

Selection and Trans-Species Polymorphism of Major Histocompatibility Complex Class II Genes in the Order Crocodylia

Weerachai Jaratlerdsiri¹, Sally R. Isberg^{1,2}, Damien P. Higgins³, Lee G. Miles¹, Jaime Gongora^{1*}

1 Faculty of Veterinary Science, University of Sydney, Sydney, New South Wales, Australia, **2** Centre for Crocodile Research, Noonamah, Northern Territory, Australia, **3** Faculty of Veterinary Science, University of Sydney, Sydney, New South Wales, Australia

Abstract

Major Histocompatibility Complex (MHC) class II genes encode for molecules that aid in the presentation of antigens to helper T cells. MHC characterisation within and between major vertebrate taxa has shed light on the evolutionary mechanisms shaping the diversity within this genomic region, though little characterisation has been performed within the Order Crocodylia. Here we investigate the extent and effect of selective pressures and trans-species polymorphism on MHC class II α and β evolution among 20 extant species of Crocodylia. Selection detection analyses showed that diversifying selection influenced MHC class II β diversity, whilst diversity within MHC class II α is the result of strong purifying selection. Comparison of translated sequences between species revealed the presence of twelve trans-species polymorphisms, some of which appear to be specific to the genera *Crocodylus* and *Caiman*. Phylogenetic reconstruction clustered MHC class II α sequences into two major clades representing the families Crocodylidae and Alligatoridae. However, no further subdivision within these clades was evident and, based on the observation that most MHC class II α sequences shared the same trans-species polymorphisms, it is possible that they correspond to the same gene lineage across species. In contrast, phylogenetic analyses of MHC class II β sequences showed a mixture of subclades containing sequences from Crocodylidae and/or Alligatoridae, illustrating orthologous relationships among those genes. Interestingly, two of the subclades containing sequences from both Crocodylidae and Alligatoridae shared specific trans-species polymorphisms, suggesting that they may belong to ancient lineages pre-dating the divergence of these two families from the common ancestor 85–90 million years ago. The results presented herein provide an immunogenetic resource that may be used to further assess MHC diversity and functionality in Crocodylia.

Citation: Jaratlerdsiri W, Isberg SR, Higgins DP, Miles LG, Gongora J (2014) Selection and Trans-Species Polymorphism of Major Histocompatibility Complex Class II Genes in the Order Crocodylia. PLoS ONE 9(2): e87534. doi:10.1371/journal.pone.0087534

Editor: William Barendse, CSIRO, Australia

Received: May 28, 2013; **Accepted:** December 30, 2013; **Published:** February 4, 2014

Copyright: © 2014 Jaratlerdsiri et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The authors have no external support or funding to report. This research was conducted using the authors' own funding.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: jaime.gongora@sydney.edu.au

Introduction

The Major Histocompatibility Complex (MHC) contains highly variable multi-gene families, which play a key role in the adaptive immune response within vertebrates. MHC characterisation and comparison across major vertebrate taxa have provided valuable insights into vertebrate evolution since the MHC evolved from primitive species, such as sharks, approximately 472–584 million years ago (MYA) [1–4]. Comparison of closely related species has revealed that changes in MHC gene content, number and organisation often occur in antigen presentation genes and this is relevant for an investigation of rapid MHC diversification, and gene loss and gain in a short evolutionary timeframe [5–7]. However, such comparative studies in non-avian reptiles, specifically the Order Crocodylia, are limited. Here we investigate to what extent diversification processes have played a role in the evolution of MHC class II α and β nucleotide sequences across 20 extant species of Crocodylia.

MHC genes are grouped into three classes. One of these, class II, is involved primarily in the presentation of extracellular peptides to helper T-cells [8]. The molecules encoded by MHC class II genes consist of α and β chains, each of which is organised

into two extracellular domains ($\alpha 1/\alpha 2$ and $\beta 1/\beta 2$ respectively), one transmembrane and one cytoplasmic tail domain [9]. Typically, MHC class II α and II β genes are organised into four to six exons: exon 1 encoding the leader sequence, exons 2 and 3 encoding the two extracellular domains, and exon 4 encoding the transmembrane domain with some variable exons (exons 5 and 6) encoding for the cytoplasmic tail and the 3'-untranslated region ([10,11], reviewed in [12]).

MHC class II genes are polymorphic and polygenic, thus generating diversity for loading and presentation of antigens from a wide range of microorganisms [13]. The high degree of polymorphism of these genes is attributed partially to diversifying selection, whereby genotypes conferring a fitness advantage on the organism are maintained [14]. This could lead to more MHC variation within [7,15] and across species [16,17] to cope with different antigenic challenges. This selection is reflected in the predominance of nonsynonymous (dN) over synonymous (dS) substitution rates; where the ratio ($\omega = dN/dS$) is expected to be more than one. Some of the polymorphism persists longer than expected among species under neutrality because γ new species generally inherit an appreciable endowment of MHC variants

from their ancestors, *ii*) some particular sequence polymorphisms at nucleotide or amino acid level can persist after speciation and sequence divergence, and *iii*) new MHC variants are slow to accumulate [18–20]. The retention of nucleotide and amino acid variants across related species has been defined as ‘trans-species polymorphism (TSP)’ [20]. However, if there is low nonsynonymous, relative to synonymous, polymorphism between MHC variants amongst species, whereby functional constraints in antigen loading and presentation make MHC sequences invariable with ω close to zero, it is considered a signature of strong purifying selection [17].

In non-avian reptiles (lepidosaurs, testudines, and crocodylians), studies on MHC class II genes are limited to a few taxa. One study of MHC class II β exon 3 (*DABI-DAB5* genes) in Galapagos marine iguanas [21] identified a greater number of gene loci than in other vertebrates suggesting independent duplication events from a small number of ancestral genes [21]. The only species of Crocodylia for which studies related to the MHC class II have been undertaken thus far are the Chinese alligator (*Alligator sinensis*) and Nile crocodile (*Crocodylus niloticus*), both of which showed some diversity of MHC genes at the population level [22,23]. Additional MHC class II studies across species of Crocodylia are required to understand the evolutionary mechanisms that have driven and maintained diversity within this class of genes. Although there is an ongoing initiative to sequence representative genomes from each of three extant families of Crocodylia (Alligatoridae, Crocodylidae and Gavialidae) [24], further data on the MHC for other species of Crocodylia still needs to be generated for comparative analyses. To address this gap, we investigated to what extent diversifying selection and trans-species polymorphism have played a role in the evolution of MHC class II α and β genes across 20 species of Crocodylia. Overall, we identified ancient trans-species polymorphisms within and between families of the Order Crocodylia, and detected a greater role of diversifying selection in the MHC class II β relative to the MHC class II α .

Results

Characterisation of MHC Class II α and β

Eighteen and eleven sequences of MHC class II α exons 2 and 3, respectively, were identified among species of the Order Crocodylia (Table 1; Figure S1). Single sequences per specimen and species were obtained for both exons 2 and 3. Whilst these data might suggest that only a single locus is present for each species, the presence of additional undetected gene copies within these species cannot be excluded. Seventy-two sequences (260 bp) of MHC class II β exon 3 were identified among 20 species of Crocodylia (Table 1; Figure S2). Between one and four sequences per individual within a species were identified, suggesting that at least two loci were being amplified in the current study. More details regarding characterisation of MHC class II α and β are described in Appendix S1 and Appendix S2.

Trans-species Polymorphism and Non-functional Sequences

Four, seven and forty-two translated sequences were identified for MHC class II α exons 2 and 3 and II β exon 3, respectively. Based on identical amino acid sequences between species, twelve trans-species polymorphisms (named TSP1 through TSP12) were identified among all the amino acid sequences: two in each of the MHC class II α exons and eight in the MHC class II β exon 3 (Table 2; Figures S1 and S2). TSP1 from MHC class II α exon 2 was detected in 13 species from six genera of Crocodylia, while TSP3 and TSP4 from MHC class II α exon 3 were observed only

among species from the genus *Crocodylus* (Table 2). TSP5 from MHC class II β exon 3 was identified in 11 of the 20 species studied. Finally, TSP6–12 were observed in two to six species each.

In total, 11 non-functional sequences were found among the MHC class II exons studied (Figures S1 and S2). Interestingly, a single base pair (bp) deletion at amino acid (aa) site 12, resulting in a break in the open reading fragment, appears to be shared by three MHC class II α exon 2 sequences (*Crrh-DA01*, *Crmo-DA02* and *Caya-DA01*) from *C. rhombifer*, *C. moreletii* and *C. yacare*. In addition, stop codons at aa site 31 or 48 were observed among three MHC class II α exon 3 sequences (*Crin-DA01*, *Oste-DA03* and *Cacr-DA02*) from *C. intermedius*, *O. tetraspis* and *C. crocodylus* (Figure S1). Despite the same individual being investigated for each species, it is difficult to conclude that the exons identified belong to a single gene due to the possibility of polygeny [8]. Apparently, the species in which the non-functional sequences of MHC class II α were identified did not have additional sequences with intact open reading fragments; it appears likely that those sequences may have been missed during the screening of clones as most species were found to possess intact MHC class II α coding sequences. For the MHC class II β exon 3, five non-functional sequences from four species of Crocodylidae (*Crrh-DB02*, *Crmo-DB01*, *Crmo-DB01* and *Crin-DB02*) and a single species of Alligatoridae (*Meni-DB02*) were identified, based on the presence of stop codons. However, other sequences from these five species displayed functional putative reading fragments with most containing conserved peptide interacting sites and cysteine bridges (C-C).

Effect of Diversifying Selection on MHC Class II α and β

Bayesian analyses of MHC class II β gene sequences among species of Crocodylia showed evidence of diversifying selection on their exon 2 with high dN/dS ratios (Figure 1). Figure 1B suggests that seven amino acid sites have been selected for diversification with highly significant support (posterior probability >0.99). Of these sites (*Cacr-B1* as a reference; Figure S3), exon 2 sites 47W (dN/dS = 7.241; CI = 2.879–14.634), 50Y (7.002; 2.881–17.725), 51K (5.314; 2.369–13.262) and 54E (5.057; 2.315–13.06) corresponded to sites in the expected peptide-binding residues (PBR) of the human MHC class II β molecule [25], and exon 2 sites 11I (4.493; 1.995–10.564), 48M (6.361; 2.511–13.707), and 49E (4.537; 1.874–10.808) were situated within a distance of two amino acid positions of the PBR (Z-test of $H_A: dN > dS$ at all the PBR in Figure S3, Test statistic = 1.716; *P*-value = 0.044). The sites under diversification, as identified in our study, resemble those PBR sites in humans that are polymorphic, yielding a wide variety of motif specificities for antigen binding so as to handle the immune response to any pathogen, regardless of how peculiar the pathogenic protein might be [25].

In contrast, the Bayesian analysis of MHC class II α sequences did not detect sites under diversifying selection in the antigen presentation region of exon 2 (Figure 1A). Although a single exon 3 site, 92V (*Crpo-DA01* as a reference; Figure S4) had mean dN/dS equal to 3.528, with the posterior probability equal to 0.96, exon 2 did not show significant posterior probability of diversifying selection. In addition, 140 sites of exons 2 and 3 (out of 143) revealed low dN/dS values ranging from 0.160 to 0.970 (Z-test of $H_A: dN < dS$, Test statistic = 2.401; *P*-value = 0.009), which is of interest for comparisons of MHC genes between species. The data analysed herein highlight a predominant role of purifying selection acting on amino acid sites of MHC class II α , especially peptide-binding sites, consistent with the high number of synonymous substitutions and highly conserved amino acid sites involved in forming the MHC-peptide complex described in Figure S1. Moreover, our analyses have shown that recombination does not

Table 1. Summary of 20 species of Crocodylia investigated in the current study for MHC class II α exons 2 and 3 plus MHC class II β exon 3, their assigned gene prefixes, and numbers of sequences per species or per individual.

Species ^a	Gene prefix	Class II β exon 3		Class II α exon 2		Class II α exon 3		GenBank accession numbers
		NS (NC) ^b	SPI (N) ^c	NS (NC)	SPI (N)	NS (NC)	SPI (N)	
Crocodylidae								
Freshwater crocodile (<i>Crocodylus johnsoni</i>)	<i>Crjo</i>	4 (9)	1 and 3 (2)	1 (4)	1 (1)	NA ^d	NA	GU126804-07, GU126954
Philippine crocodile (<i>Crocodylus mindorensis</i>)	<i>Crmi</i>	2 (4)	2 (1)	1 (4)	1 (1)	1 (4)	1 (1)	GU126810-11, GU126950, GU126959
Nile crocodile (<i>Crocodylus niloticus</i>)	<i>Crni</i>	3 (8)	1 and 2 (2)	1 (4)	1 (1)	NA	NA	GU126822-23, GU126825, GU126929
American crocodile (<i>Crocodylus acutus</i>)	<i>Crac</i>	5 (9)	2 and 3 (2)	1 (4)	1 (1)	1 (4)	1 (1)	GU126827-30, GU126832, GU126934, GU126960
Mugger crocodile (<i>Crocodylus palustris</i>)	<i>Crpa</i>	2 (4)	2 (1)	1 (4)	1 (1)	1 (4)	1 (1)	GU126833, GU126835, GU126942, GU126958
Dwarf crocodile (<i>Osteolaemus tetraspis</i>)	<i>Oste</i>	5 (9)	1 and 4 (2)	1 (4)	1 (1)	1 (4)	1 (1)	GU126836, GU126839-41, GU126843, GU126936, GU126961
Siamese crocodile (<i>Crocodylus siamensis</i>)	<i>Crsi</i>	3 (8)	1 and 3 (2)	1 (4)	1 (1)	NA	NA	GU126944, GU126846-48
Slender-snouted crocodile (<i>Mecistops cataphractus</i>)	<i>Meca</i>	4 (10)	1 and 4 (2)	1 (4)	1 (1)	NA	NA	GU126849, GU126851, GU126855- 56, GU126931
Orinoco crocodile (<i>Crocodylus intermedius</i>)	<i>Crin</i>	4 (8)	2 (2)	NA	NA	1 (4)	1 (1)	GU126890-91, GU126893, GU126895, GU126957
Cuban crocodile (<i>Crocodylus rhombifer</i>)	<i>Crrh</i>	4 (8)	1 and 3 (2)	1 (4)	1 (1)	NA	NA	GU126896-97, GU126899, GU126900, GU126951
New Guinea crocodile (<i>Crocodylus novaeguineae</i>)	<i>Crno</i>	2 (4)	2 (1)	1 (4)	1 (1)	NA	NA	GU126909, GU126911, GU126953
Saltwater crocodile (<i>Crocodylus porosus</i>)	<i>Crpo</i>	4 (10)	1 and 3 (2)	1 (4)	1 (1)	1 (4)	1 (1)	GU126912-13, GU126915, GU126919, GU126967
Morelet's crocodile (<i>Crocodylus moreletii</i>)	<i>Crmo</i>	5 (9)	2 and 3 (2)	1 (4)	1 (1)	1 (4)	1 (1)	GU126920-21, GU126923, GU126927-28, GU126938, GU126962
Alligatoridae								
American alligator (<i>Alligator mississippiensis</i>)	<i>Almi</i>	4 (10)	1 and 3 (2)	1 (4)	1 (1)	NA	NA	GU126813, GU126815-17, GU126940
Chinese alligator (<i>Alligator sinensis</i>)	<i>Alsi</i>	4 (10)	2 and 3 (2)	NA	NA	1 (4)	1 (1)	GU126880-83, GU126963
Cuvier's dwarf caiman (<i>Paleosuchus palpebrosus</i>)	<i>Papa</i>	3 (8)	0 ^e and 3 (2)	1 (4)	1 (1)	NA	NA	GU126858-59, GU126861, GU126941
Spectacled caiman (<i>Caiman crocodylus</i>)	<i>Cacr</i>	2 (8)	0 and 2 (2)	1 (4)	1 (1)	1 (4)	1 (1)	GU126865, GU126867, GU126945, GU126966
Broad-snouted caiman (<i>Caiman latirostris</i>)	<i>Cala</i>	4 (9)	1 and 3 (2)	1 (4)	1 (1)	1 (4)	1 (1)	GU126871, GU126873, GU126877- 78, GU126948, GU126964
Yacare caiman (<i>Caiman yacare</i>)	<i>Caya</i>	5 (8)	2 and 3 (2)	1 (4)	1 (1)	NA	NA	GU126903-05, GU126907-08, GU126955
Black caiman (<i>Melanosuchus niger</i>)	<i>Meni</i>	3 (8)	1 and 3 (2)	1 (4)	1 (1)	1 (4)	1 (1)	GU126887-89, GU126939, GU126965

^aCommon names and their scientific names for the Order Crocodylia in brackets.

^bNC in brackets indicates the number of clone inserts sequenced from which number of sequences within each species (NS) were identified. The NS values also include pseudogenes, and are discussed in the results section.

^cSPI indicates number of sequence(s) per individual with the number of individuals (N) examined in brackets. The SPI and N values are discussed in the results section.

^dNot applicable due to unsuccessful results from a process of cloning.

^eZero means none of the sequences observed in an individual after eliminating sequence artefacts described below.

doi:10.1371/journal.pone.0087534.t001

Table 2. Number of species possessing a particular TSP across families and genera of Crocodylia.

Family		Alligatoridae									
Crocodylidae											
Genus ^c											
TSP ^a	N ^b	Cro	Mec	Ost	All	Cai	Pal	Mel	MHC sequence ^d		
II α exon 2											
TSP1	13	7	1	0	1	2	1	1	Crmi-DA01, Crac-DA01, Crpa-DA01, Crsi-DA01, Crmo-DA02, Crpo-DA01, Mecca-DA01, Almi-DA01, Papa-DA01, Cacr-DA01, Cala-DA01, Meni-DA01		
TSP2	3	2	0	0	0	1	0	0	Crrh-DA01, Crmo-DA02, Caya-DA01		
II α exon 3											
TSP3	2	1	0	1	0	0	0	0	Oste-DA03, Crin-DA01		
TSP4	4	4	0	0	0	0	0	0	Crmi-DA02, Crac-DA02, Crpa-DA02, Crpo-DA01		
II β exon 3											
TSP5	11	8	1	1	1	0	0	0	Crjo-DB04, Crac-DB04, Crpa-DB03, Crsi-DB04, Crin-DB04&06, Crrh-DB04, Crpo-DB04, Crmo-DB04, Oste-DB04&08, Mecca-DB07, Alsi-DB04		
TSP6	6	1	1	0	2	1	1	0	Crmi-DB03, Mecca-DB03&08, Papa-DB03, Cala-DB03, Alsi-DB03, Almi-DB03		
TSP7	4	3	0	1	0	0	0	0	Oste-DB01, Crac-DB01, Crpa-DB01, Crin-DB01		
TSP8	3	0	0	0	0	2	1	0	Papa-DB02, Cacr-DB02, Caya-DB02		
TSP9	3	2	0	0	0	0	0	1	Crin-DB02, Crrh-DB02, Meni-DB02		
TSP10	2	0	0	0	0	2	0	0	Cala-DB01, Caya-DB01		
TSP11	2	2	0	0	0	0	0	0	Crjo-DB01, Crpo-DB01		
TSP12	2	2	0	0	0	0	0	0	Crjo-DB02, Crac-DB02		

^aTSPs that are translated from nucleotide sequences (last column), and that are observed among species with the same or different genera of Crocodylia.

^bTotal number of species possessing the correspondent TSP.

^cNumber of species within the following genera that share the same TSP: Crocodylus (Cro), Mecistops (Mec), Osteolaemus (Ost); Alligator (All); Caiman (Cai); Paleosuchus (Pal); Melanosuchus (Mel).

^dNucleotide sequences that are deduced into the same TSP.

doi:10.1371/journal.pone.0087534.t002

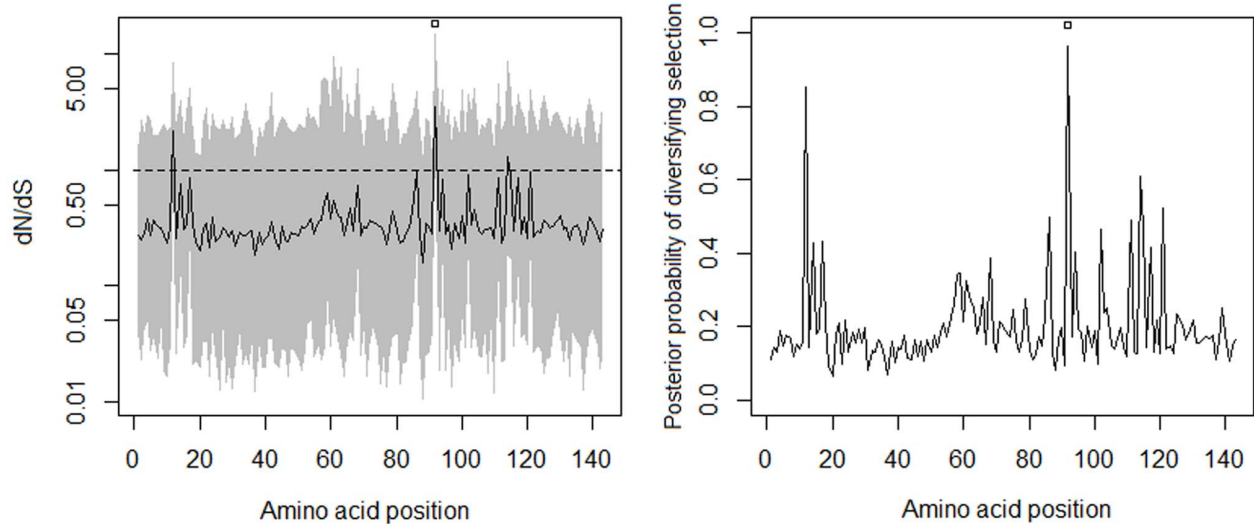
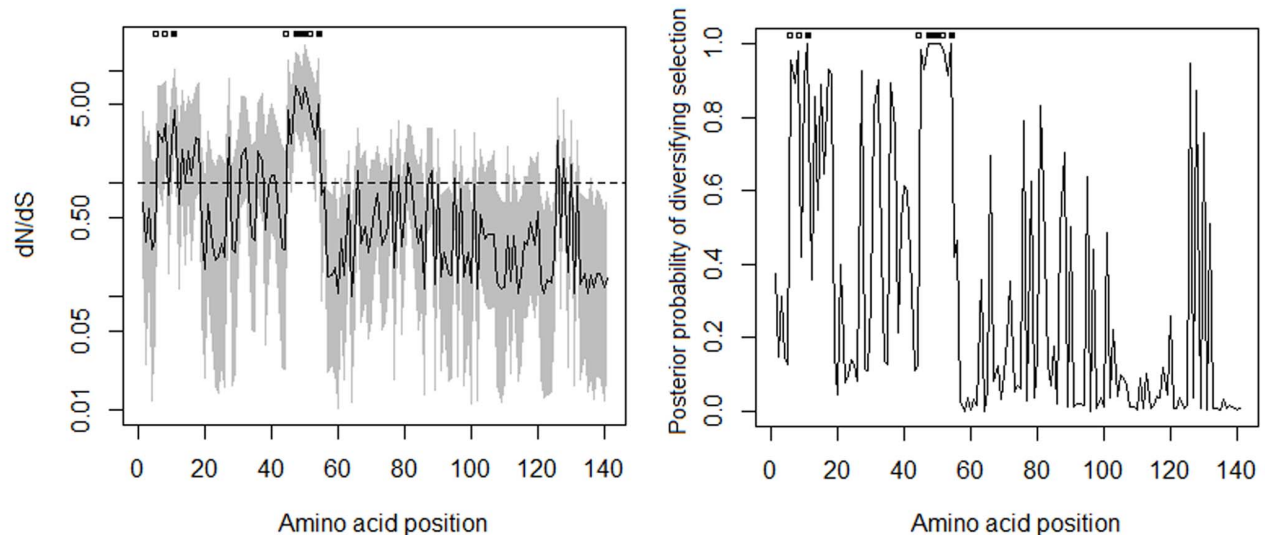
(A) MHC class II α exons 2 and 3**(B) MHC class II β exons 2 and 3**

Figure 1. Selection detection analyses of MHC class II α and β across species of Crocodylia. Tests were performed by Bayesian inference of (A) MHC class II α exons 2 and 3 sequences, and (B) MHC class II β exons 2 and 3 sequences. Graphs show spatial change in dN/dS ratios (ω) and the posterior probabilities of diversifying selection across amino acid positions. Lines on the left-sided graphs present estimates of dN/dS ratios across amino acid positions; grey areas indicate 95% highest posterior probability dense intervals; and dashed lines show dN/dS values equal to one. Amino acid positions show significant (posterior probabilities >0.95) and highly significant diversifying selection (posterior probabilities >0.99), displayed by open squares and dark squares, respectively. doi:10.1371/journal.pone.0087534.g001

play a major role in generating diversity among the MHC class II sequences in Crocodylia (described more in Appendix S3), thereby dispensing with any likelihood of bias caused by high recombination rates.

Phylogenetic Reconstruction of MHC Class II α

Bayesian analyses of MHC class II α exons 2 and 3 sequences clustered into two major clades corresponding to the families Crocodylidae and Alligatoridae (Figures 2). Maximum likelihood was consistent, with the exception of two MHC class II α exon 2 sequences from Alligatoridae, which clustered as sister clades of those from Crocodylidae (Figure S5-A). The expected likelihood

weights (ELW) test showed that the topology of this, and the Bayesian tree, did not differ significantly (confidence tree set = 0.79 and 0.21). However, clustering of both exons within the major Crocodylidae clade did not show a clear phylogenetic signal that would permit us to determine whether subclustering was according to gene or species. Within the major Alligatoridae clade, sequences at both exons appear to cluster into two subclades representing the genera *Caiman/Melanosuchus* and *Alligator/Paleosuchus* (posterior probability (PP) = 0.9 at exon 2; 0.6 at exon 3). Furthermore, phylogenetic analyses of these MHC class II α sequences, as well as those from other vertebrates, showed the Crocodylia exons clustered as a sister clade of bird *DRA* (*Anpl-DRA*).

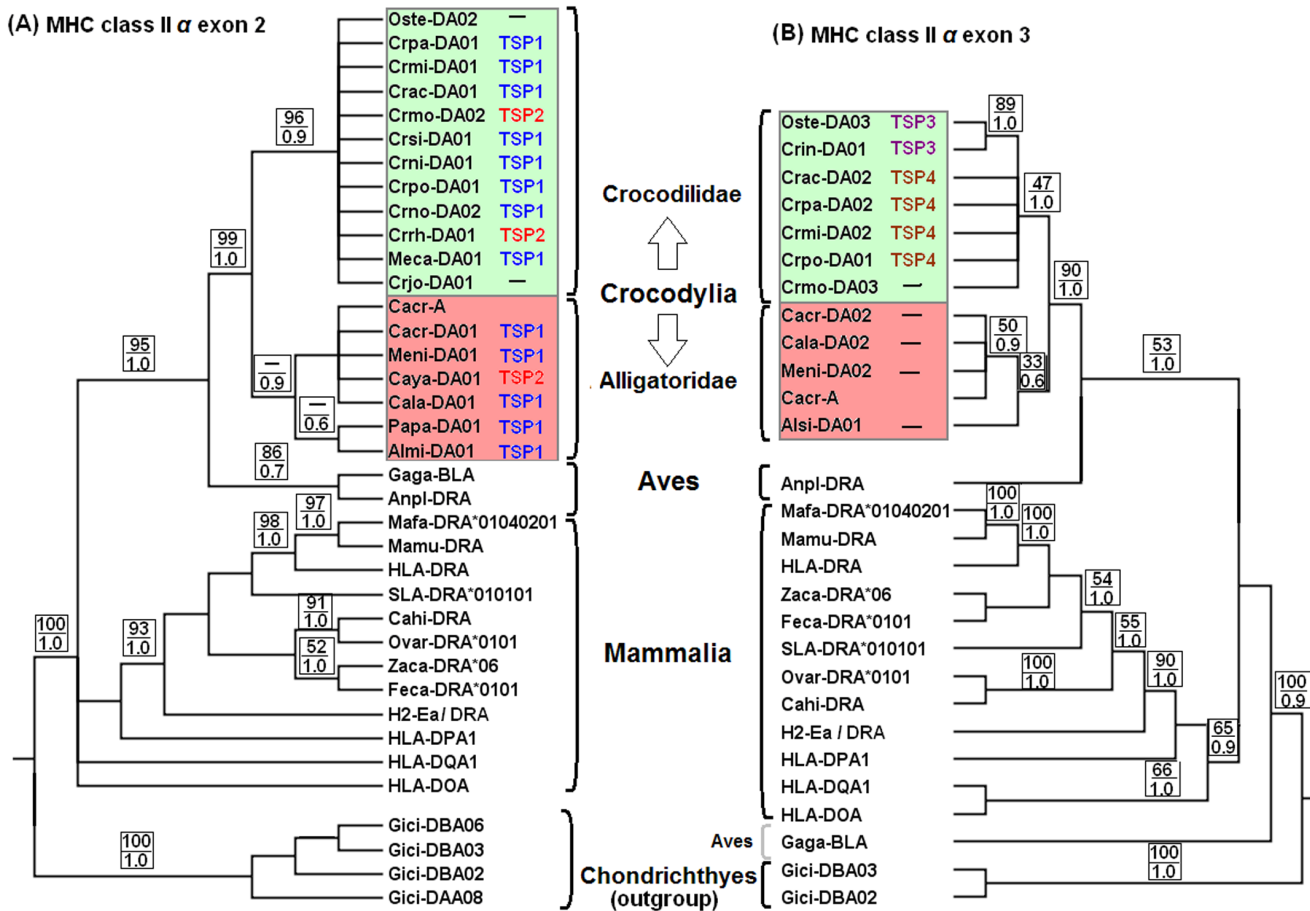


Figure 2. Bayesian phylogenetic trees of (A) MHC class II α exon 2, and (B) exon 3. Sequences of MHC class II α among different species of Crocodylia, Aves and Mammalia are analysed using the shark *Gici* sequences as an outgroup. Brackets in the middle show vertebrate groups to which the MHC class II α sequences belong. Sequences generated in the current study are from two families of Crocodylia: Crocodylidae (pale green colour) and Alligatoridae (pale red colour). TSPs associated with these sequences are described immediately after the sequence names. Support on branches is indicated by bootstrap values (BV) for maximum likelihood (above) and posterior probabilities (PP) for Bayesian analysis (below), as both analyses provide significantly similar trees described in the text. Branches with posterior probabilities below 0.5 are collapsed. doi:10.1371/journal.pone.0087534.g002

Despite clustering into two major clades, some of the amino acid TSPs, such as TSP1 and TSP2, were observed among MHC class II α variants in both clades. In particular, TSP1 contained an intact open reading fragment and was observed among the sequences of eight and five species of Crocodylidae and Alligatoridae, respectively (Table 2). This is likely to suggest that these sequences are orthologous, with the same functionality in the host's immunity.

Phylogenetic Reconstruction of MHC Class II β

Bayesian and maximum-likelihood analyses of MHC class II β sequences from Crocodylia showed two clades (Clades 1 and 2), using tuatara sequences as outgroups (Figures 3 and S6). These two clades were differentiated by ten diagnostic sites (Figure S7). The Bayesian tree showed seven subclades in Clade 1 (1A–1G). The number of these subclades was greater than, and topology different to, those detected by the maximum-likelihood analysis, but the ELW test showed that these two trees did not differ significantly (confidence tree set = 0.67 and 0.33). Clades and subclades consisted of sequences from either/both Crocodylidae and/or Alligatoridae. Some individuals had more than one MHC class II β locus amplified (Tables S1 and S2) and this could explain why, in the Bayesian tree, some MHC sequences from the same individual split into different clades/subclades (Table S2). For example, Subclade 1A clustered together one to three sequences

per individual for six Crocodylidae species, with the remaining sequences (1–2 per individual) clustered into Subclade 1C. Similarly, Subclade 1B clustered together two sequences per individual for *P. palpebrosus* and *M. niger*, while the remaining sequences (1 per individual and species) clustered into Clade 2; although one from *M. niger* is expected to be pseudogenetic (containing a stop codon). The resultant topology within each clade and subclade was comprised of different species of Crocodylia, consistent with the notion that an orthologous relationship exists for a gene among related species, as described by Burri et al. [5].

Phylogenetic analyses across higher taxa showed that orthologous relationships were also present in MHC class II β genes among major lineages within Mammalia. For instance, *DRB1* sequences from different mammalian lineages clustered together, as did *DOB* and *DQB1* sequences (Figure 3). In contrast, when chicken, quail and pheasant sequences are compared (Figure 3), MHC class II β sequences appear to cluster by species rather than gene lineages in birds.

In the light of Bayesian analysis, a clear accumulation of TSP from different species of Crocodylia, into a number of subclades, occurred (Figure 3). For instance, the sequences *Alsi-DB04* and *Meni-DB02* from Alligatoridae clustered within subclades containing mainly Crocodylidae sequences. A similar situation was

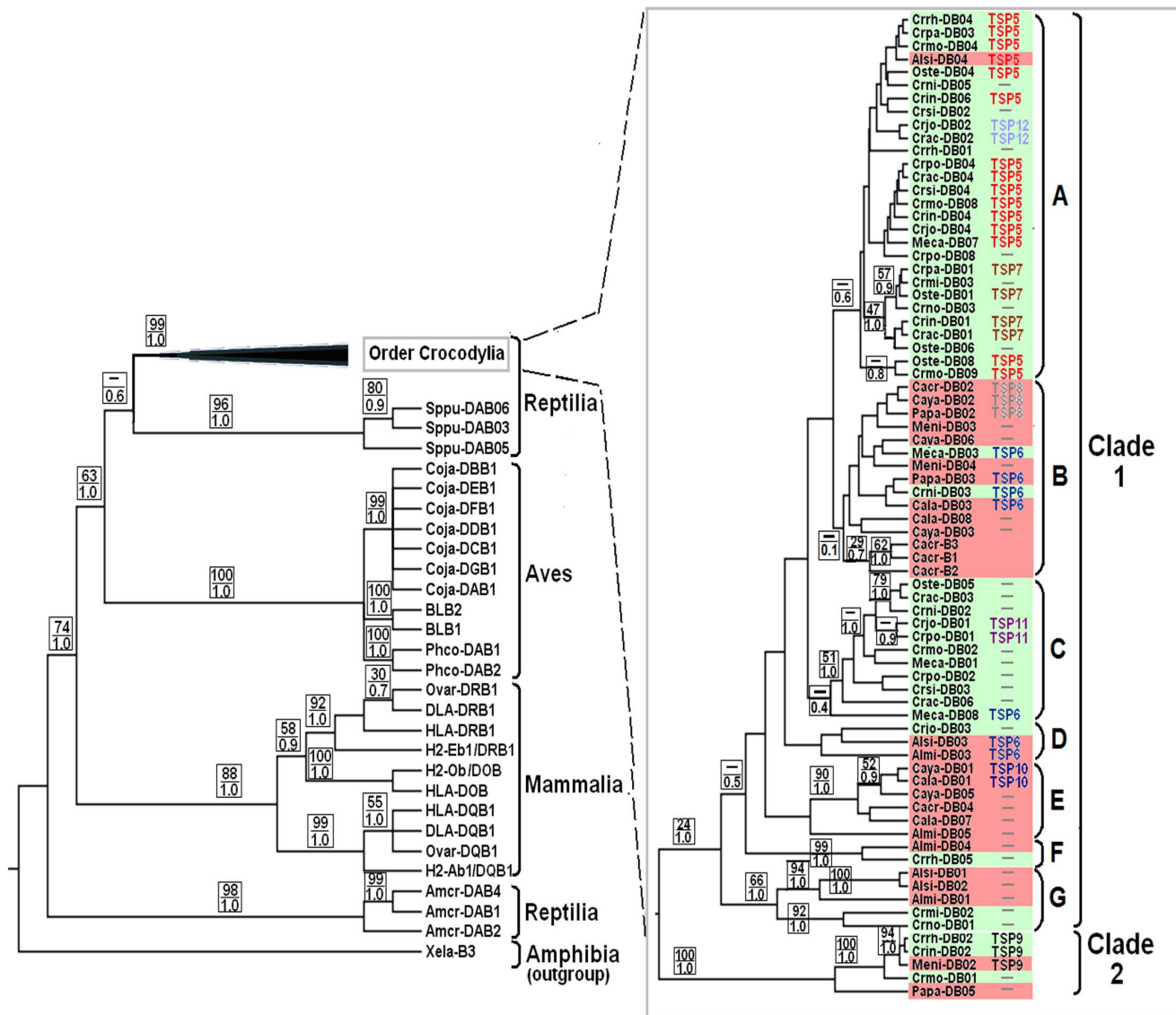


Figure 3. Bayesian phylogenetic trees of MHC class II β exon 3. This tree was constructed using sequences of MHC class II β among different species of Crocodylia, Aves and Mammalia using the amphibian sequence as an outgroup. Sequences generated in the current study are from two families of Crocodylia: Crocodylidae (pale green colour) and Alligatoridae (pale red colour). Brackets on the left show vertebrate groups to which the MHC class II β sequences belong, and those on the right show Clades 1 and 2 of the MHC sequences and seven subclades (A–G) for Clade 1. The two clades were defined on the basis of their monophyletic groupings and high posterior probability (PP = 1.0). TSP5–12 corresponding to particular MHC sequences, which are described above, follow immediately after the sequence name. Support on branches is indicated by bootstrap values (BV) for maximum likelihood (above) and posterior probabilities (PP) for Bayesian analysis (below), as both analyses provide significantly similar trees described in the text.
doi:10.1371/journal.pone.0087534.g003

observed for the sequences *Cmi-DB03* and *Meca-DB03* from Crocodylidae, which clustered with those from Alligatoridae. These examples support the conclusion that the TSPs identified in the present study play a role in complicating MHC class II evolution within and among families of Crocodylia.

Discussion

MHC Class II α and β are Subjected to Differential Selective Pressures

The current study has demonstrated that diversifying selection is driving and maintaining diversity of MHC class II β genes at the PBR within the Order Crocodylia. It could be suggested that a

range of pathogenic challenges among species of Crocodylia (reviewed in [26]) and their demographic distributions around the world (Figure S8) have a role in this diversification process. In further support of this scenario, the identification of TSPs at MHC class II β shows preferential retention of some ancestral polymorphisms in Crocodylia, whereby specific allelic polymorphisms are maintained, thereby increasing diversity in the gene pool [27]. For instance, the distribution of TSP5 and TSP6 among the genera (*Crocodylus*, *Osteolaemus*, *Mecistops*, *Alligator*, *Caiman* and/or *Paleosuchus*) and between the families (Alligatoridae and Crocodylidae) indicates that MHC class II β polymorphism may have been present before these two families diverged from the common ancestor 85–90 million years ago [28]. However, a definitive

conclusion of TSP across species of Crocodylia will require genomic sequence data from each species to avoid overestimation by assessing entire TSP exons (from the point of transcription to the polyadenylation signal). Analysis of intron data for MHC class II β genes will also permit confirmation of an ancient lineage of TSP exons as discussed above (plus differentiation between TSP and convergent evolution), particularly if the phylogenetic clustering of intron sequence data is consistent with that of exon sequence data in this study [29,30].

In contrast to mechanisms seen in MHC class II β evolution, purifying selection appears to play a predominant role in the conservation of MHC class II α genes and their PBR. Low interspecific MHC variability, and TSPs of the MHC class II α exons, suggest that the Order Crocodylia has experienced purifying selection, whereby the slow-evolving MHC genes have been maintained by removing nonsynonymous nucleotide substitutions that are deleterious to immune protein function, consistent with that seen in birds [31] and mammals [17,32]. In addition, the identification of only a single MHC class II α locus and at least two MHC class II β loci per species in the preliminary diversity analyses suggests that the α chain of the MHC molecule (II α) has maintained some conservation of protein structure necessary for interactions with different polymorphic β chains (II β). Primate species also show strong selection of MHC class II α genes (*DRA*), against any corrupt rearrangements or nonsilent mutants, in order to maintain their functional abilities to bind with all MHC class II *DRB* molecules available [33–36].

Based on the difference in selective pressures between MHC genes described above, MHC class II β genes appear to be more suitable markers for future diversity and disease-resistance association studies. They have more polymorphic sites than the MHC class II α (Appendices S1 and S2), especially those under diversifying selection ($\omega > 1$), where nonsynonymous substitutions occur at the amino acid level. As the MHC class II β gene encodes for an extracellular domain (β chain) of the MHC molecule directly involved in antigen presentation, these selected sites might have been mediated by pathogen challenges (reviewed in [37,38]). Although false detection of selected sites among MHC sequences having a high level of recombination is possible [39], such a bias appears unlikely in the present case since the recombination rates measured in the current study are low, and the selected sites identified have shown strong statistical support.

Further Sequence Subdivision of MHC Class II β Over MHC Class II α

MHC class II α and β in Crocodylia appear to cluster by gene rather than by species. The two major clades of MHC class II α , representing taxonomic families, could be interpreted as belonging to the same gene lineage, which may have emerged before Crocodylia and Alligatoridae diverged from the common ancestor 85–90 MYA [28]. Although a level of sequence divergence was observed between these two major clades, marked levels of conservation among sequences within clades were also observed. This is unexpected for some genera of Crocodylia (e.g. *Crocodylus* and *Osteolaemus*) and Alligatoridae (e.g. *Alligator* and *Paleosuchus*) that diverged approximately 22 and 65 MYA, respectively [28], particularly given that the MHC class II α exons studied herein encode for an extracellular domain of the MHC molecule (directly involved in the presentation of antigens). A possible explanation for the sequence conservation is that these MHC class II α exons may have been influenced by a preferential retention of similar variants among species within families, as an advantageous way to preserve a specific biological function [40]; and/or that mutation rates across the genome might be slow

among species of Crocodylia. Supporting the latter, the Order Crocodylia has been shown to possess a significantly lower mutation rate than other vertebrate species, using mitochondrial DNA and the nuclear RAG-1 gene [41,42]. However, further investigation may be warranted as a number of diversity studies, albeit using microsatellites, have shown moderate levels of mean heterozygosity ($H_E = 0.040\text{--}0.941$) between populations of *C. porosus* [43], *A. mississippiensis* [44] and *C. latirostris* [45] and high sequence divergence (nucleotide diversity = 0.152) at a population of *C. porosus* using MHC class I markers [46].

Furthermore, numerous phylogenetic subdivisions of MHC class II β exon 3 and their orthology of sequences from both Crocodylia and Alligatoridae may suggest a greater number of gene lineages in this exon, relative to the MHC class II α exons described above. This is consistent with the varying numbers of loci that have been observed for both MHC class II α and β genes in other vertebrates [47–49], including birds (related taxa to Crocodylia) [10,50]. These phylogenetic patterns could also be interpreted as gene lineages that have evolved divergently to each other, so as to represent different selective pressures and/or perform novel functions in immunity [6,51]. All these results are in line with the evolution of MHC class II α discussed above. This is because the gene copy number and divergence of MHC class II β is a key factor for the MHC protein assembly of wide motif specificities to antigen binding and subsequent immune responses to pathogens [25]. Alternatively, it is possible that MHC class II β genes have different evolution when their orthologous relationships have been masked, especially in some bird lineages, due to recent gene duplication events and frequent genetic exchange between genes within a species (recombination or gene conversion) [5,52–54].

Comparisons of the present results with a recent MHC class I study in the same taxa [55] suggest that the MHC in Crocodylia has undergone a differential pattern of evolution within and between gene classes. For instance, in contrast with MHC class II α genes, those for MHC class I seem to have been subject to independent events of duplication which have led to further gene-lineage diversity in Crocodylia than in Alligatoridae. This is not surprising given that MHC class I molecules appear to be more diverse in structure and immune function than class II counterparts [56].

Conclusions

From an evolutionary perspective, MHC class II α appears to be well conserved among species of the Order Crocodylia, while MHC class II β appears to have undergone a process of diversification. In this respect, diversification selection appears to have played a larger role in the evolution of MHC class II β relative to that of MHC class II α . These findings suggest marked degrees of differential evolution between the two MHC class II chains. In addition, the MHC class II β appears to be the most informative immunogenetic resource for future studies assessing population diversity for species of Crocodylia and for association studies between MHC and disease. Those studies will help refine estimates of variation in gene copy number and content among species and their implication in disease outbreaks in farmed and wild populations of Crocodylia.

Materials and Methods

Species Collection

We aimed to obtain DNA samples from all living species of Crocodylia. However, samples from only twenty extant species

representing seven genera were obtained as follows: *Caiman* (3 spp.), *Melanosuchus* (1 spp.), *Paleosuchus* (1 spp.), *Alligator* (2 spp.), *Crocodylus* (11 spp.), *Mecistops* (1 spp.), *Osteolaemus* (1 spp.) (Table 1). Most of these samples were not purpose-collected for this study but rather came from materials sourced from previous studies [57,58] so were obtained opportunistically. Some of them were originally collected under the University of Florida Institutional Animal Care and Use protocol number E423 and the University of Sydney Animal Ethics permit number N00/5–2009/3/5057.

Primers, PCR, Cloning, and Sequencing

Two primer sets were generated (Table 3) to amplify the MHC class II α exons 2 and 3 (which encode $\alpha 1$ and $\alpha 2$ domains respectively), while published primers [22] were used to amplify MHC class II β exon 3 (which encodes the $\beta 2$ domain). The former two sets were designed using the regions conserved between the spectacled caiman MHC class II α exons 2 and 3 (AF256650), MHC class II α sequences from human (*HLA-DRA*, NM_019111) and mallard (*Anpl-DRA*, AY905539) sequences available in GenBank. Similarly, MHC class II β exon 3 primers targeted conserved regions among caiman, chicken and frog sequences. The purpose of these primers was to amplify as many MHC class II gene loci as possible across all species of Crocodylia. MHC class II specificities of those primers were verified using BLASTN search in GenBank. The search performed against the genome drafts of *Alligator mississippiensis*, *Crocodylus porosus* and *Gangeticus gangeticus* ([24]; Green et al. unpublished data) showed significant matches with putative MHC class II sequences in Crocodylia (Appendix S4), suggesting that the primers would result in an unbiased amplification of the presumed targets. Those genome resources are publicly available after this study was finalised. The exons targeted here were chosen because they correspond to the extracellular domains involved in the presentation of antigens and have been used to assess the evolution of MHC class II genes in other studies [59–61].

The three gene fragments were amplified using the same Polymerase Chain Reaction (PCR) reaction conditions in 50- μ l reaction containing: 10 mM Tris-HCl, 50 mM KCl pH 8.3, 1.5 mM MgCl₂, 160 μ M each dNTP, 0.8 μ M each primer, 0.5 U of 40:1 mixture of Taq polymerase and Phusion[®] High-Fidelity DNA Polymerase (Finnzymes, Vantaa Finland), and 20–350 ng of genomic DNA. PCR cycling was as follows: 5 mins at 94°C, followed by 35 cycles of 94°C for 30 s, 61°C (for MHC class IIB primers) or 67°C (for MHC class II α primers) for 30 s, and 72°C for 40 s, with a final extension of 72°C for 10 mins. To enhance the efficiency of cloning, pre-cloning poly-A addition was performed as a final step by adding 0.25 U *Taq* polymerase to

each PCR reaction and incubating at 72°C for 10 min. PCR products were gel purified using a UltraClean[®] GelSpin[®] Kit (MO BIO Laboratories, Inc., Carlsbad California), and were subsequently cloned using a TOPO TA Cloning[®] Kit (Invitrogen Australia Pty Limited, Mulgrave Victoria). Clones with positive inserts were confirmed for the expected insert size by *i*) PCR with the same conditions and primers and *ii*) cutting the inserts from the clones using the restriction enzyme *EcoRI* (TAKARA BIO INC., Otsu Shiga); and then visualisation in 1.7% agarose gels. Positive clones from the PCR or restriction enzyme digestion were randomly selected and four to six clone inserts per individual were sequenced using both plasmid M13 forward and reverse primers at the Australian Genome Research Facility (AGRF). We estimated to what extent the number of positive clones per individual sequenced here provided information on the variation of these genes at the individual level, using the simulated statistical model based on the probability $f(r, m, n)$ for diploid genomes in the program NegativeMultinomial (<http://www.lirmm.fr/caroux/Bioinformatics/NegativeMultinomial/>). This analysis showed that the number of positive clones used here accounted for up to 31.25% of total MHC variation in each individual. We considered this variation sufficient to retrieve representative variants for each species and preliminarily reconstruct the evolutionary history of these genes in Crocodylia, contrasting with MHC diversity studies where more clones with positive inserts are recommended to be analysed [62].

Two individuals per species were used for the amplification of MHC class II β exon 3 except for *C. mindorensis*, *C. palustris* and *C. novaeguineae* for which a single individual was used due to the sample availability. A single individual per species was used for MHC class II α exons 2 and 3 analysis because preliminary screening showed them to be highly conserved amongst the families Crocodylidae and Alligatoridae (97–100%). Also, samples from two and nine species (out of 20) characterised for the MHC class II α exons 2 and 3, respectively, generated non-MHC or waste sequences although a number of positive clones were sequenced. These species were marked as 'NA' in Table 1.

The criteria for categorising an insert sequence as a true MHC variant for downstream analyses were as follows: forward and reverse strands sequenced were consistent; the sequence was present in two or more of the clones analysed per individual; and/or the sequence was detected in more than one individual within and/or across species. To further filter out amplification and recombination artefacts potentially arising during PCR and cloning, the following criteria were applied to discard sequences: unique sequences that differed by less than 3 bp from a redundant sequence of the same PCR product as recommended by Edwards

Table 3. Primer sets for MHC class II α exons 2 and 3 plus MHC class II β exon 3 used in the current study and their annealing temperatures in the cycling PCR.

Set	Primer ^a	Sequence (5'→3')	Product	Expected product size (bp)	T _A (°C) ^b	Reference
1	II $\alpha 1$ -F	AACGATGAGATCTTCCATGTGG	II α exon 2	171	67	Current study
	II $\alpha 1$ -R	GATCTGGGACCGTGTGCG				
2	II $\alpha 2$ -F	GTGTTTTCGGAGGACCTGTG	II α exon 3	240	67	Current study
	II $\alpha 2$ -R	CAGCCCCAGTGCTCCAC				
3	M2-U	CTCAGTGAAGCCCAAGGTG	II β exon 3	260	61	Liu et al. (2007)
	M2-D	GGCTGCTGTGCTCCACCTGG				

^aForward (F) and reverse primers (R).

^bAnnealing temperature.

doi:10.1371/journal.pone.0087534.t003

et al. [63] and Kloch et al. [64]; and sequences within an individual showing recombination signal from recombination test as described below (RDP3 Beta 34) [65]. More specifically, individuals with more than two sequences were checked for recombination and, if new sequences arose from a combination of the other sequences in the same individual, they were removed from the dataset. True sequences were then named using the gene prefixes of the species followed by DA (an abbreviation for MHC class II α) or DB (an abbreviation for MHC class II β) and then the identification number (Table 1), as recommended by Klein et al. [66].

Datasets and Molecular Diversity Analyses

Three nucleotide sequence datasets were generated in the current study: *i*) a MHC class II α exon 2 dataset consisting of sequences from 18 species of Crocodylia; *ii*) a MHC class II α exon 3 dataset consisting of sequences from 11 species; and *iii*) a MHC class II β exon 3 dataset consisting of sequences from 20 species. This difference in numbers of species among datasets was the result of difficulties in retrieving those genes, as explained above. Forward and reverse nucleotide sequences were overlapped using the BioEdit Sequence Alignment Editor (Ibis Therapeutics, Carlsbad California) to generate a consensus sequence. To assess whether the retrieved sequences show identity to target loci among closely and distantly related vertebrate taxa, basic local alignment searches were performed in BLAST (<http://www.ncbi.nlm.nih.gov/>). Once confirmed, nucleotide sequences were translated into amino acids based on a standard genetic code in the MEGA 5.0 program [67] and published translated amino acid sequences from caiman MHC class II α (AF256650) and β (AF256651, AF256652, and AF277661). Alignments of nucleotide and deduced amino acid (aa) sequences were generated using MUSCLE 3.6 [68]. Indels (insertion-deletion) in the alignment were excluded from downstream analyses. In order to compare degrees of polymorphism within each dataset, molecular diversity indices were calculated, including a pairwise difference between sequences, and numbers of synonymous and nonsynonymous sites. The pairwise difference compares the number of nucleotide substitutions in each pair of MHC sequences using MEGA 5.0, while the number of synonymous and nonsynonymous sites was counted using DnaSP [69].

Identification of Trans-species Polymorphisms and Non-functional Sequences

Trans-species polymorphisms (TSPs) in each MHC class II dataset were identified when identical amino acid sequences were found among two or more different species [65,70,71]. This TSP definition considers solely at the sequence level, not the evolutionary level where this polymorphism is a result of diversifying (balancing) selection [20]; however, selection was also examined as described below in Selection detection tests. Comparisons between these TSPs and phylogenetic trees based on nucleotide sequences were made in order to better visualise the sorting and/sharing of those TSPs within and among clades. Non-functional sequences were identified by the presence of stop codons and/or deletions. Putative functional MHC exons were identified when the entire exon showed a continuous open reading fragment and when conserved amino acid sites (N and C termini and disulfide bridge-forming cysteine), known to be implicated in forming the MHC-antigen binding complex, were present [56].

Selection Detection Tests

In order to assess whether diversifying selection was present in each translated amino acid site of MHC genes studied herein, an estimate of a dN/dS ratio (ω value) performing Bayesian Inference was calculated on the following two alignment datasets: the MHC class II α alignment and MHC class II β alignment (Figures S3 and S4). Each dataset consisted of MHC sequences generated in the current study and those available in GenBank. Missing data, deletions, and stop codons (unknown characters) were treated as gaps to allow information from full-length sequences to be retained. Selection tests of the MHC class II α dataset from different species of Crocodylia were investigated using an alignment that contained 28 MHC class II α sequences generated herein and a caiman MHC class II α sequence available in GenBank (AF256650). In addition, the MHC class II β dataset from 20 species of Crocodylia was tested for selection using an input alignment that contained 18 GenBank exon 2 sequences (FJ886734–FJ886741 from the Nile crocodile and AY491421–AY491430 from Chinese alligator), three GenBank sequences of exons 2 and 3 (AF256651, AF256652 and AF277661 from the spectacled caiman), and 72 exon 3 sequences generated in the current study.

An analysis of Bayesian Inference was used to estimate ω of MHC class II α and β alignments, using omegaMap version 0.5 [72]. This method is able to infer sites under diversifying selection in the presence of recombination [15], which may be overestimated by CODEML. OmegaMap measures the selection parameter (ω) and the recombination rate ($\rho = 4N_c c$), both of which are allowed to vary along the sequence alignment or amino acid positions. Two independent runs of omegaMap were conducted as follows: 5×10^5 Markov chain Monte Carlo (MCMC) iterations, 5×10^4 0000 burn-in iterations, 10^2 thinning iterations, codon frequency 1/61, the number of orderings equal to ten, and the ω model set to be independent. Results from both omegaMap runs that matched within an acceptable degree of error were subsequently interpreted, and graphics were created using R version 2.11.1 (<http://www.r-project.org>).

Recombination Tests

In addition to possible technical reasons that could result in recombinant MHC sequences (artefacts) as described above, there are mechanisms of evolution that could result in natural events of recombination [73,74]. In order to test any impact of recombination between MHC class II sequences from Crocodylia, the recombination rate across sequence length (ρ) and mean number of nucleotide substitutions in each alignment dataset (θ) were estimated and subsequently compared using omegaMap. An additional test of recombination was performed on three sequence datasets of MHC class II α exon 2, MHC class II α exon 3, and MHC class II β exon 3 using RDP3 Beta 34, as the test is able to identify possible recombinant sequences from different species of Crocodylia [75]. Default options were set with 10000 permutations and the cut-off of $p = 0.05$, and disentangle overlapping signals plus the Bonferroni correction for multiple comparisons were used. This approach allows recombinant sequences to be identified by comparing results from different recombination detection algorithms implemented in the program, using the following criteria: the recombinant sequence is detected by at least two algorithms; has a high consensus score of more than 60; and has a recombining portion from different parental sequences.

Phylogenetic Inference of MHC Class II α and β

In order to assess orthologous relationships of MHC class II α and β sequences among species of Crocodylia and identify TSPs

with very similar nucleotide sequences from two different species, phylogenetic analyses were performed using maximum likelihood in PHYML [76] and Bayesian method in BEAST version 1.5.4 [77]. The two methods were then compared to trace back the final relationships of all the MHC sequences using expected likelihood weights (ELW) with the 95% confidence tree set in TREE-PUZZLE program [78]. PHYML is found to generate a tree with high accuracy and speed, while BEAST allows enlargement of MCMC steps to average over tree space, and then makes a tree reliable with high posterior probability. The best fit model of molecular evolution [79] was selected using ModelGenerator version 0.85, according to both Bayesian Information Criterion (BIC) [80] and Akaike Information Criterion (AIC) [81]. Support values on branches for the maximum-likelihood and Bayesian methods were assessed with 10^4 nonparametric bootstraps and 5×10^8 MCMC steps (sampling every 10^4 steps and 5×10^4 burn-in steps) respectively.

Phylogenetic analyses described above were performed separately on the following three nucleotide sequence datasets as described in Table S3: *i*) a MHC class II α exon 2 dataset consisting of 19, 2 and 12 sequences from Crocodylia, Aves and Mammalia respectively as well as 4 sequences from Chondrichthyes, which was used as an outgroup; *ii*) a MHC class II α exon 3 dataset consisting of 12 sequences from Crocodylia, and the same sequences from Aves, Mammalia, and Chondrichthyes as the dataset described above; and *iii*) a MHC class II β exon 3 dataset consisting of 75 sequences from Crocodylia, 6 from other Reptilia, 11 from Aves, 10 from Mammalia, and a single sequence from Amphibia, which was used as an outgroup. The best-fitting model was selected for phylogenetic reconstruction as described above. The HKY model with a gamma distribution parameter (alpha) of 1.23 was used for the MHC class II α exon 2 dataset; HKY model with gamma distribution (alpha = 0.59) for the MHC class II α exon 3 dataset; and the TRN model with gamma distribution (alpha = 0.74) for the MHC class II β exon 3 dataset.

Supporting Information

Figure S1 Amino acid alignments of MHC class II α sequences within Crocodylia. Variable positions are relative to the sequence at the top. The first column contains the names of MHC sequences from two families of Crocodylia: Crocodylidae (pale green colour) and Alligatoridae (pale red colour). The second column presents the amino acid alignment in letters. Dots represent amino acid identity to the top sequence; X letters represent unknown amino acids due to single-base deletions; asterisks represent stop codons; and numbers above the alignments represent the order of amino acid positions. Sites in boxes with closed triangles indicate conserved residues of antigen N and C termini on the peptide-binding region of the MHC class II α exon 2 alignment, as described in Kaufman et al. (1994); and sites in boxes linked with a line indicate the cysteine bridge (C-C) observed in the MHC class II α exon 3 alignment. Background colours in the alignments indicate degrees of amino acid identity: 100% in blue; 80–100% in yellow; and below 80% in white. The end of the alignments shows GenBank accession numbers. Trans-species polymorphisms (TSPs) are represented by numbers immediately after each MHC sequence. The same TSP is assigned with the same sequential number. (PDF)

Figure S2 Amino acid alignment of MHC class II β sequences across 20 species of Crocodylia. Variable positions are relative to the sequence at the top. The first column contains the names of MHC sequences from two families of

Crocodylia: Crocodylidae (pale green colour) and Alligatoridae (pale red colour). The second column presents the amino acid alignment in letters. Dots represent amino acid identity to the top sequence; X letters represent unknown amino acids due to single-base deletions; asterisks represent stop codons; and numbers above the alignments represent the order of amino acid positions. Sites 24 and 80 in boxes linked with a line indicate the cysteine bridge (C-C); and sites 42–66 in the red box indicate a CD4+ binding region. Background colours in the alignments indicate degrees of amino acid identity: 100% in blue; 80–100% in yellow; and below 80% in white. The end of the alignment shows GenBank accession numbers. Trans-species polymorphisms (TSPs) are represented by numbers immediately after each MHC sequence. The same TSP is assigned with the same sequential number. (PDF)

Figure S3 Amino acid alignment of MHC class II β exons 2 and 3 used for selection detection tests. The first column contains the names of MHC sequences. The second column presents the amino acid alignment in letters. Question marks represent unknown amino acids, and numbers above the alignments represent the order of amino acid positions. Sites in boxes with open triangles indicate potential peptide contact residues on the peptide binding region of the HLA-DRB1 molecule based on crystallography models (Bondinas et al. 2007), and those in boxes with closed triangles indicate conserved residues of antigen N and C termini on the peptide-binding region of the MHC class II β molecule (Kaufman et al. 1994). (PDF)

Figure S4 Amino acid alignment of MHC class II α exons 2 and 3 used for selection detection tests. The first column contains the names of MHC sequences. The second column presents the amino acid alignment in letters. Question marks represent unknown amino acids, and numbers above the alignments represent the order of amino acid positions. Sites in boxes with open triangles indicate potential peptide contact residues on the peptide binding region of the HLA-DRA molecule based on crystallography models (Bondinas et al. 2007), and those in boxes with closed triangles indicate conserved residues of antigen N and C termini on the peptide-binding region of the MHC class II α molecule (Kaufman et al. 1994). (PDF)

Figure S5 Maximum-likelihood trees of (A) MHC class II α exon 2, and (B) exon 3. Sequences of MHC class II α among different species of Crocodylia, birds and mammals are analysed using the shark *Gici* sequences as an outgroup. Brackets in the middle show vertebrate groups to which the MHC class II α sequences belong. Bootstrap values (BV) over 50 are provided on branches. ELW tests show that these two trees did not differ significantly from their Bayesian trees of MHC class II α exon 2 sequences described in Fig. 2A (confidence tree set = 0.79 and 0.21), and MHC class II α exon 3 sequences described in Fig. 2B (confidence tree set = 0.62 and 0.38). (PDF)

Figure S6 Maximum-likelihood analysis of MHC class II β exon 3. Sequences of MHC class II β among different species of Crocodylia, Aves and Mammalia are analysed using the amphibian sequence as an outgroup. Brackets on the right show vertebrate groups to which the MHC class II β sequences belong. Clades were defined on the basis of their monophyletic groupings and high posterior probabilities. Bootstrap values (BV) over 50 are provided on branches, and some below 50 are selectively present. (PDF)

Figure S7 Variable nucleotide sites of MHC class II β exon 3 sequences within Crocodylia. Variable nucleotide positions are relative to the MHC class II β sequence from *C. rhombifer* (*Crrh-DB04*). Dots represent identical bases to the first sequence. The first column contains the names of MHC class II β sequences. Brackets on the right show Clades 1 and 2 of the MHC sequences, as have been explained in the text. Sites highlighted in red indicate fixed differences between Clade 1 and Clade 2 sequences.

(PDF)

Figure S8 Distribution map of the Order Crocodylia showing the number of species per country. This map does not show the actual distribution within each country, but the detailed distribution and list of species in each country can be obtained from <http://crocodylian.com/cnhc/cnhc.html>. This website also contains a list of primary references supporting this distribution map.

(PDF)

Table S1 List of exon 3 sequences of MHC class II β across 20 species of Crocodylia investigated in the current studies (Up to two individuals per species studied).

(PDF)

Table S2 Summary of MHC class II β exon 3 sequences observed within a representative from each species of Crocodylia, where more than three sequences corresponding to at least two loci have been identified in this study.

(PDF)

Table S3 List of MHC class II α and β sequences available in GenBank used in three datasets for phylogenetic analyses of MHC class II sequences among major vertebrate classes.

(PDF)

Appendix S1 Characterisation of MHC class II α exons 2 and 3 within Crocodylia.

References

- Belov K, Deakin JE, Papenfuss AT, Baker ML, Melman SD, et al. (2006) Reconstructing an Ancestral Mammalian Immune Supercomplex from a Marsupial Major Histocompatibility Complex. *PLoS Biol* 4: 317–328.
- Kaufman J, Milne S, Gobel TW, Walker BA, Jacob JP, et al. (1999) The chicken B locus is a minimal essential major histocompatibility complex. *Nature* 401: 923–925.
- Sambrook JG, Figueroa F, Beck S (2005) A genome-wide survey of Major Histocompatibility Complex (MHC) genes and their paralogues in zebrafish. *BMC Genomics* 6: 152.
- Kulski JK, Shiina T, Anzai T, Kohara S, Inoko H (2002) Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man. *Immunol Rev* 190: 95–122.
- Burri R, Hirzel HN, Salamin N, Roulin A, Fumagalli L (2008) Evolutionary patterns of MHC class II B in owls and their implications for the understanding of avian MHC evolution. *Mol Biol Evol* 25: 1180–1191.
- Nei M, Gu X, Sitnikova T (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc Natl Acad Sci USA* 94: 7799–7806.
- Alcaide M, Edwards SV, Negro JJ (2007) Characterization, Polymorphism, and Evolution of MHC Class II B Genes in Birds of Prey. *J Mol Evol* 65: 541–554.
- Murphy KP (2012) *Janeway's Immunobiology*. New York: Garland Science.
- Hughes AL, Yeager M (1998) Histocompatibility complex loci of vertebrates. *Annu Rev Genet* 32: 415–435.
- Shiina T, Shimizu S, Hosomichi K, Kohara S, Watanabe S, et al. (2004) Comparative genomic analysis of two avian (quail and chicken) MHC regions. *J Immunol* 172: 6751–6763.
- Hurt P, Walter L, Sudbrak R, Klages S, Muller I, et al. (2004) The genomic sequence and comparative analysis of the rat major histocompatibility complex. *Genome Res* 14: 631–639.
- Kulski JK, Inoko H (2005) Major Histocompatibility Complex (MHC) Genes. *Encyclopedia of Life Sciences*: John Wiley & Sons, Ltd.
- Murphy K, Travers P, Walport M (2008) *Immunobiology*. New York and London: Garland Science.
- Hughes AL (1999) *Adaptive evolution of genes and genomes*. New York: Oxford University Press.
- Alcaide M, Edwards SV, Negro JJ, Serrano D, Tella JL (2008) Extensive polymorphism and geographical variation at a positively selected MHC class II B gene of the lesser kestrel (*Falco naumanni*). *Mol Ecol* 17: 2652–2665.
- Gómez D, Conejers P, Marshall SH, Consuegra S (2010) MHC evolution in three salmonid species: a comparison between class II alpha and beta genes. *Immunogenetics* 62: 531–542.
- Furlong RF, Yang Z (2008) Diversifying and Purifying Selection in the Peptide Binding Region of *DRB* in Mammals. *J Mol Evol* 66: 384–394.
- Hughes AL, Yeager M (1997) Comparative evolutionary rates of introns and exons in murine rodents. *J Mol Evol* 45: 125–130.
- Takahata N (1990) A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. *Proc Natl Acad Sci USA* 87: 2419–2423.
- Klein J (1987) Origin of major histocompatibility complex polymorphism: The trans-species hypothesis. *Human Immunol* 19: 155–162.
- Glaberman S, Moreno MA, Caccone A (2009) Characterization and evolution of MHC class II B genes in Galápagos marine iguanas (*Amblyrhynchus cristatus*). *Dev Comp Immunol* 33: 939–947.
- Liu H, Wu X, Yan P, Jiang Z (2007) Polymorphism of Exon 3 of MHC Class II B Gene in Chinese Alligator (*Alligator sinensis*). *J Genet Genomics* 34: 918–929.
- Li E, Yan P, X W (2010) Sequence variation at exon 2 of MHC class II B gene in Nile crocodile (*Crocodylus niloticus*). *Acta Laser Biology Sinica* 19: 804–810.

(PDF)

Appendix S2 Characterisation of MHC class II β exon 3 within Crocodylia.

(PDF)

Appendix S3 Effect of recombination at MHC class II α and β .

(PDF)

Appendix S4 BLASTN searches of MHC class II primers used in this study against genome drafts of *Alligator mississippiensis* (v0.2.1), *Crocodylus porosus* (v0.2) and *Gangeticus gavalis* (v0.2). These genome resources are available in GenBank and can be accessed with authors' permission (St John et al. 2012). Good hits were observed between the primers and putative MHC class II sequences from those species, and this is likely to suggest an unbiased amplification of the presumed targets using the current primers. All the MHC class II sequences on the genomes were annotated and are expected to be published in another manuscript (Jaratlerdsiri et al. unpublished data).

(PDF)

Acknowledgments

We thank Dr Travis Glenn, Dr Kent Vliet, Robert Godshalk, Mitch Eaton and Dr Matthew Shirley, who kindly provided us with many of the DNA samples included in this investigation. We are grateful to Porosus Pty. Ltd. for providing the Australian saltwater and Johnston's crocodile samples. Many thanks go to Dr Matthew Shirley, Dr Camilla Whittington and Dr Karma Nidup for proofreading and providing advice on manuscript drafts as well as Dr Simon Ho for invaluable comments on phylogenetic analyses.

Author Contributions

Conceived and designed the experiments: JG WJ. Performed the experiments: WJ. Analyzed the data: WJ JG. Contributed reagents/materials/analysis tools: SRI DPH LGM WJ JG. Wrote the paper: WJ JG SRI DPH LGM. Initiated and directed research: JG. Revised the article critically for important interpretation: SRI DPH JG. Final approval of the version to be published: WJ SRI DPH LGM JG.

24. St John JA, Braun EL, Isberg SR, Miles LG, Chong AY, et al. (2012) Sequencing three crocodylian genomes to illuminate the evolution of archosaurs and amniotes. *Genome Biol* 13: 415.
25. Bondinas GP, Moustakas AK, Papadopoulos GK (2007) The spectrum of HLA-DQ and HLA-DR alleles, 2006: a listing correlating sequence and structure with function. *Immunogenetics* 59: 539–553.
26. Jacobson ER (2007) Parasites and Parasitic Diseases of Reptiles. In: Jacobson ER, editor. *Infectious Diseases and Pathology of Reptiles*. Boca Raton, FL: CRC Press.
27. Takahata N, Nei M (1990) Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics* 124: 967–978.
28. Oaks JR (2011) A time-calibrated species tree of Crocodylia reveals a recent radiation of the true crocodyles. *Evolution* 65: 3285–3297.
29. Gustafsson K, Andersson L (1994) Structure and polymorphism of horse MHC class II DRB genes: convergent evolution in the antigen binding site. *Immunogenetics* 39: 355–358.
30. Kriener K, O'huigin C, Tichy H, Klein J (2000) Convergent evolution of major histocompatibility complex molecules in humans and New World monkeys. *Immunogenetics* 51: 169–178.
31. Strand T, Westerdahl H, Höglund J, V Alatalo R, Siitari H (2007) The Mhc class II of the Black grouse (*Tetrao tetrix*) consists of low numbers of B and Y genes with variable diversity and expression. *Immunogenetics* 59: 725–734.
32. Ballingall KT, McKeever DJ (2005) Conservation of promoter, coding and intronic regions of the non-classical MHC class II DYA gene suggests evolution under functional constraints. *Anim Genet* 36: 237–239.
33. Gongora R, Figueroa F, O'Huigin C, Klein J (1997) *HLA-DRB9* - Possible Remnant of an Ancient Functional *DRB* Subregion. *Scand J Immunol* 45: 504–510.
34. Lekutis C, Letwin NL (1995) Biochemical and Molecular Characterization of Rhesus Monkey Major Histocompatibility Complex Class II DR. *Hum Immunol* 43: 72–80.
35. Aarnink A, Estrade L, Apoil P-A, Kita YF, Saitou N, et al. (2010) Study of cynomolgus monkey (*Macaca fascicularis*) DRA polymorphism in four populations. *Immunogenetics* 62: 123–136.
36. Hughes AL, Nei M (1989) Nucleotide substitution at major histocompatibility complex class II loci: Evidence for overdominant selection. *Proc Natl Acad Sci USA* 86: 958–962.
37. Sommer S (2005) The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Front Zool* 2: 16–33.
38. Pierney SB, Oliver MK (2006) The evolutionary ecology of the major histocompatibility complex. *Heredity* 96: 7–21.
39. Anisimova M, Nielsen R, Yang Z (2003) Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. *Genetics* 164: 1229–1236.
40. Gangoso L, Alcaide M, Grande JM, Muñoz J, Talbot SL, et al. (2012) Colonizing the world in spite of reduced MHC variation. *J Evol Biol* 25: 1438–1447.
41. Eo SH, DeWoody JA (2010) Evolutionary rates of mitochondrial genomes correspond to diversification rates and to contemporary species richness in birds and reptiles. *Proc R Soc Lond B Biol Sci* 277: 3587–3592.
42. Hugall AF, Foster R, Lee MSY (2007) Calibration choice, rate smoothing, and the pattern of tetrapod diversification according to the long nuclear gene RAG-1. *Syst Biol* 56: 543–563.
43. Isberg SR, Chen Y, Barker SG, Moran C (2004) Analysis of Microsatellites and Parentage Testing in Saltwater Crocodyles. *J Hered* 95: 445–449.
44. Davis LM, Glenn TC, Strickland DC, Guillette LJ, Elsey RM, et al. (2002) Microsatellite DNA analyses support an east-west phylogeographic split of American alligator populations. *J Exp Zool* 294: 352–372.
45. Villela PMS, Coutinho LL, Piña CI, Verdade LM (2008) Macrogeographic Genetic Variation in Broad-Snouted Caiman (*Caiman latirostris*). *J Exp Zool* 309A: 628–636.
46. Jaratlerdsiri W, Isberg SR, Higgins DP, Gongora J (2012) MHC class I of saltwater crocodyles (*Crocodylus porosus*): polymorphism and balancing selection. *Immunogenetics* 64: 825–838.
47. Alfonso C, Karlsson L (2000) Nonclassical MHC class II molecules. *Annu Rev Immunol* 18: 113–142.
48. Ballingall KT, Ellis SA, MacHugh ND, Archibald SD, McKeever DJ (2004) The DY genes of the cattle MHC: expression and comparative analysis of an unusual class II MHC gene pair. *Immunogenetics* 55: 748–755.
49. Stone RT, Muggli-Cockett NE (1990) Partial nucleotide sequence of a novel bovine major histocompatibility complex class II beta chain gene, *BoLA-DIB*. *Anim Genet* 21: 353–360.
50. Chaves LD, Krueth SB, Reed KM (2009) Defining the turkey MHC: sequence and genes of the B locus. *J Immunol* 183: 6530–6537.
51. Pigiucci M (2008) Is evolvability evolvable? *Nat Rev Genet* 9: 75–82.
52. Edwards SV, Hess CM, Gasper J, Garrigan D (1999) Toward an evolutionary genomics of the avian Mhc. *Immunol Rev* 167: 119–132.
53. Miller HC, Lambert DM (2004) Gene duplication and gene conversion in class II MHC genes of New Zealand robins (Petroicidae). *Immunogenetics* 56: 178–191.
54. Hess CM, Edwards SV (2002) The evolution of the major histocompatibility complex in birds. *BioScience* 52: 423–431.
55. Jaratlerdsiri W, Isberg SR, Higgins DP, Ho SY, Salomonsen J, et al. (2014) Evolution of MHC class I in the Order Crocodylia. *Immunogenetics* 66: 53–65.
56. Kaufman J, Salomonsen J, Flajnik MF (1994) Evolutionary conservation of MHC class I and class II molecules-different yet the same. *Semin Immunol* 6: 411–424.
57. Miles LG, Lance SL, Isberg SR, Moran C, Glenn TC (2009) Cross-species amplification of microsatellites in crocodylians: assessment and applications for the future. *Conserv Genet* 10: 935–954.
58. Jaratlerdsiri W, Rodríguez-Zárate CJ, Isberg SR, Damayanti CS, Miles LG, et al. (2009) Distribution of endogenous retroviruses in crocodylians. *J Virol* 83: 10305–10308.
59. Kriener K, O'huigin C, Klein J (2001) Independent origin of functional MHC class II genes in humans and new world monkeys. *Hum Immunol* 62: 1–14.
60. Slade RW, Mayer WE (1995) The Expressed Class II a-Chain Genes of the Marsupial Major Histocompatibility Complex Belong to Eutherian Mammal Gene Families. *Mol Biol Evol* 12: 441–450.
61. Kamath PL, Getz WM (2011) Adaptive molecular evolution of the Major Histocompatibility Complex genes, *DRA* and *DQA*, in the genus *Equus*. *BMC Evol Biol* 11: 128.
62. Galan M, Guivier E, Caraux G, Charbonnel N, Cosson JF (2010) A 454 multiplex sequencing method for rapid and reliable genotyping of highly polymorphic genes in large-scale studies. *BMC Genomics* 11: 296.
63. Edwards SV, Grahn M, Potts WK (1995) Dynamics of MHC evolution in birds and crocodylians: amplification of class II genes with degenerate primers. *Mol Ecol* 4: 719–729.
64. Kloch A, Babik W, Bajer A, Siński E, Radwan J (2010) Effects of an MHC-DRB genotype and allele number on the load of gut parasites in the bank vole *Myodes glareolus*. *Mol Ecol* 19: 255–265.
65. Anmarkrud JA, Johnsen A, Bachmann L, Lifjeld JT (2010) Ancestral polymorphism in exon 2 of bluethroat (*Luscinia svecica*) MHC class II B genes. *J Evol Biol* 23: 1206–1217.
66. Klein J, Bontrop RE, Dawkins RL, Erlich HA, Gyllenstein UB, et al. (1990) Nomenclature for the major histocompatibility complexes of different species: a proposal. *Immunogenetics* 31: 217–219.
67. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* 24: 1596–1599.
68. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32: 1792–1797.
69. Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.
70. Ottová E, Šimková A, Martín J-F, de Belloq JG, Gelnar M, et al. (2005) Evolution and trans-species polymorphism of MHC class IIβ genes in cyprinid fish. *Fish Shellfish Immunol* 18: 199–222.
71. Klein J, Sato A, Nikolaidis N (2007) MHC, TSP and the Origin of Species: From Immunogenetics to Evolutionary Genetics. *Annu Rev Genet* 41: 281–304.
72. Wilson DJ, McVean G (2006) Estimating Diversifying Selection and Functional Constraint in the Presence of Recombination. *Genetics* 172: 1411–1425.
73. Jeffreys A, May C (2004) Intense and highly localized gene conversion activity in human meiotic crossover hot spots. *Nature* 36: 151–156.
74. Martinsohn JT, Sousa AB, Guethlein LA, Howard JC (1999) The gene conversion hypothesis of MHC evolution: a review. *Immunogenetics* 50: 168.
75. Martin DP, Williamson C, Posada D (2005) RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* 21: 260–262.
76. Guindon S, Gascuel O (2003) A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52: 696–704.
77. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7: 214.
78. Schmidt HA, von Haeseler A (2003) Maximum-Likelihood Analysis Using TREE-PUZZLE. In: Baxevanis AD, Davison DB, Page RDM, Stormo G, Stein L, editors. *Current Protocols in Bioinformatics*. New York: Wiley and Sons. 25093–25097.
79. Keane TM, Creevey CJ, Pentony MM, Naughton TJ, O'McInerney J (2006) Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol Biol* 6: 29–45.
80. Schwarz G (1978) Estimating the dimension of a model. *Ann Stat* 6: 461–464.
81. Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Aut Control* 19: 716–723.