JOURNAL OF Evolutionary Biology

doi: 10.1111/jeb.12798

Evidence of gene orthology and trans-species polymorphism, but not of parallel evolution, despite high levels of concerted evolution in the major histocompatibility complex of flamingo species

M. A. F. GILLINGHAM*†‡§, A. COURTIOL‡, M. TEIXEIRA*, M. GALAN¶, A. BECHET† & F. CEZILLY*

*Equipe Ecologie Evolutive, UMR CNRS 6282 Biogéosciences, Université de Bourgogne, Dijon, France
†Centre de Recherche de la Tour du Valat, Arles, France
‡Department of Evolutionary Genetics, Leibniz Institute for Zoo and Wildlife Research, Berlin, Germany
§Institute of Evolutionary Ecology and Conservation Genomics, University of Ulm, Ulm, Germany
¶UMR CBGP (INRA/IRD/Cirad/Montpellier SupAgro), INRA EFPA, Montferrier-sur-Lez Cedex, France

Keywords:

ancestral polymorphism;
antigen-binding site;
concerted evolution;
convergent evolution;
flamingos;
gene orthology;
major histocompatibility complex genes;
parallel evolution;
trans-species polymorphism.

Abstract

The major histocompatibility complex (MHC) is a cornerstone in the study of adaptive genetic diversity. Intriguingly, highly polymorphic MHC sequences are often not more similar within species than between closely related species. Divergent selection of gene duplicates, balancing selection maintaining trans-species polymorphism (TSP) that predate speciation and parallel evolution of species sharing similar selection pressures can all lead to higher sequence similarity between species. In contrast, high rates of concerted evolution increase sequence similarity of duplicated loci within species. Assessing these evolutionary models remains difficult as relatedness and ecological similarities are often confounded. As sympatric species of flamingos are more distantly related than allopatric species, flamingos represent an ideal model to disentangle these evolutionary models. We characterized MHC Class I exon 3, Class IIB exon 2 and exon 3 of the six extant flamingo species. We found up to six MHC Class I loci and two MHC Class IIB loci. As all six species shared the same number of MHC Class IIB loci, duplication appears to predate flamingo speciation. However, the high rate of concerted evolution has prevented the divergence of duplicated loci. We found high sequence similarity between all species regardless of codon position. The latter is consistent with balancing selection maintaining TSP, as under this mechanism amino acid sites under pathogen-mediated selection should be characterized by fewer synonymous codons (due to their common ancestry) than under parallel evolution. Overall, balancing selection maintaining TSP appears to result in high MHC similarity between species regardless of species relatedness and geographical distribution.

Introduction

The multigene family of major histocompatibility complex (MHC) genes plays a central role in adaptive immunity of vertebrates (Janeway *et al.*, 2005). Classical MHC genes are among the most polymorphic genes of vertebrate species, both in terms of allelic diversity and sequence diversity (Klein, 1986; Potts *et al.*, 1994; Steele & Lloyd, 2015). Therefore, MHC genes have become an important model to elucidate how genetic diversity is maintained through evolutionary time (Sommer, 2005; Spurgin & Richardson, 2010). MHC diversity is consensually thought to be predominantly selected by pathogen-mediated selection, although sex-

Correspondence: Mark Gillingham, Institute of Evolutionary Ecology and Conservation Genomics, Albert-Einstein Allee 11, D-89069 Ulm, Germany.
Tel.: 0049-731-5022660; fax: 0049-731-5022683; e-mail: mark.gillingham@uni-ulm.de

ual selection may also play a role (Sommer, 2005; Spurgin & Richardson, 2010). Whether the current diversity results from the maintenance of an initial diversity over relatively long-evolutionary timescales or from recurrent selection pressures regenerating eroded diversity remains nonetheless difficult to determine.

In addition to high sequence polymorphism in regions coding for antigen-binding sites (ABS) in MHC genes, an intriguing observation is that sequences are often not more similar within species than between closely related species (Klein, 1986, 1987; Nei *et al.*, 1997; Klein *et al.*, 1998; van Oosterhout, 2009). Understanding this pattern requires to assess the influence of four different mechanisms susceptible to influence sequence similarity both within and between species: (i) prespeciation gene duplication followed by divergent selection; (ii) concerted evolution; (iii) long-lasting balancing selection maintaining ancestral trans-species polymorphism (TSP); and (iv) parallel evolution.

Gene duplication, frequent in MHC evolution, is an important mechanism that may lead to higher similarity between MHC sequences of different species than within species. Indeed, when duplication predates speciation and each gene duplicate is under divergent selection across long-evolutionary timescales (which can be traced across orders, i.e. gene orthology), then higher sequence similarity is expected within loci rather than between the duplicated loci, regardless of species identity (Nei *et al.*, 1997; Gu & Nei, 1999; Nei & Rooney, 2005).

Concerted evolution is defined as the process which leads to higher sequence similarity between sequences from paralogous loci within one species than between sequences of orthologous loci of different species, even though gene duplication occurred prior to speciation (Liao, 1999). The main mechanism driving concerted evolution is frequent recombination and gene conversion events between closely related loci, which results in frequent genetic exchange between loci (Hess & Edwards, 2002; Burri *et al.*, 2008, 2010). In order for the divergence of duplicated loci to occur, the rate of concerted evolution needs to be low. Therefore, phylogenetic reconstruction of MHC genes that has undergone an ancient prespeciation duplication event with low rates of concerted evolution should result in gene clustering by gene loci rather than by species (Nei *et al.*, 1997; Gu & Nei, 1999; Nei & Rooney, 2005). In contrast, high concerted evolution will disrupt the divergence of duplicated loci and phylogenetic reconstruction of MHC sequences should then cluster by species or genus (Nei *et al.*, 1997; Hess & Edwards, 2002; Nei & Rooney, 2005; Burri *et al.*, 2008, 2010).

Two additional mechanisms can generate high sequence similarities between species. First, the TSP hypothesis predicts that ancient allele lineages are maintained beyond speciation events by strong pathogen-mediated balancing selection at ABS (Klein, 1987). If pathogen-mediated balancing selection maintaining TSP is the main driver of MHC sequence similarity between species, then the genealogy of ABS sequences should be strongly intermixed, independently of species, following phylogenetic reconstruction. The second mechanism is parallel evolution, which occurs when similar pathogen selection pressures between species (e.g. sympatric species) select for functionally similar MHC alleles (Yeager & Hughes, 1999). Under parallel evolution, sympatric species exposed to similar pathogens are expected to have more sequence similarity at ABS codons under similar pathogen-mediated selection than allopatric species (Yeager & Hughes, 1999). In this case, ABS codon similarity occurs independently of balancing selection maintaining ancestral lineages and, therefore, allelic genealogy of codons under positive selection should cluster allelic lineages from species that share the same habitats, independently of phylogenetic inertia. However, phylogenetic analyses from the highly polymorphic short sequences of ABS MHC exons, with high rates of concerted evolution and positive selection, biases phylogenetic interpretation (Schierup & Hein, 2000). Alternative methods to phylogenetic reconstruction of MHC ABS sequences are therefore needed to disentangle mechanisms of MHC evolution (Edwards *et al.*, 1999; Kriener *et al.*, 2001; Burri *et al.*, 2008, 2010; Lenz *et al.*, 2013).

Here, we study the evolution of the MHC by characterizing the MHC Class I exon 3, MHC Class IIB exon 2 and MHC Class IIB exon 3 for all six extant flamingo species. Although we know little about how pathogen load varies between and within flamingo species, it is reasonable to expect historical contrasting differences in pathogen communities across the two sides of the Atlantic Ocean. Furthermore, dispersal limitation and autocorrelation of environmental variables between closely situated habitats is predicted to reduce differences in pathogen communities (Steinitz *et al.*, 2006; Soininen *et al.*, 2007). The latter therefore implies that there should be similar pathogen-mediated selection pressure between sympatric flamingo species. An important beneficial feature of the flamingo model in respect to the understanding of MHC evolution is that the phylogeny of the flamingo species does not match with their cross-Atlantic geographical distribution (Torres *et al.*, 2014; Fig. 1). Flamingos therefore present an ideal model to test hypotheses of MHC macro-evolution by investigating MHC sequence similarity across sympatric and allopatric flamingo species. Examining the two main classes of the classical MHC molecules allowed us to investigate the MHC sequence similarity of loci which may be under different selection pressures. Using these markers, we simultaneously consider all of the four aforementioned mechanisms (divergent selection of ancient gene duplicates, concerted evolution, balancing selection maintaining TSP and parallel evolution) to account for the high level of sequence

similarity observed between all of the six species. In addition to traditional phylogenetic analyses, we complemented our analyses using: (i) the comparison of highly polymorphic regions with more conserved regions of the MHC as this has been shown to reveal gene orthology independently from the effect of pathogen-mediated balancing selection (Edwards *et al.*, 1999; Kriener *et al.*, 2001; Burri *et al.*, 2008, 2010); (ii) the comparison of the codon structure of amino acids under strong positive selection. Indeed, we can exploit the redundancy of the genetic code to draw different predictions of codon structure under balancing selection maintaining TSP and parallel evolution. If sequence similarity is due to balancing selection maintaining TSP, amino acid sites under strong positive selection should be characterized by fewer synonymous codons (due to their common ancestry) than under parallel evolution. Indeed under balancing selection maintaining TSP, amino acids shared between alleles should present a high frequency of identical codons regardless of species sympatry. In contrast, under parallel evolution, as a result of codon redundancy in the genetic code, the proportion of identical codons for amino acids shared between species should be the same, regardless of whether codons are under positive selection or not (Lundberg & McDevitt, 1992; Lenz *et al.*, 2013).

Materials and methods

Model organism and samples

The six extant flamingo species appear to form a single genus distributed into two distinct clades (Torres *et al.*, 2014). One subclade includes an old world species, the greater flamingo (*Phoenicopterus roseus*), along with two new world species, the closely related American flamingo (*Phoenicopterus ruber*) and the Chilean flamingo (*Phoenicopterus chilensis*). The second subclade includes the remaining old world species, the lesser flamingo (*Phoeniconaias minor*), along with the other two new world species, the Andean flamingo (*Phoenicoparrus andinus*) and James's flamingo (*Phoenicoparrus jamesi*) (Torres *et al.*, 2014).

For greater flamingos, we used 127 blood and feather samples collected from chicks between 1998 and 2009 from four breeding colonies which cover the distribution of the species across the Mediterranean basin (Table 1): Garaet Ezzemoul, northern Algeria (35°53′N, 06°30′E); Camargue, southern France (43°25′N, 04°38′E); Fuente de Piedra, southern Spain (37°06′N, 04°45′W); and finally, Gediz delta, western Turkey (38°32′N, 26°52′E). A fresh liver sample was also collected and frozen from a deceased adult greater flamingo in the Camargue. In addition to the greater flamingos sampled, three individuals from each of the other five extant species of flamingos were sampled (Table 1). Blood samples were collected from three captive American flamingos and three captive Chilean flamingos, at the 'Parc des Oiseaux' (Villars-les-Dombes, France). Three feather samples from lesser flamingos were collected on the ground at Lake Nakuru in Kenya (02°22′S, 36°05′E). Four feathers were collected on the ground at the Salar de Surire (18°51′S, 69°00′W) in Chile, two belonging to two Andean flamingos, and two to two James's flamingos. Finally, one feather from an Andean flamingo and one from a James's flamingo were collected from captive

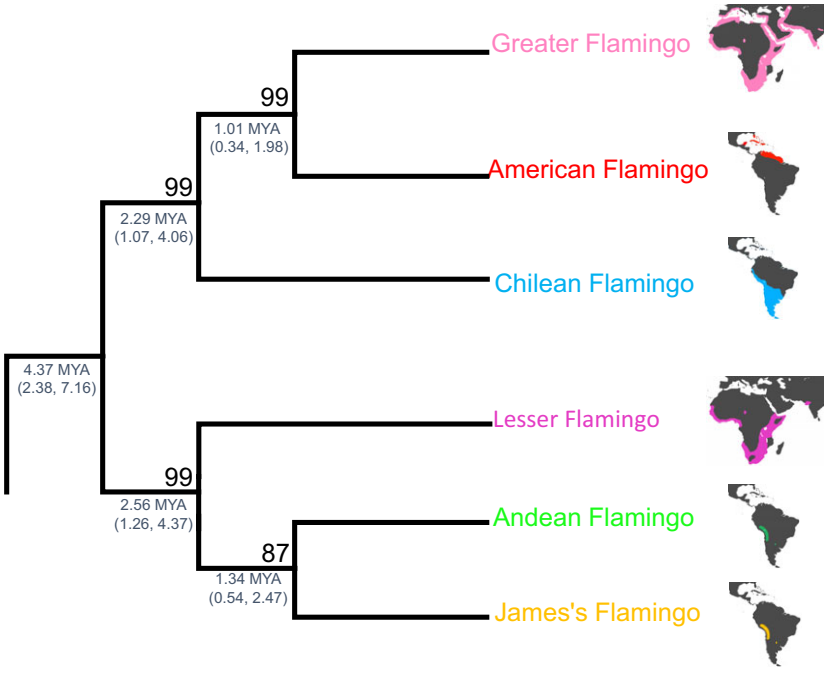

Fig. 1 Cladogram showing the relationships of the extant flamingo species adapted from Torres *et al.* (2014) (copyright pending) and the approximate geographical distribution of each flamingo species. The flamingos fall out in two subclades with high support, which is not congruent with their geographic distribution. Bootstrap support values (maximum likelihood) are shown above each node and ages for divergences (95% confidence intervals) within Phoenicopteridae based on 12 loci with two fossil calibrations are shown below each node (Torres *et al.*, 2014).

Table 1 Summary of samples used for MHC Class I exon 3 and MHC Class IIB exon 2 analysis.

Species	Population	Year of sampling	Type of sample	*N*	Nb. of PCR replicates	Nb. of PCRs	Mean read coverage of MHC Class I exon 3	Mean read coverage of MHC Class IIB exon 2
Greater flamingo (*Phoenicopterus roseus*)	Algeria	2008 & 2009	Feather	31	1 Duplicate	32	1535.59 (± 80.11)	1368.03 (± 130.99)
	France	1998	Blood	32	1 Duplicate	33	1572.7 (± 90.33)	972.33 (± 70.93)
	Spain	2007	Blood	32	1 Duplicate	33	1534.64 (± 69.14)	1115.39 (± 80.36)
	Turkey	2007 & 2009	Blood (2007) & feather (2009)	31	1 Duplicate	32	1463.78 (± 79.77)	1016.94 (± 111.09)
	France* (reference sample)	2008	Blood	1	1 Duplicate	2	940 (± 128)	671 (± 171)
	France* (liver sample)	2010	Liver	1	Triplicate of cDNA and gDNA	6	1895.67 (± 195.45)	1277 (± 170.8)
American flamingo (*Phoenicopterus ruber*)	Cuba	2010	Blood (zoo samples)	3	1 Duplicate	4	1846.75 (± 108.48)	967.25 (± 352.47)
Chilean flamingo (*Phoenicopterus chilensis*)	Unknown	2010	Blood (zoo samples)	3	1 Duplicate	4	1766 (± 240.52)	1138 (± 180.95)
Lesser flamingo (*Phoeniconaias minor*)	Kenya	2010	Feather	3†	1 Duplicate	4	841.5 (± 350.65)	492.5 (± 347.6)
Andean flamingo (*Phoenicoparrus andinus*)	Chile	2010	Feather (including one zoo sample)	3	1 Duplicate	4	1680 (± 750.39)	1138.25 (± 626.67)
James's flamingo (*Phoenicoparrus jamesi*)	Chile	2010	Feather (including one zoo sample)	3	1 Duplicate	4	1128.25 (± 432.95)	728 (± 296.65)
Total		1998–2010		143	11	158	1524.22 (42.02)	1089.08 (± 48)

N, number of individuals. Mean read coverage is the mean number of sequences passing quality filtering criteria according to the standard amplicon pipeline from 454 Genome Sequencer FLX System Software. Values in brackets are standard errors of the mean.
*Samples used to optimize primers.
†Identical MHC genotypes between the three feather samples would suggest that all samples came from the same individual.

individuals at the Wildfowl & Wetlands Trust (Slimbridge, UK). Blood samples were obtained from punctuation of the brachial vein and stored in preservation buffer (Seutin *et al.*, 1991), whereas feathers were stored in silica gel (Sigma , St. Louis, MO, USA).

We identified the species of the feather samples found on the ground in Kenya and Chile by PCR amplifying and Sanger sequencing a 722-bp part of the cytochrome oxydase domain I (COI), (primers: 'Pr-COIa-F' 5′-ACGCTTCAACACTCAGCCAT-3′, and 'Pr-COIa-R' 5′-TAATTCCAAAGCCTGGTAGG-3′) using identical methods as described by Geraci *et al.* (2012). The resulting sequences were BLAST searched in the GenBank database (99–100% of identity on 98–100% of coverage). The mitochondrial COI marker has previously been demonstrated to fully resolve flamingo species (Torres *et al.*, 2014).

DNA and RNA extraction

Genomic DNA from blood, feather and liver samples was extracted using a standard phenol–chloroform method, and quality was UV assessed (Spectramax plus 384, Molecular devices) before normalization at ca 50 ng μL^{-1}. Total RNA was extracted from the liver sample, using the QIAGEN RNeasy Animal Tissue kit following the manufacturer's protocol, and reverse transcription to complementary DNA (cDNA) was performed with the iScript select cDNA synthesis kit (BIO-RAD) and oligo(dT) primer.

Characterization of MHC Class I exon 3, MHC Class IIB exon 2 and MHC Class IIB exon 3

For MHC Class I, we chose to target exon 3 of the MHC Class I as it is the most frequently used and the better known of the two highly polymorphic regions of MHC Class I (the other being exon 2) known to code for antigen-presenting regions (Alcaide *et al.*, 2009). For MHC Class IIB, we chose to target to exon 2 and 3. The former is a highly polymorphic region which codes for antigen-presenting regions, whereas the latter is a more conserved region of the MHC which codes for the CD4-binding site involved in T lymphocyte binding. Inferring gene duplication based on highly polymorphic regions of the MHC from phylogenetic reconstruction

can be challenging because the effects of intense pathogen-mediated selection will mask gene duplication history (Edwards *et al.*, 1999; Wittzell *et al.*, 1999; Kriener *et al.*, 2001; Burri *et al.*, 2008, 2010). Using exon 3 of the MHC Class IIB, which is a more conserved region not under pathogen-mediated balancing selection, can therefore help to infer duplication history of MHC Class IIB, as has been achieved in owls (Burri *et al.*, 2008, 2010). See the supporting information for detailed methods of MHC Class I exon 3 and MHC Class IIB (intron 1, exon 2, intron 2 and exon 3) characterization (using previously published primers and traditional cloning of amplicons), PCR conditions and design of flamingo-specific primers. Clones of MHC Class I exon 3 (48 clones), MHC Class IIB exon 2 (24 clones) and MHC Class IIB exon 3 (48 clones) from two independent PCRs for each species of flamingo were sent for Sanger sequencing. Sequences that were identical between two independent PCRs were assigned as alleles.

Barcoded primer design, amplicon pooling and 454 GS-FLX Titanium pyrosequencing

We chose to target 204–207 bp of exon 3 of MHC Class I and 165 bp of exon 2 of MHC Class IIB for 454 Titanium pyrosequencing. Primers FLACI-EX3-F and FLACI-EX3-R for exon 3 MHC Class I, and FLA-CII-EX2-R and FLACII-EX2-R for exon 2 MHC Class IIB (see supporting information for primer sequence) were modified to include a 7-bp barcode and the 30-bp adapters at the 5′ end of the primers for the 454 Titanium sequencing reagent series as described in Galan *et al.* (2012). Each barcode differed from one another by at least 3 bp to minimize the risk of assignment error due to PCR or sequencing errors in the barcode. As summarized in Table 1, for each MHC marker (MHC Class I exon 3 and MHC Class IIB exon 2) we pooled 158 amplicons which included the following: several internal controls of replicate amplicons from a random sample from each population; triplicate amplicons from the genomic DNA of the liver sample; triplicate amplicons from the cDNA of the liver sample; the duplicate of the greater flamingo sample sequenced by Sanger method during the optimization of the primers; and a duplicate from a random sample for each of the remaining five species of flamingos. For each MHC marker, amplification was confirmed by the visualization of 3 μL of PCR products on a 1.5% agarose gel and all 158 amplicons were pooled in equal proportions. Five microlitre of each of the two pools was then visualized once again on a 1.5% agarose gel. Samples were subsequently processed by Beckman Coulter Genomics (Danvers, MA, USA) which included eliminating non-specific PCR products by running the two pools a microfluidic electrophoresis Pippin Prep (Sage Science, Beverly, MA, USA) and selection fragments of the expected size (± 50 base pairs). In addition following emulsion PCR (emPCR), the two pools were sequenced on a 454 Genome Sequencer FLX (Roche, Basel, Switzerland) on two independent Titanium picotiter plates. Only reads that passed the initial GS run processor filtering were used in subsequent analyses. Details of assigning barcoded reads to individuals using the SEquence Sorter & AMplicon Explorer (SESAME) software (Meglécz *et al.*, 2011), allele validation, allele amplification efficiencies and minimum number of reads per amplicon are presented in the supporting information and followed a protocol adapted from Sommer *et al.* (2013).

Positive selection

We tested for positive selection of MHC Class I exon 3, MHC Class IIB exon 2 and MHC Class IIB exon 3 alleles by testing for an excess of nonsynonymous substitutions (d_N) over synonymous substitutions (d_S) in amino acid sites ($\omega = d_N/d_S$ ratio) using CODEML, which is included in the PAML 4.7 package (Yang, 2007). We investigated the following models: M1a (nearly neutral), M2a (positive selection), M7 ($\beta : \omega$ ratio varies among sites according to a β distribution) and M8 ($\beta + \omega$: similar to M7 but with an additional site class that allows $\omega > 1$). We built a phylogenetic tree for CODEML analysis using maximum-likelihood (ML) estimation of phylogenetic relationships in PHYML 3.0 (Guindon *et al.*, 2010) with the substitution model with the most support as estimated by JMODELTEST 2.1.4 (Posada, 2008). Both M2a and M8 assume that amino acid sites may undergo positive selection, whereas their respective counterpart M1a and M7 consider a neutral scenario. Models M7 and M8 are thought to perform better under the occurrence of frequent recombination than M1a and M2a models implemented in CODEML but are thought to be less conservative (Anisimova *et al.*, 2003; Yang, 2007). For comparative purposes, we therefore perform a likelihood ratio test on both sets of models (M1a vs. M2a and M7 vs. M8). Because positive selection was detected, we then inferred which amino acid sites were under significant positive selection through 1100 permutations using the Bayes Empirical Bayes procedure (BEB). This analysis therefore precludes the need to assume that sites that are ABS, the sites predicted to be under balancing selection, are the same between chickens (the only birds for which ABS are known for MHC Class I exon 3) or mammals (for MHC Class IIB exon 2) and flamingos. The sequence of one MHC Class I exon 3 allele, which was found to have a premature stop codon (see Results), was excluded from this analysis.

Recombination analyses

As recently recommended, we used a set of several methods to detect signals of recombination (Mansai & Innan, 2010; Martin *et al.*, 2011). We tested for a correlation between distance and linkage disequilibrium, which is consistent with evidence of recombination, by running PERMUTE (implemented within the OMEGAMAP package; Wilson & McVean, 2006). Further evidence of recombination in the alignment was tested using the RDP3 package (Martin *et al.*, 2010) to apply the methods RDP (Martin & Rybicki, 2000), MAXCHI (Smith, 1992), CHIMAERA (Posada & Crandall, 2001), GENECONV (Padidam *et al.*, 1999) and SISCAN (Gibbs *et al.*, 2000) to detect recombination breaking points.

Phylogenetic analyses

The phylogenetic relationship between MHC Class I exon 3, MHC Class IIB exon 2 and MHC Class IIB exon 3 of flamingo species was visualized using the software SPLITS TREE 4.13.1 (Huson & Bryant, 2006) using Neighbor-Net networks based on Kimura's two-parameter model (which is the only available substitution model available for this analysis in Splits Tree). Neighbor-Net networks allow us to visualize the complex evolutionary histories of genes in the presence of reticulate events such as hybridization, horizontal gene transfer and/or recombination. They are therefore ideal to investigate relationships between MHC alleles whose evolution is characterized by gene duplication and loss as well as interlocus recombination events. However, the simple Kimura's two-parameter model is unlikely to be appropriate for the complex evolutionary history of the MHC and we therefore complemented phylogenetic analysis with Bayesian phylogenetic reconstructions, which was achieved using MrBayes 3.2.4 (Ronquist & Huelsenbeck, 2003) with the implemented substitution model that had the most support as estimated by JMODELTEST 2.1.4 (Posada, 2008). We partitioned the coding sequence, and each position was assumed to have different rates of evolution. We conducted two independent runs consisting four chains and 5×10^7 generations. We evaluated convergence by investigating the average standard deviation of split frequencies between runs and the potential scale reduction factor. Finally, posterior distributions were examined in TRACER 1.6 (Rambaut *et al.*, 2014) and the first half of the topologies was discarded as burn-in.

Burri *et al.* (2008), using the more conserved sequences of exon 3 of MHC Class IIB from 14 owl species, found that alleles clustered according to two gene lineages rather than species following phylogenetic reconstruction. Such clustering of genes may indicate early gene duplication and high sequence similarity between species by descent or recent parallel duplication followed by functional convergence. To distinguish these hypotheses, Burri *et al.* (2010) contrasted phylogenetic trees derived from third codon positions to those derived from full sequences. As third codon position of sequences can be considered nearly neutral, similar sequence topology between the two analyses suggests early duplication prior to speciation rather than recent duplication. We applied a similar strategy to investigate gene duplication history of the MHC Class IIB. For MHC Class IIB exon 3, we included sequences of avian species from the public domain database of NCBI and compared the Bayesian phylogenies of complete sequences with that of the Bayesian phylogenies of only the third codon position.

For MHC Class I exon 3 and MHC Class IIB exon 2, we first produced a Neighbor-Net network of full sequences and a second Neighbor-Net network using nucleotides that were under positive selection only, as these are the sites that are predicted to be ABS and to be under pathogen-mediated balancing selection (Bjorkman *et al.*, 1987). Finally, in order to compare flamingo MHC alleles to that of other avian species, we reconstructed a Bayesian phylogenetic tree for each loci using sequences of avian species from the NCBI database (see supporting information for list of species and GenBank accession numbers).

Codon usage analysis

Under balancing selection maintaining TSP, amino acid sites under strong positive selection should be characterized by fewer synonymous codons (due to their common ancestry) than under parallel evolution. Therefore, to distinguish between balancing selection maintaining TSP and parallel evolution, we assessed the proportion of identical codons encoding the same amino acids at positively selected sites between pairs of species following Lenz *et al.*'s (2013) study. Then, we assumed that under parallel evolution identical amino acids have the same proportion of identical codons at amino acid sites independently of them undergoing strong positive selection or not. We therefore drew codons with replacement from within the pool of all codons present in the complete exon of each species, to simulate new sequences synonymous to those observed at the peptide level. We measured codon similarity between species using these simulated data as described above, and we reiterated this process (simulation + measurement) 9999 times to generate the distribution of similarity measurements under the hypothesis of parallel evolution. Computing the proportion of simulated similarity values greater or equal to observed values, thereby provide the *P*-value of the test of parallel evolution. This analysis was implemented in R 3.0 (R Core Team, 2014) using functions provided by the package SEQINR (Charif & Lobry, 2007), which allows for straightforward sequence manipulations.

Results

Characterization of MHC Class I and MHC Class IIB

In total, 105 different MHC Class I exon 3 (accession numbers: KU054001–KU054105) and 136 MHC Class IIB exon 2 alleles (accession numbers: KU054106–KU054241; Table 2) were identified by 454 sequencing (see supporting information for details of 454 sequencing output) across all six species of flamingos (see Table 1 for coverage of each species). This included all 22 MHC Class I exon 3 and 15 MHC Class IIB exon 2 alleles found by cloning and Sanger sequencing (only sequences found in two independent PCRs were retained) from one sample of each flamingo species and the cDNA sample from the liver of a greater flamingo. However, for these samples, 15 MHC Class I exon 3 and 22 MHC Class IIB exon 2 alleles were identified by 454 sequencing that were not identified by cloning and Sanger sequencing. Allele calling repeatability was higher for MHC Class IIB exon 2 (100%) than for MHC Class I exon 3 (88%) (see supporting information for more details of allele calling repeatability). However, as in this study we are interested in sequence similarity and we did not perform any analysis based on allele frequency, allele dropout that is responsible for imperfect repeatability (Sommer *et al.*, 2013; Burri *et al.*, 2014) should not have any adverse effects on our results. We also identified 27 MHC Class IIB exon 3 alleles (accession numbers: KU053963–KU053985 & KU053997–KU054000) through traditional cloning and sequencing methods based on six genomic samples from each species of flamingos and a cDNA sample of greater flamingo. Note that MHC Class IIB exon 3 was not characterized using 454 sequencing. In addition, we also identified 11 sequences of intron 1 (accession numbers: KU053986–KU053996) and 23 sequences of intron 2 of MHC Class IIB (accession numbers: KU053963–KU053985). As flamingo introns showed little to no sequence similarity with other avian species, we focused on the functional markers (i.e. exons) of the MHC to infer MHC evolution (see supporting information for more details of MHC Class IIB introns).

Table 2 Summary statistics of MHC Class I exon 3 and MHC Class IIB exon 2 variation for the six species of flamingos sampled in this study.

Species	*N*	*m*	*S*	Nb. of alleles per individual	Nb. of private alleles
(a) MHC Class I exon 3					
Greater flamingo	128	86	89	3–11	67
American flamingo	3	13	60	6–8	1
Chilean flamingo	3	10	62	8–9	1
Lesser flamingo	3*	10	55	7–10	4
Andean flamingo	3	13	68	3–8	9
James's flamingo	3	11	65	4–6	4
Total	143	105	98	3–11	86
(b) MHC Class IIB exon 2					
Greater flamingo	128	113	68	1–4	106
American flamingo	3	10	46	3–4	7
Chilean flamingo	3	6	35	4	5
Lesser flamingo	3*	2	18	2	1
Andean flamingo	3	5	43	2–4	4
James's flamingo	3	8	48	2–4	5
Total	143	136	71	1–4	128
(c) MHC Class IIB exon 3†					
Greater flamingo	2	8	23	4	4
American flamingo	1	4	20	4	2
Chilean flamingo	1	3	13	3	3
Lesser flamingo	1	4	23	4	2
Andean flamingo	1	4	15	4	4
James's flamingo	1	4	20	4	3
Total	7	23	31	3–4	18

N, number of individuals; *m*, number of alleles; *S*, number of segregating sites.

*Identical MHC genotypes between the three feather samples would suggest that all samples came from the same individual.

†Data of MHC Class IIB exon 3 from cloning and sequencing of two individuals of greater flamingos (1 genomic DNA sample and 1 cDNA sample) and one individual from each of the other flamingo species.

The maximal number of alleles within a single individual was found in the greater flamingo with 11 MHC Class I exon 3 and 4 MHC Class IIB alleles (for both exon 2 and 3), which suggests at least six and two loci, respectively (Table 2). Results from our limited sampling of the other five flamingo species suggested a similar number of loci between species, with the maximum number of MHC Class I exon 3 alleles found to be ten for the lesser flamingo, nine for the Chilean flamingo, eight for the American and the Andean flamingos, and six for the James's flamingo (Table 2a). A maximum of four MHC Class IIB (for both exon 2 and 3) alleles were found within an individual for all flamingo species, except for the lesser flamingo that showed two alleles for exon 2 and three for exon 3 (Table 2b). As expected, the number of segregating sites was much lower for MHC Class IIB exon 3 than for MHC Class IIB exon 2 (Table 2c). The high divergence and number of alleles captured within individuals suggest that our primers efficiently amplified alleles across MHC loci (Figs 2 and 3).

All 105 MHC Class I exon 3 alleles and all of the 136 MHC Class IIB exon 2 are potentially functional except for one of the 105 MHC Class I exon 3 alleles (belonging to an Andean flamingo sample: SOSU11-05; accession number: KU054028), as no other alleles presented a premature stop codon. However, comparison between the genomic DNA and cDNA amplicons of the liver sample of the greater flamingo sampled in 2010 revealed that only five of eight MHC Class I exon 3 alleles identified in the genomic DNA amplicon were also detected in the cDNA sample. The latter suggests that, at least in the greater flamingo, a minimum of

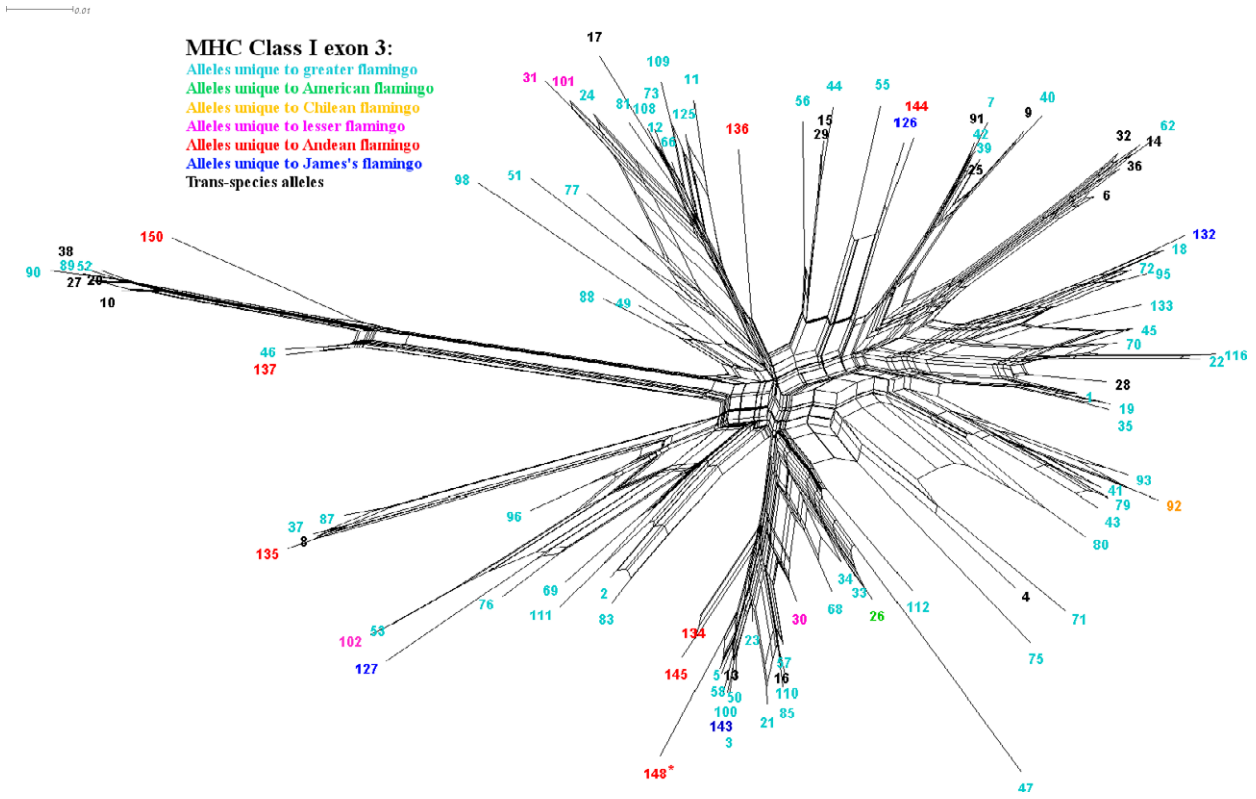

Fig. 2 The Neighbor-Net network constructed from the 105 MHC Class I exon 3 alleles (accession numbers: KU054001–KU054105) from the six flamingo species. Allele 148 (*) is a suspected pseudogene since it has a premature stop codon within the exon.

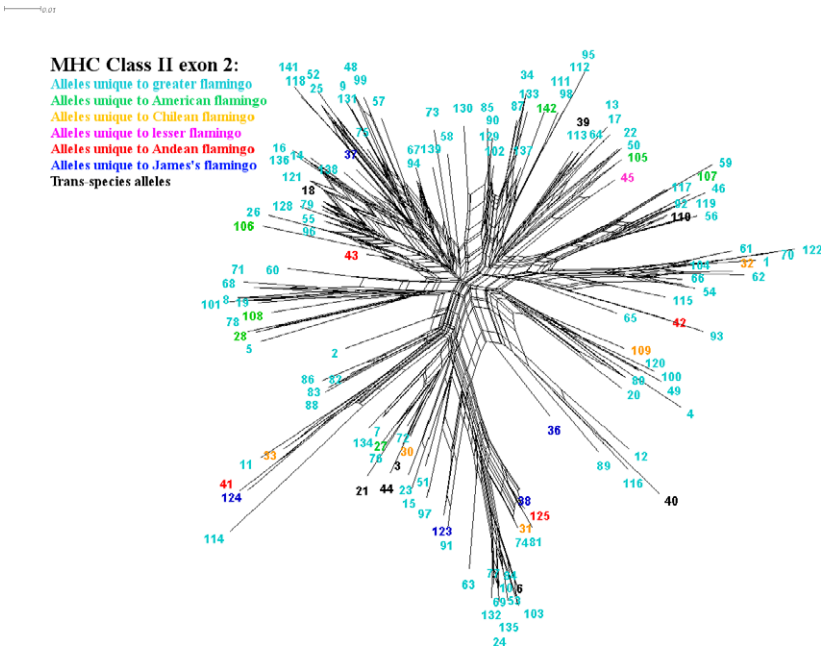

Fig. 3 The Neighbor-Net network constructed from the 136 MHC Class IIB exon 2 alleles (accession numbers: KU054106–KU054241) from the six flamingo species.

two MHC Class I exon 3 loci amplified may not be expressed although more extensive analyses are needed to confirm this. In contrast, all four MHC Class IIB exon 2 alleles that were identified in the genomic DNA amplicon were also detected in the cDNA sample, suggesting that both MHC Class IIB exon 2 loci amplified are expressed.

Evidence for positive selection

Table 3 summarizes the results of the ML analyses in CODEML. Fifty-one of the 70 amino acid positions (73%) of MHC Class I exon 3 were polymorphic, and five codon positions were found to be under significant positive selection by both M2a and M8 models (BEB > 0.95 posterior probabilities; Table 3a). Twenty of the 55 amino acid positions (36%) of MHC Class IIB exon 2 were polymorphic, and seven codon positions were found to be under significant positive selection by both M2a and M8 models (BEB > 0.95 posterior probabilities; Table 3b), with an additional codon found to be under significant positive selection by model M8. Finally, none of the codons of MHC Class IIB exon 3 were under significant positive selection and positive selection models were not better supported than neutral models (Table 3c). Therefore, we confirmed that exon 3 of MHC Class IIB was relatively conserved between flamingo species as found in other bird species and is therefore ideal to infer gene duplication history while controlling for positive selection (Burri *et al.*, 2008).

Evidence of recombination

We found significant evidence for recombination in all exons investigated. Indeed, there was a significant negative correlation between linkage disequilibrium rate estimates and physical distance between sites in MHC Class I exon 3, MHC Class IIB exon 2 and MHC Class IIB exon 3 (permutation test implemented in OMEGAMAP: $r = -0.08$, $P < 0.001$; $r = -0.10$, $P = 0.002$; and $r = -0.78$, $P = 0.001$, respectively). However, we could

Table 3 Log-likelihood values and parameter estimates of models testing for positive selection at the MHC Class I exon 3, MHC Class IIB exon 2 and MHC Class IIB exon 3 alleles of all six flamingo species.

Models	lnL	Estimates of parameters	Positively selected sites†
(a) MHC Class I exon 3			
M1a (neutral)	−2559.722	$p_0 = 0.955$, $p_1 = 0.045$* $\omega_0 = 0.051$, $\omega_1 = 1$	Not allowed
M2a (selection)	−2503.316	$p_0 = 0.937$, $p_1 = 0.058$, $p_2 = 0.004$ $\omega_0 = 0.071$, $\omega_1 = 1$, $\omega_2 = 5.151$	9G**, **15S****, **54E****, **57Q****, **58W****
M7 (β)	−2569.595	$p = 0.019$, $q = 0.132$	Not allowed
M8 (β and ω)	−2506.075	$p_0 = 0.996$, $p_1 = 0.004$, $p = 0.120$, $q = 0.845$, $\omega_s = 5.04$	9G**, **15S****, **54E****, **57Q****, **58W****
(b) MHC Class IIB exon 2			
M1a (neutral)	−3318.089	$p_0 = 0.909$, $p_1 = 0.091$ $\omega_0 = 0.024$, $\omega_1 = 1$	Not allowed
M2a (selection)	−3203.963	$p_0 = 0.816$, $p_1 = 0.164$, $p_2 = 0.020$ $\omega_0 = 0.030$, $\omega_1 = 1$, $\omega_2 = 4.716$	2Y**, **23Y****, 33D**, 43I**, **46D****, **47E****, 53T**, **54F****
M7 (β)	−3311.609	$p = 0.012$, $q = 0.077$	Not allowed
M8 (β and ω)	−3204.072	$p_0 = 0.984$, $p_1 = 0.016$, $p = 0.011$, $q = 0.035$, $\omega_s = 4.84$	2Y**, **23Y****, 33D**, 43I**, **46D****, **47E****, **50S***, 53T**, **54F****
(c) MHC Class IIB exon 3			
M1a (neutral)	−655.744	$p_0 = 0.828$, $p_1 = 0.171$ $\omega_0 = 0.008$, $\omega_1 = 1$	Not allowed
M2a (selection)	−655.744	$p_0 = 0.828$, $p_1 = 0.029$, $p_2 = 0.142$ $\omega_0 = 0.008$, $\omega_1 = 1$, $\omega_2 = 1$	None
M7 (β)	−655.748	$p = 0.010$, $q = 0.053$	Not allowed
M8 (β and ω)	−655.738	$p_0 = 0.977$, $p_1 = 0.023$, $p = 0.013$ $q = 0.074$, $\omega_s = 1.388$	None

Values p and q are parameters of the beta distribution. $\omega = d_N/d_S$. p_n is the proportion of sites falling into the ω_n.
Sites found to be under significant positive selection according to Bayes Empirical Bayes (BEB) at the 95% (*) and 99% (**) confidence interval level are also shown.
†Site positions for MHC Class I exon 3 sequences according to the 204–207 bp sequence positions (accession numbers: KU054001–KU054027 & KU054029–KU054105) and for MHC Class IIB exon 2 sequences according to the 163–165 bp sequence positions obtained from 454 sequencing (accession numbers: KU054106–KU054241). Site position for MHC Class IIB exon 3 sequences according to the 207 bp of exon 3 obtained through cloning and traditional sequencing (accession numbers: KU053963–KU053985 & KU053997–KU054000). The MHC Class I exon 3 sequence with a stop codon (accession number: KU054028) was excluded from the analysis. In bold are sites that are known to be antigen biding in chickens for MHC Class I exon 3 (Wallny *et al.*, 2006) and in humans for MHC Class IIB exon 2 (Brown *et al.*, 1993).

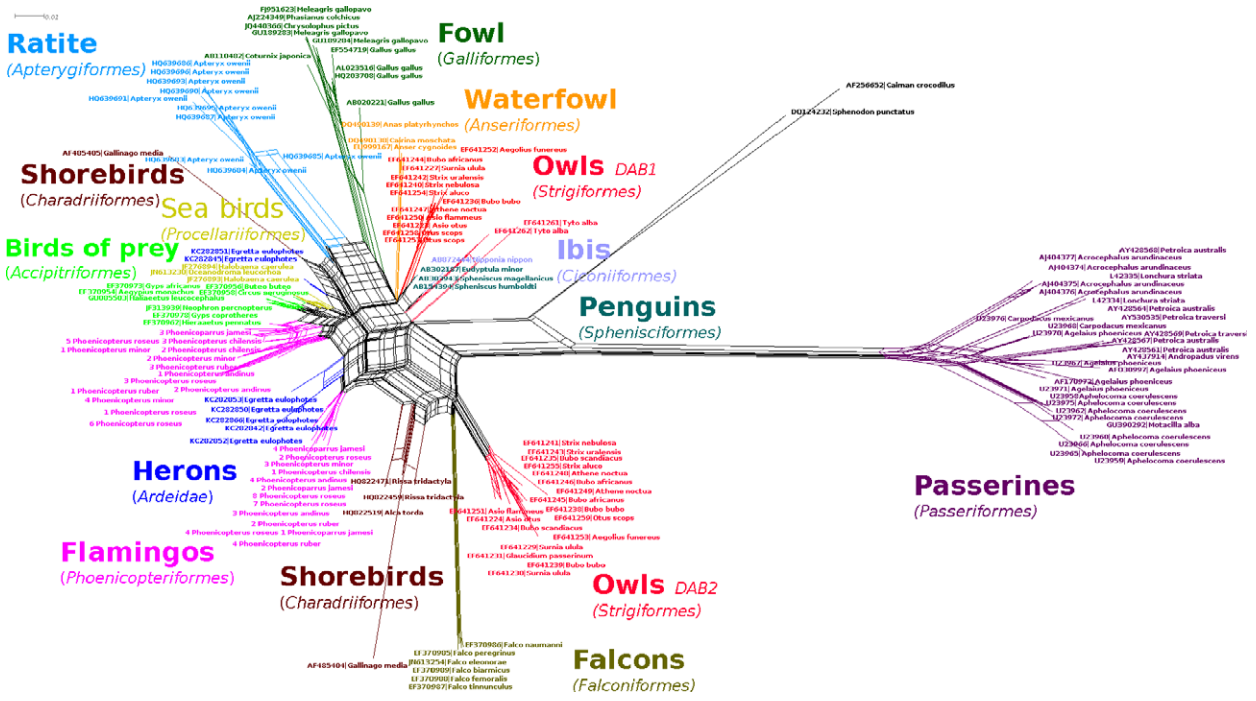

Fig. 4 The Neighbor-Net network constructed from the 27 MHC Class IIB exon three (accession numbers: KU053963–KU053985 & KU053997–KU054000) alleles from the six flamingo species and sequences isolated in other avian species (list of species and accession number within figure) from full sequences.

not detect the location of recombination breaking points for any of the MHC markers (RDP, MAXCHI, CHIMAERA, GENECONV and SISCAN).

Evidence of gene orthology masked by concerted evolution between flamingo species

Neighbor-Net network analysis of the full codon sequences of MHC Class IIB exon 3 showed that flamingo sequences tended to cluster into two lineages and each lineage tended to cluster closely (but not within) to the two orthologous genes previously identified in owls and charadriiformes (Burri *et al.*, 2008, 2010; Fig. 4). Furthermore, flamingo MHC Class IIB exon 3 sequences were segregated by distantly related species of birds of prey, petrels and herons (Fig. 4). Both clusters of flamingo sequences hold alleles from all species which is a strong indicator of gene orthology.

However, some flamingo individuals had more than three alleles within a cluster of alleles (Fig. 4), suggesting that sequences did not cluster into loci. Furthermore, Bayesian phylogenetic reconstruction of the full codon sequence of avian MHC Class IIB exon 3 sequences did not reveal the same two clusters of flamingo sequences identified by the Neighbor-Net network analysis and basal nodes had low statistical support with distantly related species of birds of prey, petrels and herons (Fig. 5a). Bayesian phylogenetic reconstruction of the third codon positions retained the two ancient lineages previously identified in owls (DAB1 and DAB2; Burri *et al.*, 2010) but sequences from flamingos tended to remain clustered together (although once again basal nodes had low statistical support; Fig. 5b).

Visual inspection of sequence alignment (Fig. 6) revealed that 5 of the 16-nt sites located near the 5′ end of the MHC Class IIB exon 3 sequence that showed strong signals of orthology in owls were polymorphic in flamingos (Burri *et al.*, 2008, 2010). Four of these 5-nt sites were in positions 13–16 and were identical to the polymorphisms found between *DAB1* and *DAB2* in owls. The latter tentatively suggests that the two lineages identified in flamingos may be the ancient duplicated avian *DAB1* and *DAB2* lineages, but the high rate of concerted evolution has homogenized these two lineages within flamingos and has prevented divergence.

Evidence for balancing selection maintaining TSP but not parallel evolution of antigen-presenting regions

The Neighbor-Net analyses show no clustering for both MHC Class I exon 3 allele lineages and MHC Class IIB exon 2 allele lineages into sympatric distribution of flamingo species (Figs 2 and 3). The results remained

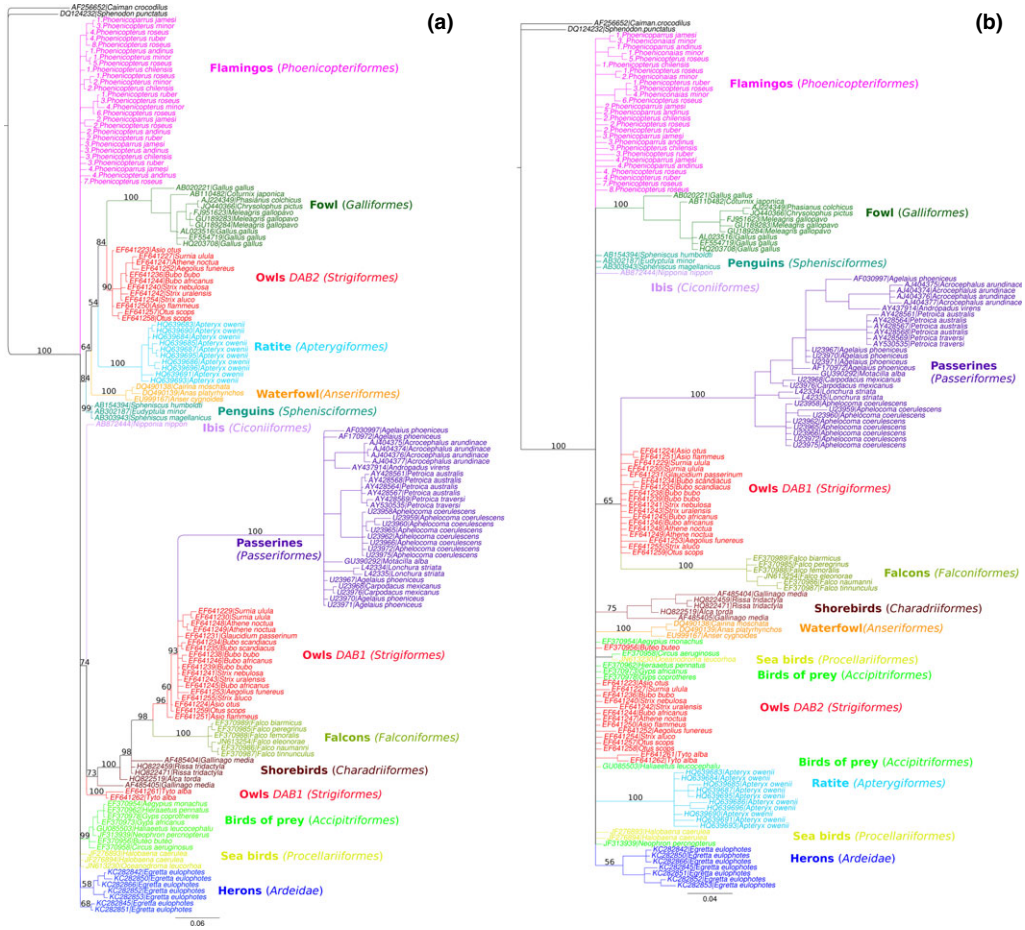

Fig. 5 Consensus Bayesian topology (50% majority consensus) constructed from the 27 MHC Class IIB exon three alleles (accession numbers: KU053963–KU053985 & KU053997–KU054000) from the six flamingo species and sequences isolated in other avian species (list of species and accession number within figure) from full sequences (a: GTR + G + I) and third codon positions (b, a: HKY + G). Only supports for avian orders and families are provided.

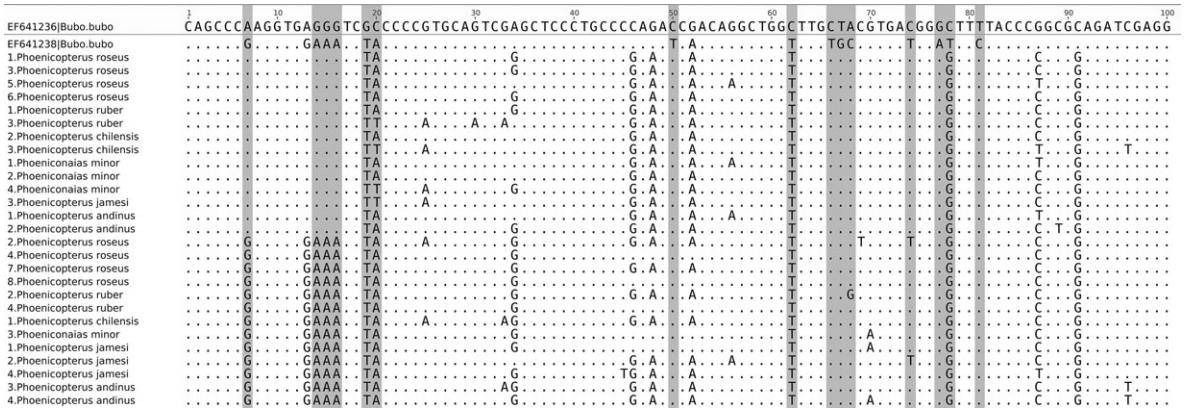

Fig. 6 Nucleotide sequence alignment of the first 100-nt sites of MHCIIB exon 3 of flamingo species in comparison with a DAB1 and DAB2 sequence of the Eurasian eagle owl (*Bubo bubo*). Sites that were identified as strong signals of locus signature by Burri *et al.* (2008) are highlighted in grey.

highly similar when Neighbor-Net analysis was carried out on nucleotide positions under positive selection only (Figs S3 and S4). Bayesian phylogenetic reconstruction of both ABS exons of flamingos and other distantly related avian species had low statistical support at basal nodes, although sequences from flamingos tended to cluster together (Fig. S5). We also found that 19 of the MHC Class I exon 3 of 105 alleles identified and 8 of 136 MHC Class IIB exon 2 alleles presented the same nucleotide sequences between flamingo species (Table 4). The latter supports balancing selection maintaining TSP because it is unlikely that long identical sequences may result from parallel evolution.

To formally investigate this possibility, we studied the similarity of codons encoding the same amino acid in positively selected regions between flamingo species. We found an average (± SD) of codon similarity between species of 89.22% (± 1.69) for MHC Class I exon 3 and of 99.70% (± 1.12) for MHC Class IIB exon 2 (see Table 4 for pairwise comparison between all six flamingo species). These values were higher than all similarity values obtained while simulating the hypothesis of parallel evolution, therefore rejecting this hypothesis in favour of balancing selection maintaining TSP (all $P < 0.0001$; see Fig. S6).

Discussion

In this study, we first characterized the MHC Class I exon 3, MHC Class IIB exon 2 and MHC Class IIB exon 3 of all flamingo species. We then used the relatively conserved region of MHC Class IIB exon 3 to infer duplication orthology. Finally, we used the contrasting patterns of phylogeny and sympatric distribution of the six extant flamingo species to investigate the relative role of parallel evolution and balancing selection maintaining TSP in generating high sequence similarity at the ABS sites of the MHC between closely related species. We found evidence of up to six MHC Class I loci and two MHC Class IIB loci in flamingo species. Comparison between Neighbor-Net networks of full codon sequences and sequence alignments of the conserved region of the MHC Class IIB exon 3 suggests that gene duplication of MHC Class IIB occurred prior to flamingo speciation. However, high rates of interlocus concerted evolution are preventing divergence of loci even in this conserved region of the MHC. Finally, sequence analysis of alleles of ABS coding regions of the MHC Class I and MHC Class IIB strongly supports balancing selection maintaining TSP rather than parallel evolution at these sites.

Characterization of the MHC in flamingo species

We found a variable number of gene duplications between MHC Class I and MHC Class IIB in flamingos suggesting that patterns of gene duplication cannot be generalized between MHC classes. Using 454 sequencing, we identified up to 11 MHC Class I exon 3 alleles in greater flamingos and four MHC Class IIB exon 2 alleles within an individual in all flamingo species, indicating the presence of at least six and two different MHC loci, respectively. Similarly to the blue petrel (*Halobaena caerulea*) both MHC Class IIB loci were expressed but not all of six MHC Class I loci were (in the greater flamingo, at least three of six MHC Class I exon 3 loci were expressed compared to at least five of eight MHC Class I exon 3 loci for the blue petrel; Strandh *et al.*, 2011).

Repeatability of MHC Class I exon 3 allele calling was below 99.9% between replicates in our study as reported in other studies which amplify a large number of MHC loci in a single PCR (Galan *et al.*, 2010; Zagalska-Neubauer *et al.*, 2010; Sepil *et al.*, 2012; Strandh *et al.*, 2012). We therefore cannot exclude the possibility that more extensive analyses of MHC Class I exon 3 in flamingo species may reveal more than the six loci we have identified in this study. Indeed, MHC characterization by PCR may selectively amplify particular genes in multigene families (Wagner, 2005; Sommer *et al.*, 2013; Burri *et al.*, 2014) biasing our estimated absolute values of MHC loci. In contrast, for MHC Class IIB exon 2, we are confident that there are only two duplicated loci based on the following lines of evidence. First, repeatability of allele calling for MHC Class IIB exon 2 was 100%, with no individual found to have more than four alleles. Second, the independent amplification, cloning and sequencing of MHC Class IIB exon 3 also revealed no more than four alleles per individual for each individual per flamingo species. To unravel the

Table 4 Matrix of the proportion of identical amino acids and the number of identical full sequences between species (in brackets) for MHC Class I exon 3 (below diagonal) and Class IIB exon 2 (above diagonal).

	Greater, %	American, %	Chilean, %	Lesser, %	Andean, %	James's, %
Greater		97.52 (3)	98.82 (1)	100 (1)	97.76 (0)	97.37 (2)
American	90.56 (12)		98.20 (0)	100 (0)	97.77 (0)	96.72 (0)
Chilean	90.99 (9)	89.74 (5)		100 (0)	99.11 (0)	98.27 (0)
Lesser	88.92 (6)	87.80 (4)	88.02 (3)		100 (0)	97.37 (0)
Andean	88.48 (4)	87.10 (2)	88.22 (2)	86.23 (2)		98.70 (1)
James's	91.86 (7)	91.24 (3)	91.34 (2)	89.18 (2)	88.57 (2)	

complex history of MHC duplication in birds, a PCR-based approach such as the one we used here should ideally be complemented by Southern blots (Westerdahl *et al.*, 1999; Wittzell *et al.*, 1999; Silva & Edwards, 2009) and/or genome characterization (Balakrishnan *et al.*, 2010) as these labour-intensive methods allow for higher confidence in the inference of MHC loci number.

The number of duplication events for MHC Class I in flamingos is larger than those reported for a large number of nonpasserine species (1–4 loci; fowl (*Gallus gallus*), Kaufman *et al.*, 1999; mallard duck (*Anas platyrhynchos*), Moon *et al.*, 2005; birds of prey, Alcaide *et al.*, 2009; turkey (*Meloeagris gallopavo*), Chaves *et al.*, 2010; red-billed gull (*Larus scopulinus*), Cloutier *et al.*, 2011) but relatively similar to what has been found in blue petrels (8 loci; Strandh *et al.*, 2011) and red knots (*Calidris canutus*) (6 loci; Buehler *et al.*, 2013). For the MHC Class IIB, the number of loci we identified is in line with most nonpasserine species (1–3 loci; Kaufman *et al.*, 1999; Wittzell *et al.*, 1999; Alcaide *et al.*, 2007, 2014; Strand *et al.*, 2007; Burri *et al.*, 2008; Hughes *et al.*, 2008; Kikkawa *et al.*, 2009; Juola & Dearborn, 2012; Strandh *et al.*, 2012; Promerová *et al.*, 2013). Overall, flamingos present a complex gene duplication history and an associated loss of gene function, with substantial differences with other avian groups.

Evidence of gene orthology masked by concerted evolution between flamingo species

As found in most vertebrates, there was no evidence that the MHC Class IIB exon 3, which codes for the more conserved CD4-binding site, was under significant positive selection. This exon was therefore relatively more conserved than the other exons, which precludes this region from the strong biases during phylogenetic analyses resulting from strong signals of positive selection (Schierup & Hein, 2000). However, we did detect strong signals of recombination and/or conversation events, which may significantly bias terminal branch length during phylogenetic analyses (Schierup & Hein, 2000). That we have failed to identify the locations of recombination breaking points may be explained by a lack of statistical power of the tests we used (Mansai & Innan, 2010), especially in the light of the small size of fragments used in this study (Martin *et al.*, 2010). We could not assign alleles to loci; therefore, significant recombination events found in this study may be the result of interlocus and/or intralocus recombination (and cannot be interpreted as conclusive evidence of concerted evolution). Therefore, conclusions of gene orthology from phylogenies should be considered with caution and in conjunction with other lines of evidence.

Under a scenario of divergent selection, we would expect phylogenetic reconstruction to segregate sequences into loci, even when considering the short fragment lengths used in this study (see Burri *et al.*, 2010 for simulations). On the other hand, under a scenario of high rates of concerted evolution disrupting divergent selection as a result of frequent interlocus genetic exchange (via recombination and gene conversion), we would expect phylogenetic reconstruction to cluster sequences by species or genus. Bayesian phylogenetic reconstruction of full codon MHC Class IIB exon 3 sequences revealed low statistical support of basal nodes with no evidence of alleles segregating into separate lineages. Furthermore, all flamingo sequences tended to cluster together. Bayesian phylogenetic reconstruction of MHC sequences in flamingo species is therefore consistent with high rates of concerted evolution disrupting divergence of orthologs. We are nonetheless confident that MHC Class IIB duplication occurred prior to flamingo speciation around 3–6 million years ago (Torres *et al.*, 2014), based on the following lines of evidence. First, we found the same number of gene duplications of MHC Class IIB across all six extant flamingo species. Second, allele sequences tended to cluster in two lineages with at least one sequence from each species present in each cluster of the full codon Neighbor-Net network, suggesting slight divergence between loci. However, close inspection of the full codon Neighbor-Net network revealed that some individuals had up to three alleles within a single cluster, which is further evidence of concerted evolution homogenizing the two MHC loci in flamingo species.

Although phylogenetic reconstruction of MHC Class IIB exon 3 of third codon positions did retain the two ancient lineages in owls identified by Burri *et al.* (2010), we found no evidence of a such pattern for other avian species. We therefore cannot exclude the possibility that any pattern in allele clustering between flamingos and other distantly related birds may result from functional convergence rather than the maintenance of orthologous genes across avian orders. Elucidation of the functional similarities between exon 3 of flamingos with that of distantly related species would require analysis of the structure of the MHC Class IIB protein (Burri *et al.*, 2010). Furthermore, characterization of the MHC in grebes, the sister clade of flamingos (van Tuinen *et al.*, 2001; Chubb, 2004; Ericson *et al.*, 2006; Hackett *et al.*, 2008; McCormack *et al.*, 2013), would help to disassociate these hypotheses and further refine the duplication history of MHC Class IIB in flamingos. Nonetheless, we believe that functional convergence is unlikely because none of the positions segregating allele lineages in MHC Class II exon 3 were found to be under positive selection. Furthermore, identical polymorphisms (at all three codon positions), which were found to be strong signals of gene orthology in distantly related species of owls, were identified in flamingo sequences tentatively suggesting that the

two gene lineages identified in flamingos may be of the ancient orthologous *DAB1* and *DAB2* genes.

Regardless of whether flamingo orthologous genes belong to the *DAB1* and *DAB2* avian lineages, our results therefore suggest that the divergence of orthologous genes following early gene duplication across avian orders, as observed in mammals, owls and charadriiformes, appears to be the exception rather than the rule in birds (see also Edwards *et al.*, 1995; Wittzell *et al.*, 1999; Miller & Lambert, 2004; Kikkawa *et al.*, 2005). The latter is explained by the higher rates of concerted evolution homogenizing loci in most birds (Hess & Edwards, 2002), including in the more conserved region of the exon 3 MHC Class IIB.

Evidence of balancing selection maintaining TSP but not of parallel evolution in the MHC of flamingos

Phylogenetic analyses of MHC alleles at MHC Class I exon 3 and MHC Class IIB exon 2 revealed no clustering by flamingo species distribution which argues for balancing selection maintaining TSP and against parallel evolution. However once again, phylogenetic analyses from short fragments, with high rates of concerted evolution and positive selection, biases phylogenetic interpretation (Schierup & Hein, 2000). Other lines of evidence must therefore be used to disassociate between balancing selection maintaining TSP and parallel evolution.

Despite our limited sampling of flamingo species other than *P. roseus*, we found a high number of MHC alleles that were identical between flamingo species. We confirmed that balancing selection maintaining TSP was more likely than parallel evolution by also performing an analysis of codon usage that could be employed irrespective of the relationship between relatedness and habitat distribution. This analysis was based on sites under positive selection, which we detected both at exon 3 of MHC Class I and exon 2 of MHC Class IIB, as found in most vertebrate. We found that amino acid sites under strong positive selection at the ABS of both MHC Class I and MHC Class IIB showed significantly higher codon similarity of identical amino acid sites than expected under parallel evolution. As in most MHC studies not conducted on model organisms, we could not assign alleles to loci in our study which may bias the latter result. Nonetheless, pooling loci to identify sites under positive selection is likely to be conservative rather than leading to false positives, as different selection patterns between the pooled sites should reduce the overall selection signal.

So far, evidence of parallel evolution influencing MHC evolution is limited. It has been suggested to be responsible for high sequence similarity between sympatric species of racoons and skunks (Srithayakumar *et al.*, 2012). In addition, Schwensow *et al.* (2010) found that two alleles with high sequence similarity between two sympatric species of lemurs were negatively associated with locally prevalent *Ascaris* infection, which may also suggest parallel evolution. However, both of these examples are limited by the fact that in these biological systems relatedness is confounded with ecological similarities. In flamingos, sympatric species are more distantly related and some closely related species (Fig. 1) are cross-Atlantic allopatric species which inhabit contrasting habitats. We therefore expect flamingo species to be an ideal model to observe parallel evolution in the MHC while controlling for the effects of common ancestry. However, and consistent with Lenz *et al.*'s (2013) study on sticklebacks, our results support ancestral TSPs that pre-date the first speciation event of the ancestor of extant flamingos (3–6 Ma; Torres *et al.*, 2014), whereby MHC alleles are maintained across long-evolutionary timescales by balancing selection despite the contrasting environments encountered by some allopatric species.

Conclusion

Our study suggests gene orthology within the recently diverged species of flamingos. However, high rates of concerted evolution is homogenizing polymorphism between loci, preventing divergence within species even in conserved regions of the MHC, and thus, in turn preventing the clear identification of the ancient orthologs identified in other distantly related birds. MHC alleles at ABS have been maintained over long-evolutionary timescales by balancing selection despite the distinct geographical ranges covered by allopatric species and the high rates of concerted evolution. There is currently limited empirical evidence of parallel evolution at the MHC (Lenz *et al.*, 2013), and instead, our study supports the hypothesis that balancing selection maintaining ancestral alleles that predate speciation is the main mechanism that results in sequence similarity at the MHC between closely related species in vertebrates.

Acknowledgments

We thank Christine Dubreuil for her invaluable assistance in the laboratory. Salins Group kindly granted access to the flamingo colony in the Camargue. This study was funded by the TOTAL foundation, the Conseil Régional de Bourgogne, the MAVA foundation and the Centre National de la Recherche Scientifique (CNRS). Mark Gillingham was supported by a postdoctoral grant from the Conseil Régional de Bourgogne and a DFG grant (DFG Gi 1065/2-1). We are deeply grateful to Luc Hoffmann and the late Alan Johnson for the instigation of the long-term study on the Greater flamingo and we warmly thank all the people who participated in collecting genetic samples of greater

flamingos: Juan Amat, Boudjema Samraoui, Özge Balkız, Christophe Germain and Antoine Arnaud. We further like to thank Rebecca Lee at the Wildfowl & Wetlands Trust (Slimbridge, UK), the Parc des Oiseaux (Villars-les-Dombes, France) and Jean-François Perrot-Minnot and Yves Kaiser for samples of other flamingo species. We thank Sylvain Piry for SESAME software assistance. Finally, we thank Retto Burri, Lewis Spurgin and two anonymous reviewers for their comments which have substantially improved the manuscript.

References

Alcaide, M., Edwards, S.V. & Negro, J.J. 2007. Characterization, polymorphism, and evolution of MHC class II B genes in birds of prey. *J. Mol. Evol.* **65**: 541–554.

Alcaide, M., Edwards, S.V., Cadahía, L. & Negro, J.J. 2009. MHC class I genes of birds of prey: isolation, polymorphism and diversifying selection. *Conserv. Genet.* **10**: 1349–1355.

Alcaide, M., Muñoz, J., Martínez-de la Puente, J., Soriguer, R. & Figuerola, J. 2014. Extraordinary MHC class II B diversity in a non-passerine, wild bird: the Eurasian Coot Fulica atra (Aves: Rallidae). *Ecol. Evol.* **4**: 688–698.

Anisimova, M., Nielsen, R. & Yang, Z. 2003. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. *Genetics* **164**: 1229–1236.

Balakrishnan, C.N., Ekblom, R., Völker, M., Westerdahl, H., Godinez, R., Kotkiewicz, H. *et al.* 2010. Gene duplication and fragmentation in the zebra finch major histocompatibility complex. *BMC Biol.* **8**: 29.

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.

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

Buehler, D.M., Verkuil, Y.I., Tavares, E.S. & Baker, A.J. 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., 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. *Mol. Biol. Evol.* **25**: 1180–1191.

Burri, R., Salamin, N., Studer, R.A., Roulin, A. & Fumagalli, L. 2010. Adaptive divergence of ancient gene duplicates in the avian MHC class II β. *Mol. Biol. Evol.* **27**: 2360–2374.

Burri, R., Promerová, M., Goebel, J. & Fumagalli, L. 2014. PCR-based isolation of multigene families: lessons from the avian MHC class IIB. *Mol. Ecol. Resour.* **14**: 778–788.

Charif, D. & Lobry, J.R. 2007. SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In: *Structural Approaches to Sequence Evolution: Molecules, Networks, Populations* (U. Bastolla, M. Porto, H. Eduardo Roman, M. Vendruscolo eds), pp. 207–232. Springer Verlag, New York.

Chaves, L.D., Faile, G.M., Krueth, S.B., Hendrickson, J.A. & Reed, K.M. 2010. Haplotype variation, recombination, and gene conversion within the turkey MHC-B locus. *Immunogenetics* **62**: 465–477.

Chubb, A.L. 2004. New nuclear evidence for the oldest divergence among neognath birds: the phylogenetic utility of ZENK (i). *Mol. Phylogenet. Evol.* **30**: 140–151.

Cloutier, A., Mills, J.A. & Baker, A.J. 2011. Characterization and locus-specific typing of MHC class I genes in the red-billed gull (*Larus scopulinus*) provides evidence for major, minor, and nonclassical loci. *Immunogenetics* **63**: 377–394.

Edwards, S., Grahn, M. & Potts, W. 1995. Dynamics of Mhc evolution in birds and crocodilians: amplification of class II genes with degenerate primers. *Mol. Ecol.* **4**: 719–729.

Edwards, S.V., Hess, C.M., Gasper, J. & Garrigan, D. 1999. Toward an evolutionary genomics of the avian Mhc. *Immunol. Rev.* **167**: 119–132.

Ericson, P.G.P., Anderson, C.L., Britton, T., Elzanowski, A., Johansson, U.S., Källersjö, M. *et al.* 2006. Diversification of Neoaves: integration of molecular sequence data and fossils. *Biol. Lett.* **2**: 543–547.

Galan, M., Guivier, E., Caraux, G., Charbonnel, N. & Cosson, J.-F. 2010. A 454 multiplex sequencing method for rapid and reliable genotyping of highly polymorphic genes in large-scale studies. *BMC Genom.* **11**: 296.

Galan, M., Pagès, M. & Cosson, J.-F. 2012. Next-Generation Sequencing for Rodent Barcoding: Species Identification from Fresh. *Degraded and Environmental Samples. PLoS ONE* **7**: e48374.

Geraci, J., Béchet, A., Cézilly, F., Ficheux, S., Baccetti, N., Samraoui, B., et al. 2012. Greater flamingo colonies around the Mediterranean form a single interbreeding population and share a common history. *J. Avian Biol.* **43**: 341–354.

Gibbs, M.J., Armstrong, J.S. & Gibbs, A.J. 2000. Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* **16**: 573–582.

Gu, X. & Nei, M. 1999. Locus specificity of polymorphic alleles and evolution by a birth-and-death process in mammalian MHC genes. *Mol. Biol. Evol.* **16**: 147–156.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**: 307–321.

Hackett, S.J., Kimball, R.T., Reddy, S., Bowie, R.C.K., Braun, E.L., Braun, M.J. *et al.* 2008. A phylogenomic study of birds reveals their evolutionary history. *Science* **320**: 1763–1768.

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

Hughes, C.R., Miles, S. & Walbroehl, J.M. 2008. Support for the minimal essential MHC hypothesis: a parrot with a single, highly polymorphic MHC class II B gene. *Immunogenetics* **60**: 219–231.

Huson, D.H. & Bryant, D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**: 254–267.

Janeway, C.A., Travers, P., Walport, M. & Shlomchik, M.J. 2005. *Immunobiology: The Immune System in Health and Disease*. Garland Science, London.

Juola, F.A. & Dearborn, D.C. 2012. Sequence-based evidence for major histocompatibility complex-disassortative mating in a colonial seabird. *Proc. Biol. Sci.* **279**: 153–162.

Kaufman, J., Milne, S., Göbel, T.W., Walker, B.A., Jacob, J.P., Auffray, C. *et al.* 1999. The chicken B locus is a minimal

essential major histocompatibility complex. *Nature* **401**: 923–925.

Kikkawa, E.F., Tsuda, T.T., Naruse, T.K., Sumiyama, D., Fukuda, M., Kurita, M. *et al.* 2005. Analysis of the sequence variations in the Mhc DRB1-like gene of the endangered Humboldt penguin (*Spheniscus humboldti*). *Immunogenetics* **57**: 99–107.

Kikkawa, E.F., Tsuda, T.T., Sumiyama, D., Naruse, T.K., Fukuda, M., Kurita, M. *et al.* 2009. Trans-species polymorphism of the Mhc class II DRB-like gene in banded penguins (genus *Spheniscus*). *Immunogenetics* **61**: 341–352.

Klein, J. 1986. *Natural History of the Major Histocompatibility Complex*. Wiley, New York.

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

Klein, J., Sato, A., Nagl, S. & O'hUigín, C. 1998. Molecular trans-species polymorphism. *Annu. Rev. Ecol. Syst.* **29**: 1–21.

Kriener, K., O'hUigin, C. & Klein, J. 2001. Independent origin of functional MHC class II genes in humans and new world monkeys. *Hum. Immunol.* **62**: 1–14.

Lenz, T.L., Eizaguirre, C., Kalbe, M. & Milinski, M. 2013. Evaluating patterns of convergent evolution and trans-species polymorphism at MHC immunogenes in two sympatric stickleback species. *Evolution* **67**: 2400–2412.

Liao, D. 1999. Concerted evolution: molecular mechanism and biological implications. *Am. J. Hum. Genet.* **64**: 24–30.

Lundberg, A.S. & McDevitt, H.O. 1992. Evolution of major histocompatibility complex class II allelic diversity: direct descent in mice and humans. *Proc. Natl. Acad. Sci. USA* **89**: 6545–6549.

Mansai, S.P. & Innan, H. 2010. The power of the methods for detecting interlocus gene conversion. *Genetics* **184**: 517–527.

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

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.

Martin, D.P., Lemey, P. & Posada, D. 2011. Analysing recombination in nucleotide sequences. *Mol. Ecol. Resour.* **11**: 943–955.

McCormack, J.E., Harvey, M.G., Faircloth, B.C., Crawford, N.G., Glenn, T.C. & Brumfield, R.T. 2013. A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. *PLoS One* **8**: e54848.

Meglécz, E., Piry, S., Desmarais, E., Galan, M., Gilles, A., Guivier, E. *et al.* 2011. SESAME (SEquence Sorter & AMplicon Explorer): genotyping based on high-throughput multiplex amplicon sequencing. *Bioinformatics* **27**: 277–278.

Miller, H.C. & Lambert, D.M. 2004. Gene duplication and gene conversion in class II MHC genes of New Zealand robins (Petroicidae). *Immunogenetics* **56**: 178–191.

Moon, D.A., Veniamin, S.M., Parks-Dely, J.A. & Magor, K.E. 2005. The MHC of the duck (*Anas platyrhynchos*) contains five differentially expressed class I genes. *J. Immunol.* **175**: 6702–6712.

Nei, M. & Rooney, A.P. 2005. Concerted and birth-and-death evolution of multigene families. *Annu. Rev. Genet.* **39**: 121–152.

Nei, M., Gu, X. & Sitnikova, T. 1997. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc. Natl. Acad. Sci. USA* **94**: 7799–7806.

van Oosterhout, C. 2009. Trans-species polymorphism, HLA-disease associations and the evolution of the MHC. *Commun. Integr. Biol.* **2**: 408–410.

Padidam, M., Sawyer, S. & Fauquet, C.M. 1999. Possible emergence of new geminiviruses by frequent recombination. *Virology* **265**: 218–225.

Posada, D. 2008. jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* **25**: 1253–1256.

Posada, D. & Crandall, K.A. 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc. Natl. Acad. Sci. USA* **98**: 13757–13762.

Potts, W.K., Manning, C.J. & Wakeland, E.K. 1994. The role of infectious disease, inbreeding and mating preferences in maintaining MHC genetic diversity: an experimental test. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **346**: 369–378.

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, Austria.

Rambaut, A., Suchard, M.A., Xie, D. & Drummond, A.J. 2014. Tracer v1.6. http://beast.bio.ed.ac.uk/Tracer.

Ronquist, F. & Huelsenbeck, J.P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.

Schierup, M.H. & Hein, J. 2000. Consequences of recombination on traditional phylogenetic analysis. *Genetics* **156**: 879–891.

Schwensow, N., Dausmann, K., Eberle, M., Fietz, J. & Sommer, S. 2010. Functional associations of similar MHC alleles and shared parasite species in two sympatric lemurs. *Infect. Genet. Evol.* **10**: 662–668.

Sepil, I., Moghadam, H.K., Huchard, E. & Sheldon, B.C. 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.

Seutin, G., White, B.N. & Boag, P.T. 1991. Preservation of avian blood and tissue samples for DNA analyses. *Can. J. Zool.* **69**: 82–90.

Silva, M.C. & Edwards, S.V. 2009. Structure and evolution of a new avian MHC class II B gene in a sub-Antarctic seabird, the thin-billed prion (Procellariiformes: *Pachyptila belcheri*). *J. Mol. Evol.* **68**: 279–291.

Smith, J.M. 1992. Analyzing the mosaic structure of genes. *J. Mol. Evol.* **34**: 126–129.

Soininen, J., McDonald, R. & Hillebrand, H. 2007. The distance decay of similarity in ecological communities. *Ecography* **30**: 3–12.

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

Sommer, S., Courtiol, A. & Mazzoni, C.J. 2013. MHC genotyping of non-model organisms using next-generation sequencing: a new methodology to deal with artefacts and allelic dropout. *BMC Genom.* **14**: 542.

Spurgin, L.G. & Richardson, D.S. 2010. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proc. Biol. Sci.* **277**: 979–988.

Srithayakumar, V., Castillo, S., Mainguy, J. & Kyle, C.J. 2012. Evidence for evolutionary convergence at MHC in two

broadly distributed mesocarnivores. *Immunogenetics* **64**: 289–301.
Steele, E.J. & Lloyd, S.S. 2015. Soma-to-germline feedback is implied by the extreme polymorphism at IGHV relative to MHC. *BioEssays* **37**: 557–569.
Steinitz, O., Heller, J., Tsoar, A., Rotem, D. & Kadmon, R. 2006. Environment, dispersal and patterns of species similarity. *J. Biogeogr.* **33**: 1044–1054.
Strand, T., Westerdahl, H., Höglund, J., V Alatalo, R. & Siitari, H. 2007. The Mhc class II of the Black grouse (*Tetrao tetrix*) consists of low numbers of B and Y genes with variable diversity and expression. *Immunogenetics* **59**: 725–734.
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.
Strandh, M., Westerdahl, H., Pontarp, M., Canbäck, B., Dubois, M.-P., Miquel, C. *et al.* 2012. Major histocompatibility complex class II compatibility, but not class I, predicts mate choice in a bird with highly developed olfaction. *Proc. Biol. Sci.* **279**: 4457–4463.
Torres, C.R., Ogawa, L.M., Gillingham, M.A., Ferrari, B. & van Tuinen, M. 2014. A multi-locus inference of the evolutionary diversification of extant flamingos (Phoenicopteridae). *BMC Evol. Biol.* **14**: 36.
van Tuinen, M., Butvill, D.B., Kirsch, J.A. & Hedges, S.B. 2001. Convergence and divergence in the evolution of aquatic birds. *Proc. Biol. Sci.* **268**: 1345–1350.
Wagner, A. 2005. Distributed robustness versus redundancy as causes of mutational robustness. *BioEssays* **27**: 176–188.
Wallny, H.-J., Avila, D., Hunt, L.G., Powell, T.J., 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.
Westerdahl, H., Wittzell, H. & von Schantz, T. 1999. Polymorphism and transcription of Mhc class I genes in a passerine bird, the great reed warbler. *Immunogenetics* **49**: 158–170.
Wilson, D.J. & McVean, G. 2006. Estimating diversifying selection and functional constraint in the presence of recombination. *Genetics* **172**: 1411–1425.
Wittzell, H., Bernot, A., Auffray, C. & Zoorob, R. 1999. Concerted evolution of two Mhc class II B loci in pheasants and domestic chickens. *Mol. Biol. Evol.* **16**: 479–490.
Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**: 1586–1591.
Yeager, M. & Hughes, A.L. 1999. Evolution of the mammalian MHC: natural selection, recombination, and convergent evolution. *Immunol. Rev.* **167**: 45–58.
Zagalska-Neubauer, M., Babik, W., Stuglik, M., Gustafsson, L., Cichoń, M. & Radwan, J. 2010. 454 sequencing reveals extreme complexity of the class II Major Histocompatibility Complex in the collared flycatcher. *BMC Evol. Biol.* **10**: 395.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Figure S1 Schematic illustration of primer position (represented by arrows) used in this study for amplification of MHC Class I and MHC Class IIB. Boxes represent exons.

Figure S2 Phylogenetic reconstruction by maximum likelihood (TrN + G; 1000 bootstraps) of the 23 MHC Class II intron 2 flamingo alleles (accession numbers: KU053963–KU053985).

Figure S3 The Neighbor-Net network constructed from the positively selected sites of the 105 MHC Class I exon 3 alleles (accession numbers: KU054001–KU054105) from the six flamingo species.

Figure S4 The Neighbor-Net network constructed from the positively selected sites of the 136 MHC Class II exon 2 alleles (accession numbers: KU054106–KU054241) from the six flamingo species.

Figure S5 Consensus Bayesian topology (50% majority consensus) constructed from the 105 MHC Class I exon 3 alleles from the six flamingo species and sequences isolated in other avian species (see below for list of species and accession number) with a *Ctenosaura clarki* sequence used as an outgroup (a: HKY + G) and from the 136 MHC Class II exon 2 alleles from the six flamingo species and sequences isolated in other avian species (see below for list of species and accession number) with a *Homo sapiens* sequence used as an outgroup (b: GTR + G). Only supports for avian orders and families are provided.

Figure S6 Expected frequency distribution of the proportion of identical antigen binding sites (ABS) codons at MHC Class I exon 3 (below diagonal) and MHC Class IIB exon 2 (above diagonal) under parallel evolution between each species of flamingos.

Received 27 December 2014; revised 12 November 2015; accepted 15 November 2015