MBE Advance Access published May 12, 2010

Research Article

# Adaptive divergence of ancient gene duplicates in the avian MHC class II B

Reto Burri *†, Nicolas Salamin ‡§, Romain A. Studer ‡§, Alexandre Roulin ‡ & Luca Fumagalli *

* Laboratory for Conservation Biology, Department of Ecology and Evolution

University of Lausanne, Switzerland

‡ Department of Ecology and Evolution, University of Lausanne, Switzerland

§ Swiss Institute of Bioinformatics, Lausanne, Switzerland

† Corresponding author:

reto.burri@unil.ch

Tel. ++41 21 692 41 74

Fax ++41 21 692 41 65



*Keywords*: gene duplication, major histocompatibility complex, multigene family evolution, neofunctionalization, positive selection

*Running head*: Avian MHC class II β evolution

*Abbreviations*: MHCII, major histocompatibility complex class II; PP, posterior probability

## Abstract

Gene duplication and neofunctionalization are known to be important processes in the evolution of phenotypic complexity. They account for important evolutionary novelties that confer ecological adaptation, such as the major histocompatibility complex (MHC), a multigene family crucial to the vertebrate immune system. In birds, where 2 major MHC class II β (MHCIIβ) exon 3 lineages have been recently characterized, 2 hypotheses for the evolutionary history of MHCIIβ lineages were proposed. These lineages could have arisen either by (i) an ancient duplication and subsequent divergence of one paralog, or (ii) recent parallel duplications followed by functional convergence. Here we compiled a data set consisting of 63 MHCIIβ exon 3 sequences from 6 avian orders to distinguish between these hypotheses and to understand the role of selection in the divergent evolution of the 2 avian MHCIIβ lineages. Based on phylogenetic reconstructions and simulations we show that a unique duplication event preceding the major avian radiations gave rise to 2 ancestral MHCIIβ lineages that were each likely lost once later during avian evolution. Maximum likelihood estimation shows that following the ancestral duplication positive selection drove a radical shift from basic to acidic amino acid composition of a protein domain facing the α-chain in the MHCII αβ-heterodimer. Structural analyses of the MHCII αβ-heterodimer highlight that 3 of these residues are potentially involved in direct interactions with the α-chain, suggesting that the shift following duplication may have been accompanied by coevolution of the interacting α- and β-chains. These results provide new insights into the long-term evolutionary relationships among avian MHC genes and open interesting perspectives for comparative and population genomic studies of avian MHC evolution.

## Introduction

Gene duplication and neofunctionalization are important evolutionary processes that contribute to phenotypic complexity (Lynch and Conery 2000; Conant and Wolfe 2008). The evolution of the vertebrate adaptive immune system, in particular the major histocompatibility complex (MHC), is a well-known example of an important ecological adaptation that arose in large part through evolutionary novelties following genetic and genomic duplications (Kumánovics, Takada, and Lindahl 2003; Nei and Rooney 2005). Genes of the MHC are involved in pathogen recognition and defence and thereby directly linked to individual health and fitness (Klein 1986). They represent the most polymorphic genetic system known in vertebrates to date (Gaudieri et al. 2000; Geraghty et al. 2002). The evolutionary arms-race between pathogens and hosts has driven the evolution of exceptional allelic diversity at the MHC (Apanius et al. 1997; Bernatchez and Landry 2003; Piertney and Oliver 2006). The human population for example exhibits up to 500 alleles at a single locus (Robinson et al. 2003) that enable a fine tuning of recognition mechanisms to face rapid, continuous pathogen evolution. Polymorphism in the MHC is however not restricted to allelic variation. The molecular evolution of the MHC involves frequent gene duplication and gene loss that result in vast rearrangements and pronounced variation in gene number and genomic organization among organisms (Kulski et al. 2002; Kelley, Walter, and Trowsdale 2005). Whether or not differences in MHC architecture may be connected to an organism's potential to efficiently evolve immune responses has not been studied so far, but patterns observed in chicken (*Gallus gallus*) and Japanese quail (*Coturnix japonica*) suggest this to be the case. Indeed, the tight linkage of genes arising from compact MHC architecture and a low number of genes appears to be responsible for the strong association of particular MHC haplotypes with disease resistance and susceptibility in these species (Kaufman 1999; Kaufman et al. 1999a; Shimizu et al. 2004; Shaw et al. 2007).

Compared to mammals, the chicken MHC stands out with an unparalleled architectural simplicity (Kaufman et al. 1999b). While mammalian MHCs are usually highly complex and

redundant gene regions that cover up to several million base pairs (3.6 Mbp in humans) and harbour hundreds of genes (Kumánovics, Takada, and Lindahl 2003), the 'minimal essential MHC' of the chicken contains 19 genes spread over merely 92 kb (Kaufman et al. 1999b) in the B-region plus at least 2 each of the non-classical class I and class II genes in the unlinked Y-region (Briles et al. 1993; Miller et al. 1994). This comparably stream-lined organization is in line with the observation that birds tend to have more compact genomes than mammals (Bonneaud, Burnside, and Edwards 2008). In chicken, the tight linkage of genes resulting from the simple and compact MHC architecture appears to have facilitated the coevolution of MHC genes within haplotypes, eventually leading to the tightest associations of specific MHC haplotypes with infectious disease resistance known today (Kaufman et al. 1999a; Kaufman 2000). Reduced recombination between genes (but see also Hosomichi et al. 2008) in a compact genomic architecture was proposed to facilitate the coevolution of genes within MHC haplotypes (Kaufman 1999), although this link has not been generalized to other bird species so far. The general simplicity of avian MHCs is also questioned by studies in Japanese quail (Shiina et al. 2004; Hosomichi et al. 2006) and passerines (reviewed in Hess and Edwards 2002), which revealed the presence of more genes and longer introns than in chicken. However, an increasing number of studies in various other bird orders report low numbers of genes (Tsuda et al. 2001; Kikkawa et al. 2005; Alcaide, Edwards, and Negro 2007; Strand et al. 2007; Alcaide et al. 2008; Burri et al. 2008a; Hughes, Miles, and Walbroehl 2008). Definite conclusions on the ancestral state of MHC architecture can not be drawn before the relationships of MHC genes among avian orders have been established. The phylogenetic relationships among avian MHC genes though are notoriously difficult to reconstruct (Hess and Edwards 2002). The high rates of gene duplication and gene loss inherent to MHC evolution blur the signals of gene history over the long term. In birds, elevated rates of concerted evolution further complicate the accurate reconstruction of gene history by masking the orthologous relationships of genes among species (Edwards, Wakeland, and Potts 1995; Edwards et al. 1999; Hess and Edwards 2002). The first

insights into the phylogenetic relationships of MHC genes among avian orders were reported only recently (Burri et al. 2008b). Based on data from exon 3, the phylogenetic relationships of 2 MHC class II β (MHCIIβ) gene lineages (*DAB1* and *DAB2*) in owls (Strigiformes) have been traced back across and even beyond an entire avian order to about 100 million years ago (mya) (Burri et al. 2008b). Most intriguingly, the 2 owl MHCIIβ lineages appeared to form 2 polyphyletic gene lineages. This suggests that the modern avian MHCIIβ genes are either descendant from ancient lineages that emerged by duplication early during avian evolution ('early duplication hypothesis'), or that their similarity is a result of parallel duplications followed by functional convergence ('convergence hypothesis').

To distinguish between these two alternative evolutionary hypotheses, we investigated whether (i) according to the 'convergence hypothesis' a common selective pressure in multiple taxa could have led to convergent evolution of exon 3 in parallel MHCIIβ duplicates at a recent evolutionary timescale, or whether (ii) divergent selection underpinned the divergence of exon 3 in 2 ancient MHCIIβ lineages such as predicted by the 'early duplication hypothesis'. To this end we contrasted phylogenetic relationships among genes inferred from functional variation (codon sequences and amino acid sequences) to the relationships inferred from third codon positions. Under the 'convergence hypothesis', the phylogenetic relationships are expected to be different among these analyses, with third positions revealing the neutral gene history and functional variation representing the functional relationships that evolved under convergent selection. Under the 'early duplication hypothesis' the gene history inferred from third codon positions is expected to match the phylogenetic relationships inferred from codon and amino acid sequences. Furthermore, we studied the role of positive selection on the diversification of the avian MHCIIβ exon 3. Based on extensive phylogenetic analyses and computer simulations we show that a duplication event preceded the major avian radiations about 100 mya. Subsequent to duplication, positive selection drove the adaptive divergence, including a shift from basic to acidic composition in part of one paralog. Based on the analysis of the three-dimensional

structure of the MHCII protein we show how molecular coevolution between the interacting chains in the MHCII αβ-heterodimer may have been implicated in the divergence of gene duplicates.

## Material and Methods

### *Sequence data*

As previously proposed by other authors (Edwards et al. 1999) and demonstrated in several taxa (Wittzell et al. 1999; Miller and Lambert 2004; Burri et al. 2008b), exon 3 and the 3'-UTR trace the evolutionary relationships among MHCIIβ genes better than the regions involved in pathogen recognition. Their evolutionary history is less distorted by the effects of positive selection and gene conversion that play an important role in the evolution of the pathogen binding region encoded by exon 2 (Hughes and Yeager 1998; Yeager and Hughes 1999; Burri et al. 2008b). In order to reconstruct the gene history of avian MHCIIβ genes, we compiled a data set containing avian MHCIIβ exon 3 sequences (216 bp of in total 270 bp of this exon) from all available bird species in the GenBank database (N=63). When several loci per species were available, a single sequence was kept if these loci were previously shown to cluster in the same lineage (Table 1; Burri et al. 2008b). Reptilian sequences from the tuatara (Sphenodon punctatus, DQ124232, Miller, Belov, and Daugherty 2005) and the spectacled caiman (*Caiman crocodilus*, AF256652) were used as outgroups in the phylogenetic reconstructions. Sequences from the barn owl (*Tyto alba*) and the great snipe (*Gallinago media*, a shorebird) were not included in the phylogenetic analyses because they are considered likely recombinants and thus can not be placed accurately in phylogenetic reconstructions (Burri et al. 2008b).

Sequence assembly and translation into amino acid sequences prior to alignment were performed in MEGA 4.0 (Tamura et al. 2007). Amino acid sequences were aligned using the ClustalW algorithm (Thompson, Higgins, and Gibson 1994) and translated back to nucleotide

sequences. An amino acid alignment is provided in Figure S1. The amino acid residue numbering reported hereafter follows Bondinas et al. (2007). The mean number of synonymous (S) and non-synonymous sites (N) between gene lineages was calculated in MEGA 4.0 using the method of Nei and Gojobori with Jukes-Cantor correction (Nei and Gojobori 1986).

### *Phylogenetic reconstructions*

The amino acid and codon sequences of MHCIIβ exon 3 (216 bp) were used to reconstruct the functional relationships between avian MHCIIβ genes. For the phlyogenetic reconstruction of gene history, only third codon positions were included, in order to exclude the potential influence of selection. We analyzed only the third codon positions of exon 3 because intronic sequences of the avian MHCIIβ are too divergent to be alignable. The third positions of exon 2 were not included because of the presumably high rates of recombination and gene conversion acting on this region (Burri et al. 2008b), which distort its phylogenetic signal (Posada and Crandall 2002).

The amino acid substitution model was evaluated using ProtTest 1.4 (Abascal, Zardoya, and Posada 2005). For nucleotide data the best fitting nucleotide substitution model for each codon position was evaluated using MrModeltest 2.2 (Nylander 2004) according to the Akaike information criterion (Akaike 1974; Posada and Buckley 2004). All phylogenetic reconstructions were performed with MrBayes 3.1 (Ronquist and Huelsenbeck 2003) and RAxML (Stamatakis 2006). For coding sequences, each codon position was entered as a separate partition to account for their respective rates of evolution. Bayesian analyses included 2 independent runs each with 4 chains of $5 \cdot 10^7$ generations. Convergence was evaluated using the average standard deviation of split frequencies between runs and the potential scale reduction factor. Posterior distributions were examined in Tracer 1.4 (Rambaut and Drummond 2007). The first half of the topologies was discarded as burnin. For maximum likelihood analyses, $10^4$ bootstrap replicates were performed. Consensus topologies were obtained by the extended majority rule consensus in the CONSENSE program from the PHYLIP 3.68 package (Felsenstein 2005).

Bayesian phylogenetic reconstructions based on short sequences, such as the third codon positions of MHCIIβ data, can be sensitive to the chosen priors. We therefore performed prior sensitivity analyses by selecting different parameter values for exponential and uniform branch lengths priors (see Table 2 for parameter values). The exponential branch length prior is closer to an uninformative prior, and Ronquist and Huelsenbeck (2003) advise against using uniform branch length priors, because of their close-to-random substitution probabilities and the large prior probability they put on long branches. However, small exponential priors may underestimate the probability of convergence on long branches (2007). To account for this bias and allow for higher probabilities of convergence, we included uniform branch length priors as well as high mean expected branch lengths in the prior sensitivity analyses (for parameters see Table 2). This permitted us to bias the analyses in favour of the 'convergence hypothesis' and to obtain conservative estimates of the posterior probability of the 'early duplication hypothesis'. Additionally, all analyses were performed with the HKY+I+G and GTR+I+G nucleotide substitution models to control for potential saturation effects.

To estimate the time point of gene duplication and the number of losses of the *DAB1* and *DAB2* lineages, we traced their presence and absence on the orders' phylogeny extracted from Hackett *et al.* (2008) using parsimony ancestral state reconstruction in Mesquite 2.6 (Maddison and Maddison 2009).

### *Phylogenetic hypothesis testing and phylogenetic simulations*

The duplication and convergence hypotheses make different predictions about the monophyletic relationships of specific genes and species groups. These predictions can be incorporated into phylogenetic hypothesis testing by constraining the possible topologies to form a single monophyletic group for the owl *DAB1* and *DAB2* genes (for the 'convergence hypothesis'), or the monophyly of the newly arisen gene lineage of owl *DAB1* plus passerines and falcons (for the 'early duplication hypothesis'). To distinguish which of the evolutionary

hypotheses better explains the patterns observed in the avian MHCIIβ, we evaluated their Bayesian posterior probabilities by filtering the Bayesian 95% credible posterior topologies of reconstructions based on third codon positions with the alternative monophyly criteria indicated above. Additionally, we performed approximately unbiased (AU) tests of phylogenetic tree selection (Shimodaira 2002) using CONSEL (Shimodaira and Hasegawa 2001) to identify which hypothesis better explains the data.

Finally, computer simulations were performed to test whether codons evolving under positive selection and convergent evolution could bias the phylogenetic inference towards one of the 2 hypotheses. Trees including the 6 bird orders were created to represent the gene history according to the 'convergence hypothesis' (Figure S2) by following recent phylogenetic studies (Ericson et al. 2006; Brown, Payne, and Mindell 2007; Hackett et al. 2008). The relationships within the owl genes and within the falcon and passerine clades were left unresolved, leaving only a backbone topology. The relationships within these 3 clades and the branch lengths were inferred using PAUP 4.0 beta 10 (Swofford 2003) by forcing the heuristic search to keep the corresponding clades. The program EVOLVER from the PAML 4.0 package (Yang 2007) was modified to allow convergent evolution of codons, and sequences of 72 codons were simulated. Positive selection or convergent evolution was simulated to happen along the branch subtending the owl *DAB1* lineage and along the branch leading to the falcon+passerine clade (Figure S2). The number of sites under positive or convergent evolution was changed from 2 to 30 by steps of 2, and the ω parameter was changed to 5, 25, 50 and 75 for the positive selection simulations. The simulated data sets were analysed with PhyML 3.0 (Guindon and Gascuel 2003) with a GTR+G model of evolution. Each simulation was replicated 100 times for the full simulated coding sequences. The proportion of trees finding the topologies unbiased by convergent selection including the 'owl *DAB1* + owl *DAB2*' clade, and the proportion of topologies biased by selection including an 'owl *DAB1* + falcons + passerines' clade were evaluated in R using the ape package (Paradis, Claude, and Strimmer 2004).

***Maximum-likelihood analyses of positive selection***

To evaluate whether positive selection was involved in the evolution of the avian MHCIIβ and to identify codon sites that underwent adaptive evolution, we conducted maximum likelihood analyses of positive selection as implemented in the CodeML software contained in the PAML 4.0 package (Yang 2007). Three different codon models (Yang and Nielsen 2002; Zhang, Nielsen, and Yang 2005) were optimized on the topology obtained from Bayesian reconstructions based on third codon positions. The nearly-neutral site model M1a allows ω ($d_N/d_S$) to vary among amino acid positions, but keeps it constant over the topology. Only neutral evolution ($\omega_1=1$) and purifying selection ($0<\omega_0<1$) are allowed in this model. The alternative branch-site model A of positive selection allows ω to vary both among sites and branches. Sites in pre-defined foreground branches are allowed to evolve under positive selection ($\omega_2>1$), while the background branches evolve neutrally ($\omega_1=1$) or under purifying selection ($0<\omega_0<1$). The neutral branch-site model A' is identical to the branch-site model of selection, but does not allow for positive selection ($\omega_2=1$).

The branch-site models A and A' require an *a priori* definition of the branches which positive selection may act on. Because we were interested in the role of positive selection in lineage divergence, the branch separating the *DAB1* from the *DAB2* lineage was fixed as the foreground branch. Analyses were conducted using the highest posterior probability Bayesian topology, the maximum likelihood consensus topology as well as a topology compiled from phylogenetic studies (Figure S3). For the latter, the relationships among orders and families were extracted from recent avian phylogenies (Ericson et al. 2006; Brown, Payne, and Mindell 2007; Hackett et al. 2008). Only the orders included in our analyses were considered and the duplication introduced at the node leading to the neoavian lineages. Relationships within orders and families were retrieved from phylogenetic studies conducted in the respective taxa (Wink and Heinrich 1999; Lerner and Mindell 2005; Ksepka, Bertelli, and Giannini 2006; Kimball and Braun 2008; Treplin et al. 2008). Analyses of positive selection have been shown to be insensitive to minor

errors in tree topologies (Yang, Wong, and Nielsen 2005). Nonetheless, to exclude the possibility that the results were due to a bias in the topology, the analyses were repeated using 10 random topologies from reconstructions from third codon positions filtered for *DAB1* monophyly. The branch-site model A of positive selection was tested against the neutral models M1a ($\chi^2$, d.f.=2) and A' ($\chi^2$, d.f.=1) using likelihood ratio tests. Significance of sites under positive selection was estimated using Bayes Empirical Bayes inference (Yang, Wong, and Nielsen 2005). Significantly higher likelihood of model A over A' accounts for the possibility of relaxed purifying selection mimicking positive selection (Yang, Wong, and Nielsen 2005).

***Homology modeling, electrostatic charge computation and visualization***

To have a more precise idea of the functional changes of the divergent and positively selected residues, we performed structural studies on the ancestral amino acid sequences of each order and for each ancestral node back to the point of gene duplication. Although we excluded the great snipe (the only shorebird) from phylogenetic analyses (see above), it was included in the reconstruction of ancestral sequences for two reasons. First, the inclusion of the great snipe did not negatively affect these analyses, and second, it is the only species apart from owls possessing genes of both lineages. All sequences encompassed 72 of the 90 codons of exon 3, except for the great snipe where only 55 codons were available. The ancestral sequences were reconstructed based on the compiled phylogeny as described above (Figure S3), and nucleotide-based analyses of codon sequences were performed using the BaseML software contained in the PAML 4.0 package (Yang 2007). This procedure estimates parameters under a nucleotide substitution model (HKY with different parameters for each codon position; option GC, model=4, Mgene=4), but the reconstructed nucleotide triplets are treated as codons to infer the most likely amino acid encoded (Yang 2007).

The three-dimensional structures of the amino acid sequences were reconstructed by homology modelling based on the human MHCII structure (PDB code: 2ICW). The three-dimensional

models were obtained by homology using MODELLER 9v6 (Sali and Blundell 1993). After homology modeling, slight optimizations and the addition of hydrogen atoms were performed using the PDB2PQR software release 1.2.1 (Dolinsky et al. 2004) and the implemented force field *amber*. Finally, the electrostatic potential maps were computed using the Adaptive Poisson-Boltzmann Solver (APBS) release 1.0.0 (Baker et al. 2001) by solving the nonlinear (full) Poisson-Boltzmann equation (option npbe) under implicit water solvent (dielectric constant sdie=80). The visualizations of the electrostatic potential maps were performed in the molecular viewer PyMOL (DeLano 2002). To identify the positions of the divergent and positively selected residues in the MHCII dimer, we retrieved the MHCII α- (183 amino acid residues) and MHCII β-chain sequences (192 amino acid residues) of chicken from GenBank (accession numbers ABU88979.1 and ABU89001.1, respectively). We modeled the interaction surface of this MHCII αβ-heterodimer directly based on the human MHCII structure (PDB: 2ICW). To find the residues participating in the interaction between the $\alpha$- and $\beta$-chains monomers of chicken, we used the Protein Interfaces, Surfaces and Assemblies Service (PISA) at the European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html; Krissinel and Henrick 2007)

## Results/Discussion

### *Early gene duplication during avian MHC evolution*

We adopted 3 different approaches to distinguish between the competing 'convergence hypothesis' and 'early duplication hypothesis' for the evolution of avian MHCIIβ exon 3 lineages. We (i) contrasted the functional phylogenetic relationships inferred from amino acid and codon sequences with the relationships revealed by third codon positions that can be considered as nearly-neutral, (ii) evaluated the plausibility of both hypotheses using Bayesian and maximum likelihood phylogenetic hypothesis testing, and (iii) performed computer simulations to test the presence of phylogenetic bias towards a specific hypothesis.

The ‘convergence hypothesis’ predicts that positive selection can produce functionally similar proteins in different lineages. In this case the phylogenetic relationships inferred from amino acid or codon sequences are biased (Edwards 2009), reflecting functional similarity arisen by selection, while the third codon positions are expected to trace the neutral gene history (Christin et al. 2007). Convergence thus is predicted to produce conflicting phylogenetic signals between functional and neutral genetic variation (Christin et al. 2007; Castoe et al. 2009). By contrast, under the ‘early duplication hypothesis’ signals of functional and neutral genetic variation are predicted to be consistent. In the avian MHCIIβ, our phylogenetic reconstructions based on the amino acid or codon sequences revealed a monophyletic clade (hereafter named *DAB1* lineage) including passerines, falcons and the owl *DAB1* gene (Figure 1A). Consistent with predictions derived from the ‘early duplication hypothesis’, the same evolutionary relationships were also recovered from third codon positions (Figure 1B).

The monophyly of the *DAB1* lineage was consistent among all analyses and highly significant for functional data, but the low number of third codon positions available weakened the statistical support for *DAB1* monophyly in third codon positions (Figure 1). Bayesian phylogenetic hypothesis testing showed that the ‘early duplication hypothesis’, involving a monophyletic *DAB1* lineage, was significantly better supported by third codon positions (median $PP_{dup}$=0.5789) than a scenario of convergent evolution involving a monophyletic owl MHCIIβ clade (median $PP_{con}$=0.0029). This result was not dependant on branch length prior distributions, and the ‘convergence hypothesis’ could be significantly rejected under all prior settings ($PP_{con}$<0.05, Table 2). Maximum likelihood testing of the alternative hypotheses confirmed this conclusion (AU test, $p_{con}$=0.012). Furthermore, in Bayesian analyses extreme priors had the tendency to increase the probability of the ‘convergence hypothesis’, while for the ‘early duplication hypothesis’ posterior probabilities were more stable (Table 2). This suggests that the posterior probability was preferentially driven by the likelihood of the data and that the weight of the data is larger for the ‘early duplication hypothesis’.

The phylogenetic relationships congruent with the 'convergence hypothesis' were used to simulate DNA data under different levels of positive selection and convergent evolution. These simulations showed that if the owl *DAB1* genes and the genes of falcons and passerines evolved independently according to the 'convergence hypothesis', divergent positive selection alone is not able to introduce sufficient bias into the reconstruction to group the owl *DAB1* genes with falcon and passerine sequences (Figure 2). Simulations of convergent evolution showed that with only 3 convergent residues, such as observed in our data (Figures S1 and S4; residues 101, 120, 121), the probability to bias the phylogenetic reconstructions was lower than 0.09 (Figure 2C). This probability increased to 0.6 with 12 convergent residues and to 0.9 with 30 residues (Figure 2C). This suggests that molecular convergence can introduce a bias in the phylogenetic reconstructions of the avian MHCIIβ, which may lead to misinterpreting the similarities between the owl, falcons and passerines DAB1 genes as a sign of their common ancestry. However, our simulations show that this is only the case when a large amount, far larger than observed in our data, of convergent evolution is present. More importantly, our simulations also demonstrate that only phylogenetic reconstructions based on full coding sequences will be biased whereas those based on third codon positions, on which we base our argument against convergent evolution, were far less affected (Figure 2).

Based on these results, we conclude that the 'convergence hypothesis' that would have implied the action of common selective pressures on exon 3 in a number of species acting on a recent evolutionary timescale is highly unlikely, and that the two avian MHCIIβ lineages recovered by our phylogenetic analyses (Figure 1) resulted from an ancestral duplication. While previous studies that reported orthologous relationships among avian MHC genes either included only closely related species (Wittzell et al. 1999; Miller and Lambert 2004; Strand et al. 2007) or did not exclude convergent evolution (Burri et al. 2008b), the present study provides compelling evidence for avian MHC orthologs that have been maintained over a greater time span of avian evolution. As the evolutionary relationships revealed from exon 3 suffer much less

from the distorting effects of diversifying selection and gene conversion than regions directly interacting with pathogens (see Materials and Methods), we expect the phylogenetic relationships inferred from exon 3 to represent the evolutionary history of MHCIIβ genes. More data from other exons, UTRs, other MHC genes, and ultimately genomic structure are awaited to get deeper insights into the long-term evolutionary history of the avian MHC.

### *Birth-death evolution in the avian MHC*

Today's avian *DAB1* and *DAB2* lineages appear to be the result of a single duplication event. This implies that *DAB1* and *DAB2* emerged early during avian evolution, with the duplication event predating the major avian radiations that took place around 100 mya (Hedges and Kumar 2009). Our results from the avian MHCIIβ closely resemble the patterns of birth-death evolution observed in the mammalian MHC (Ohno 1970; Nei and Hughes 1992), namely (i) the presence of a single *DAB* lineages in most species, which is due to independent losses of either *DAB* lineage during avian radiation, and (ii) the divergence of paralogs that followed duplication (see below). To reconstruct the patterns of birth and death in the *DAB* lineages, we estimated the number of losses of either gene lineage using ancestral state reconstruction. Our results show that a birth-death scenario of the evolution of the *DAB* lineages requires a single duplication event preceding the major avian radiations about 100 mya, followed by independent losses of *DAB1* about 50-80 mya in birds of prey, and of *DAB2* about 70-110 mya in the ancestor of falcons and passerines (Figure 3) (Brown, Payne, and Mindell 2007). These timescales closely match those observed in the mammalian MHCIIβ, where orthologs of the *DR*, *DQ*, *DP* and *DO* isoforms are found both in humans and rodents, which diverged about 90 mya (Hedges and Kumar 2009).

These results suggest that the processes of birth and death in the long-term evolution of the MHC occurred over similar timescales in birds and mammals. The small number of genes found in older taxa, like chicken (Kaufman et al. 1999b), black grouse (Strand et al. 2007) and

penguins (Tsuda et al. 2001; Kikkawa et al. 2005) may indeed reflect the ancestral state of birds. However, results from ancestral state reconstruction show that in at least part of the younger bird taxa MHC evolution involved reductions of gene number subsequent to duplication (Figure 3). Notably, the small number of genes present in falcons and birds of prey (Alcaide, Edwards, and Negro 2007; Alcaide et al. 2008) appear to have evolved through secondary reductions of gene numbers.

Clearly, these results are based on the assumption that none of the avian *DAB* lineages remained undetected in the analyzed species. Two major problems may lead to missing lineages. First, gene conversion, resulting in non-independent sequence evolution, among genes of the *DAB1* and *DAB2* lineages appears to have led to mixed lineage-specific signatures in barn owl and great snipe (Burri et al. 2008a; Burri et al. 2008b). While in these 2 cases, traces of both lineages can still be observed, gene conversion events of longer sequences could have removed all lineage-specific phylogenetic signals, thus precluding the detection of both gene lineages in other taxa. Based on the current data, we can not exclude this possibility. Second, in most bird species, isolation and characterization of the MHC rely on PCR-based approaches (Table 1), which may indeed miss whole gene lineages (Wagner et al. 1994). This is particularly true for pseudo-genes that may show elevated levels of divergence and in the presence of copy number variation between individuals, which has been demonstrated in Japanese quail (Hosomichi et al. 2006; Hosomichi et al. 2008). If *DAB1* was present in birds of prey or *DAB2* in passerines and falcons, the number of inferred gene losses would be reduced, while the presence of *DAB1* in basal taxa would shift the birth of the new lineage to a more ancient timescale. Indeed, if both lineages were present in basal avian taxa, this could suggest orthology of avian *DAB* lineages with mammalian MHC isoforms. Whereas sequence features of the only isolated avian MHCIIα sequence from chicken support orthology with the mammalian DR isoform (Salomonsen et al. 2003), evidence of orthologous relationships with mammalian isoforms from MHCIIβ sequences is very limited (Salomonsen et al. 2003). Avian MHCIIβ sequences usually cluster among birds

and do not intermix with mammalian sequences in phylogenetic reconstructions; a pattern that we could also confirm by phylogenetic analyses of codon and amino acid sequences as well as third codon positions when sequences of the human MHCIIβ isoforms were included in our set of avian MHC sequences (data not shown).

A comprehensive picture of birth-death evolution in avian MHC will require confirming the number of *DAB* lineages present in various avian taxa as well as the origins of the avian MHCIIβ lineages that we propose here. This could be achieved by isolation and phylogenetic analyses of avian MHCIIα sequences and in-depth studies of MHC architecture including information on intron structure and pseudo-genes.

### *Adaptive divergence between paralogs*

According to Ohno's (1970) model of multigene family evolution, subsequent to duplication one paralog retains the original gene function, while the others are free to evolve new function (neofunctionalization). In the avian MHC, the amino acid divergence between *DAB1 and DAB2* (Figure 1A) suggests that adaptive evolution involving positive Darwinian selection generated functional divergence at the origin of the *DAB1* lineage. The hallmark of positive selection is an excess of non-synonymous (amino acid altering) substitutions relative to synonymous substitutions (which retain the encoded amino acid). In the avian MHCIIβ, 155 sites are non-synonymous and 49 are synonymous between the *DAB1* and *DAB2* lineages. Maximum likelihood analyses revealed that the non-synonymous substitution rate ($d_N$) significantly exceeds the synonymous substitution rate ($d_S$, $\omega=d_N/d_S>1$) at a number of sites on the branch leading to the *DAB1* lineage, thereby supporting a significant impact of positive selection in the evolution of *DAB1* (branch site models A vs. A', $p<0.01$, $\omega=52$, with the origin of *DAB1* fixed as foreground branch; Table S1). Two amino acid residues (100 and 118) were identified to be under positive selection with high posterior probability (PP>0.95, Table S1), while 6 other residues, hereafter referred to as divergent residues, are potentially under positive selection (PP>0.50, Table S1).

These results imply that adaptive amino acid divergence occurred in *DAB1* following duplication, while *DAB2* remained similar to the ancestral avian MHCIIβ sequence. Six of the 8 divergent or positively selected residues resulted in substitutions that changed the physiochemical properties of the amino acids (Figures S4 and 4A). This reinforces the idea of an ancestral specialisation through neofunctionalization of the *DAB1* lineage in the avian MHCIIβ. It appears that, as proposed by Ohno's (1970) models of multigene family evolution, MHCIIβ duplication in an avian ancestor relaxed one of the MHCIIβ paralogs (*DAB1*) from evolutionary constraints, allowing it to diverge, while the other paralog (*DAB2*) maintained the original MHC function. Nonetheless, an important fraction of recent bird species carry only the derived *DAB1* paralog, which suggests that the functional divergence of *DAB1*, at least in the form we can observe today, did not impair the original MHC function.

To gain a more precise idea of the potential functional changes between the proteins encoded by the *DAB1* and *DAB2* genes, we studied their three-dimensional structure. All but one divergent residue (residue 101) are located on the protein surface and form an electrostatic patch (Figure 4). Electrostatic charges differ substantially between the *DAB1* and *DAB2* molecules, with predominantly acidic residues in *DAB1* versus basic residues in *DAB2* (Figure 4A). The *DAB2* lineage exhibits the same physiochemical properties as the proteins inferred to be ancestral, which include proteins of wildfowl and penguins (Figure 4A; the latter 2 proteins are not shown as they are identical to the ancestral protein). This illustrates the physiochemical changes involved in the functional evolution of the avian MHCIIβ, and reconfirms the conclusion of adaptive divergence of *DAB1* and conserved evolution of *DAB2*.

At the cell-surface, functional MHCII proteins are expressed as heterodimers composed of one α- and one β-chain. Given the radical change from basic residues in *DAB2* to acidic ones in the *DAB1* β-chain, we were interested in locating the divergent region in the mature αβ-heterodimer. Analysis of the three-dimensional structure of the chicken αβ-heterodimer shows that the divergent residues in the β-chain are situated in a region facing the α-chain (Figure 4B).

Moreover, the interacting α- and β-chains in the chicken αβ-heterodimer show complementary physiochemical properties in this region (Figure 4B). Detailed analyses of the nature of interactions between the chains show that 3 residues (βR100, βY118, βT120) engage in hydrogen bonds between the α- and β-chains' residues (Figure 4C). Analyses using the PISA interface show that the NH1 group of the arginine βR100 is able to form salt bridges with the carboxyl group of the two glutamic acids αE95 (3.29Å) and αE101 (3.89Å) (Figure 4C). The hydroxyl group of the tyrosine βY118 forms a hydrogen bond with the carboxyl groups of the proline αP93 (3.79Å) and alanine αA94 (2.94Å). Similarly, threonine βT120 forms a hydrogen bond of 2.81Å with the carboxyl group of alanine αA94. Interestingly, residue β112 is involved in the formation of homodimers of heterodimers (tetramers composed of 2 hetero-dimers) during T-cell presentation in the human DQ (Bondinas, Moustakas, and Papadopoulos 2007). No MHCIIα sequence of species possessing the *DAB1* lineage has been isolated to date (including in the ongoing genome sequencing of the passerine model species zebra finch, *Taeniopygia guttata*), and for the time being we can only hypothesize about how the changed physiochemical properties in the *DAB1* β-chain affect the αβ-interaction and the stability of the αβ-heterodimer. For the murine *I-A* MHCII gene and the human DQ isoform, it was shown that mismatching of the interacting α- and β-chains leads to weakened stability and reduced cell-surface expression of αβ-heterodimers (e.g. Germain, Bentley, and Quill 1985; Braunstein and Germain 1987; Sant, Braunstein, and Germain 1987; Sant et al. 1991; Kwok et al. 1993). Residues that are critical determinants constraining stable cell-surface expression of mouse I-A and human DQ molecules were mainly pinpointed to the 5' region of exon 2 (Sant, Braunstein, and Germain 1987; Kwok et al. 1993; Bondinas, Moustakas, and Papadopoulos 2007), but additional residues in the 3' region of exon 3 - the region divergent between avian *DAB1* and *DAB2* - appears to play a role for heterodimer stability in the human DQ (Kwok et al. 1993). In mice and primates, the α- and β-chains of I-A and DQ heterodimers are supposed to co-evolve (Braunstein and Germain 1987; Kwok et al. 1993; Sauermann 1998). Based on this observation, we hypothesize that the radical

change in physiochemical properties in the divergent patch between avian *DAB1* and *DAB2* has been accompanied by coevolution of the interacting residues of the α-chains, conferring the interacting chains with complementery physiochemical properties to guarantee heterodimer stability and efficient cell-surface expression. In species carrying *DAB* genes from both lineages we thus would predict the presence of 2 *DAA* genes encoding α-chains with physiochemical properties matching the respective β-chains. Coevolution in this specific case may however not be assisted by recombinational proximity of the MHCIIα and MHCIIβ genes, such as suggested by Kaufman (1999) for other genes within MHC haplotypes, because the chicken MHCIIα and MHCIIβ genes are not tightly linked (Salomonsen et al. 2003) and synteny is usually highly conserved among birds (e.g. Backstrom et al. 2008).

## Conclusions and perspectives

In this study we provide new insights into the long-term evolutionary history of the avian MHC, and provide a hypothesis of how molecular coevolution may have underpinned the divergence and neofunctionalization of an ancestral MHCIIβ duplicate. According to our results, 2 avian MHCIIβ lineages emerged by a unique duplication event preceding the major avian radiations. Subsequently, the *DAB1* lineage underwent neofunctionalization by diverging from the ancestral lineage through positive selection. This divergence includes a shift from basic to acidic composition in a region of the protein that holds residues potentially involved in the interaction with the MHCII α-chain and in the formation of homodimers of heterodimers during T-cell binding. Given the positions of the divergent residues in the MHCII αβ-heterodimer, we thus hypothesize that this divergence involved the coevolution of MHCIIβ-encoded β-chains with their interacting α-chains.

Our findings are of particular interest to the understanding of the evolutionary history of the avian MHC and, by providing essential information on long-term gene history in the avian MHC,

open promising perspectives for advances in the understanding of the evolution of avian MHC architecture and avian disease resistance. Isolation of avian MHCIIα genes will have to confirm our hypothesis of αβ-chain coevolution. A comparative approach integrating a broad taxon sampling of birds may show whether possessing only ancestral *DAB2*, only derived *DAB1*, or both lineages provides taxa with adaptive advantages. From a population genomic perspective on multigene family evolution, it will be interesting to see whether selection is responsible for the simultaneous presence of ancestrally and secondarily reduced simple avian MHC architectures, or whether neutral processes such as differing effective population sizes, population dynamics or species histories may be sufficient to explain the varying degrees of complexity observed in avian MHC architectures today.

## Supplementary material

Supplementary table S1 and figures S1-S4 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

## Acknowledgements

We thank Seraina Klopfstein and Pascal-Antoine Christin for many helpful discussions. Seraina Klopfstein, Marc Robinson-Rechavi, Scott V. Edwards, John Wang and three anonymous reviewers provided precious comments on an earlier version of this manuscript. The computations were performed at the Vital-IT (http://www.vital-it.ch) Center for high-performance computing of the Swiss Institute of Bioinformatics. This study was supported by Swiss National Science Foundation grants 3100A0-109852/1 to LF, PPOOA-102913 to AR and 3100A0-116412 to NS.

## References


Abascal, F., R. Zardoya, and D. Posada. 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics **21**:2104-2105.

Aguilar, A., S. V. Edwards, T. B. Smith, and R. K. Wayne. 2006. Patterns of Variation in MHC Class II {beta} Loci of the Little Greenbul (Andropadus virens) with Comments on MHC Evolution in Birds. J Hered **97**:133-142.

Akaike, H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control **19**:716-723.

Alcaide, M., S. Edwards, and J. Negro. 2007. Characterization, Polymorphism, and Evolution of MHC Class II B Genes in Birds of Prey. Journal of Molecular Evolution **65**:541-554.

Alcaide, M., S. V. Edwards, L. Cadahía, and J. J. Negro. 2008. MHC class I genes of birds of prey: isolation, polymorphism and diversifying selection. Conservation Biology.

Apanius, V., D. Penn, P. R. Slev, L. R. Ruff, and W. K. Potts. 1997. The nature of selection on the major histocompatibility complex. Critical Reviews in Immunology **17**:179-224.

Backstrom, N., N. Karaiskou, E. H. Leder, L. Gustafsson, C. R. Primmer, A. Qvarnstrom, and H. Ellegren. 2008. A Gene-Based Genetic Linkage Map of the Collared Flycatcher (Ficedula albicollis) Reveals Extensive Synteny and Gene-Order Conservation During 100 Million Years of Avian Evolution. Genetics **179**:1479-1495.

Baker, N. A., D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci U S A **98**:10037-10041.

Bernatchez, L., and C. Landry. 2003. MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? Journal of Evolutionary Biology **16**:363-377.

Bondinas, G., A. Moustakas, and G. Papadopoulos. 2007. The spectrum of HLA-DQ and HLA-DR alleles, 2006: a listing correlating sequence and structure with function. Immunogenetics **59**:539-553.

Bonneaud, C., J. Burnside, and S. V. Edwards. 2008. High-Speed Developments in Avian Genomics. BioScience **58**:587-596.

Braunstein, N. S., and R. N. Germain. 1987. Allele-specific control of Ia molecule surface expression and conformation: implications for a general model of Ia structure-function relationships. Proceedings of the National Academy of Sciences of the United States of America **84**:2921-2925.

Briles, W. E., R. M. Goto, C. Auffray, and M. M. Miller. 1993. A polymorphic system related to but genetically independent of the chicken major histocompatibility complex. Immunogenetics **37**:408-414.

Brown, J. W., R. B. Payne, and D. P. Mindell. 2007. Comment. Nuclear DNA does not reconcile 'rocks' and 'clocks' in Neoaves: a comment on Ericson et al. Biology Letters **3**:257-259.

Burri, R., H. Niculita-Hirzel, A. Roulin, and L. Fumagalli. 2008a. Isolation and characterization of major histocompatibility complex (MHC) class II B genes in the Barn owl (Aves: Tyto alba ). Immunogenetics **60**:543-550.

Burri, R., H. Niculita-Hirzel, N. Salamin, A. Roulin, and L. Fumagalli. 2008b. Evolutionary Patterns of MHC Class II B in Owls and Their Implications for the Understanding of Avian MHC Evolution. Molecular Biology and Evolution **25**:1180-1191.

Castoe, T. A., A. P. J. de Koning, H.-M. Kim, W. Gu, B. P. Noonan, G. Naylor, Z. J. Jiang, C. L. Parkinson, and D. D. Pollock. 2009. Evidence for an ancient adaptive episode of convergent molecular evolution. Proceedings of the National Academy of Sciences **106**:8986-8991.

Christin, P.-A., N. Salamin, V. Savolainen, M. R. Duvall, and G. Besnard. 2007. C4 Photosynthesis Evolved in Grasses via Parallel Adaptive Genetic Changes. Current Biology **17**:1241-1247.

Conant, G. C., and K. H. Wolfe. 2008. Turning a hobby into a job: How duplicated genes find new functions. Nat Rev Genet **9**:938-950.

DeLano, W. L. 2002. The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA.

Dolinsky, T. J., J. E. Nielsen, J. A. McCammon, and N. A. Baker. 2004. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Reserach **32**:W665-667.

Edwards, S., J. Gasper, and M. March. 1998. Genomics and polymorphism of Agph-DAB1, an Mhc class II B gene in red- winged blackbirds (Agelaius phoeniceus). Molecular Biology and Evolution **15**:236-250.

Edwards, S. V. 2009. Natural selection and phylogenetic analysis. Proceedings of the National Academy of Sciences **106**:8799-8800.

Edwards, S. V., C. M. Hess, J. Gasper, and D. Garrigan. 1999. Toward an evolutionary genomics of the avian Mhc. Immunological Reviews **167**:119-132.

Edwards, S. V., E. K. Wakeland, and W. Potts. 1995. Contrasting histories of avian and mammalian Mhc genes revealed by class II B sequences from songbirds. Proceedings of the National Academy of Sciences of the United States of America **92**:12200-12204.

Ekblom, R., M. Grahn, and J. Höglund. 2003. Patterns of polymorphism in the MHC class II of a non-passerine bird, the great snipe (*Gallinago media*). Immunogenetics **54**:734-741.

Ericson, P., C. Anderson, T. Britton, A. Elzanowski, U. Johansson, M. Källersjö, J. Ohlson, T. Parsons, D. Zuccon, and G. Mayr. 2006. Diversification of Neoaves: integration of molecular sequence data and fossils. Biology Letters **2**:543-547.

Felsenstein, J. 2005. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.

Gasper, J. S., T. Shiina, H. Inoko, and S. V. Edwards. 2001. Songbird genomics: Analysis of 45 kb upstream of a polymorphic Mhc class II gene in red-winged blackbirds (Agelaius phoeniceus). Genomics **75**:26-34.

Gaudieri, S., R. L. Dawkins, K. Habara, J. K. Kulski, and T. Gojobori. 2000. SNP Profile within the Human Major Histocompatibility Complex Reveals an Extreme and Interrupted Level of Nucleotide Diversity. Genome Research **10**:1579-1586.

Geraghty, D. E., R. Daza, L. M. Williams, Q. Vu, and A. Ishitani. 2002. Genetics of the immune response: identifying immune variation within the MHC and throughout the genome. Immunological Reviews **190**:69-85.

Germain, R. N., D. M. Bentley, and H. Quill. 1985. Influence of allelic polymorphism on the assembly and surface expression of class II MHC (Ia) molecules Cell **43**:233-242.

Guindon, S., and O. Gascuel. 2003. A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. Syst Biol **52**:696-704.

Hackett, S. J., R. T. Kimball, S. Reddy, R. C. K. Bowie, E. L. Braun, M. J. Braun, J. L. Chojnowski, W. A. Cox, K.-L. Han, J. Harshman, C. J. Huddleston, B. D. Marks, K. J. Miglia, W. S. Moore, F. H. Sheldon, D. W. Steadman, C. C. Witt, and T. Yuri. 2008. A Phylogenomic Study of Birds Reveals Their Evolutionary History. Science **320**:1763-1768.

Hedges, S. B., and S. Kumar. 2009. The Timetree of Life. Oxford University Press, Oxford.

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

Hess, C. M., J. Gasper, H. E. Hoekstra, C. E. Hill, and S. V. Edwards. 2000. MHC Class II Pseudogene and Genomic Signature of a 32-kb Cosmid in the House Finch (Carpodacus mexicanus). Genome Res. **10**:613-623.

Hosomichi, K., M. M. Miller, R. M. Goto, Y. Wang, S. Suzuki, J. K. Kulski, M. Nishibori, H. Inoko, K. Hanzawa, and T. Shiina. 2008. Contribution of Mutation, Recombination, and Gene Conversion to Chicken Mhc-B Haplotype Diversity. J Immunol **181**:3393-3399.

Hosomichi, K., T. Shiina, S. Suzuki, M. Tanaka, S. Shimizu, S. Iwamoto, H. Hara, Y. Yoshida, J. Kulski, H. Inoko, and K. Hanzawa. 2006. The major histocompatibility complex (Mhc)

class IIB region has greater genomic structural flexibility and diversity in the quail than the chicken. BMC Genomics **7**:322.

Hughes, A. L., and M. Yeager. 1998. Natural selection at major histocompatibility complex loci of vertebrates. Annual Review of Genetics **32**:415-435.

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

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

Kaufman, J. 2000. The simple chicken major histocompatibility complex: life and death in the face of pathogens and vaccines. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences **355**:1077-1084.

Kaufman, J., J. Jansen, I. Shaw, B. Walker, S. Milne, S. Beck, and J. Salomonsen. 1999a. Gene organisation determines evolution of function in the chicken MHC. Immunological Reviews **167**:101-117.

Kaufman, J., S. Milne, T. W. F. Gobel, B. A. Walker, J. P. Jacob, C. Auffray, R. Zoorob, and S. Beck. 1999b. The chicken B locus is a minimal essential major histocompatibility complex. Nature **401**:923-925.

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

Kikkawa, E. F., T. T. Tsuda, T. K. Naruse, D. Sumiyama, M. Fukuda, M. Kurita, K. Murata, R. P. Wilson, Y. LeMaho, M. Tsuda, J. K. Kulski, and H. Inoko. 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., T. T. Tsuda, D. Sumiyama, T. K. Naruse, M. Fukuda, M. Kurita, R. P. Wilson, Y. LeMaho, G. D. Miller, M. Tsuda, M. Koichi, J. K. Kulski, and H. Inoko. 2009. Trans-

species polymorphism of the *Mhc* class II *DRB*-like gene in banded penguins (genus *Spheniscus*). Immunogenetics **61**:341-352.

Kimball, R. T., and E. L. Braun. 2008. A multigene phylogeny of Galliformes supports a single origin of erectile ability in non-feathered facial traits. Journal of Avian Biology **39**:438-445.

Klein, J. 1986. Natural history of the major histocompatibility complex. Wiley, New York.

Kolaczkowski, B., and J. W. Thornton. 2007. Effects of Branch Length Uncertainty on Bayesian Posterior Probabilities for Phylogenetic Hypotheses. Molecular Biology and Evolution **24**:2108-2118.

Krissinel, E., and K. Henrick. 2007. Inference of macromolecular assemblies from crystalline state. Journal of Molecular Biology **372**:774-797.

Ksepka, D. T., S. Bertelli, and N. P. Giannini. 2006. The phylogeny of the living and fossil Sphenisciformes (penguins). Cladistics **22**:412-441.

Kulski, J. K., T. Shiina, T. Anzai, S. Kohara, and H. Inoko. 2002. Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man. Immunological Reviews **190**:95-122.

Kumánovics, A., T. Takada, and K. F. Lindahl. 2003. Genomic Organization of the Mammalian MHC. Annual Review of Immunology **21**:629-657.

Kwok, W. W., S. Kovats, P. Thurtle, and G. T. Nepom. 1993. HLA-DQ allelic polymorphisms constrain patterns of class II heterodimer formation. J Immunol **150**:2263-2272.

Lerner, H. R. L., and D. P. Mindell. 2005. Phylogeny of eagles, Old World vultures, and other Accipitridae based on nuclear and mitochondrial DNA. Molecular Phylogenetics and Evolution **37**:327-346.

Lynch, M., and J. S. Conery. 2000. The Evolutionary Fate and Consequences of Duplicate Genes. Science **290**:1151-1155.

Maddison, W. P., and D. R. Maddison. 2009. Mesquite: a modular system for evolutionary analysis. Version 2.6 http://mesquiteproject.org.

Miller, H. C., K. Belov, and C. Daugherty, H. . 2005. Characterization of MHC class II genes from an ancient reptile lineage, Sphenodon (tuatara). Immunogenetics **V57**:883-891.

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

Miller, M. M., R. Goto, A. Bernot, R. Zoorob, C. Auffray, N. Bumstead, and W. E. Briles. 1994. Two Mhc class I and two Mhc class II genes map to the chicken Rfp-Y system outside the B complex. Proceedings of the National Academy of Sciences of the United States of America **91**:4397-4401.

Nei, M., and T. Gojobori. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Molecular Biology and Evolution **3**:418-426.

Nei, M., and A. L. Hughes. 1992. Balanced polymorphism and evolution by the birth-and-death process in the MHC loci. Pp. 27-38 *in* K. Tsuji, M. Aizawa, and T. Sasazuki, eds. 11th Histocompatibility Workshop and Conference. Oxford University Press, Oxford.

Nei, M., and A. P. Rooney. 2005. Concerted and birth-and-death evolution of multigen families. Annual Review of Genetics **39**:121-152.

Nylander, J. A. A. 2004. MrModeltest v2. Evolutionary Biology Centre, Uppsala University.

Ohno, S. 1970. Evolution by Gene Duplication. Springer, New York.

Paradis, E., J. Claude, and K. Strimmer. 2004. APE: analyses of phylogenetics and evolution in R language Bioinformatics **20**:289-290.

Piertney, S. B., and M. K. Oliver. 2006. The evolutionary ecology of the major histocompatibility complex. Heredity **96**:7-21.

Posada, D., and T. Buckley. 2004. Model Selection and Model Averaging in Phylogenetics: Advantages of Akaike Information Criterion and Bayesian Approaches Over Likelihood Ratio Tests. Systematic Biology **53**:793-808.

Posada, D., and K. A. Crandall. 2002. The Effect of Recombination on the Accuracy of Phylogeny Estimation. Journal of Molecular Evolution **54**:396-402.

Rambaut, A., and A. J. Drummond. 2007. Tracer v1.4, Available from http://beast.bio.ed.ac.uk/Tracer

Robinson, J., M. J. Waller, P. Parham, N. de Groot, R. Bontrop, L. J. Kennedy, P. Stoehr, and S. G. Marsh. 2003. IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. Nucleic Acids Reserach **31**:311 - 314.

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

Sali, A., and T. L. Blundell. 1993. Comparative protein modelling by satisfaction of spatial restraints. Journal of Molecular Biology **234**:779-815.

Salomonsen, J., D. Marston, D. Avila, N. Bumstead, B. Johansson, H. Juul-Madsen, G. D. Olesen, P. Riegert, K. Skjødt, O. Vainio, M. V. Wiles, and J. Kaufman. 2003. The properties of the single chicken MHC classical class II a chain (B-LA) gene indicate an ancient origin for the DR/E-like isotype of class II molecules. Immunogenetics **55**:605-614.

Sant, A. J., N. S. Braunstein, and R. N. Germain. 1987. Predominant role of amino-terminal sequences in dictating efficiency of class II major histocompatibility complex alpha beta dimer expression. Proceedings of the National Academy of Sciences of the United States of America **84**:8065-8069.

Sant, A. J., L. R. Hendrix, J. E. Coligan, W. L. Maloy, and R. N. Germain. 1991. Defective intracellular transport as a common mechanism limiting expression of inappropriately paired class II major histocompatibility complex alpha/beta chains. J. Exp. Med. **174**:799-808.

Sauermann, U. 1998. DQ-haplotype analysis in rhesus macaques: implications for the evolution of these genes. Tissue Antigens **52**:550-557.

Shaw, I., T. J. Powell, D. A. Marston, K. Baker, A. van Hateren, P. Riegert, M. V. Wiles, S. Milne, S. Beck, and J. Kaufman. 2007. 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 **178**:5744-5752.

Shiina, T., S. Shimizu, K. Hosomichi, S. Kohara, S. Watanabe, K. Hanzawa, S. Beck, J. K. Kulski, and H. Inoko. 2004. Comparative Genomic Analysis of Two Avian (Quail and Chicken) MHC Regions. Journal of Immunology **172**:6751-6763.

Shimizu, S., T. Shiina, K. Hosomichi, S. Takahashi, T. Koyama, T. Onodera, J. K. Kulski, and H. Inoko. 2004. MHC class IIB gene sequences and expression in quails ( Coturnix japonica) selected for high and low antibody responses. Immunogenetics **56**:280-291.

Shimodaira, H. 2002. An Approximately Unbiased Test of Phylogenetic Tree Selection. Systematic Biology **51**:492-508.

Shimodaira, H., and M. Hasegawa. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics **17**:1246-1247.

Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics **22**:2688-2690.

Strand, T., H. Westerdahl, J. Höglund, R. V. Alatalo, and H. Siitari. 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.

Swofford, D. L. 2003. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Tamura, K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. Molecular Biology and Evolution **24**:1596-1599.

Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Reserach **22**:4673 - 4680.

Treplin, S., R. Siegert, C. Bleidorn, H. S. Thompson, R. Fotso, and R. Tiedemann. 2008. Molecular phylogeny of songbirds (Aves: Passeriformes) and the relative utility of common nuclear marker loci. Cladistics **24**:328-349.

Tsuda, T., M. Tsuda, T. Naruse, H. Kawata, A. Ando, T. Shiina, M. Fukuda, M. Kurita, I. LeMaho, J. Kulski, and H. Inoko. 2001. Phylogenetic analysis of penguin (*Spheniscidae*) species based on sequence variation in MHC class II genes. Immunogenetics **53**:712-716.

Vincek, V., D. Klein, R. Graser, F. Figueroa, C. O'hUigin, and J. Klein. 1995. Molecular cloning of major histocompatibility complex class II B gene cDNA fom the Bengalese finch *Lonchura striata*. Immunogenetics **42**:262-267.

Wagner, A., N. Blackstone, P. Cartwright, M. Dick, B. Misof, P. Snow, G. Wagner, J. Bartels, M. Murtha, and J. Pendleton. 1994. Surveys of gene families using polymerase chain reaction: PCR selection and PCR drift. Systematic Biology **43**:250-261.

Westerdahl, H., H. k. Wittzell, and T. r. von Schantz. 2000. Mhc diversity in two passerine birds: no evidence for a minimal essential Mhc. Immunogenetics **52**:92-100.

Wink, M., and P. Heinrich. 1999. Molecular evolution and systematics of owls (Strigiformes). Pp. 39-57 *in* C. König, F. Weick, and J. H. Beckin, eds. Owls of the World. Pica Press.

Wittzel, H., T. von Schantz, R. Zoorob, and C. Auffray. 1994. Molecular characterisazion of three *Mhc* class II B haplotypes in the ring-necked pheasant. Immunogenetics **39**:395-403.

Wittzell, H., A. Bernot, C. Auffray, and R. Zoorob. 1999. Concerted evolution of two Mhc class II B loci in pheasants and domestic chickens. Molecular Biology and Evolution **16**:479-490.

Yang, Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Molecular Biology and Evolution **24**:1586-1591.

Yang, Z., W. S. W. Wong, and R. Nielsen. 2005. Bayes Empirical Bayes Inference of Amino Acid Sites Under Positive Selection. Molecular Biology and Evolution **22**:1107-1118.

Yang, Z. H., and R. Nielsen. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Molecular Biology and Evolution **19**:908-917.

Yeager, M., and A. L. Hughes. 1999. Evolution of the mammalian MHC: natural selection, recombination, and convergent evolution. Immunological Reviews **167**:45-58.

Zhang, J. Z., R. Nielsen, and Z. H. Yang. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Molecular Biology and Evolution **22**:2472-2479.

**Table 1** Bird species used in phylogenetic analyses and ancestral sequence reconstruction. The number of known MHCIIB genes, gene nomenclature from GenBank, methods of MHC analyses for each species and relevant references are provided.

| Order / Family | Species | Accession Number | Gene nomenclature | No. Genes | Methods MHC analysis | References |
|---|---|---|---|---|---|---|
| **Wildfowl** Galloanseres | *Anas platyrhynchos* | AF390589 | - | - | - | Chan et al., unpublished |
| | *Cairina moschata* | DQ490138 | - | - | - | Mannes & Schulz, unpublished |
| | *Coturnix japonica* | AB265805 | CojaII-14 | 7 | cosmid (180 kb) | Shiina et al (2004) |
| | *Gallus gallus* | AL023516 | B-LB1 | 4, 2 of which pseudo-genes | cosmid (92 kb) | Kaufman et al. (1999b), Miller et al.(1994) |
| | *Numida meleagris* | EF643464 | B-LB2 | - | - | Kumar et al., unpublished |
| | *Phasianus colchicus* | AJ224349 | DAB1 | 2 | RFLP, cDNA library | Wittzel et al. (1994; 1999) |
| **Penguins** Sphenisciformes | *Eudyptula minor* | AB302187 | - | 1 | PCR | Tsuda et al. (2001) |
| | *Spheniscus demersus* | AB301946 | - | 1 | PCR | Kikkawa et al. (2009) |
| | *Spheniscus magellanicus* | AB325529 | - | 1 | PCR | Kikkawa et al. (2009) |
| | *Spheniscus mendiculus* | AB302191 | - | 1 | PCR | Kikkawa et al. (2009) |
| **Shorebirds** Charadriiformes | *Gallinago media* | AF485404, AF485405 | - | >3 | PCR, RFLP | Ekblom et al. (2003) |
| **Birds of prey** Accipitridae | *Accipiter gentilis* | EF370953 | *Acge*-DRB | 1 | PCR | Alcaide et al. (2007) |
| | *Aegypius monachus* | EF370954 | *Aemo*-DRB | 1 | | |
| | *Aquila chrysaetos* | EF370955 | *Aqch*-DRB | 2 | | |
| | *Buteo buteo* | EF370956 | *Buteo*-DRB | 1 | | |
| | *Circaetus gallicus* | EF370957 | *Ciga*-DRB | 2 | | |
| | *Circus aeruginosus* | EF370958 | *Ciae*-DRB | 2 | | |
| | *Elanus caeruleus* | EF370959 | *Elca*-DRB | 2 | | |
| | *Gypaetus barbatus* | EF370891 | *Gyba*-DRB | 1 | | |
| | *Gyps africanus* | EF370973 | *Gyaf*-DRB | 2 | | |
| | *Gyps coprotheres* | EF370978 | *Gyco*-DRB | 2 | | |
| | *Harpyhaliaeetus coronatus* | EF370961 | *Haco*-DRB | 2 | | |
| | *Hieraaetus pennatus* | EF370962 | *Hipe*-DRB | 2 | | |
| | *Milvus milvus* | EF370963 | *Mimi*-DRB | 1 | | |
| | *Neophron percnopterus* | EF370964 | *Nepe*-DRB | 2 | | |
| | *Pandion halieatus* | EF370965 | *Paha*-DRB | 1 | | |
| **Owls** Strigiformes | *Aegolius funereus* | EF641252, EF641253 | *Aefu*-DAB1, -DAB2 | 2 | RT-PCR (Southern blot [b]) | Burri et al. (2008b), Alcaide et al. (2007) |
| | *Asio flammeus* | EF641250, EF641251 | *Asfl*-DAB1, -DAB2 | 2 | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | *Asio otus* | EF641223, EF641224 | *Asot*-DAB1, -DAB2 | 2 [a] | | |
| | *Athene noctua* | EF641247, EF641248 | *Atno*-DAB1, -DAB2 | 4 [a] | | |
| | *Bubo africanus* | EF641244, EF641245 | *Buaf*-DAB1, -DAB2 | 2 | | |
| | *Bubo bubo* | EF641236, EF641238 | *Bubu*-DAB1, -DAB2 | 3 [a] | | |
| | *Bubo scandiacus* | EF641233, EF641235 | *Busc*-DAB1, -DAB2 | 2 | | |
| | *Glaucidium passerinum* | EF641231 | *Glpa*-DAB1 | 1 | | |
| | *Otus scops* | EF641257, EF641259 | *Otsc*-DAB1, -DAB2 | 3 [a] | | |
| | *Strix aluco* | EF641254, EF641256 | *Stal*-DAB1, -DAB2 | 3 [a] | | |
| | *Strix nebulosa* | EF641240, EF641241 | *Stne*-DAB1, -DAB2 | 2 | | |
| | *Strix uralensis* | EF641242, EF641243 | *Stur*-DAB1, -DAB2 | 2 | | |
| | *Surnia ulula* | EF641226, EF641230 | *Suul*-DAB1, -DAB2 | 3 | | |
| **Falcons** Falconidae | *Falco biarmicus* | EF370989 | *Fabi*-DRB | 1 | PCR | Alcaide et al. (2007) |
| | *Falco femoralis* | EF370988 | *Fafe*-DRB | 1 | | |
| | *Falco naumanni* | EF370986 | *Fana*-DRB | 1 | | |
| | *Falco peregrinus* | EF370985 | *Fape*-DRB | 1 | | |
| | *Falco tinnunculus* | EF370987 | *Fati*-DRB | 1 | | |
| **Passerines** Passeriformes | *Acrocephalus arundinaceus* | AJ404376 | - | 4, 9 | cDNA library, RFLP | Westerdahl et al. (2000) |
| | *Agelaius phoeniceus* | AF030997 | *Agph*-DAB1 | 3, 4-6 | cosmid (45 kb), Southern blot | Edwards et al. (1998), Edwards et al. (1999), Gasper et al. (2001) |
| | *Andropadus virens* | AY437914 | - | 8, 1 of which a putative pseudo-gene | PCR, RT-PCR | Aguilar et al. (2006) |
| | *Aphelocoma coerulescens* | U23958 | - | 3, 4-6 | RT-PCR, Southern blot | Edwards et al. (1995), Edwards et al. (1999) |
| | *Carpodacus mexicanus* | U23976 | - | 2, 1 of which a pseudo-gene | RT-PCR, cosmid (32 kb) | Edwards et al. (1995), Hess et al. (2000) |
| | *Lonchura striata* | L42335 | - | 1 | RT-PCR | Vincek et al. (1995) |
| | *Petroica australis* | AY428564 | - | 4, 3-8 | RT-PCR, RFLP | Miller & Lambert (2004) |
| | *Petroica traversi* | AY428569 | - | 2, 3 | RT-PCR, RFLP | Miller & Lambert (2004) |

[a] Number of genes inferred from joint data of Alcaide et al. (2007) and Burri et al. (2008b).

[b] In Barn owl (*Tyto alba*), which for reasons indicated in the Material and Methods section was not included in phylogenetic analyses, Southern blots indicate 2 gene copies (Burri et al. 2008a).

**Table 2** Bayesian posterior probabilities of the early duplication and convergence scenarios reconstructed from third codon positions under different nucleotide substitution models and branch length priors.

| Model | Branch length prior | | -lnL | Posterior probabilities (PP) | | |
|---|---|---|---|---|---|---|
| | form | mean | | $PP_{dup}$ | $PP_{con}$ | $PP_{dup}/PP_{con}$ |
| HKY+I+G | exponential | 0.025 | 851.00 | 0.6720 | 0.0017 | 394 |
| | | 0.05 | 872.47 | 0.5933 | 0.0013 | 462 |
| | | 0.1 | 888.92 | 0.5353 | 0.0016 | 326 |
| | | 1 | 905.00 | 0.6424 | 0.0029 | 223 |
| | | 2 | 907.95 | 0.6610 | 0.0026 | 251 |
| | | 10 | 910.59 | 0.7000 | 0.0021 | 339 |
| | | 100 | 953.66 | 0.6219 | 0.0181 | 34 |
| GTR+I+G | exponential | 0.025 | 850.16 | 0.6363 | 0.0029 | 219 |
| | | 0.05 | 866.07 | 0.5757 | 0.0023 | 251 |
| | | 0.1 | 887.30 | 0.5792 | 0.0012 | 492 |
| | | 1 | 903.64 | 0.6485 | 0.0018 | 363 |
| | | 2 | 916.49 | 0.6644 | 0.0018 | 363 |
| | | 10 | 910.34 | 0.6805 | 0.0028 | 245 |
| | | 100 | 950.20 | 0.5748 | 0.0163 | 35 |
| HKY+I+G | uniform | 0-0.1 | 916.30 | 0.5546 | 0.0171 | 32 |
| | | 0-1 | 960.25 | 0.5761 | 0.0203 | 28 |
| | | 0-2 | 965.02 | 0.5202 | 0.0280 | 19 |
| | | 0-10 | 963.30 | 0.5785 | 0.0226 | 26 |
| GTR+I+G | uniform | 0-0.1 | 912.89 | 0.4915 | 0.0124 | 40 |
| | | 0-1 | 959.12 | 0.5674 | 0.0151 | 38 |
| | | 0-2 | 958.04 | 0.5424 | 0.0183 | 30 |
| | | 0-10 | 966.98 | 0.5360 | 0.0237 | 23 |

**Figure 1** Phylogenetic relationships in the avian MHCIIβ exon 3. Branch colours indicate avian MHCIIβ lineages. Green, *DAB1*; blue, *DAB2*. **A** Phylogenetic relationships revealed by amino acid and codon sequences (216 bp) of avian MHCIIβ exon 3 (Bayesian extended majority consensus tree calculated from the 95% HPD credible set). Only supports for avian orders and families and for among-order relationships >50% are provided. Branch lengths were re-estimated in PAUP for codon sequences (GTR+G). Numbers above branches represent Bayesian posterior probabilities and ML bootstrap supports from phylogenetic reconstructions based on codon sequences. Numbers below branches represent the same supports from phylogenetic reconstructions based on amino acid sequences. *, consult Table 2 for sensitivity analyses of the Bayesian posterior probability of this node. **B** Phylogenetic relationships inferred from third codon positions (72 bp) of avian MHCIIβ exon 3 (Bayesian extended majority consensus tree calculated from the 95% HPD credible set; HKY+I+G, exponential branch length prior with mean expected branch length μ=0.1). Numbers above branches represent Bayesian posterior probabilities and ML bootstrap supports, respectively. Branch lengths were re-estimated in PAUP (HKY+I+G). In both analyses barn owl and great snipe (a shorebird) sequences were excluded due to reasons described in the Material and Methods section.

**Figure 2** Results of computer simulations of convergent evolution. Shown are the proportion of simulation replicates showing the unbiased simulated topology with monophyly of the owls' *DAB1* and *DAB2* corresponding to the gene history under the 'convergence hypothesis' (**A**, **B**), and the proportion of replicates showing the topology biased by selection with monophyly of a lineage including the owl *DAB1*, falcons and passerines (**C**, **D**) under positive selection (black lines) and convergence (grey line). Proportions from reconstructions from full coding sequences (**A**, **C**) and from third codon positions from the simulated sequences excusively (**B**, **D**) are shown.

**Figure 3** Proposed evolutionary history of avian MHCIIβ exon 3.

**Figure 4** Three-dimensional analyses of MHCII protein structure. **A** Evolution of the divergent protein domain between *DAB1* (green branches) and *DAB2* (blue branches) along the avian radiation. The positions of the amino acid sites within the domain are indicated for the proteins inferred to be ancestral to each *DAB* lineage. Blue color of the protein domain indicates basicity; red indicates acidity. Note that residue 98, although not exhibiting the signs of positive selection in our analyses, is indeed divergent between the *DAB* lineages (see also Burri et al. 2008b). **B** Three-dimensional structure of the chicken MHCII αβ-heterodimer. The divergent protein domain is indicated in color. The α-chain is given in green, the β-chain in grey. **C** Cartoon representation of the inferred charge-charge interactions between the chicken MHCII α- and β-chains in the divergent protein domain. An overview of the αβ-heterodimer is given at the upper left. Hydrogen bonds are depicted by dashed blue lines. Residues implicated in hydrogen bonds are depicted in stick representation. Residue numbering follows Bondinas et al. (2007).

**Figure S1** Amino acid alignment of MHCIIβ exon 3 sequences used in phylogenetic reconstructions. Numbering follows Bondinas et al. (2007).

**Figure S2** Topology used in computer simulations of positive selection and convergent evolution. Positive selection or convergent evolution were simulated to happen along the branches indicated in red, i.e. branches subtending the owl *DAB1* lineage and the branch leading to the falcon+passerine clade.

**Figure S3** Topology compiled from phylogenetic studies used in maximum likelihood analyses of positive selection and in ancestral sequence reconstructions. For the latter only the relationships between orders were used (shorebirds, which are not indicated in the figure insert into the branch following the penguin split).

**Figure S4** Ancestral amino acid and nucleotide sequences at sites significantly and potentially under positive selection in the avian MHCIIβ exon 3 for each order/family. Orders in blue indicate *DAB2*, orders in green *DAB1*. Table: Deep blue, ancestral reptilian amino acid; light blue, derived reptilian/ancestral avian amino acid; green, *DAB1* apomorphy; red, order-/family-specific apomorphy. Grey shading indicates nucleotide sites involved in divergence between *DAB1* and *DAB2* as indicated in Burri et al. (2008b). Residue numbering follows Bondinas et al. (2007).

Figure 1

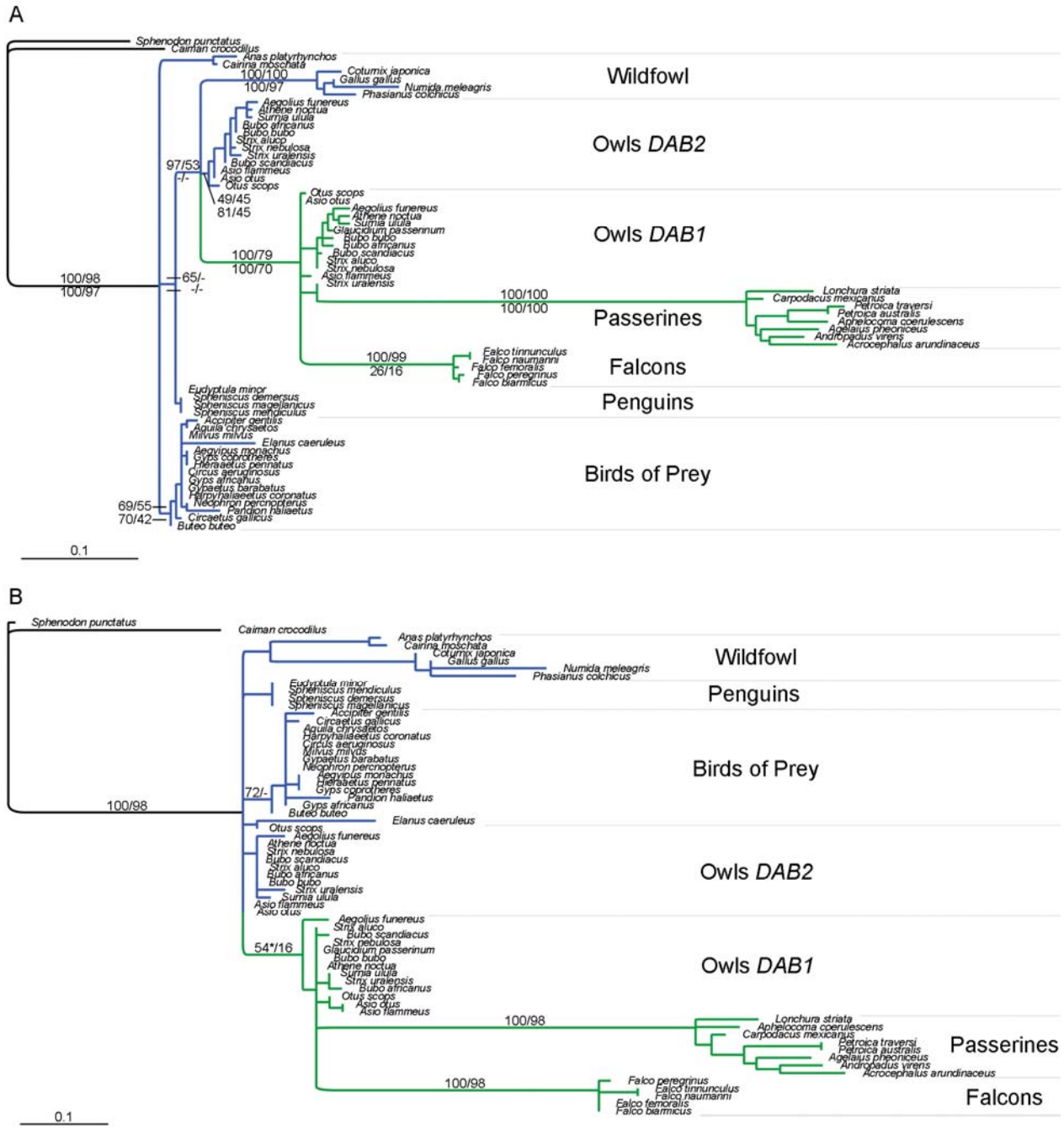

Figure 2

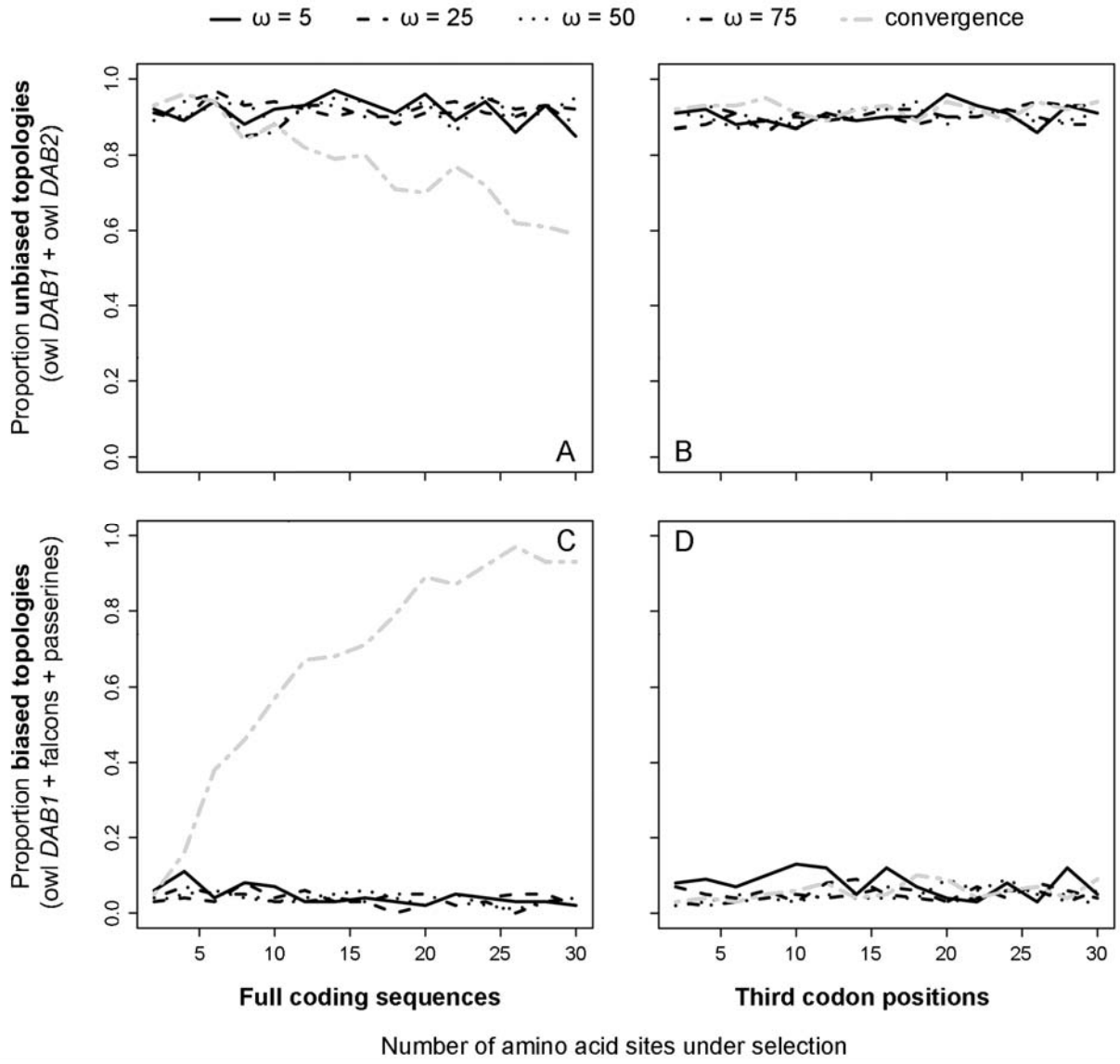

Figure 3

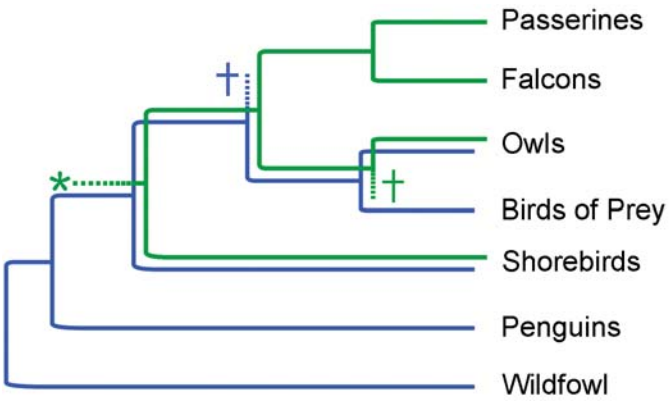

Figure 4