

RESEARCH ARTICLE

Open Access



New data from basal Australian songbird lineages show that complex structure of MHC class II β genes has early evolutionary origins within passerines

Shandiya Balasubramaniam^{1,2*}, Rebecca D. Bray³, Raoul A. Mulder², Paul Sunnucks⁴, Alexandra Pavlova⁴ and Jane Melville¹

Abstract

Background: The major histocompatibility complex (MHC) plays a crucial role in the adaptive immune system and has been extensively studied across vertebrate taxa. Although the function of MHC genes appears to be conserved across taxa, there is great variation in the number and organisation of these genes. Among avian species, for instance, there are notable differences in MHC structure between passerine and non-passerine lineages: passerines typically have a high number of highly polymorphic MHC paralogs whereas non-passerines have fewer loci and lower levels of polymorphism. Although the occurrence of highly polymorphic MHC paralogs in passerines is well documented, their evolutionary origins are relatively unexplored. The majority of studies have focussed on the more derived passerine lineages and there is very little empirical information on the diversity of the MHC in basal passerine lineages. We undertook a study of MHC diversity and evolutionary relationships across seven species from four families (Climacteridae, Maluridae, Pardalotidae, Meliphagidae) that comprise a prominent component of the basal passerine lineages. We aimed to determine if highly polymorphic MHC paralogs have an early evolutionary origin within passerines or are a more derived feature of the infraorder Passerida.

Results: We identified 177 alleles of the MHC class II β exon 2 in seven basal passerine species, with variation in numbers of alleles across individuals and species. Overall, we found evidence of multiple gene loci, pseudoalleles, trans-species polymorphism and high allelic diversity in these basal lineages. Phylogenetic reconstruction of avian lineages based on MHC class II β exon 2 sequences strongly supported the monophyletic grouping of basal and derived passerine species.

Conclusions: Our study provides evidence of a large number of highly polymorphic MHC paralogs in seven basal passerine species, with strong similarities to the MHC described in more derived passerine lineages rather than the simpler MHC in non-passerine lineages. These findings indicate an early evolutionary origin of highly polymorphic MHC paralogs in passerines and shed light on the evolutionary forces shaping the avian MHC.

Keywords: Trans-species polymorphism, Gene duplication, Concerted evolution, Convergent evolution, Birth-and-death model, Accordion model, Passerida, Corvida

* Correspondence: sbalasubramaniam@museum.vic.gov.au

¹Department of Sciences, Museum Victoria, Melbourne, VIC 3001, Australia

²School of BioSciences, The University of Melbourne, Melbourne, VIC 3010, Australia

Full list of author information is available at the end of the article



Background

The major histocompatibility complex (MHC) is a complex multigene family that regulates the function of the adaptive branch of the vertebrate immune system. The MHC is comprised of many different types of genes, such as classical and non-classical MHC genes, as well as non-MHC genes such as natural killer cells and tapasin [1]. Classical and non-classical MHC genes may be distinguished from each other in that the former are polymorphic and highly expressed, whereas the latter exhibit lower levels of expression and are often monomorphic [2]. MHC genes may also be grouped as class I, II or III genes based on the types of molecules they encode. Class I genes encode receptors that are presented on the surfaces of most nucleated cells and primarily facilitate immune responses to intracellular pathogens, whereas class II genes are only found on a subset of cells and are associated with immune responses to extracellular pathogens [3]. Class III genes encode molecules involved in the complement component of the immune response rather than the adaptive response [2].

The high levels of polymorphism characteristic of classical MHC class I and II genes may be largely attributed to the peptide-binding region (PBR), which displays great diversity of alleles and extensive sequence variation among different alleles [3, 4]. The genes of this region encode proteins that form the molecular groove where peptides are bound and then presented to T-cells, resulting in the appropriate immune response being triggered [1]. Each allele typically responds to a category of potential antigens, so individuals and populations with greater MHC diversity may be better able to cope with a range of infections [5, 6].

A number of models of multigene family evolution, such as the birth-and-death model [7] and the accordion model [5], have been proposed to explain the generally high levels of polymorphism and the large and variable number of classical MHC loci among individuals and species. According to the birth-and-death model, new genes are created via repeated duplication (birth) and then either retained in the genome or eventually lost (death) after becoming non-functional through deleterious mutations [7]. The accordion model, on the other hand, posits that the number of MHC genes expands and contracts in response to fluctuating selection pressure [5]. These models are not mutually exclusive, and both models predict the redundancy that occurs in the genome of most species, whereby multiple duplicated gene loci exist in both class I and II gene regions [6]. One explanation for this redundancy is that once alleles are generated, they may remain in the genome even in the absence of selection pressure [6]. Alleles may be maintained over long evolutionary time scales through speciation events, which may then resolve as trans-species polymorphisms (TSP) in

phylogenetic reconstructions, where similar alleles are present in groups that diverged over millions of years ago such as rodents and primates [10, 11]. On a smaller scale, genetic variation in the MHC may also be generated among alleles and between loci via the processes of point mutation, recombination and gene conversion, and maintained through the influence of balancing selection.

Among birds, the most studied and well-described MHC is that of the chicken (*Gallus gallus*), which only contains one dominantly expressed molecule of class I and one of class II [6, 7]. The chicken MHC is small and densely arranged; a compact organisation of genes with short introns, small physical size, lack of redundancy, and low overall number of classical class I and II genes with few pseudogenes, has led to the chicken MHC being described as a “minimal essential MHC” [7, 8]. The small size and simplicity of the chicken MHC allows for the co-evolution of genes as haplotypes over considerable periods of time, with no recombination between class I and II genes having been detected over thousands of experiments [7, 9]. The result of this is that the most striking associations between MHC haplotype and resistance or susceptibility to disease have been characterised in chickens [10, 11].

Similarly simple genetic organisation and small numbers of expressed classical MHC loci have been noted in a few galliform species, such as the pheasant (*Phasianus colchicus*) [12], black grouse (*Tetrao tetrix*) [23], and grey partridge (*Perdix perdix*) [13], although the quail (*Coturnix japonica*) has been found to have higher levels of gene duplication [14]. Variable levels of MHC complexity have been characterised among other avian groups, such as penguins, seabirds, raptors, rails, cranes, and waders [15–20], and passerine species typically show multiple loci with extensive gene duplication, evidence of recombination, high levels of polymorphism, and the presence of pseudogenes [21–25]. This suggests the minimal essential model is not applicable to all avian species, as the passerine MHC appears to be larger and more complex than that described in chickens. The passerine MHC, along with a number of non-passerine lineages, has been characterised as having high rates of concerted evolution and/or recent duplications [2, 12], leading to few reports of orthologous relationships among birds [12, 26–29]. Thus, although the overall function of the MHC is conserved across birds, there appear to be considerable differences in the genomic organisation of the MHC between lineages [30].

With a few exceptions [27, 31–33], research on the MHC in passerines has focussed on species within the more derived infraorder Passerida, whereas very little empirical information exists on the structure and complexity of the MHC in the Corvida (*sensu* [34]), a paraphyletic basal clade within the passerines (Fig. 1). It remains

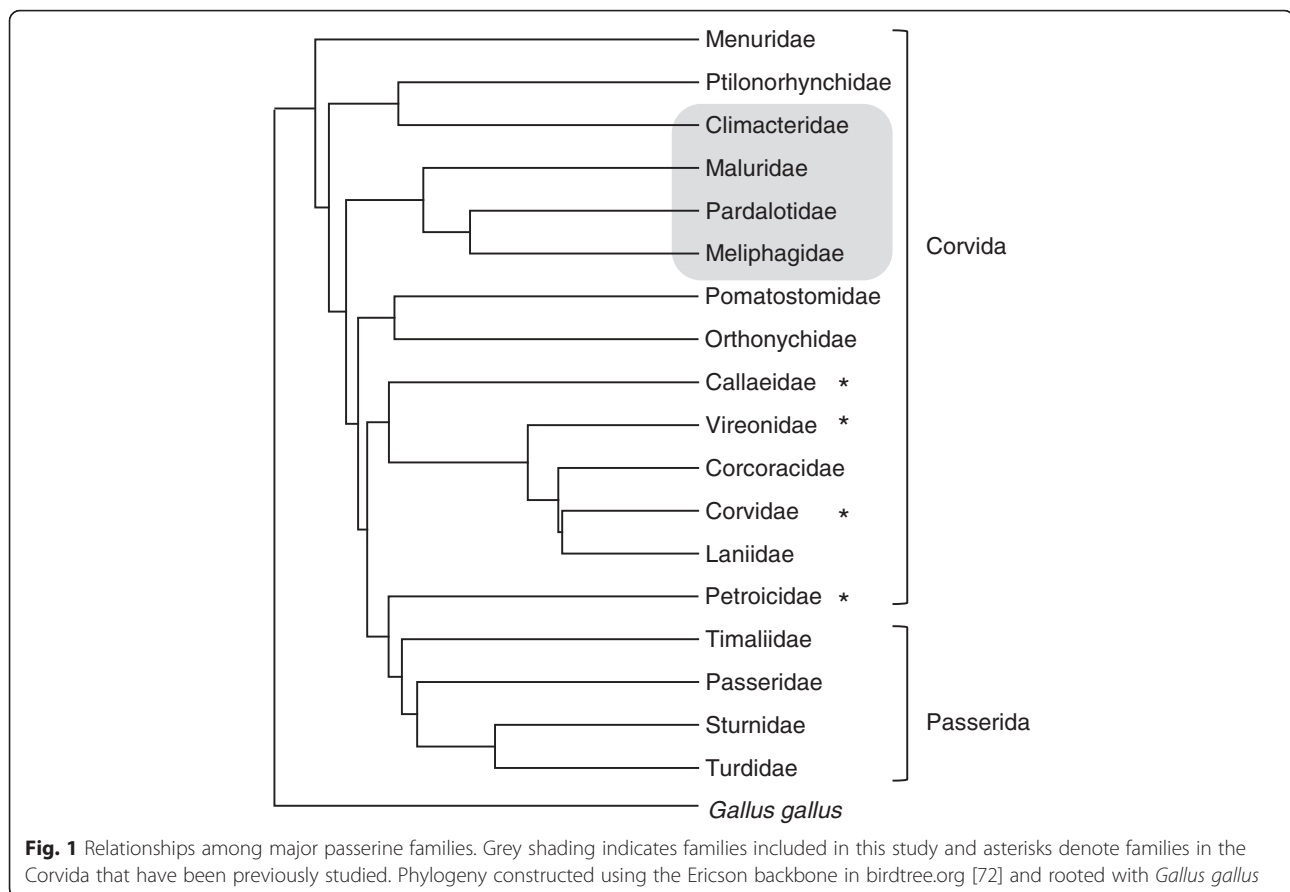


Fig. 1 Relationships among major passerine families. Grey shading indicates families included in this study and asterisks denote families in the Corvida that have been previously studied. Phylogeny constructed using the Ericson backbone in birdtree.org [72] and rooted with *Gallus gallus*

unclear whether a complex MHC is characteristic of all passerines or is a more derived trait occurring in the Passerida and the few species of the Corvida that have been studied (Fig. 1). Here, we used cloning and next-generation 454 sequencing techniques to characterise the polymorphism and allelic diversity of classical MHC class II β exon 2. We focussed on seven Australian species from four prominent basal passerine families (Table 1, Fig. 1) to test if there is increasing complexity of the MHC structure from basal to derived passerines: if there is such a progressive increase in MHC complexity, basal passerine lineages should show a more simple MHC structure similar to that described in non-passerine lineages. To better understand the evolutionary origins of complex MHC structure in passerines, we also reconstructed phylogenetic relationships among MHC sequences in these species.

Results and discussion

In order to assess levels of polymorphism and allelic diversity in the MHC class II genes of basal passerines, we chose the following species: Brown Treecreeper *Climacteris picumnus*, Superb Fairy-wren *Malurus cyaneus*, Spotted Pardalote *Pardalotus punctatus*, Striated Pardalote *P. striatus*, White-plumed honeyeater *Lichenostomus*

penicillatus, Fuscous honeyeater *L. fuscus*, and Yellow-tufted Honeyeater *L. melanops*. These seven species are from four families (Climacteridae, Maluridae, Pardalotidae, Meliphagidae) that comprise a prominent component of the Corvida. We amplified a 159 bp region that includes the highly diverse PBR of the MHC class II β exon 2 using cloning and 454 sequencing methods. The high level of coverage provided by 454 sequencing makes it a convenient method for assessing genetic diversity in highly polymorphic, multilocus systems, such as the MHC [25, 35, 36]. We used 454 sequencing and cloning with Sanger sequencing complementarily to assess levels of genetic diversity in seven basal passerine species.

The two methods resulted in different numbers of alleles being identified in each species. We isolated 98 MHC class II β exon 2 sequences from 21 individuals using cloning with Sanger sequencing (see Additional file 1), with the total number of alleles detected in each species ranging from 7 in the Spotted Pardalote to 19 in the Striated Pardalote. Of the 98 sequences isolated, 17 were found in more than one species, resulting in 81 unique alleles across the seven species. Sample sizes were the same in every species and all samples were from the same study region, so the variation in allele numbers across species is unlikely to be a product of sampling design. Within

Table 1 Sample information and MHC allelic diversity

Family	Species	Species prefix	N _A (range)	N _{loci}	N _{ps}
Climacteridae	Brown treecreeper <i>Climacteris picumnus</i>	Clpi	30 (8–17)	9	0
Meliphagidae	Yellow-tufted honeyeater <i>Lichenostomus melanops</i>	Lime	39 (11–14)	7	2
	White-plumed honeyeater <i>Lichenostomus penicillatus</i>	Lipe	19 (6–11)	6	0
	Fuscous honeyeater <i>Lichenostomus fuscus</i>	Lifu	24 (6–10)	5	1
Maluridae	Superb fairy-wren <i>Malurus cyaneus</i>	Macy	33 (2–14)	7	2
Pardalotidae	Spotted pardalote <i>Pardalotus punctatus</i>	Papu	26 (5–11)	6	3
	Striated pardalote <i>Pardalotus striatus</i>	Past	39 (7–13)	7	1

N_A - total number of alleles per species; range - range of alleles detected in individuals of each species; N_{loci} - putative number of loci per species; N_{ps} - number of pseudoalleles

species, the number of alleles identified in each individual ranged from 5 to 14.

454 sequencing was conducted on 13 samples in total (see Additional file 1), across all seven study species. Due to small volumes of available blood samples, different individuals were sequenced using 454 compared to Sanger sequencing, and from a different location in the study area (see Methods). Using 454 sequencing, we obtained 16,257 reads of the 159-bp region showing a complete match to the forward and reverse primers. After a stepwise filtering procedure, we were left with 66 % of the original reads (10,783), with coverage depth ranging from 114 to 570 reads per allele (see Additional file 2). This remaining 66 % of reads comprised a total of 146 alleles, of which 12 were found in more than one species. Of the remaining 134 unique alleles recovered through 454 sequencing, 38 were also identified through cloning. Similar to the sequences isolated through cloning, numbers of alleles varied widely across species and individuals, ranging from 11 in the White-plumed Honeyeater to 26 in the Brown Treecreeper. The isolation of more than two alleles from each individual indicates the presence of multiple MHC class II β loci in each of these species, from a minimum of five loci in the Fuscous Honeyeater to nine loci in the Brown Treecreeper. The conservative filtering process we used would have potentially excluded a larger number of true alleles from individuals with higher coverage compared to individuals with lower coverage, so levels of MHC allelic diversity may be underestimated in some species.

Despite fewer individuals having been screened through 454 sequencing compared to cloning, larger numbers of alleles were identified in total and in each individual (see Additional file 1). Although 454 sequencing has the advantage of avoiding some artefacts associated with bacterial cloning, it is still vulnerable to artefacts arising during PCR and has been shown to produce a higher percentage of sequencing errors than Sanger sequencing [37–40]. In contrast, Sanger sequencing probably represents a smaller proportion of the allelic diversity of a species, compared with 454 sequencing, with a

prevalence of locally common alleles while less common (or rare) alleles are missed in cloning. Our use of both these methods suggests 454 sequencing is a much more efficient, although possibly less accurate, way to assess levels of allelic diversity in species and populations.

Polymorphism and allelic diversity of MHC loci

We identified a total of 177 alleles of the MHC class II β exon 2 across seven species. Numbers of alleles varied widely across species and individuals, from 19 to 39 alleles per species and 2 to 17 alleles per individual (Table 1). Based on the maximum number of alleles in an individual, we inferred a minimum of 5–9 loci per species, a range which falls within the middle of the spectrum for the number of MHC class II β loci described in passerines ($N = 3–20$; [22]). Frameshift mutations and stop codons in sequences from five of the seven species suggest pseudoalleles are present in the dataset (Table 1), a level similar to that found in the infraorder Passerida [41, 42]. All seven species displayed high levels of allelic diversity, with high intraspecific nucleotide and amino acid distances compared to other passerines ([43, 44]; Table 2). The highest levels of diversity were observed at the PBR in all species, a pattern again consistent with Passerida MHC [2].

Phylogenetic relationships among MHC alleles

Bayesian reconstruction of the phylogenetic relationships among MHC class II β exon 2 alleles from the seven species largely reflected taxonomic relationships at the genus level. The majority of alleles clustered in one of three well-supported clades (≥ 90 % posterior probability; Fig. 2), with alleles from honeyeaters (Meliphagidae), pardalotes (Pardalotidae), and treecreepers (Climacteridae) grouping separately to each other. The phylogenetic network showed a similar pattern of clustering, with alleles from honeyeaters, pardalotes and treecreepers falling into three distinct clusters (Fig. 2). Sequences from the Superb Fairy-wren (Maluridae) were the exception to this general pattern, where approximately half of the sequences (15/33 sequences; 45 %) were basal to the

Table 2 Mean nucleotide and amino acid distances

Species	Nucleotide			Amino acid		
	All sites	PBR	non-PBR	All sites	PBR	non-PBR
Clpi	0.147 (0.019)	0.266 (0.054)	0.110 (0.021)	0.245 (0.046)	0.574 (0.125)	0.154 (0.045)
Lime	0.229 (0.023)	0.455 (0.072)	0.167 (0.022)	0.423 (0.062)	0.845 (0.202)	0.315 (0.059)
Lipe	0.253 (0.026)	0.513 (0.080)	0.184 (0.025)	0.409 (0.064)	0.890 (0.193)	0.288 (0.058)
Lifu	0.259 (0.025)	0.539 (0.072)	0.188 (0.025)	0.421 (0.062)	0.909 (0.183)	0.302 (0.059)
Macy	0.318 (0.028)	0.521 (0.078)	0.263 (0.030)	0.536 (0.071)	0.939 (0.197)	0.433 (0.071)
Papu	0.134 (0.018)	0.332 (0.068)	0.077 (0.014)	0.229 (0.046)	0.624 (0.158)	0.123 (0.034)
Past	0.141 (0.019)	0.338 (0.064)	0.084 (0.015)	0.246 (0.048)	0.641 (0.166)	0.141 (0.036)

Mean nucleotide and amino acid distances among putative alleles from MHC class II β exon 2. Nucleotide distances are corrected for multiple substitutions with the Kimura 2-parameter model and amino acid distances are corrected using expectations from the Poisson distribution. Standard errors are based on 1000 bootstrap replicates and given in parentheses. *PBR* peptide binding region

main Pardalotidae-Meliphagidae lineage in the Bayesian tree but did not form a monophyletic clade, and other Maluridae sequences were intermingled within Climacteridae, Pardalotidae and Meliphagidae.

The basal position of the Superb Fairy-wren sequences may result from recombination generating a pattern of mixed ancestry, which causes recombinant sequences to fall outside the parental clusters [45]. The intermingling of other Superb Fairy-wren sequences within clades in the phylogenetic tree containing other genera may indicate a trans-species mode of evolution, a pattern which has been widely documented in MHC class II β sequences of both passerines and non-passerines [28, 46–50]. Trans-species polymorphism (TSP) is a mechanism by which identical alleles occur in related species by being passed on from ancestral to descendant species [51]. This can be difficult to differentiate from convergent evolution, which can produce similar alleles in different species through independent evolutionary pathways. Being sympatric, the seven species in this study may display signatures of convergent evolution if they are subject to similar parasite or pathogen pressure. To differentiate between signatures of convergent evolution and TSP, we compared phylogenetic reconstructions based on non-synonymous and synonymous substitutions only. If the species are subject to similar selective pressures, the phylogeny based on non-synonymous substitutions should show mixed clustering of sequences across families whereas the phylogeny based on synonymous substitutions should show alleles grouping according to family or species. If, however, the similarity of alleles across species is a result of TSP, the two phylogenies should show similar patterns of allelic clustering across families [48, 52]. A comparison of phylogenies based on non-synonymous and synonymous substitutions showed the latter pattern of clustering, with the majority of clusters in both phylogenies showing alleles from more than one family (see Additional file 3). Mixed clusters predominantly comprised alleles from honeyeaters, pardalotes and fairy-wrens, whereas alleles

from treecreepers clustered intraspecifically for the most part. Based on our analysis of exon 2, the phylogenetic signatures support a TSP explanation of the clustering pattern rather than one of convergent evolution. Although this result suggests the presence of TSP, it must be noted that distinguishing between phylogenetic patterns generated by TSP and other non-TSP patterns and processes is problematic when the loci are unknown. The evolution of exon 2 is complex and it is possible that, for example, rates of allele sharing based only on this region are an overestimation, thereby leading to incorrect inferences of TSP [52, 53]. Comparison of phylogenies based on intron and exon regions, as well as other related species with allopatric distributions, and identification of alleles associated with individual loci, would clarify the roles of convergent evolution, TSP, and other processes in generating these patterns of clustering.

Fifteen alleles in the dataset occurred in at least two different species, and there was variation among genera in levels of TSP. In 11 instances, alleles occurred across genera between fairy-wrens and another species. In all other instances TSP occurred within genera, either between the two species of pardalotes or three species of honeyeaters. Treecreepers did not show any instances of TSP with other study species, which may be related to treecreepers being in a separate evolutionary lineage to the other study species (Climacteridae; Fig. 1). Retention of alleles through TSP between treecreepers and the other basal lineages in this study would therefore need to occur over longer evolutionary timeframes of at least 60 million years [54].

Inspection of amino acid alignments revealed seven sites with residues unique to each of the Climacteridae, Pardalotidae and Meliphagidae, distinguishing sequences in the three families from each other (see Additional file 4). Fairy-wren sequences, being intermingled within these groups, showed the residues particular to whichever group they fell in with. This pattern of clustering by species is not uncommon in birds and could be explained by either a)

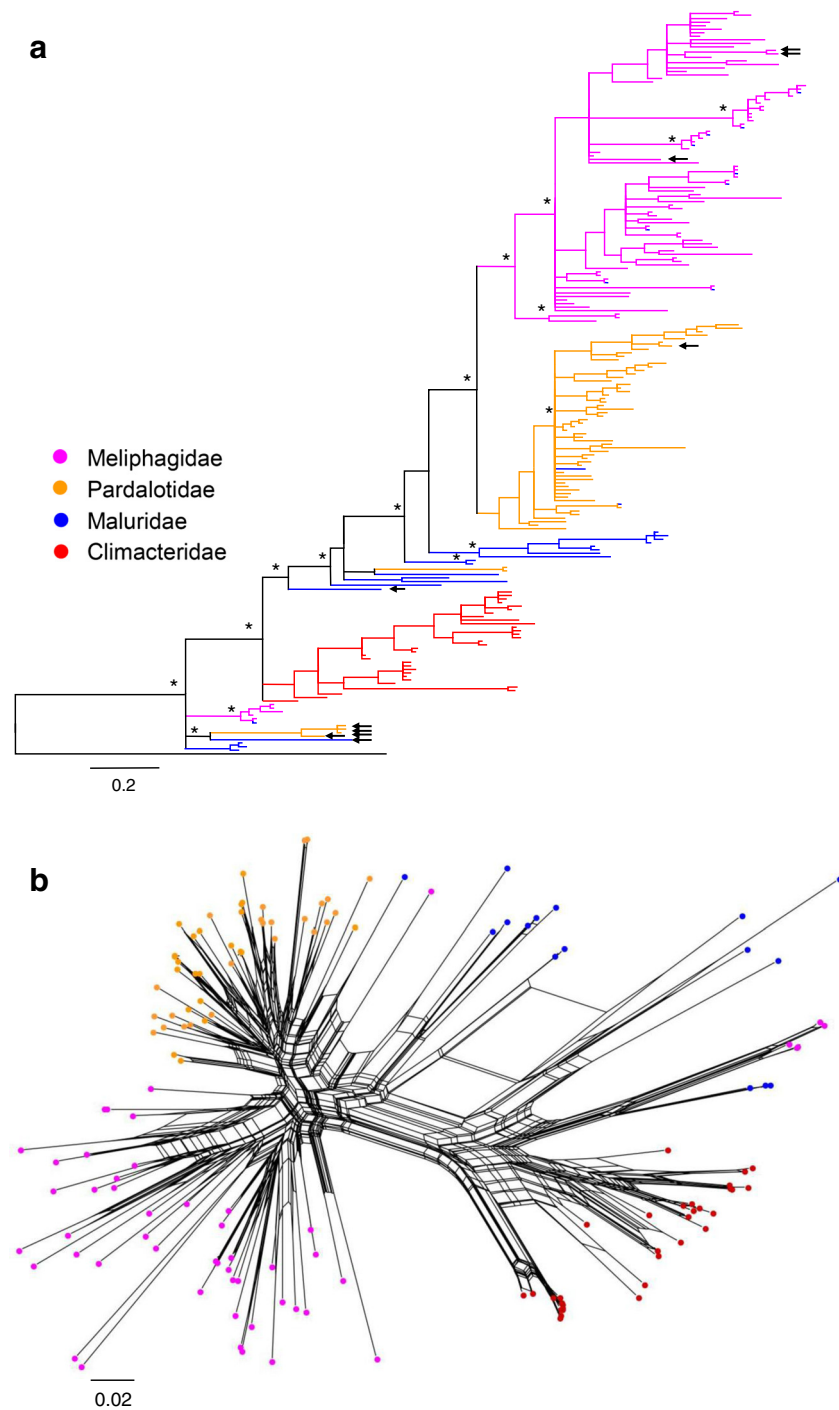


Fig. 2 Relationships of MHC class II β exon 2 sequences among basal passerine species. **a** Bayesian phylogeny of MHC alleles from the seven basal passerine species in this study, rooted with *Crocodylus niloticus*; asterisks denote major branches with posterior probabilities > 90 %, and pseudoalleles are indicated by arrows. **b** Neighbour-net of MHC class II β exon 2 sequences from seven basal passerine species. Sequences are colour coded by family

recent duplication, where genes diverge following speciation and then duplicate to form similar copies of the gene, or b) concerted evolution, where duplication occurs prior to speciation but the duplicated genes

become homogenised through gene conversion [2]. Both these processes result in intraspecific loci being more similar to each other than to orthologous loci in other species. Among birds, patterns of orthology may

be obscured because of rapid gene duplication and homogenisation of MHC class II genes [33].

To assess the relationships among MHC class II β exon 2 sequences in the wider phylogenetic context, we constructed a phylogeny including species from a range of avian taxa (Fig. 3). All passerine species formed a well-supported monophyletic clade (100 % posterior probability) that was also strongly supported by topology testing, using a Bayesian stepping stone approach (Bayes Factor 75.35). All species from the non-passerine taxa (Accipitridae, Apterygidae, Ardeidae, Galliformes, Procellariiformes, Spheniscidae and Strigidae) formed a polytomy at the base of the phylogeny. Within the passerine clade, most sequences from honeyeaters, pardalotes and fairy-wrens grouped together, albeit in a clade without strong posterior probability support (Group A), whereas treecreeper sequences fell separately to these sequences. The exceptions to this were one group of pseudoalleles and two groups of sequences that had either aspartic acid (D) or alanine (A) instead of glutamic acid (E) in the first position (see Additional file 4); these three groups of sequences instead clustered separately from the other sequences in this study in a well-supported clade (Group B). The clustering of pseudoalleles and putatively functional sequences (Group B) may either indicate the latter are actually non-functional alleles or that the pseudoalleles evolved from these alleles. There was no apparent structuring according to passerine taxonomic relationships, because Corvida and Passerida sequences were intermingled in the passerine clade. There is some evidence that rates of diversifying and homogenising forces may vary between lineages of birds [28], which could explain the pattern here where some sequences cluster according to species whereas other clusters comprise sequences from different species. The grouping of basal passerine species from this study with species from both the Corvida and Passerida suggests a complex MHC class II β structure may be common throughout the passerine order.

Conclusions

In this study we detected evidence of multiple gene loci, high levels of polymorphism and allelic diversity, and the presence of pseudogenes among classical MHC class II β exon 2 of seven passerine species. Our analyses strongly support a monophyletic grouping of passerines from the Corvida and Passerida, signifying an early evolutionary origin of complex MHC class II β structure in the Passeriformes. Phylogenetic analyses based on non-synonymous and synonymous substitutions supported a trans-species explanation for the mixed clustering of alleles across the different families in this study. However, we cannot rule out the possibility that convergent evolution may have also played a role in generating these

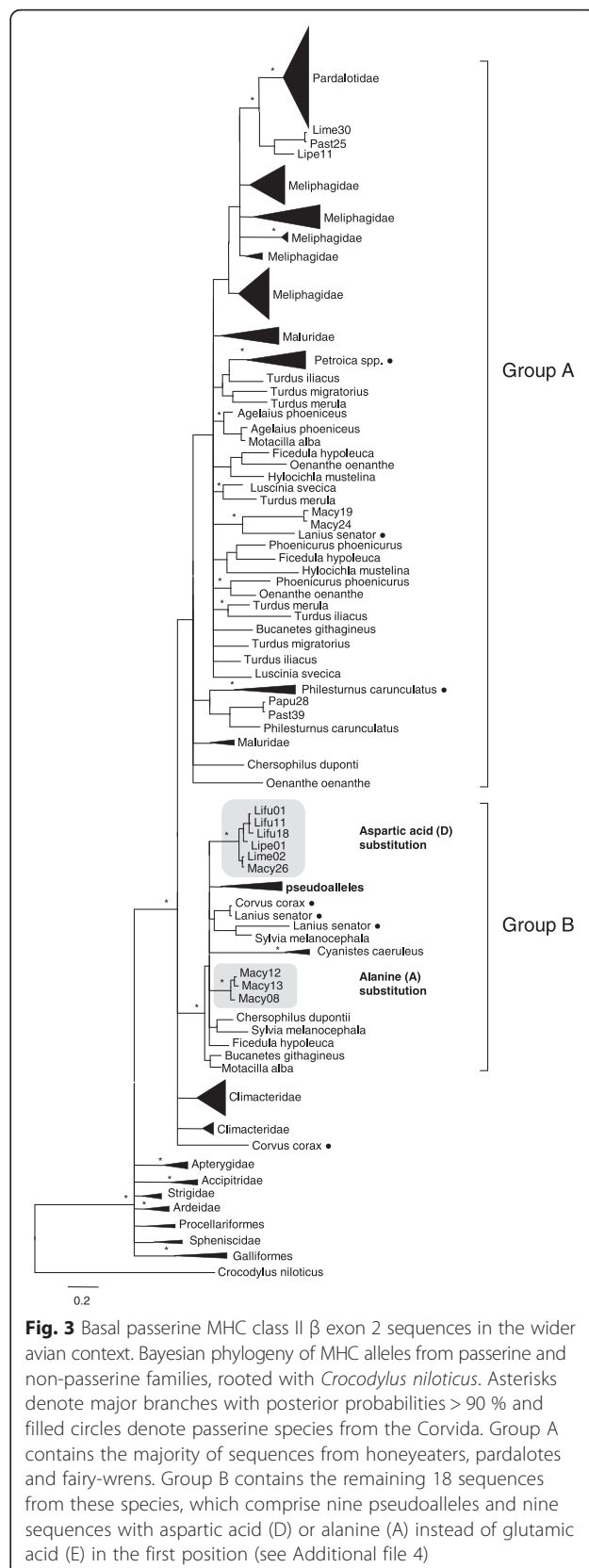


Fig. 3 Basal passerine MHC class II β exon 2 sequences in the wider avian context. Bayesian phylogeny of MHC alleles from passerine and non-passerine families, rooted with *Crocodylus niloticus*. Asterisks denote major branches with posterior probabilities > 90 % and filled circles denote passerine species from the Corvida. Group A contains the majority of sequences from honeyeaters, pardalotes and fairy-wrens. Group B contains the remaining 18 sequences from these species, which comprise nine pseudoalleles and nine sequences with aspartic acid (D) or alanine (A) instead of glutamic acid (E) in the first position (see Additional file 4)

phylogenetic signatures. Phylogenetic analyses based on sequences from other MHC class II exon and intron regions may clarify the relative roles of TSP and convergent evolution in generating these signatures. A comparison of 454 sequencing and cloning methods suggested the former is a much more efficient way to assess levels of allelic diversity in species and populations. Continued characterisation of this gene region, as well as other non-coding regions of the MHC, in species from different phylogenetic levels may improve our understanding of the rates of gene conversion, recombination, diversification and homogenisation occurring in the avian MHC. Such research will also provide insights into the evolutionary significance of the disparity in MHC complexity between passerine and non-passerine species.

Methods

Sample collection and study design

We collected blood samples from one to three individuals of seven Australian passerine species from Victoria, Australia, as described in Amos et al. ([55]; see Additional file 1). These seven species (Brown Treecreeper *Climacteris picumnus*, Superb Fairy-wren *Malurus cyaneus*, Spotted Pardalote *Pardalotus punctatus*, Striated Pardalote *P. striatus*, White-plumed honeyeater *Lichenostomus penicillatus*, Fuscous honeyeater *L. fuscus*, Yellow-tufted Honeyeater *L. melanops*) are from four families (Climacteridae, Maluridae, Pardalotidae, Meliphagidae) that comprise a prominent component of basal passerine lineages.

As part of a pilot study we tested previously published MHC class II β primers on the study species (Additional file 5), from which we identified a single set of primers that consistently amplified across all the study species. These primers amplified a 159 bp region of the MHC class II β exon 2 and were used for all our genetic work. To assess levels of polymorphism and allelic diversity in the seven study species, we amplified the 159 bp region of the MHC class II β exon 2 using cloning and 454 sequencing methods. This 159 bp fragment includes the highly diverse PBR as well as fit within the length restrictions for 454 sequencing. The PBR and non-PBR were inferred assuming functional congruence to the human HLA-DR1 molecule [56].

DNA extraction, PCR, cloning, and sequencing

Blood samples were digested overnight using Proteinase K and DNA was extracted with standard phenol-chloroform protocols for Sanger sequencing [57], while DNA was extracted from blood samples using a salting-out protocol for 454 sequencing [58]. The resulting DNA was suspended in Tris-EDTA. For Sanger sequencing the degenerate primers 326 and 325 [59] were used to amplify a 159 bp region of the MHC class II β gene, spanning the majority of exon 2. For 454 sequencing we amplified the same 159 bp region of

the MHC class II β gene using HPLC-purified fusion primers. The forward fusion primer (5'-CCATCTCAT CCCTGCGTGTCTCCGACTCAGNNNNNNNNNNNG **AGTGYCAYTAYYTNAAYGGYAC**-3') comprised the 454 GS FLX Titanium Primer A, a 10 bp Multiplex Identifier (MID) tag (indicated with Ns) to differentiate among individuals, and the 326 primer sequence (in bold). The reverse fusion primer (5'-CCTATCCCCTG TGTGCCTTGGCAGTCTCAGNNNNNNNNNNNGTAG **TTGTGNCKGCAGTANSTGTCCAC**-3') similarly comprised the 454 GS FLX Titanium Primer B, a 10 bp MID tag (indicated with Ns) and the 325 primer sequence (in bold).

For both Sanger and 454 sequencing standard 25 μ l PCRs were conducted with approximately 25 ng of genomic DNA. Sanger sequencing PCRs: 1 x Taq buffer (Promega), 2.5 mM MgCl₂, 0.2 mM of each dNTP, 0.5 units platinum *Taq* polymerase (Invitrogen), 0.02 mg/ml BSA (Sigma-Aldrich) and 0.4 mM of each primer. 454 sequencing PCRs: 1 x GoTaq Colourless Master Mix (Promega), 2 mM MgCl₂, and 0.4 mM of each primer. A negative control was included in each PCR and a touch-down protocol was used for all amplifications. The thermal profile consisted of an initial 5 min denaturation at 95 °C, followed by 10 cycles denaturation at 95 °C for 30 s, annealing at 60–44 °C for 30 s, with the annealing temperature decreasing by 4 °C every 2 cycles, and extension at 72 °C for 90 s. This was followed by 35 cycles of denaturation at 95 °C for 45 s, annealing at 45 °C for 45 s, and extension at 72 °C for 90 s, followed by a final extension at 72 °C for 10 min. PCR products were visualised on 1.2 % agarose gels to confirm amplification.

Cloning of PCR products for Sanger sequencing was undertaken by ligating PCR product into the pCR[®]II-TOPO bacterial plasmid (TOPO TA Cloning[®] Kit Dual Promoter, Invitrogen) and transformed into TOP10F['] chemically competent cells following the manufacturer's protocol for the TOPO TA Cloning[®] Kit Dual Promoter (Invitrogen). Recombinant clones were detected by blue/white screening and selected clones were suspended in 30 μ l of ddH₂O for a minimum of 1 h. Prior to sequencing, clones were screened for the expected insert size using 2 μ l of bacterial water in a PCR containing M13 forward and reverse primers. For each individual, at least 20 amplified clones of the expected insert size were purified with 1 μ l ExoSAP-IT (USB), and sequenced commercially (Macrogen, Korea) using M13 primers in Sanger sequencing. To validate each allelic sequence, DNA from each individual was amplified and cloned twice. Cloned sequences were retained if they occurred in at least two independent PCRs. The retained sequences were edited, assembled and aligned using Geneious v. 6 [60].

PCR products for 454 sequencing were purified using the Agencourt AMPure XP purification kit (Beckman

Coulter) according to the manufacturer's instructions. Purified products were pooled in equimolar concentrations and sequenced commercially (Macrogen, Korea) on an eighth plate on a 454 GS-FLX run.

Bioinformatics following 454 sequencing

Following 454 sequencing, we used the jMHC software [61] to extract sequences and assign reads to individuals. Only reads showing a complete match to the forward and reverse primers were retained, ensuring that all assigned reads covered the whole amplicon. At the level of the whole dataset, we retained only sequences of exactly 159 bp which occurred in at least two independent PCRs, each represented by at least three reads [36]. Remaining sequences were assigned to individuals based on MID tags and aligned in Geneious v. 6 [60]. At the individual level, we discarded any alleles with coverage lower than 10 % of the allele with the highest coverage within each individual in order to remove less reliable sequence variants [62]. Despite the potential for excluding true alleles in individuals with very high coverage, we believe this conservative approach provides a sufficiently comprehensive assessment of the levels of diversity across species to assess patterns of MHC evolution in passerines.

Allelic diversity and phylogenetic analyses

We calculated intraspecific nucleotide and amino acid diversity using the Kimura 2-parameter and Poisson models respectively in MEGA v.5 and estimated standard errors through 1000 bootstrap replicates [63]. We reconstructed the evolutionary relationships among MHC class II β exon 2 sequences from a) the basal passerine species in this study, and b) a wider range of avian taxa, including species from Accipitridae, Apterygidae, Ardeidae, Galliformes, Procellariiformes, Spheniscidae, Strigidae, and non-Australian Passeriformes (see Additional file 6). We estimated the best-fit model of evolutionary change using jModelTest2 ([GTR + I; [64, 65]) based on Akaike's Information Criterion and constructed a Bayesian phylogeny in MRBAYES v. 3.2 [66] with a Nile crocodile sequence (*Crocodylus niloticus* FJ886734) as an outgroup. Trees were sampled every 1,000 generations over 50 million generations of two simultaneous runs, with one cold and three heated Markov Coupled Monte-Carlo chains. The first 25 % of trees were discarded as burn-in and the remaining used to construct a consensus tree which was visualised in FigTree v. 1.4 [67]. We evaluated the monophyly of the passerine sequences in our dataset by comparing estimates of marginal likelihood in natural log units for positively and negatively constrained topologies [68] in MRBAYES v. 3.2. We performed topology testing, using a Bayesian stepping stone approach for 5 million generations and assessed support for the constrained topology over the negative

constraint using Bayes Factors [69]. To assess phylogenetic relationships based on putatively neutral and adaptive genetic variation, we constructed Neighbour-Joining trees in MEGA v. 5 [63] using the Nei-Gojobori method with Jukes-Cantor correction based on a) non-synonymous (d_N) and b) synonymous substitutions (d_S). Bootstrap tests of trees were conducted using 5,000 replicates. As exon 2 of the MHC class II β genes has been demonstrated to undergo high rates of recombination and gene conversion [2, 70, 71], we complementarily utilised phylogenetic networks in addition to phylogenetic trees to visualise relationships among MHC sequences. We used the neighbour-net algorithm based on Jukes-Cantor distances in Splitstree v. 4 [45] to examine the relationships among MHC sequences from the seven basal passerine species.

Ethics approvals

Samples were collected under permits from the Victorian Department of Environment and Primary Industries (numbers 10004294 under the *Wildlife Act* 1975 and the *National Parks Act* 1975, and NWF10455 under section 52 of the *Forest Act* 1958), the Australian Bird and Bat Banding Schemes and under approval and monitoring of Monash University ethics processes (BSCI/2007/07).

Availability of supporting data

The sequence data supporting the results of this article are available in the Figshare digital repository and can be accessed at doi:10.4225/49/571353BCF0125.

Additional files

Additional file 1: MHC class II β exon 2 alleles identified through cloning and 454 sequencing methods. Sample size, range and total number of alleles per species, and putative number of loci per species are given. (XLSX 8 kb)

Additional file 2: Relationship between depth of coverage (total number of reads) obtained through 454 sequencing and number of alleles for each individual. (PDF 166 kb)

Additional file 3: Neighbour-joining tree estimated from a) non-synonymous and b) synonymous substitutions of MHC class II β exon 2. The seven species in this study are coloured according to family and the tree rooted with *Crocodylus niloticus*. (PDF 76 kb)

Additional file 4: Amino acid alignment of 159 nucleotide sites from MHC class II β exon 2. Amino acid residues distinguishing Meliphagidae, Pardalotidae and Climacteridae are highlighted in grey. (PDF 1154 kb)

Additional file 5: Primer sequences trialled in this study. Standard International Union of Biochemistry (IUB) codes are used for degenerate primers. (XLSX 9 kb)

Additional file 6: Avian taxa included in Bayesian phylogeny. GenBank accession numbers and family or order classification are given for species included in the Bayesian phylogeny of avian taxa (Fig. 2). (XLSX 10 kb)

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

SB, RM, PS, AP and JM conceived and developed the study. SB and RB were involved in data collection. SB, RB, PS, AP and JM analysed and interpreted the data. All authors read, advised on revisions, and approved the final manuscript.

Acknowledgements

We thank all the members of the Birds Linkage team, especially A. Lill and N. Takeuchi, and numerous volunteers for assistance with fieldwork.

Funding

This work was supported by the Australian Research Council (LP0776322), Department of Environment and Primary Industries, Museum Victoria, Parks Victoria, the North Central and Goulburn Broken Catchment Management Authorities, University of Melbourne (SB) and the Australian Academy of Sciences (JM).

Author details

¹Department of Sciences, Museum Victoria, Melbourne, VIC 3001, Australia. ²School of BioSciences, The University of Melbourne, Melbourne, VIC 3010, Australia. ³Terrestrial Vertebrates, Western Australian Museum, Perth, WA 6986, Australia. ⁴School of Biological Sciences, Monash University, Melbourne, VIC 3800, Australia.

Received: 10 October 2015 Accepted: 10 May 2016

Published online: 21 May 2016

References

- Klein J. Natural history of the major histocompatibility complex. New York: Wiley; 1986.
- Hess CM, Edwards SV. The Evolution of the Major Histocompatibility Complex in Birds. *BioScience*. 2002;52:423.
- Hughes AL, Yeager M. Natural selection at major histocompatibility complex loci of vertebrates. *Annu Rev Genet*. 1998;32:415–35.
- Garrigan D, Hedrick PW. Perspective: detecting adaptive molecular polymorphism: lessons from the MHC. *Evol Int J Org Evol*. 2003;57:1707–22.
- Klein J, Ono H, Klein D, O'huigin C. The Accordion Model of Mhc Evolution. In: Gergely J, Benczúr M, Erdei A, Falus A, Füst G, Medgyesi G, editors. *Prog. Immunol*, vol. VIII. Berlin Heidelberg: Springer; 1993. p. 137–43.
- Kaufman J, Völk H, Walny HJ. A "minimal essential Mhc" and an "unrecognized Mhc": two extremes in selection for polymorphism. *Immunol Rev*. 1995;143:63–88.
- Kaufman J, Milne S, Göbel TW, Walker BA, Jacob JP, Auffray C, et al. The chicken B locus is a minimal essential major histocompatibility complex. *Nature*. 1999;401:923–5.
- Shaw I, Powell TJ, Marston DA, Baker K, van Hateren A, Riegert P, et al. Different evolutionary histories of the two classical class I genes BF1 and BF2 illustrate drift and selection within the stable MHC haplotypes of chickens. *J Immunol*. 2007;178:5744–52.
- Skjødt K, Koch C, Crone M, Simonsen M. Analysis of chickens for recombination within the MHC (B-complex). *Tissue Antigens*. 1985;25:278–82.
- Plachy J, Pink JR, Hála K. Biology of the chicken MHC (B complex). *Crit Rev Immunol*. 1992;12:47–79.
- Schat K. Immunity in Marek's disease and other tumors. In: Toivanen A, Toivanen P, editors. *Avian Immunol. Basis Pract*. Boca Raton: CRC Press; 1987.
- Wittzell H, Bernot A, Auffray C, Zoorob R. Concerted evolution of two Mhc class II B loci in pheasants and domestic chickens. *Mol Biol Evol*. 1999;16:479–90.
- Promerová M, Králová T, Bryjová A, Albrecht T, Bryja J. MHC class IIB exon 2 polymorphism in the grey partridge (*Perdix perdix*) is shaped by selection, recombination and gene conversion. *PLoS ONE*. 2013;8:e69135.
- Hosomichi K, Shiina T, Suzuki S, Tanaka M, Shimizu S, Iwamoto S, et al. The major histocompatibility complex (Mhc) class IIB region has greater genomic structural flexibility and diversity in the quail than the chicken. *BMC Genomics*. 2006;7:322.
- Kikkawa EF, Tsuda TT, Naruse TK, Sumiyama D, Fukuda M, Kurita M, et al. Analysis of the sequence variations in the Mhc DRB1-like gene of the endangered Humboldt penguin (*Spheniscus humboldti*). *Immunogenetics*. 2005;57:99–107.
- Alcaide M, Edwards SV, Negro JJ. Characterization, polymorphism, and evolution of MHC class II B genes in birds of prey. *J Mol Evol*. 2007;65:541–54.
- Silva MC, Edwards SV. 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*. 2009;68:279–91.
- Eklom R, Grahn M, Hoglund J. Patterns of polymorphism in the MHC class II of a non-passerine bird, the great snipe (*Gallinago media*). *Immunogenetics*. 2003;54:734–41.
- Alcaide M, Muñoz J, Martínez-de la Puente J, Soriguer R, Figuerola J. Extraordinary MHC class II B diversity in a non-passerine, wild bird: the Eurasian Coot *Fulica atra* (Aves: Rallidae). *Ecol. Evol*. 2014;4:688–98.
- Kohyama TI, Akiyama T, Nishida C, Takami K, Onuma M, Momose K, et al. Isolation and characterization of major histocompatibility complex class II B genes in cranes. *Immunogenetics*. 2015;67:705–10.
- Balakrishnan CN, Eklom R, Völker M, Wester Dahl H, Godinez R, Kotkiewicz H, et al. Gene duplication and fragmentation in the zebra finch major histocompatibility complex. *BMC Biol*. 2010;8:29.
- Bollmer JL, Dunn PO, Whittingham LA, Wimpee C. Extensive MHC class II B gene duplication in a passerine, the common Yellowthroat (*Geothlypis trichas*). *J Hered*. 2010;101:448–60.
- Sato A, Tichy H, Grant PR, Grant BR, Sato T, O'huigin C. Spectrum of MHC class II variability in Darwin's finches and their close relatives. *Mol Biol Evol*. 2011;28:1943–56.
- Promerová M, Albrecht T, Bryja J. Extremely high MHC class I variation in a population of a long-distance migrant, the Scarlet Rosefinch (*Carpodacus erythrinus*). *Immunogenetics*. 2009;61:451–61.
- Sepil I, Moghadam HK, Huchard E, Sheldon BC. Characterization and 454 pyrosequencing of Major Histocompatibility Complex class I genes in the great tit reveal complexity in a passerine system. *BMC Evol Biol*. 2012;12:68.
- Strand T, Wester Dahl H, Höglund J, V Alatalo R, Siitari H. 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*. 2007;59:725–34.
- Miller HC, Lambert DM. Gene duplication and gene conversion in class II MHC genes of New Zealand robins (Petroicidae). *Immunogenetics*. 2004;56:178–91.
- Burri R, Hirzel HN, Salamin N, Roulin A, Fumagalli L. Evolutionary patterns of MHC class II B in owls and their implications for the understanding of avian MHC evolution. *Mol Biol Evol*. 2008;25:1180–91.
- Burri R, Salamin N, Studer RA, Roulin A, Fumagalli L. Adaptive divergence of ancient gene duplicates in the avian MHC class II β . *Mol Biol Evol*. 2010;27:2360–74.
- Kelley J, Walter L, Trowsdale J. Comparative genomics of major histocompatibility complexes. *Immunogenetics*. 2005;56:683–95.
- Alcaide M, Liu M, Edwards SV. Major histocompatibility complex class I evolution in songbirds: universal primers, rapid evolution and base compositional shifts in exon 3. *PeerJ*. 2013;1:e86.
- Sutton JT, Robertson BC, Grueber CE, Stanton J-A, Jamieson IG. Characterization of MHC class II B polymorphism in bottlenecked New Zealand saddlebacks reveals low levels of genetic diversity. *Immunogenetics*. 2013;65:619–33.
- Edwards SV, Wakeland EK, Potts WK. Contrasting histories of avian and mammalian Mhc genes revealed by class II B sequences from songbirds. *Proc Natl Acad Sci U S A*. 1995;92:12200–4.
- Sibley C, Ahlquist J. Phylogeny and classification of the birds of the world. New Haven: Yale University Press; 1990.
- Babik W, Taberlet P, Ejsmond MJA, Radwan J. New generation sequencers as a tool for genotyping of highly polymorphic multilocus MHC system. *Mol Ecol Resour*. 2009;9:713–9.
- Galan M, Guivier E, Caraux G, Charbonnel N, Cosson J-F. A 454 multiplex sequencing method for rapid and reliable genotyping of highly polymorphic genes in large-scale studies. *BMC Genomics*. 2010;11:296.
- Moore MJ, Dhingra A, Soltis PS, Shaw R, Farmerie WG, Folta KM, et al. Rapid and accurate pyrosequencing of angiosperm plastid genomes. *BMC Plant Biol*. 2006;6:17.
- Longeri M, Zanotti M, Damiani G. Recombinant DRB sequences produced by mismatch repair of heteroduplexes during cloning in *Escherichia coli*. *Eur J Immunogenet*. 2002;29:517–23.
- Lenz TL, Becker S. Simple approach to reduce PCR artefact formation leads to reliable genotyping of MHC and other highly polymorphic loci—implications for evolutionary analysis. *Gene*. 2008;427:117–23.
- Burri R, Promerová M, Goebel J, Fumagalli L. PCR-based isolation of multigene families: lessons from the avian MHC class IIB. *Mol Ecol Resour*. 2014;14:778–88.

41. Babik W, Durka W, Radwan J. Sequence diversity of the MHC DRB gene in the Eurasian beaver (*Castor fiber*). *Mol Ecol*. 2005;14:4249–57.
42. Oppelt C, Wutzler R, von Holst D. Characterisation of MHC class II DRB genes in the northern tree shrew (*Tupaia belangeri*). *Immunogenetics*. 2010;62:613–22.
43. Bonneaud C, Sorci G, Morin V, Wester Dahl H, Zoorob R, Wittzell H. Diversity of Mhc class I and II B genes in house sparrows (*Passer domesticus*). *Immunogenetics*. 2004;55:855–65.
44. Anmarkrud JA, Johnsen A, Bachmann L, Lifjeld JT. Ancestral polymorphism in exon 2 of bluethroat (*Luscinia svecica*) MHC class II B genes: MHC class II diversity in bluethroats. *J Evol Biol*. 2010;23:1206–17.
45. Huson DH, Bryant D. Application of Phylogenetic Networks in Evolutionary Studies. *Mol Biol Evol*. 2006;23:254–67.
46. Seddon JM, Ellegren H. A temporal analysis shows major histocompatibility complex loci in the Scandinavian wolf population are consistent with neutral evolution. *Proc R Soc Lond B Biol Sci*. 2004;271:2283–91.
47. Radwan J, Kawalko A, Wójcik JM, Babik W. MHC-DRB3 variation in a free-living population of the European bison, *Bison bonasus*. *Mol Ecol*. 2007;16:531–40.
48. Eimes JA, Townsend AK, Sepil I, Nishiumi I, Satta Y. Patterns of evolution of MHC class II genes of crows (*Corvus*) suggest trans-species polymorphism. *PeerJ* [Internet]. 2015;3. Available from: <https://peerj.com/articles/853>. Accessed 17 May 2016.
49. Strandh M, Lannefors M, Bonadonna F, Wester Dahl H. Characterization of MHC class I and II genes in a subantarctic seabird, the blue petrel, *Halobaena caerulea* (Procellariiformes). *Immunogenetics*. 2011;63:653–66.
50. Miller HC, Bowker-Wright G, Kharkrang M, Ramstad K. Characterisation of class II B MHC genes from a ratite bird, the little spotted kiwi (*Apteryx owenii*). *Immunogenetics*. 2011;63:223–33.
51. Klein J, Sato A, Nagl S, O'hUigin C. Molecular Trans-Species Polymorphism. *Annu Rev Ecol Syst*. 1998;29:1–C1.
52. Kriener K, O'hUigin C, Tichy H, Klein J. Convergent evolution of major histocompatibility complex molecules in humans and New World monkeys. *Immunogenetics*. 2000;51:169–78.
53. Takahata N, Satta Y. Selection, convergence, and intragenic recombination in HLA diversity. *Genetica*. 1998;102–103:157–69.
54. Barker FK, Cibois A, Schikler P, Feinstein J, Cracraft J. Phylogeny and diversification of the largest avian radiation. *Proc Natl Acad Sci U S A*. 2004;101:11040–5.
55. Amos JN, Harrisson KA, Radford JQ, White M, Newell G, Nally RM, et al. Species- and sex-specific connectivity effects of habitat fragmentation in a suite of woodland birds. *Ecology*. 2014;95:1556–68.
56. Brown J, Jardetzky T, Gorga J, Stern L, Urban R, Strominger J, et al. Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature*. 1993;364:33–9.
57. Sambrook J, Fritsch E, Maniatis T. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, New York: Harbor Laboratory Press; 1989.
58. Harrisson K, Pavlova A, Amos JN, Takeuchi N, Lill A, Radford JQ, et al. Fine-scale effects of habitat loss and fragmentation despite large-scale gene flow for some regionally declining woodland bird species. *Lands Ecol*. 2012;27:813–27.
59. Edwards S, Grahn M, Potts W. Dynamics of Mhc evolution in birds and crocodylians: amplification of class II genes with degenerate primers. *Mol Ecol*. 1995;4:719–30.
60. Drummond A, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, et al. Geneious [Internet]. 2011. Available from: <http://www.geneious.com/>. Accessed 17 May 2016.
61. Stuglik MT, Radwan J, Babik W. jMHC: software assistant for multilocus genotyping of gene families using next-generation amplicon sequencing. *Mol Ecol Resour*. 2011;11:739–42.
62. Stiebens VA, Merino SE, Chain FJJ, Eizaguirre C. Evolution of MHC class I genes in the endangered loggerhead sea turtle (*Caretta caretta*) revealed by 454 amplicon sequencing. *BMC Evol Biol*. 2013;13:95.
63. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 2011;28:2731–9.
64. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol*. 2003;52:696–704.
65. Durrin D, Taboada G, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods*. 2012;9:772.
66. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 2012;61:539–42.
67. Rambaut A. FigTree [Internet]. 2009. Available from: <http://tree.bio.ed.ac.uk/software/figtree>. Accessed 17 May 2016.
68. Xie W, Lewis PO, Fan Y, Kuo L, Chen M-H. Improving Marginal Likelihood Estimation for Bayesian Phylogenetic Model Selection. *Syst Biol*. 2011;60:150–60.
69. Kass RE, Raftery AE. Bayes Factors. *J Am Stat Assoc*. 1995;90:773.
70. Schaschl H, Wandeler P, Suchentrunk F, Obexer-Ruff G, Goodman SJ. Selection and recombination drive the evolution of MHC class II DRB diversity in ungulates. *Heredity*. 2006;97:427–37.
71. Nei M, Gu X, Sitnikova T. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc Natl Acad Sci U S A*. 1997;94:7799–806.
72. Jetz W, Thomas GH, Joy JB, Hartmann K, Moores AO. The global diversity of birds in space and time. *Nature*. 2012;491:444–8.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

