleles as input. Two likelihood ratio tests of positive selection as proposed by Yang et al. (2005) were carried out comparing models M1a with M2a and models M7 with M8. Models M1a and M7 are neutral, while models M2a and M8 allow for a proportion of sites to evolve under positive selection. Likelihood ratio test statistics (i.e., $2*(lnL_b - lnL_a)$) were compared to the χ^2 distribution with two degrees of freedom. When the best-fit model was M2a or M8, sites under positive selection were determined through the Bayes empirical Bayes (BEB) approach. Input tree files used to run CodeML were generated from MrBayes under the GTR + Γ model. In order to ensure that signals of selection were not sensitive to tree topology, we used the best tree as input, and then reperformed the CodeML analysis with nine other topologies randomly chosen from the posterior distribution of topologies.

Genetic architecture

The MHC haplotype reconstruction for each individual was performed based on the allelic segregation within families. From the resulting haplotypes, we investigated linkage among MHC loci. We estimated linkage between: (i) the two MHC-I loci; (ii) the two MHC-IIB loci; and (iii) MHC classes. Because homozygote parents are uninformative regarding the occurrence of recombination, our linkage estimation was based only on heterozygous parents that transmitted a minimum of five gametes. Because a given parent can be heterozygous at one MHC class but homozygous at the other, the number of parents to assess linkage differed between analyzes involving MHC-I (103 parents, 804 gametes), MHC-IIB (57 parents, 438 gametes), and both classes (76 parents, 535 gametes).

Recombinant gametes were inferred from the rationale provided in Gaigher et al. (2016). From a fully heterozygous parent, a maximum of 16 different haplotypes are expected to be transmitted to offspring in case of free homologous recombination among all loci. If, in contrast, all MHC loci are linked, only two different haplotypic combinations should be observed in offspring; in this case, alleles at the four linked loci are generally transmitted together. Following this rationale, and assuming that allelic combinations resulted from a minimum number of recombination events, we deduced the frequency of recombinant gametes in our family data, which is indicative of the amount of linkage of the four loci.

Haplotype characterization

Firstly, we estimated the diversity combined within barn owl MHC-I and MHC-IIB haplotypes using three different genetic distances: (i) the nucleotide sequence-based p-distance; (ii) amino acid sequence-based p-distance; and (iii) amino acid functional distance. Nucleotide and amino acid distances between MHC alleles were calculated using MEGA5. Functional distances were measured as reported by Agbali et al. (2010) and Dearborn et al. (2016). Briefly, the 20 amino acids were described as numerical measures according to five physicochemical properties (Sandberg et al. 1998), which were used to calculate a Euclidean distance between each pair of amino acids. The functional distance between alleles for MHC-I and MHC-IIB loci was estimated as the mean of Euclidean distances. Then, to test whether the diversity combined within MHC-I and MHC-IIB haplotypes was higher than expected, we performed two tests: test 1 investigated whether the haplotypes observed in the population combined more diversity than a random set of the same number of haplotypes sampled from all possible haplotypes. We then investigated whether haplotypes that combine high diversity are present at elevated frequencies in the population relative to a random combination of alleles, such as expected if selection favored haplotypes combining higher than average diversity. To this end, in test 2 we tested whether the diversity observed with the population haplotype frequency distribution was higher than the one expected with a random combination of alleles, while considering the two loci's allele frequency distributions. Haplotype frequency used for test 2 was obtained from the different haplotypes that adults transmitted to offspring. For these two tests, 10^5 randomizations were run. These tests were performed independently on each MHC class and on three sequence partitions, namely the entire exon sequences, codons situated in the PBR, and codons inferred to be under positive selection. All statistical tests were performed in R 3.1.3 (R Core Team 2014).

Results

MHC-I and MHC-IIB characterization

Out of 937 individuals, 96, 79, and 83% were successfully genotyped for MHC-I, MHC-IIB DAB1, and MHC-IIB DAB2, respectively. The remaining individuals could not be genotyped mainly due to low coverage. A total of 69 MHC-I alleles, 25 MHC-IIB DAB1 alleles, and 17 MHC-IIB DAB2 alleles were identified (Fig. 1). None showed
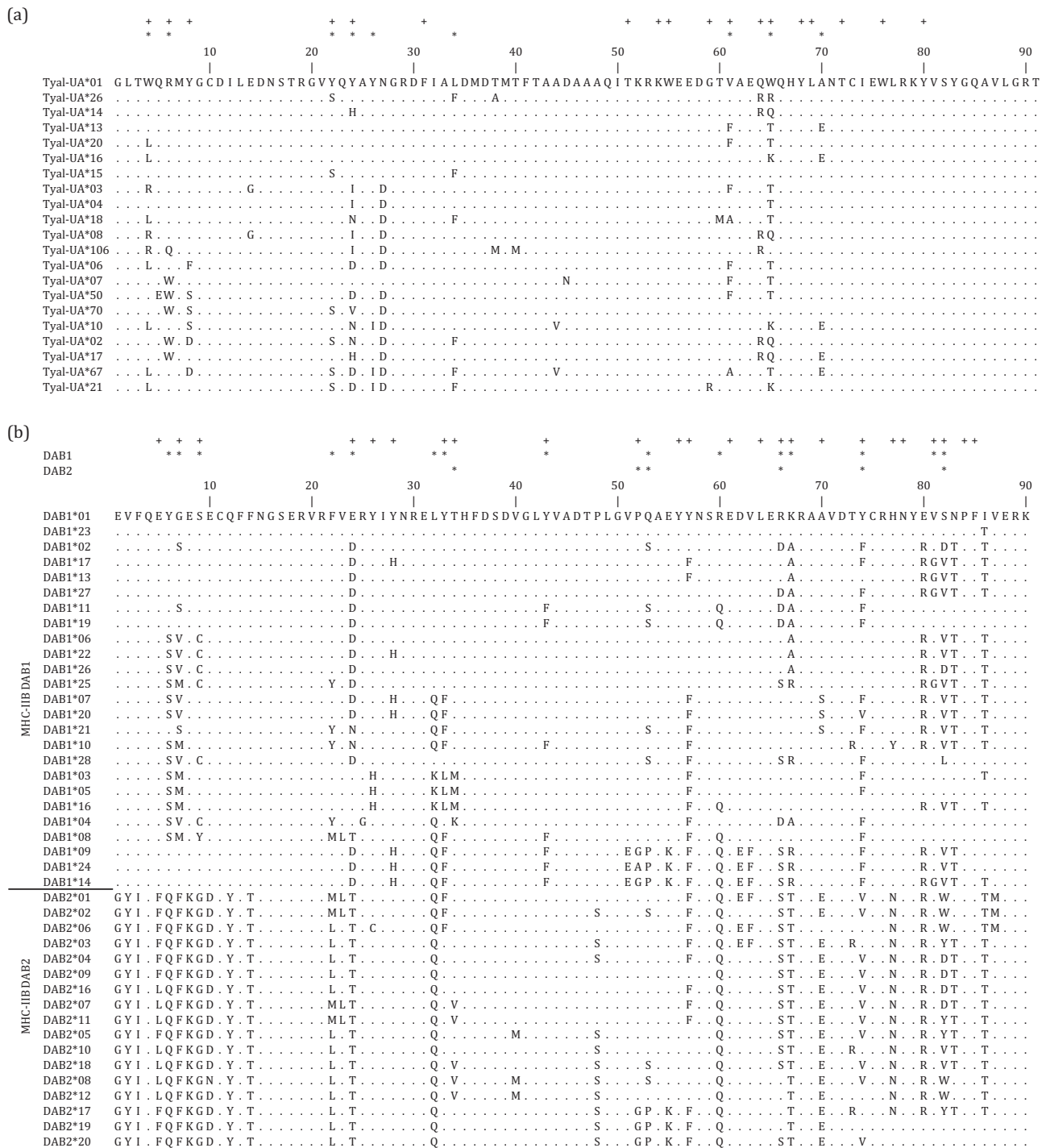
(a)

```
                +  +  +                      +  +          +                        +    + +     + +    + +    + +   +      +      +
              *   *                       *   *  *        *                          *         *         *
                    10        20        30        40        50        60        70        80        90
                    |         |         |         |         |         |         |         |         |
Tyal-UA*01  G L T W Q R M Y G C D I L E D N S T R G V Y Q Y A Y N G R D F I A L D M D T M T F T A A D A A A Q I T K R K W E E D G T V A E Q W Q H Y L A N T C I E W L R K Y V S Y G Q A V L G R T
Tyal-UA*26  . . . . . . . . . . . . . . . . . . S . . . . . . . . . . . . F . . . A . . . . . . . . . . . . . . . . . . . . . . . . R R . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Tyal-UA*14  . . . . . . . . . . . . . . . H . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . R Q . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Tyal-UA*13  . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . F . . . T . . . . E . . . . . . . . . . . . . . . . . . . . . . . .
Tyal-UA*20  . . . L . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . F . . . T . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Tyal-UA*16  . . . L . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . K . . . E . . . . . . . . . . . . . . . . . . . . . . . .
Tyal-UA*15  . . . . . . . . . . . . . . . . . . S . . . . . . . . . . F . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Tyal-UA*03  . . . R . . . . . . . . . G . . . . . . . . . I . D . . . . . . . . . . . . . . . . . . . . . . . . . . . F . . . T . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Tyal-UA*04  . . . . . . . . . . . . . . . . . . . . . . . I . D . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . T . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Tyal-UA*18  . . . L . . . . . . . . . . . . . . N . D . . . . F . . . . . . . . . . . . . . . . . . . . . . M A . . . T . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Tyal-UA*08  . . . R . . . . . . . . . G . . . . . . . . . I . D . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . R Q . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Tyal-UA*106 . . . R . Q . . . . . . . . . . . . . . . . . I . D . . . . . . . . M . M . . . . . . . . . . . . . . . . R . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Tyal-UA*06  . . . L . . F . . . . . . . . . . . . . . D . D . . . . . . . . . . . . . . . . . . . . . . . . . . . F . . . T . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Tyal-UA*07  . . . . . W . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . N . . . . . F . . . T . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Tyal-UA*50  . . . . E W . S . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . F . . . T . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Tyal-UA*70  . . . . . W . S . . . . . . . . . S . V . D . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Tyal-UA*10  . . . L . . . S . . . . . . . . N . I D . . . . . . . . . . V . . . . . . . . . . . . . . . . . . . . . . K . . . E . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Tyal-UA*02  . . . . . W . D . . . . . . . S . N . D . . . F . . . . . . . . . . . . . . . . . . . . . . . . . . . . . R Q . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Tyal-UA*17  . . . . . W . . . . . . . . . H . D . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . R Q . . . . E . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Tyal-UA*67  . . . L . . . D . . . . . . . S . D . I D . . . . F . . . . . V . . . . . . . . . . . . . . . . . . . . A . . . T . . . . E . . . . . . . . . . . . . . . . . . . . . . . . . . .
Tyal-UA*21  . . . L . . . . . . . . . . . S . D . I D . . . . F . . . . . . . . . . . . . . . . . . . . . . . . . . R . . . . K . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
```

(b) [sequence alignment figure for MHC-IIB DAB1 and DAB2 alleles]

Fig. 1 Amino acid sequences of MHC-I (**a**) and MHC-IIB alleles (**b**) of barn owl. Only the most divergent alleles are shown to illustrate the diversity in MHC-I, whereas all alleles are presented for MHC-IIB. +, residues associated to the PBR. *, sites identified to have evolved under positive selection according the M8 model (*P* > 95%) (based on the Bayes empirical Bayes approach). The evidence for positive selection at residues 5 and 63 of DAB2 was very sensitive to tree topology and consequently not considered as robust. The same applied to residue 8 of MHC-I

evidence of non-functionality, such as frameshift mutations or stop codons. All nucleotide sequences translated into unique amino acid sequences for MHC-IIB, and only four were synonymous for MHC-I. Sequence analyzes revealed that both MHC-I and MHC-IIB loci exhibited the classical characteristics of functional MHC genes: (i) high genetic

Fig. 2 Histogram of the amino acid p-distance between MHC-I (**a**) and MHC-IIB (**b**) alleles in Swiss barn owls. In (**b**), white bars represent p-distance between alleles within MHC-IIB DAB1 locus, gray within MHC-IIB DAB2, and black between MHC-IIB loci
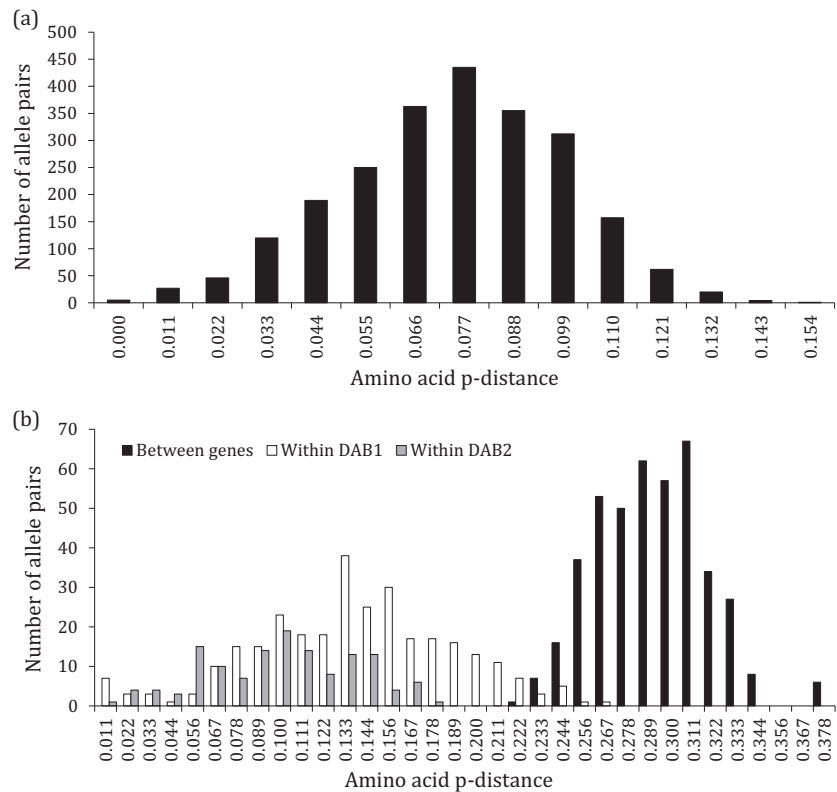
diversity mainly located in the peptide-binding regions (Fig. 1; Supplementary Table S1); (ii) evidence of positive selection (Fig. 1); and (iii) footprints of recombination and gene conversion (Supplementary Table S2). DAB1 displayed a higher diversity than DAB2 (π: DAB1, 0.071; DAB2, 0.053; Supplementary Table S1) with a different amino acid composition (Fig. 1). Our population covers a large variation of allele frequencies from very common to very rare alleles (the frequency of the most common alleles for MHC-I, MHC-IIB DAB1, and MHC-IIB DAB2 genes were 0.12, 0.26, and 0.50 respectively; Supplementary Figure S1).

In line with the monophyly of MHC-IIB loci in the phylogenetic tree (Supplementary Figure S2), MHC-IIB exon 2 is highly divergent between both loci (mean amino acid p-distance between loci, within DAB1, and within DAB2, respectively: 0.292, 0.138, and 0.099) (Fig. 2; Supplementary Figure S3). In contrast, the MHC-I tree exhibited a polytomic topology indicative of reticulate evolution of alleles not only within but also between the two loci (Supplementary Figure S2 and Figure S3). The MHC-I pairwise genetic distances revealed a unimodal distribution, with a mean amino acid p-distance of 0.075 (Fig. 2). Although assigning alleles to loci based on the MHC-I tree was impossible, this could be achieved based on family

data. Indeed, given that we observe a set of alleles combining only with another specific set of alleles, we were able to attribute alleles to loci (Supplementary Figure S4). However, this analysis reveals allele sharing among loci, for instance, Tyal-UA*01 allele occurred on the two MHC-I loci within the same haplotype (Supplementary Figure S4) (Gaigher et al. 2016).

Linkage within and between MHC classes

We inferred MHC-I/MHC-IIB haplotypes in offspring based on the pattern of allele segregation within families, and tracked recombination events to estimate linkage among MHC loci. In line with expectations of tight linkage between MHC loci, our analyzes revealed that for both classes each parent almost exclusively transmitted two different haplotypes to offspring (Fig. 3a). Within 438 analyzed gametes, no recombination event was detected between MHC-IIB loci, and for MHC-I out of 804 gametes only three showed evidence for recombination between loci. In contrast, between MHC classes eight recombination events were detected within 535 gametes (Fig. 3b). In addition, nine other recombinant gametes were detected; however due to homozygosity of parents for one locus, recombination events were impossible to locate (i.e.,

Fig. 3 Genetic linkage between MHC loci in Swiss barn owls. To facilitate the reading, all allele names were reduced to the allele number. Crosses indicate the presence of alleles in offspring. These two examples illustrate families in which both parents are heterozygotes for all loci. **a** Family with only four observed haplotypes in 8 offspring. For instance, father transmitted only two different haplotypes to offspring: DAB1*03/DAB2*05/MHC-I*02/MHC-I*12 and DAB1*07/DAB2*01/MHC-I*03/MHC-I*09. **b** Family composed of 12 offspring, in which recombination has been detected between MHC classes. Black backgrounds indicate recombinant haplotypes
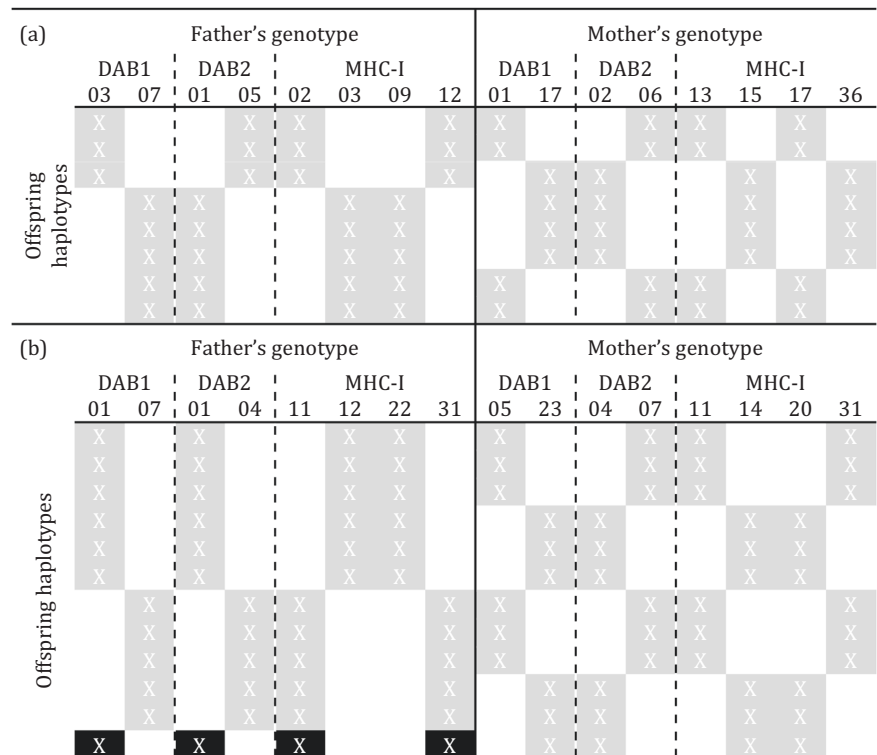
between MHC classes or between loci of the same class). In total, we found evidence for 20 recombination events, implying that MHC loci are linked (lower than 3 cM), but with a stronger linkage within than between MHC classes, and with a stronger linkage between MHC-IIB loci than between MHC-I loci. As may be expected from the latter result, the most common MHC-I alleles are found in haplotypes in combination with many different alleles (for instance Tyal-UA*01, *02, and *03 combine with 13, 12, 12 different alleles, respectively), whereas the most common MHC-IIB DAB1 alleles group with exclusively one or a few DAB2 alleles (Tyal-DAB1*01, Tyal-DAB1*10, and Tyal-DAB1*05 combine with two, two and one DAB2 alleles) (Supplementary Figure S4). This last point was supported by the strong linkage between MHC-IIB loci estimated from the likelihood ratio test ($P < 0.001$).

Haplotype characterization

A total of 111 MHC-I and 40 MHC-IIB different haplotypes were observed (Supplementary Figure S4). Across MHC classes, 210 different haplotypes were identified. Our data highlighted that only 11 and 9% of all possible allelic combinations were realized for MHC-I and MHC-IIB, respectively. In addition, our population compiles a wide variation of haplotype frequencies from common to rare

haplotypes (Supplementary Figure S4), with important amino acid divergence between alleles (Figs. 1, 2). Consequently, we took advantage of our data to first test whether the diversity combined within the MHC-I and MHC-IIB haplotypes that are observed in the population was higher than expected under a random set of all possible haplotypes. We found no support in this direction. Neither nucleotide, amino acid nor functional within-haplotype diversity in the population were significantly higher than in random sets of haplotypes, regardless of the MHC class (Test 1, Table 1). Then, we tested whether MHC haplotypes with higher frequencies combine the highest diversity, relative to an expected haplotype frequency distribution (Test 2, Table 1). The most common MHC-IIB DAB2 allele (MHC-IIB DAB2*01) displays on average the highest amino acid distance with DAB1 alleles (mean amino acid p-distance: 0.311); hence we performed the second test considering allele frequencies in the expected distribution, in order to account for processes unrelated to selection (Test 2, Table 1). We found low support in this direction, with only a significant shift for high diversity at the nucleotide level in the PBR and positively selected site (PSS) data, as well as at the amino acid level in the PSS (Test 2, Table 1). Overall, an inverse trend was observed for MHC-I haplotypes; i.e., observed haplotypes appear to have lower diversity compared to random expectations (Table 1).

Table 1 Mean within-haplotype diversity for MHC-I and MHC-IIB in Swiss barn owls

	MHC-I						MHC-IIB					
	Test 1			Test 2			Test 1			Test 2		
	Exp.	Obs.	P	Exp.	Obs.	P	Exp.	Obs.	P	Exp.	Obs.	P
Nucleotide p-distance												
Entire	0.034	0.030	1.000	0.030	0.029	0.813	0.174	0.176	0.183	0.178	0.178	0.061
PBR	0.152	0.133	0.999	0.136	0.134	0.710	0.317	0.319	0.341	0.314	0.316	**0.002**
PSS	0.224	0.195	0.999	0.195	0.188	0.918	0.455	0.463	0.135	0.446	0.450	**0.000**
Amino acid p-distance												
Entire	0.069	0.061	0.999	0.060	0.060	0.520	0.292	0.295	0.259	0.303	0.304	0.138
PBR	0.316	0.281	0.996	0.283	0.289	0.216	0.491	0.495	0.316	0.496	0.498	0.092
PSS	0.483	0.423	1.000	0.414	0.413	0.510	0.749	0.762	0.178	0.751	0.755	**0.030**
Amino acid functional distance												
Entire	0.395	0.351	0.998	0.355	0.355	0.510	1.495	1.496	0.480	1.537	1.533	0.937
PBR	2.059	1.859	0.989	1.927	1.945	0.362	2.594	2.603	0.420	2.667	2.655	0.981
PSS	3.039	2.708	0.997	2.754	2.734	0.614	3.645	3.656	0.435	3.710	3.685	0.991

Values in bold are significant ($P < 0.05$)

Test 1 investigates whether the diversity combined within haplotypes observed in the population is higher than a random set of the same number of haplotypes sampled from all possible haplotypes. Tests 2 investigates whether the diversity combined within the observed haplotype frequency distribution is higher than within a random combination of alleles, while considering the allele frequency distribution

Exp mean expected diversity, *Obs* mean observed diversity, *P* *P*-value, *Entire* entire sequence, *PBR* peptide-binding region, *PSS* positively selected site

Discussion

In the present study, we took advantage of the simple MHC organization of the barn owl and extensive family data to investigate whether tight linkage among MHC genes may favor the evolution of haplotypes that associate functionally divergent alleles, and thus grant the transmission of a high amount of MHC diversity across generations. Our analysis revealed the following main results: (i) a contrasted evolutionary dynamics between MHC classes, where on one hand the two MHC-I loci are indistinguishable due to their high sequence similarity, and on the other hand the two MHC-IIB loci are strongly divergent; (ii) a tight linkage between all MHC loci, but with a stronger linkage within than between MHC classes; and (iii) no evidence for shifts towards high within-haplotype MHC diversity at the amino acid sequence level in our population. As our dataset provided a good representation of the barn owl haplotype diversity in the study population, sample size is unlikely to explain the lack of evidence for evolution towards high-diversity haplotypes. Given the likely biological meaning of our finding, we therefore discuss how the evolution of high-diversity haplotypes in our population may be constrained by the molecular evolution of MHC genes.

Ultimately, from a functional perspective it is unlikely to matter whether two divergent MHC molecules situated at the cell surface are encoded by alleles of the same locus but on different (paternal and maternal) chromosomes, or by alleles of two paralogs linked in the same haplotype. The sole advantage of divergent alleles combined within a haplotype may therefore be that it assures the inheritance of a certain level of MHC diversity across generations. A previous study in the MHC-DRB of wild baboons suggested that selection favors haplotypes combining different sets of DRB supertypes (i.e., clusters of alleles based on their similar amino acid physicochemical properties), leading to an overall high diversity over multiple loci in individuals (Huchard et al. 2008). In contrast, here we found only low support for high within-haplotype diversity. Explanations for this finding may be fundamentally different between the two MHC classes.

For MHC-I, the evolution of high-diversity haplotypes may be constrained by high rates of recombination and gene conversion. These processes have previously been documented to shape MHC diversity especially in birds (Hess and Edwards 2002; Miller and Lambert 2004; Spurgin et al. 2011; Promerová et al. 2013; Goebel et al. 2017). In addition, allele shuffling by gene conversion between tandem duplicates is more frequent if loci are physically linked

(Ezawa et al. 2006). The high levels of linkage between barn owl MHC-I loci (i.e., few crossing over events) may therefore favor the occurrence of gene conversion and explain the sharing of alleles among duplicates. In line with this, we previously demonstrated allele sharing among barn owl MHC-I loci, as well as CNV (Gaigher et al. 2016), both of which decrease the level of divergence between loci. Barn owl MHC-I diversity therefore tends towards a homogenization across both loci, suggesting high rates of gene conversion. Our results even suggest that observed haplotypes combine lower diversity compared to random expectations. Whether this is promoted by selection remains to be addressed.

In contrast, the highly divergent evolutionary history between the two MHC-IIB loci may inherently have promoted the evolution toward high-diversity haplotypes. Here, it is important to note that, had we randomized alleles between rather than within loci, haplotypes would be significantly more diverse than by chance: the two barn owl MHC-IIB loci exhibit fixed differences in the amino acid sequence, especially within the PBR in 5' of the sequence (Burri et al. 2008). These fixed differences generate much higher allelic diversity between than within the MHC-IIB loci, and their maintenance may either be due to selection, or due to the limited rate of recombination found in the barn owl MHC-IIB. In either case, as the two loci are already divergent, an even higher level of divergence may be not be of additional advantage, as the fixed differences between duplicates may already ensure the transmission of a sufficient amount of diversity to the next generation.

In the MHC context, the evolution of high-diversity haplotypes may be promoted by the very same mechanism restricting the co-segregation of co-adapted alleles, i.e., recombination. Recombination (sensu lato) represents a major driver of MHC evolution by generating new MHC allelic combinations (see for instance Richman et al. 2003; Promerová et al. 2009; Spurgin et al. 2011), which may offer an adaptive potential against pathogens (She et al. 1991). When selection is strong enough, new high-diversity combinations of alleles can be locked, and increase in frequency in the population. At the same time however, if recombination rates are high enough to recombine divergent alleles into beneficial high-diversity haplotypes, it may be equally likely to break up such advantageous combinations. Our results therefore may suggest that in a system involved in defense against pathogens, such as the MHC, considerable flexibility—and hence recombination—may be required to parallel the dynamics of pathogens in time and space (Milinski 2006), and that the advantages of recombination surmount those of suppressed recombination to maintain high-diversity haplotypes.

To conclude, different evolutionary dynamics may govern the evolution of within-haplotype diversity and selection

for high-diversity MHC haplotypes may be weak in the studied population. Whether this reflects MHC diversity levels close to the optimum or results from constraints imposed by recombination is a topic of future investigation.

Data archiving

Previously identified MHC-I sequences are available on GenBank (accession numbers: KX189198-KX189343). MHC-IIB sequences described in this study were deposited in GenBank (accession numbers for DAB1: MG595289-MG595313; for DAB2: MG595314-MG595330). Family data, MHC genotypes and haplotypes were deposited on Dryad database (https://doi.org/10.5061/dryad.745t0).

Compliance with ethical standards

References

Agbali M, Reichard M, Bryjová A, Bryja J, Smith C (2010) Mate choice for nonadditive genetic benefits correlate with MHC dissimilarity in the rose bitterling (*Rhodeus ocellatus*). Evolution 64:1683–1696

Bernatchez L, Landry C (2003) MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? J Evol Biol 16:363–377

Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, Wiley DC (1987) The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. Nature 329:512–518

Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL et al. (1993) Three-dimensional structure of the human class-II histocompatibility antigen HLA-DR1. Nature 364:33–39

Bruen TC, Philippe H, Bryant D (2006) A simple and robust statistical test for detecting the presence of recombination. Genetics 172:2665–2681

Buehler DM, Verkuil YI, Tavares ES, Baker AJ (2013) Characterization of MHC class I in a long-distance migrant shorebird suggests multiple transcribed genes and intergenic recombination. Immunogenetics 65:211–225

Burri R, Antoniazza S, Gaigher A, Ducrest A-L, Simon C, The European Barn Owl N (2016) The genetic basis of color-related local adaptation in a ring-like colonization around the Mediterranean. Evolution 70:140–153. et al.

Burri R, Niculita-Hirzel H, Roulin A, Fumagalli L (2008) Isolation and characterization of major histocompatibility complex (MHC)

class IIB genes in the Barn owl (Aves: *Tyto alba*). Immunogenetics 60:543–550

Conant GC, Wolfe KH (2008) Turning a hobby into a job: how duplicated genes find new functions. Nat Rev Genet 9:938–950

Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. Nat Methods 9:772–772

Dearborn DC, Gager AB, McArthur AG, Gilmour ME, Mandzhukova E, Mauck RA (2016) Gene duplication and divergence produce divergent MHC genotypes without disassortative mating. Mol Ecol 25:4355–4367

Doherty PC, Zinkernagel RM (1975) Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. Nature 256:50–52

Ezawa K, OOta S, Saitou N (2006) Genome-wide search of gene conversions in duplicated genes of mouse and rat. Mol Biol Evol 23:927–940

Gaigher A, Burri R, Gharib WH, Taberlet P, Roulin A, Fumagalli L (2016) Family-assisted inference of the genetic architecture of major histocompatibility complex variation. Mol Ecol Res 16:1353–1364

Gaudieri S, Dawkins RL, Habara K, Kulski JK, Gojobori T (2000) SNP profile within the human major histocompatibility complex reveals an extreme and interrupted level of nucleotide diversity. Genome Res 10:1579–1586

Goebel J, Promerová M, Bonadonna F, McCoy KD, Serbielle C, Strandh M et al. (2017) 100 million years of multigene family evolution: origin and evolution of the avian MHC class IIB. BMC Genom 18:460

Hess CM, Edwards SV (2002) The evolution of the major histocompatibility complex in birds. Bioscience 52:423–431

Huchard E, Weill M, Cowlishaw G, Raymond M, Knapp LA (2008) Polymorphism, haplotype composition, and selection in the Mhc-DRB of wild baboons. Immunogenetics 60:585–598

Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics 111:147–164

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23:254–267

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol 16:1099–1106

Kaufman J (2000) The simple chicken major histocompatibility complex: life and death in the face of pathogens and vaccines. Philos Trans R Soc Lond B Biol Sci 355:1077–1084

Kaufman J, Milne S, Gobel TWF, Walker BA, Jacob JP, Auffray C et al. (1999) The chicken B locus is a minimal essential major histocompatibility complex. Nature 401:923–925

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

Klein J, Bontrop RE, Dawkins RL, Erlich HA, Gyllensten UB, Heise ER et al. (1990) Nomenclature for the major histocompatibility complexes of different species: a proposal. Immunogenetics 31:217–219

Klein J, Sato A (2000) The HLA system. N Engl J Med 343:702–709

Koch M, Camp S, Collen T, Avila D, Salomonsen J, Wallny H-J et al. (2007) Structures of an MHC class I molecule from B21 chickens illustrate promiscuous peptide binding. Immunity 27:885–899

Lenz TL (2011) Computational prediction of MHC II-antigen binding supports divergent allele advantage and explains trans-species polymorphism. Evolution 65:2380–2390

Lenz TL, Wells K, Pfeiffer M, Sommer S (2009) Diverse MHC IIB allele repertoire increases parasite resistance and body condition in the Long-tailed giant rat (*Leopoldamys sabanus*). BMC Evol Biol 9:269

Librado P, Rozas J (2009) DnaSPv5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452

Lighten J, van Oosterhout C, Paterson IG, McMullan M, Bentzen P (2014) Ultra-deep Illumina sequencing accurately identifies MHC class IIb alleles and provides evidence for copy number variation in the guppy (*Poecilia reticulata*). Mol Ecol Res 14:753–767

Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290:1151–1155

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

Milinski M (2006) The major histocompatibility complex, sexual selection, and mate choice. Ann Rev Ecol Evol Syst 37:159–186

Miller HC, Lambert DM (2004) Gene duplication and gene conversion in class II MHC genes of New Zealand robins (Petroicidae). Immunogenetics 56:178–191

Nowak MA, Tarczy-Hornoch K, Austyn JM (1992) The optimal number of major histocompatibility complex molecules in an individual. Proc Natl Acad Sci USA 89:10896–10899

Oliver MK, Telfer S, Piertney SB (2009) Major histocompatibility complex (MHC) heterozygote superiority to natural multi-parasite infections in the water vole (*Arvicola terrestris*). Proc R Soc Lond B Biol Sci 276:1119–1128

Penn DJ, Damjanovich K, Potts WK (2002) MHC heterozygosity confers a selective advantage against multiple-strain infections. Proc Natl Acad Sci USA 99:11260–11264

Piertney SB, Oliver MK (2006) The evolutionary ecology of the major histocompatibility complex. Heredity 96:7–21

Posada D, Crandall KA (2001) Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proc Natl Acad Sci USA 98:13757–13762

Promerová M, Albrecht T, Bryja J (2009) Extremely high MHC class I variation in a population of a long-distance migrant, the Scarlet Rosefinch (*Carpodacus erythrinus*). Immunogenetics 61:451–461

Promerová M, Králová T, Bryjová A, Albrecht T, Bryja J (2013) MHC class IIB exon 2 polymorphism in the grey partridge (*Perdix perdix*) is shaped by selection, recombination and gene conversion. PLoS ONE 8:e69135

R Core Team (2014) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. http://www.R-project.org/

Rambaut A, Suchard M, Xie D, Drummond A (2014) Tracer v1.6, available from http://tree.bio.ed.ac.uk/software/tracer/.

Richman AD, Herrera LG, Nash D (2003) Evolution of MHC class II Eβ diversity within the genus Peromyscus. Genetics 164:289–297

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574

Sandberg M, Eriksson L, Jonsson J, Sjöström M, Wold S (1998) New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids. J Med Chem 41:2481–2491

Savage AE, Zamudio KR (2011) MHC genotypes associate with resistance to a frog-killing fungus. Proc Natl Acad Sci USA 108:16705–16710

Sawyer S (1999) GENECONV: a computer package for the statistical detection of gene conversion. Distributed by the author, Department of Mathematics, Washington University in St. Louis. http://www.math.wustl.edu/~sawyer/geneconv/

Sebastian A, Herdegen M, Migalska M, Radwan J (2016) Amplisas: a web server for multilocus genotyping using next-generation amplicon sequencing data. Mol Ecol Res 16:498–510

Sepil I, Moghadam H, Huchard E, Sheldon B (2012) Characterization and 454 pyrosequencing of major histocompatibility complex class I genes in the great tit reveal complexity in a passerine system. BMC Evol Biol 12:68

She JX, Boehme SA, Wang TW, Bonhomme F, Wakeland EK (1991) Amplification of major histocompatibility complex class II gene diversity by intraexonic recombination. Proc Natl Acad Sci USA 88:453–457

Smith J (1992) Analyzing the mosaic structure of genes. J Mol Evol 34:126–129

Sommer S (2005) The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Front Zool 2:16

Spurgin LG, Richardson DS (2010) How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. Proc R Soc Lond B Biol Sci 277:979–988

Spurgin LG, van Oosterhout C, Illera JC, Bridgett S, Gharbi K, Emerson BC et al. (2011) Gene conversion rapidly generates major histocompatibility complex diversity in recently founded bird populations. Mol Ecol 20:5213–5225

Strandh M, Lannefors M, Bonadonna F, Westerdahl H (2011) Characterization of MHC class I and II genes in a subantarctic seabird, the blue petrel, *Halobaena caerulea* (Procellariiformes). Immunogenetics 63:653–666

Stutz WE, Bolnick DI (2014) Stepwise threshold clustering: a new method for genotyping MHC loci using next-generation sequencing technology. PLoS One 9:e100587

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Trowsdale J, Parham P (2004) Mini-review: defense strategies and immunity-related genes. Eur J Immunol 34:7–17

Wakeland E, Boehme S, She J, Lu C-C, McIndoe R, Cheng I et al. (1990) Ancestral polymorphisms of MHC class II genes: divergent allele advantage. Immunol Res 9:115–122

Walker BA, Hunt LG, Sowa AK, Skjødt K, Göbel TW, Lehner PJ et al. (2011) The dominantly expressed class I molecule of the chicken MHC is explained by coevolution with the polymorphic peptide transporter (TAP) genes. Proc Natl Acad Sci USA 108:8396–8401

Wallny H-J, Avila D, Hunt LG, Powell TJ, Riegert P, Salomonsen J et al. (2006) Peptide motifs of the single dominantly expressed class I molecule explain the striking MHC-determined response to Rous sarcoma virus in chickens. Proc Natl Acad Sci USA 103:1434–1439

Wegner KM, Kalbe M, Kurtz J, Reusch TBH, Milinski M (2003) Parasite selection for immunogenetic optimality. Science 301:1343–1343

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591

Yang Z, Wong WSW, Nielsen R (2005) Bayes empirical Bayes inference of amino acid sites under positive selection. Mol Biol Evol 22:1107–1118

Zagalska-Neubauer M, Babik W, Stuglik M, Gustafsson L, Cichon M, Radwan J (2010) 454 sequencing reveals extreme complexity of the class II major histocompatibility complex in the collared flycatcher. BMC Evol Biol 10:395