

Bates College

**SCARAB**

---

All Faculty Scholarship

Departments and Programs

---

2022

## Meta-analysis of major histocompatibility complex (MHC) class IIA reveals polymorphism and positive selection in many vertebrate species

Donald Dearborn

*Bates College*, [ddearbor@bates.edu](mailto:ddearbor@bates.edu)

Sophie Warren

*Bates College*

Frank Hailer

*College of Biomedical and Life Sciences*

Follow this and additional works at: [https://scarab.bates.edu/faculty\\_publications](https://scarab.bates.edu/faculty_publications)

---

### Recommended Citation

Dearborn DC, Warren S, & Hailer F. (2022). Meta-analysis of major histocompatibility complex (MHC) class IIA reveals polymorphism and positive selection in many vertebrate species. *Molecular Ecology*, 31(24), 6390–6406.

This Article is brought to you for free and open access by the Departments and Programs at SCARAB. It has been accepted for inclusion in All Faculty Scholarship by an authorized administrator of SCARAB. For more information, please contact [batesscarab@bates.edu](mailto:batesscarab@bates.edu).

# Meta-analysis of major histocompatibility complex (MHC) class IIA reveals polymorphism and positive selection in many vertebrate species

Donald C. Dearborn<sup>1,2</sup>  | Sophie Warren<sup>1</sup> | Frank Hailer<sup>3</sup> 

<sup>1</sup>Biology Department, Bates College, Lewiston, Maine, USA

<sup>2</sup>Roux Institute, Northeastern University, Portland, Maine, USA

<sup>3</sup>Organisms and Environment, School of Biosciences, Cardiff University, Cardiff, UK

## Correspondence

Donald C. Dearborn, Biology Department, Bates College, 44 Campus Ave, Lewiston, ME, USA.

Email: [ddearbor@bates.edu](mailto:ddearbor@bates.edu)

Frank Hailer, Organisms and Environment, School of Biosciences, Cardiff University, Sir Martin Evans Building, Museum Avenue, Cardiff CF10 3AX, Wales, UK. Email: [hailerf@cardiff.ac.uk](mailto:hailerf@cardiff.ac.uk)

## Present address

Sophie Warren, Department of Health Policy, London School of Economics and Political Science, London, UK

## Funding information

National Institute of General Medical Sciences, Grant/Award Number: P20GM0103423; National Institutes of Health; National Evolutionary Synthesis Center, Grant/Award Number: NSF #EF-0905606; School of Biosciences, Cardiff University

Handling Editor: David Coltman

## Abstract

Pathogen-mediated selection and sexual selection are important drivers of evolution. Both processes are known to target genes of the major histocompatibility complex (MHC), a gene family encoding cell-surface proteins that display pathogen peptides to the immune system. The MHC is also a model for understanding processes such as gene duplication and trans-species allele sharing. The class II MHC protein is a heterodimer whose peptide-binding groove is encoded by an MHC-IIA gene and an MHC-IIB gene. However, our literature review found that class II MHC papers on infectious disease or sexual selection included IIA data only 18% and 9% of the time, respectively. To assess whether greater emphasis on MHC-IIA is warranted, we analysed MHC-IIA sequence data from 50 species of vertebrates (fish, amphibians, birds, mammals) to test for polymorphism and positive selection. We found that the number of MHC-IIA alleles within a species was often high, and covaried with sample size and number of MHC-IIA genes assayed. While MHC-IIA variability tended to be lower than that of MHC-IIB, the difference was only ~25%, with ~3 fewer IIA alleles than IIB. Furthermore, the unexpectedly high MHC-IIA variability showed clear signatures of positive selection in most species, and positive selection on MHC-IIA was stronger in fish than in other surveyed vertebrate groups. Our findings underscore that MHC-IIA can be an important target of selection. Future studies should therefore expand the characterization of MHC-IIA at both allelic and genomic scales, and incorporate MHC-IIA into models of fitness consequences of MHC variation.

## KEYWORDS

allelic diversity, diversifying selection, heterodimer, meta-analysis, MHC, pathogen-mediated selection

## 1 | INTRODUCTION

Across jawed vertebrates, the genes of the major histocompatibility complex (MHC) make cell-surface proteins that display pathogen peptides to the immune system (Murphy & Weaver, 2017). The

MHC is central to host-pathogen interactions, and different MHC alleles encode proteins that can differ in their effectiveness against particular pathogens (Sepil et al., 2013; Wegner et al., 2003). This makes MHC variation a key component of host-parasite dynamics (Jeffery & Bangham, 2000; Spurgin & Richardson, 2010). In turn,

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Molecular Ecology* published by John Wiley & Sons Ltd.

fitness differences in MHC genotypes among prospective mates and future offspring can make MHC an important target of sexual selection (Kamiya et al., 2014; Milinski, 2006). More broadly, studying the MHC has also become valuable for understanding an array of evolutionary processes, including frequency-dependent selection (Phillips et al., 2018), heterozygote advantage (Penn et al., 2002), gene duplication and deletion (Nei & Rooney, 2005), copy number variation (Minias et al., 2019), gene conversion (Spurgin et al., 2011), adaptive introgression (Dudek et al., 2019), and trans-species polymorphism (Lighten et al., 2017). The MHC is therefore a key focus of a range of studies across evolution and ecology.

MHC proteins fall in two categories: class I MHC is expressed on all cell types and displays intracellularly derived peptides, whereas class II MHC is only on professional antigen-presenting cells and displays peptides from extracellularly encountered pathogens (Murphy & Weaver, 2017). Structurally, the peptide binding groove of the MHC protein is encoded by a single gene in class I MHC, whereas two different genes jointly encode the binding groove in class II MHC (Figure 1). Specifically, class II MHC is a heterodimer consisting of an alpha subunit from an MHC-IIA gene, and a beta subunit from an MHC-IIB gene. Both of the class II subunits have amino acids that directly interact with the pathogen peptide (Brown et al., 1993) and thus have the potential to influence the effectiveness or specificity of peptide binding and pathogen defence.

Across taxa, though, "...most research has focused on the second and third exons of class I genes and the second exon of class II B genes, because of their traditional consideration as primary targets of pathogen-mediated selection" (Canal et al., 2010). A large body of single-species studies has found that MHC class I and class IIB routinely exhibit high levels of allelic polymorphism (i.e., genetic variability at a given locus) within and between populations and a history of positive selection at peptide binding sites (Höglund, 2009; Piertney & Oliver, 2006; Radwan et al., 2020). Perhaps as a

result, class II meta-analyses or review papers typically focus on MHC-IIB (Bernatchez & Landry, 2003; Burri et al., 2014; Hess & Edwards, 2002; Minias et al., 2018, 2019; Winternitz et al., 2013). In contrast, MHC-IIA loci appear to have received comparatively little attention, despite their contribution to the MHC protein's peptide binding groove (Figure 1) and hence their potential importance for understanding fitness consequences of MHC variation.

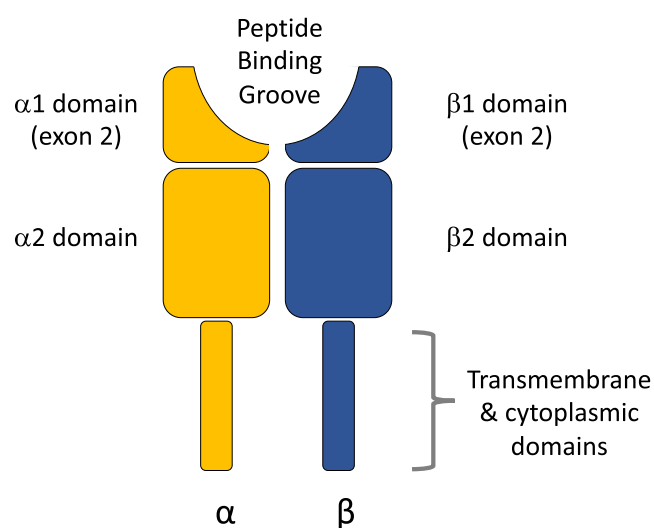
Much of the early studies on MHC-IIA focused on mammals, which led to - and then was further facilitated by - the existence of widely used sets of PCR primers and a clearer understanding of orthology and genomic organization than in other taxa (Kumánovics et al., 2003; Maccari et al., 2017; Trowsdale & Knight, 2013). This early work found evidence of polymorphism in the DQA lineage of MHC-IIA in a range of mammals (Bontrop et al., 1999; Gyllensten & Erlich, 1989; Janova et al., 2009; Scott et al., 1991; Wagner et al., 1996). More recently, a multispecies analysis of mammalian MHC-DQ genes suggested that MHC-IIA might indeed be an important target of diversifying selection: data from the DQA lineage in 10 nonhuman mammal species showed substantial evidence of positive selection within species (Amills et al., 2008), highlighting the potential for MHC-IIA to warrant greater attention in studies of MHC ecology and evolution. In the 15 years since that multispecies analyses of Amills et al. (2008), MHC-IIA sequencing efforts have expanded both within and beyond mammals in a way that makes this a valuable moment to assess the state of the field.

Our synthesis has two main aims. First, we assess the functional attention paid to MHC-IIA by examining the extent to which MHC-IIA is underrepresented, relative to MHC-IIB, in studies of sexual selection or disease association. Second, we compile sequences and associated metadata on MHC-IIA genes in nonhuman vertebrate taxa to test for two phenomena that are widely understood as fundamental to the evolution of MHC-IIB: (1) Does MHC-IIA show pronounced genetic variability (allelic polymorphism) within species? and (2) Is the existing variation in MHC-IIA consistent with a process of positive selection rather than neutral evolution and purifying selection? We address these questions by analysing existing MHC-IIA sequence data in a standardized way across a variety of taxa, including a test of three predictor variables (sample size, number of genes assayed, and taxonomic group within the vertebrates) that could contribute to differences in observed levels of MHC-IIA polymorphism or selection regime in different species. We also make direct, paired comparisons of the polymorphism and selection regime of MHC-IIA and MHC-IIB genes, for cases where population-matched data for both classes of loci are available.

## 2 | MATERIALS AND METHODS

### 2.1 | Prevalence of MHC-IIA versus MHC-IIB in studies of sexual selection or infectious disease

We began by assessing the extent to which studies of sexual selection or infectious disease associations in class II MHC might be disproportionately focused on MHC-IIB rather than MHC-IIA. To that



**FIGURE 1** Schematic illustration of the class II MHC protein. This protein is a heterodimer, in which both the alpha subunit (left; encoded by MHC-IIA loci) and beta subunit (right; encoded by MHC-IIB loci) contribute to the peptide binding groove

end, we conducted literature searches spanning the years 2000 to 2021 for nonhuman empirical studies of those topics, and we then categorized those studies based on whether they contained data on MHC-IIA, MHC-IIB, or both.

Specifically, to assess the sexual selection literature we searched Scopus in November 2021 for papers whose title, abstract, or keywords included (mhc OR histocompatibility) AND ("mate choice" OR "mating" OR "sexual select\*") AND (class AND II). This yielded 354 papers published after 1999. We then excluded review papers, papers about humans, papers that were not actually about sexual selection or mate choice, and papers that lacked class II MHC sequence or haplotype data. We also excluded six studies (4 horse, 2 Soay sheep) that used multigene haplotypes but did not analyse or interpret data at the gene level. Our full set of criteria yielded 116 papers, as listed in Section 1 of the Supporting Information.

To assess the literature on infectious disease associations with MHC-IIA or MHC-IIB alleles, we searched Scopus in November 2021 for papers whose title, abstract, or key words included (mhc OR histocompatibility) AND ("disease\*" OR "pathogen\*" OR "parasite\*" OR "infect\*") AND ("class II" OR "class-II"). This yielded 655 papers published after 1999. We then excluded review papers, papers about humans, papers about laboratory animal models of human disease, papers on noninfectious disease, and papers that lacked class II MHC sequence or haplotype data. We also excluded 12 papers (7 chicken, 4 mouse, 1 *Xenopus*) that knew or assessed multigene MHC haplotypes spanning at least one MHC-IIA gene and one MHC-IIB gene but that did not consider MHC-IIA or MHC-IIB specifically, nor did they attempt to ascertain which genes within the haplotypes were actually the source of the association between haplotype and disease state. Our full set of criteria yielded 95 papers, as listed in Section 1 of the Supporting Information.

We then assessed whether MHC-IIA was becoming more prevalent in such studies over time. To test whether MHC-IIA data was more likely to appear in more recent papers on these topics, we conducted a logistic regression analysis for the 116 mate choice papers and, separately, for the 95 disease association papers. The presence or absence of MHC-IIA data in a class II MHC paper was modelled as a binomial GLM with a logit link, with Year of publication as a continuous predictor variable:

$$\text{IncludesIIA}_i \sim \text{Binomial}(\pi_i, n_i)$$

$$E(\text{IncludesIIA}_i) \sim n_i \times \pi_i \text{ and } \text{var}(\text{IncludesIIA}_i) = n_i \times \pi_i \times (1 - \pi_i)$$

$$\text{logit}(\pi_i) = \eta_i$$

$$\eta_i = \beta_1 + \beta_2 \times \text{Year}_i$$

## 2.2 | MHC-IIA diversity and selection: Literature search

We surveyed the literature for studies of MHC-IIA polymorphism, initially by searching in Scopus for (MHC OR "major histocompatibility

complex") AND ("class II alpha" OR "class II  $\alpha$ " OR "class II A"), and then scanning those papers' references and papers that cited them, expanding forward until no new papers were detected. Papers were used in the analysis if they met the following inclusion criteria:

- Accessible sequence data of MHC-IIA alleles. If data were not available in a repository (e.g., GenBank or IPD-MHC [Maccari et al., 2017]) or in the paper itself, we emailed authors to request data.
- Sequences that spanned at least 90% of exon 2, the exon which encodes the alpha 1 (MHC-IIA) or beta 1 domain (MHC-IIB) of the protein.
- A defined sample size of at least two individuals.
- Samples drawn from a biologically defined group. This was typically wild-caught individuals from one or several populations, but alternatively could include animals from aquaculture or other farmed settings, fish bought at a specific market, or domesticated animals of particular breeds. We excluded a few studies of captive or domesticated animals where the sampled individuals were deliberately chosen for characterization based on a priori knowledge of their MHC genotypes (e.g., using only known MHC-homozygotes).
- One data set per species. This criterion came into play in 11 species for which there were multiple papers that met the other inclusion criteria. In those cases, we selected the paper that provided the most thorough characterization of MHC polymorphism (based on sample size, breadth of sampling, number of genes, and extent of exon 2 coverage). There were no cases in which data on different loci in a species were combined across studies, mainly because of differences between studies in which populations were sampled.

Because MHC characterization methodology varies so widely (O'Connor et al., 2019), we did not restrict our set of studies based on other aspects of methodology, including the number of genes that the authors estimated that they were amplifying, the number of primer pairs used to generate the sequences, the use of DNA versus RNA source material, or the sequencing or sequence characterization methodology (e.g., SSCP, Illumina, cloning and Sanger).

For papers that met our inclusion criteria for MHC-IIA data, we then searched for MHC-IIB data (in the same paper or in a different paper) from the same species. Here we applied the same inclusion criteria, with the additional requirement that the MHC-IIB data be drawn from the same population as the MHC-IIA data, to allow an apples-to-apples paired comparison.

We gave careful consideration to issues involving the number and orthology of genes. Studies of some well-characterized mammalian systems are able to distinguish between orthologues (e.g., DRA, DQA, DPA). However, most studies of nonmodel systems lack information on the existence of orthology or copy number, and lack locus-specific primers. In fact, many studies do not know the number of genes that exist in their organism, nor the number of genes or gene copies that are being amplified by their PCR primers. Instead,

the minimum number of amplified genes is inferred based on the maximum number of sequences recovered from one individual. Consequently, studies of nonmodel species routinely produce a set of MHC-IIA sequences that are from multiple genes and that cannot be sorted into genes or lineages. As a result, any analysis of genetic variability or selection history must jointly examine MHC-IIA alleles irrespective of gene lineage. For consistency, we adopted the same approach for the subset of species for which gene lineages are well characterized (typically in mammals). That is, if a study separately amplified DQA and DRA, we pooled all alleles into a single characterization of polymorphism and selection. This is consistent with a functional allelic framework across loci (rather than a gene-centred framework) to thinking about how much MHC variation exists and how selection has shaped that variation over time. Further exploration of this topic with respect to the data sets in this analysis is provided in Section 2 of the Supporting Information.

### 2.3 | MHC-IIA diversity and selection: Sequence alignment within species

For highly similar sequences, we used Sequencher 5.4 (Gene Codes Corporation, Ann Arbor, MI, USA) to separately align each species' MHC-IIA or MHC-IIB alleles, or used the original authors' alignment when available. For more divergent sequences within species, we made amino acid alignments using Muscle in MEGA X (Kumar et al., 2018). After aligning a species' MHC-IIA or MHC-IIB sequences, we verified or trimmed the sequences to exon 2, trimmed the initial base(s) and final base(s) to include only complete codons, and eliminated any alleles that had nonunique nucleotide sequences after trimming.

### 2.4 | Allelic diversity in MHC-IIA

We imported the aligned sequences of each species separately into MEGA X to measure intraspecific polymorphism in MHC-IIA. We measured each species' variability in nucleotide sequences by counting the number of unique alleles and by computing the average p-distance between sequences using pairwise deletion. Species were included in polymorphism analyses even if they had only a single allele, in which case DNA and amino acid allele numbers were set to one, and nucleotide/amino acid p-distances were recorded as zero. Nearly all MHC polymorphism studies lacked allele frequency data, and thus the p-distance calculations were conducted with a simplified data set that included exactly one copy of each allele, that is, assuming equal allele frequencies. In actuality, the allele frequencies are unlikely to all be equal. The direction and magnitude in which this would impact the estimate of p-distance would depend on the relative divergence of the high- and low-frequency alleles, which of course is unknown in the absence of allele frequency data.

We measured variability in deduced amino acid sequences using the same approach. After translating the trimmed nucleotide

sequences and removing any nonunique amino acid alleles, we counted the number of unique alleles and computed average p-distance in MEGA X using pairwise deletion. As with the above approach for nucleotide data, this makes the necessary simplifying assumption that the alleles occur at equal frequencies.

We then explored whether the number of MHC-IIA alleles might vary across species in relation to the scope of sampling and/or aspects of evolutionary history. Specifically, we ran a generalized linear model (GLM) to test for main effects of the following fixed-effects predictor variables: sample size (NumAnimals; continuous), number of genes assayed (NumGenes; continuous; as determined by the original authors of each data set), and taxonomy (Taxon; categorical: fish, amphibian, bird, mammal). The number of unique nucleotide sequences (NumAllelesDNA) was modeled as a negative binomial GLM with a log link:

$$\text{NumAllelesDNA}_i \sim \text{NegBin}(\mu_i, k)$$

$$E(\text{NumAllelesDNA}_i) = \mu_i \text{ and } \text{var}(\text{NumAllelesDNA}_i) = \mu_i + (\mu_i^2 / k)$$

$$\log(\mu_i) = \eta_i$$

$$\eta_i = \beta_1 + \beta_2 \times \text{NumAnimals}_i + \beta_3 \times \text{NumGenes}_i + \beta_4 \times \text{Taxon}_i$$

### 2.5 | Positive selection on MHC-IIA

Within species, we tested whether MHC-IIA exon 2 sequences showed signatures of positive selection. We used EasyCodeML (Gao et al., 2019) to implement site models in the CodeML program in PAML (Yang, 2007), comparing models that include different amounts of purifying selection, neutral evolution, and positive selection.

First, we built a phylogenetic tree for the trimmed, aligned MHC-IIA nucleotide sequences for each species separately. As described above, to accommodate the large number of species in which orthology or gene identity of alleles was unknown, we combined all MHC-IIA alleles from a given species into a single alignment. We used MEGA X to build a maximum likelihood tree, using a general time reversible model, with gamma-distributed rates among sites plus invariant sites, and pairwise deletion (MEGA's "use all sites") to accommodate indels. This tree and the alignment were then used as inputs for EasyCodeML.

We used the Preset Mode in EasyCodeML to run different site models, comparing the null model M7 against the alternate model M8. Null model M7 is a nearly neutral model in which the values of  $\omega$  (the ratio of nonsynonymous to synonymous substitutions, dN/dS) follow an empirically determined beta distribution bounded by 0 and 1 that describes a mix of neutral evolution and purifying selection across sites. In the alternate model M8, additional parameters allow some codons to evolve by positive selection. Specifically, a beta distribution is fit to a subset of codons that are empirically determined to be evolving via neutral evolution or purifying selection, while a different subset of codons is empirically identified as undergoing

positive selection in which  $\omega_s$  is estimated from the data as some value  $>1$  (Yang et al., 2005).

We extracted three key pieces of output data:

- the overall evidence that a model with positive selection is a better fit than a nearly neutral model, i.e. the statistic of the likelihood-ratio test:  $2(\log[L(M8)] - \log[L(M7)])$
- under M8, the proportion of codons that are estimated to be under positive selection (i.e.  $p_1$  of M8)
- under M8, the magnitude of dN/dS at those codons that are estimated to be under positive selection (i.e.  $\omega_s$  of M8)

We then explored whether the support for M8 might vary across species in relation to the scope of sampling and/or aspects of evolutionary history. Specifically, we ran a generalized linear model to test for main effects of the following fixed-effects predictor variables: sample size (NumAnimals; continuous), number of genes assayed (NumGenes; continuous), and taxonomy (Taxon; categorical: fish, amphibian, bird, mammal). The statistic of the likelihood ratio test of CodeML's model M8 as an alternative to the null hypothesis M7 (M8\_LnL) was modeled as a gamma GLM with a log link:

$$M8\_LnL_i \sim \text{Gamma}(\mu_i, \tau)$$

$$E(M8\_LnL_i) = \mu_i \text{ and } \text{var}(M8\_LnL_i) = (\mu_i / \tau)$$

$$\log(\mu_i) = \eta_i$$

$$\eta_i = \beta_1 + \beta_2 \times \text{NumAnimals}_i + \beta_3 \times \text{NumGenes}_i + \beta_4 \times \text{Taxon}_i$$

## 2.6 | MHC-IIA diversity and selection: Comparisons with MHC-IIB

For a subset of the species with MHC-IIA data, we were able to find MHC-IIB data from the same individuals or populations. This allowed a test of whether MHC-IIA and MHC-IIB differ in number of alleles or in evidence for positive selection, when controlling for the other possible sources of interspecific variation described above. Thus, for these two response variables (NumAllelesDNA and M8\_LnL) we compared MHC-IIA and MHC-IIB by running generalized linear mixed models (GLMMs). These models were similar to the GLMs outlined above except that we included two additional predictor variables: Class as a fixed effect (categorical: IIA, IIB) and Species as a random intercept to account for dependency of MHC-IIA and MHC-IIB data from the same species. This GLMM structure allowed us to ask if a species' MHC-IIA and MHC-IIB differ in diversity or selection history, when controlling for potential effects of sample size (NumAnimals), number of genes amplified (NumGenes), and taxonomic group (Taxon: mammal, bird, amphibian, fish).

The number of unique nucleotide sequences (NumAllelesDNA) was modeled as a Poisson GLMM with a log link:

$$\text{NumAllelesDNA}_{ij} \sim \text{Poisson}(\mu_{ij})$$

$$E(\text{NumAllelesDNA}_{ij}) = \mu_{ij} \text{ and } \text{var}(\text{NumAllelesDNA}_{ij}) = \mu_{ij}$$

$$\log(\mu_{ij}) = \eta_{ij}$$

$$\eta_{ij} = \beta_1 + \beta_2 \times \text{NumAnimals}_{ij} + \beta_3 \times \text{NumGenes}_{ij} + \beta_4 \times \text{Taxon}_{ij} + \beta_5 \times \text{Class}_{ij} + \text{Species}_j + \varepsilon_{ij}$$

$$\text{Species}_j \sim N(0, \sigma^2_{\text{Species}})$$

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

Evidence for positive selection - i.e. the statistic of the likelihood ratio test of CodeML's model M8 as an alternative to the null hypothesis M7 (M8\_LnL) - was modeled as a gamma GLMM with a log link:

$$M8\_LnL_i \sim \text{Gamma}(\mu_i, \tau)$$

$$E(M8\_LnL_i) = \mu_i \text{ and } \text{var}(M8\_LnL_i) = (\mu_i / \tau)$$

$$\log(\mu_i) = \eta_i$$

$$\eta_i = \beta_1 + \beta_2 \times \text{NumAnimals}_i + \beta_3 \times \text{NumGenes}_i + \beta_4 \times \text{Taxon}_i + \beta_5 \times \text{Class}_i + \text{Species}_j + \varepsilon_{ij}$$

$$\text{Species}_j \sim N(0, \sigma^2_{\text{Species}})$$

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

GLMs and GLMMs were run in R 4.1.1 (R Core Team, 2021) with the lme4 (Bates et al., 2015), MASS (Venables & Ripley, 2002), contrast (Kuhn et al., 2021), multcomp (Hothorn et al., 2008), and emmeans (Lenth, 2022) packages. Code is available on Dryad as described in the Data Availability statement. Significance of predictor variables was assessed by single-term deletion, dropping them individually from the model and testing for worse fit using a likelihood test. Differences between levels of significant categorical variables (i.e., taxon) were assessed with post hoc contrasts with unadjusted *p*-values.

GLMs and GLMMs were fitted and validated by assessing the following: over- or under-dispersion; excess zeroes in the response variable; multicollinearity among the predictors; departures from linearity in Q-Q plots; patterns in plots of residuals versus fitted values, scale versus location, and residuals versus leverage; values of Cook's Distance; and patterns in plots of residuals versus predictors.

## 3 | RESULTS

### 3.1 | Prevalence of MHC-IIA versus MHC-IIB in studies of sexual selection or infectious disease

Our literature search found that ecological studies of class II MHC have indeed heavily focused on MHC-IIB at the expense of MHC-IIA (Figure 2). In studies of sexual selection and class II MHC, the



overwhelming majority of papers included only MHC-IIB data ( $n = 106$ ), versus studies that examined MHC-IIA ( $n = 3$ ) or both IIA and IIB ( $n = 7$ ). Put another way, 97% of the 116 class II sexual selection papers examined MHC-IIB data, whereas only 9% examined MHC-IIA data. The use of MHC-IIA data in class II sexual selection studies has not increased with time (no significant effect of publication year in logistic regression: likelihood  $X^2_1 = 2.64$ ,  $p = .104$ , pseudo- $R^2 = 0.04$ ; Table S1).

Likewise, studies of disease associations have disproportionately focused solely on MHC-IIB ( $n = 78$  papers), versus those that examined MHC-IIA ( $n = 9$ ) or both IIA and IIB ( $n = 8$ ). In total, 91% of

the 95 class II disease papers examined MHC-IIB data, whereas only 18% examined MHC-IIA data. The use of MHC-IIA data in class II disease studies has not increased with time (no significant effect of publication year in logistic regression: likelihood  $X^2_1 < 0.01$ ,  $p = .99$ , pseudo- $R^2 < 0.01$ ; Table S2).

### 3.2 | MHC-IIA diversity and selection: Summary of MHC-IIA data sets

We found MHC-IIA studies of 50 vertebrate species that met our inclusion criteria for analysis of diversity (see Appendix 1 and Methods; e.g., including studies which found just one allele in a species). The median sample size of analysed individuals was 26 (range: 2–674), and the number of genes assayed ranged from 1 to 4. Mammals were the most commonly studied group ( $n = 31$ ), followed by bony fish ( $n = 13$ ), birds ( $n = 4$ ), and amphibians ( $n = 2$ ). Of those 50 data sets, 42 contained at least three alleles (the required minimum number of alleles for CodeML) and thus were also amenable to CodeML tests of positive selection. We refer readers to Section 2 of the Supporting Information for detailed comments on our decision to pool alleles across genes within species.

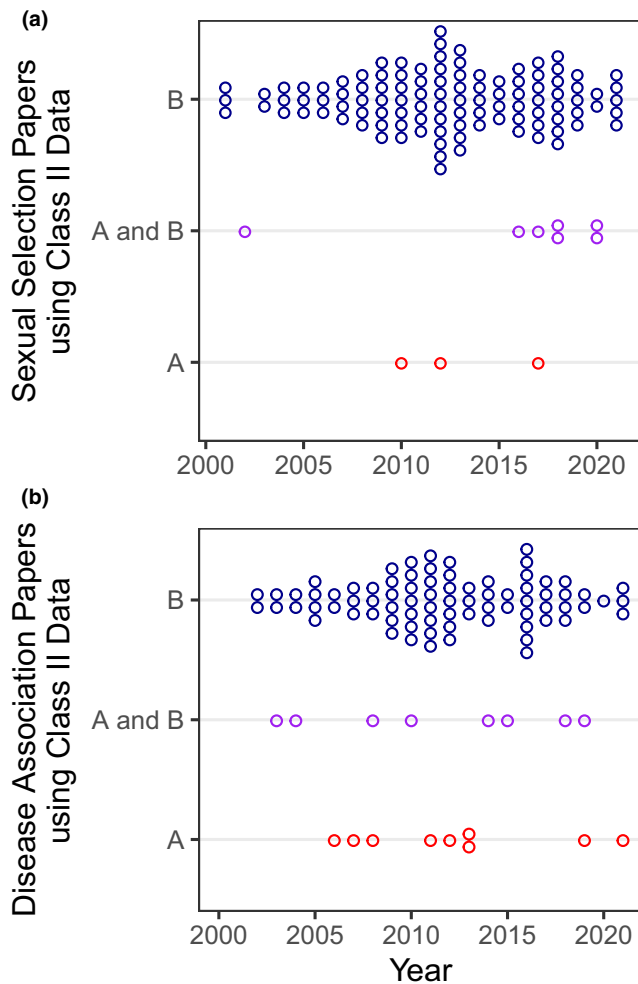
### 3.3 | Allelic diversity in MHC-IIA

The number of MHC-IIA alleles varied widely across data sets, with a median of two genes assayed and eight unique exon 2 nucleotide sequences detected per species (quartiles: 4, 15; Figure 3). We used a GLM to test three variables that might explain some of this interspecific variation in number of alleles. The number of MHC-IIA alleles reported in a species covaried positively with sample size (GLM likelihood ratio  $X^2_{df=1} = 11.1$ ,  $p = .001$ ) and with number of genes assayed (GLM likelihood ratio  $X^2_{df=1} = 6.3$ ,  $p = .012$ ), but there was no significant variation among taxonomic groups (mammal, bird, amphibian, fish; GLM likelihood ratio  $X^2_{df=3} = 6.5$ ,  $p = .088$ ). Explained deviance for the model was 30.8% (Figure 4, Table S3). A separate GLM of only mammals found significant differences in MHC-IIA diversity between the six mammalian orders represented in our data sets (Figure S1, Table S4).

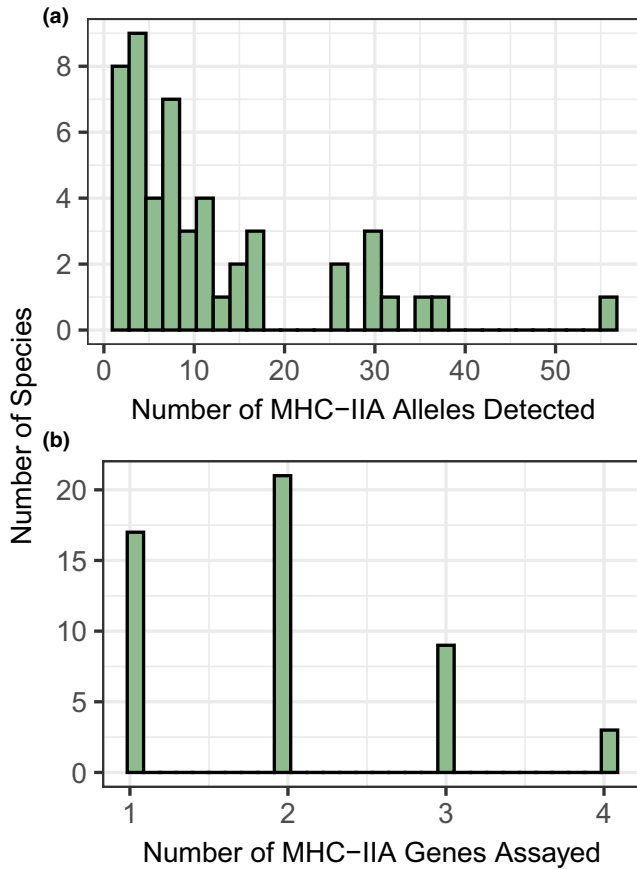
Within species, the reported MHC-IIA alleles were diverse: the mean p-distance between a species' alleles was  $0.125 \pm 0.101$  SD for nucleotide sequences, and  $0.191 \pm 0.130$  SD for inferred amino acid sequences (Figure S2). We also found that the number of amino acid sequences within a species almost perfectly mirrored the underlying number of DNA haplotypes, with a slope of 0.959 (95% CI: 0.926–0.991) and an  $r^2$  of 0.986 (Figure S3).

### 3.4 | Positive selection on MHC-IIA

MHC-IIA sequences showed evidence of positive selection in most species when assessed with CodeML. We measured the statistic of



**FIGURE 2** Prevailing use of MHC-IIB and MHC-IIA data in empirical studies of sexual selection (top) and disease (bottom) published between 2000 and 2021. Each circle represents one study. In both sets of papers, the prevalence of MHC-IIA data has not increased significantly over the past two decades (see text). (a) Studies of sexual selection, classified by whether a paper's data includes MHC-IIA ( $n = 3$ ), IIA and IIB ( $n = 7$ ), or just MHC-IIB ( $n = 106$ ). In total, 97% of the 116 class II sexual selection papers examined MHC-IIB data, whereas only 9% examined MHC-IIA data. (b) Studies of disease associations, classified by whether a paper's data includes MHC-IIA ( $n = 9$ ), IIA and IIB ( $n = 8$ ), or just MHC-IIB ( $n = 78$ ). In total, 91% of the 95 class II disease papers examined MHC-IIB data, whereas only 18% examined MHC-IIA data.



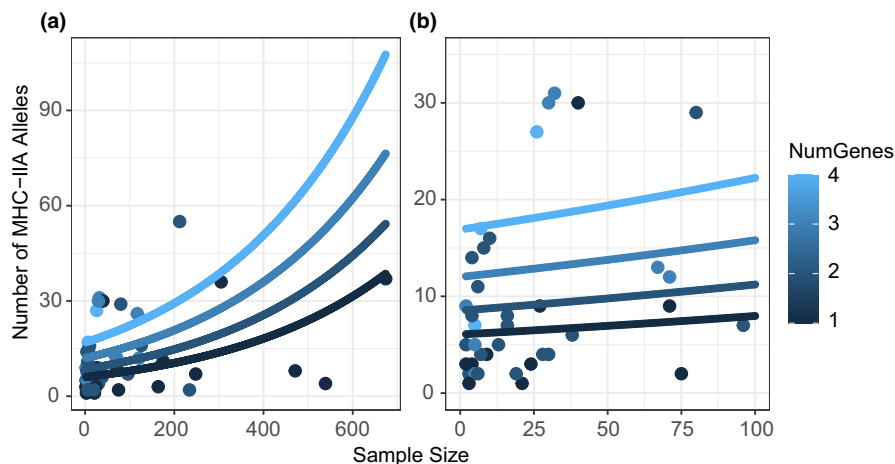
**FIGURE 3** MHC-IIA variability across the 50 analysed data sets. (a) At the allelic level, the detected number of unique MHC-IIA exon 2 nucleotide sequences ranged from 1 to 55 across 50 species, with a median of 8. (b) At the gene level, the data sets from the top panel each assayed between 1 and 4 MHC-IIA genes.

the likelihood ratio test of model M8, which incorporates positive selection at an empirically determined subset of codons, versus null model M7, which allows only purifying selection and drift. In 30 of 42 vertebrate species (71%), the evidence for M8 exceeded the critical value of  $X^2_{df=2} = 5.99$  (Figure S4). Across the 42 species, the proportion of MHC-IIA exon 2 codons estimated as being under positive selection (i.e.,  $p_1$  of M8) had a median value of 0.188 (Figure S4). For those codons classified by M8 as being under positive selection, the median value for  $\omega_s$  of M8 was 5.4 (Figure S4).

We used a GLM to test three variables that might explain some of the interspecific variation in the evidence for positive selection on MHC-IIA. The support for CodeML's model M8 differed between taxonomic groups (GLM likelihood ratio  $X^2_{df=3} = 19.9$ ,  $p = .0002$ ) but did not covary with sample size (GLM likelihood ratio  $X^2_{df=1} = 1.7$ ,  $p = .190$ ) or number of genes assayed (GLM likelihood ratio  $X^2_{df=1} = 0.3$ ,  $p = .584$ ). Explained deviance for the model was 27.5% (Table S5). Post-hoc contrasts across the four taxonomic groups found stronger evidence for positive selection in fish than in any other group ( $p = .001$  vs. mammals,  $p = .010$  vs. birds,  $p = .002$  vs. amphibians), while mammals, amphibians, and birds did not differ from each other (all  $p > .08$ ; Figure 5).

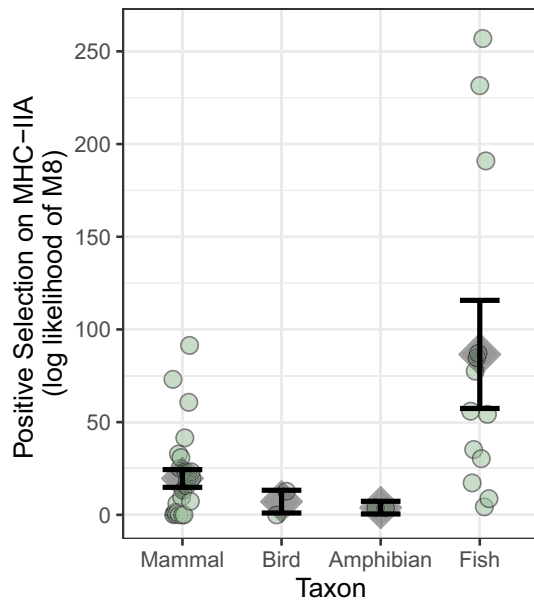
### 3.5 | MHC-IIA diversity and selection: Comparisons with MHC-IIB

Of the 50 species with MHC-IIA data, 27 also had MHC-IIB data from the same individuals or populations. Of those joint IIA / IIB data sets, 20 of 27 species had enough alleles at both IIA and IIB for CodeML analysis of both types of genes. Although we restricted our MHC-IIB data sets to those drawn from at least the same population



**FIGURE 4** Relationship between sample size and observed MHC-IIA variability in published vertebrate data sets. Observed values (dots) and GLM fitted values (lines) for number of IIA alleles in a species (i.e., number of unique exon 2 nucleotide sequences). (a) Full range of data,  $n = 50$  species. (b) Magnified view of both axes for better readability of the 38 studies with sample sizes <100. In the GLM, number of alleles covaried positively with sample size and with number of genes assayed (NumGenes) but did not differ among taxonomic groups (Taxon: mammal, bird, amphibian, fish). The fitted values shown by the lines are averaged across levels of Taxon, weighted by the relative frequency of mammals, birds, amphibians, and fish in the raw data. Sample size (x-axis) and number of genes (colour coding) vary across the ranges observed in the raw data. See Table S3 for GLM parameters.

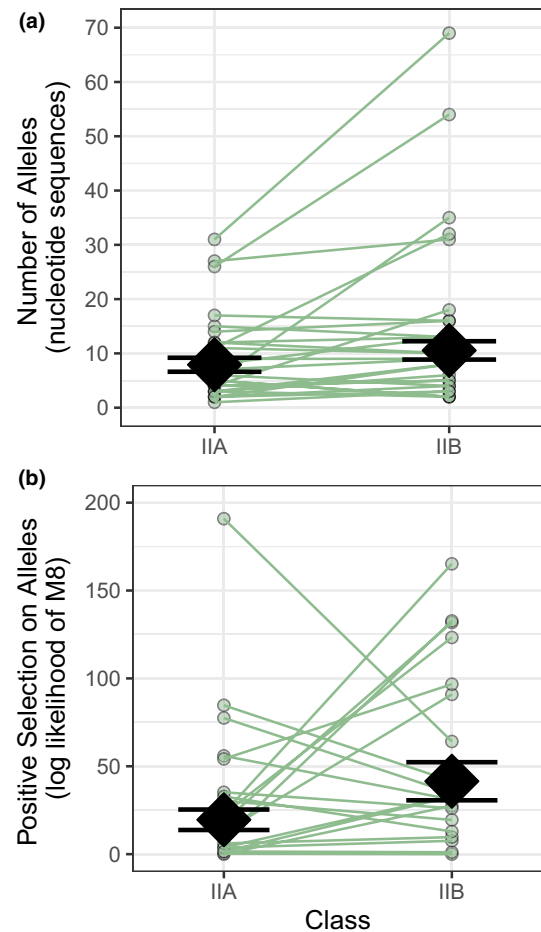




**FIGURE 5** Signals of positive selection at MHC-IIA in vertebrates. Observed values (green dots,  $n = 42$ ) and estimated marginal means (grey diamonds  $\pm$  SE) from a GLM analysing the statistic of the likelihood ratio test of CodeML's model M8 as an alternative to the null hypothesis M7 for IIA in 42 species. This evidence of positive selection on IIA was greater in fish than in other taxonomic groups, but was not significantly related to sample size or number of genes assayed. Points are horizontally dodged for clarity. Estimated marginal means by taxonomic group account for effects of other covariates in the model. See Table S5 for GLM parameters.

as the MHC-IIA data, the corresponding studies frequently differed in sample size (7 of 27 studies) or in number of genes assayed (12 of 27 studies). Thus, to test robustly for differences in polymorphism or selection in MHC-IIA versus MHC-IIB, we fitted GLMMs to compare IIA and IIB while accounting for potential effects of those other variables.

Although MHC-IIA was polymorphic, it was roughly 25% less so than MHC-IIB, as assessed in a Poisson GLMM with log link. Among 27 species with appropriate IIA and IIB polymorphism data, the number of unique exon 2 nucleotide sequences was lower in MHC-IIA than in MHC-IIB (GLMM likelihood ratio  $X^2_{df=1} = 10.9$ ,  $p = .0009$ ), when accounting for positive effects of sample size (GLMM likelihood ratio  $X^2_{df=1} = 24.1$ ,  $p = .0001$ ), number of genes assayed (GLMM likelihood ratio  $X^2_{df=1} = 20.1$ ,  $p = .0001$ ), and differences among taxonomic groups (GLMM likelihood ratio  $X^2_{df=3} = 10.0$ ,  $p = .0183$ ) (Figure 6, Figure S5, Table S6). The overall model had a pseudo- $R^2$  of 0.52 for fixed effects (NumAnimals, NumGenes, and Taxon), and a total pseudo- $R^2$  of 0.86 when including the random effect of species. More concretely, a species had an average of roughly three fewer alleles in MHC-IIA than in MHC-IIB (marginal means  $\pm$  SE:  $7.9 \pm 1.3$  for IIA, vs.  $10.6 \pm 1.7$  for IIB). Post-hoc contrasts on taxonomic groups also found that fish and amphibians both had roughly twice as many class II alleles as mammals ( $z = 3.10$ ,  $p = .002$  and  $z = 1.98$ ,  $p = .048$ , respectively; Figure S5), while no other pairwise comparisons between taxonomic groups were significant (all  $p > .16$ ).



**FIGURE 6** Comparison of MHC-IIA and MHC-IIB diversity (a) and positive selection (b). Observed values (small green dots, with lines connecting data from same species) and estimated marginal means (large black diamonds,  $\pm$  SE) from GLMMs are shown, to account for effects of other covariates. (a) Number of unique class II exon 2 nucleotide sequences in a species. See Table S6 for GLMM parameters. (b) Evidence for positive selection in a species, as measured by the statistic of the likelihood ratio test of CodeML's model M8 as an alternative to the null hypothesis M7. See Table S7 for GLMM parameters. In both GLMMs, the response variable was modeled as a function of fixed main effects of Class (IIA, IIB), sample size, number of genes assayed, and taxonomic group (mammal, bird, amphibian, fish). Species was used as a random effect to account for dependencies in IIA and IIB data sets from the same species.

Using a similar mixed-model approach, we found that signals of positive selection were stronger for MHC-IIB than for MHC-IIA. With 40 matched data sets from 20 species, we fit a GLMM to predict variation between data sets in the statistic of the likelihood ratio test of CodeML's model M8 as an alternative to the null hypothesis M7. This evidence for positive selection was roughly twice as strong for MHC-IIB than for MHC-IIA (GLMM likelihood ratio  $X^2_{df=1} = 5.0$ ,  $p = .0254$ ) when controlling for positive effects of sample size (GLMM likelihood ratio  $X^2_{df=1} = 7.6$ ,  $p = .0057$ ), number of genes assayed (GLMM likelihood ratio  $X^2_{df=1} = 12.8$ ,  $p = .0004$ ), and differences among taxonomic groups (GLMM

likelihood ratio  $X^2_{df=3} = 20.7, p = .0001$  (Figure 6, Figure S6, and Table S7). Marginal and conditional pseudo- $R^2$  for the model were both 0.63, because the random effect of species did not account for a measurable amount of variation in the response variable (unlike the GLMM of number of alleles, where species was important). Post-hoc contrasts on taxonomic groups found stronger evidence for positive selection in fish than in mammals ( $p < .0001$ ) or amphibians ( $p = .0004$ ); other pairwise contrasts were not significant (all  $p > .05$ ; Figure S6).

## 4 | DISCUSSION

As a starting point, our literature summary provided a quantitative confirmation that the genes encoding the alpha subunit of class II MHC have received very little attention, compared to MHC-IIB, in studies of functional variation such as sexual selection or disease associations (Figure 2). Furthermore, this attention to MHC-IIA has not increased markedly over the past two decades. This observed scarcity of MHC-IIA data in studies of sexual selection or disease cannot be entirely attributed to a lack of PCR primers or other foundational studies on IIA: a search for basic characterizations of MHC-IIA genes found studies of 50 different vertebrate species that met our inclusion criteria for assessing polymorphism and positive selection. Our analyses of those 50 species found that MHC-IIA is often polymorphic and shows signatures of positive selection, making the case that studies of ecology, evolution, and conservation in MHC class II should incorporate IIA much more often.

### 4.1 | Conclusions on MHC-IIA variability and patterns of selection

The first broad conclusion from our analysis of those DNA sequence data sets is that MHC-IIA is often polymorphic, with a median of eight and as many as 55 alleles reported. Not surprisingly, the extent of polymorphism varied widely across studies. This interspecific variation in the number of MHC-IIA alleles was not driven by any consistent differences between mammals, birds, amphibians, and fish (although there was variation between mammalian orders – see Figure S1, Table S4). However, the number of MHC-IIA alleles covaried positively with the number of animals sampled, and also with the number of genes assayed. The most likely explanation for the relationship with sample size is simply that greater sampling effort yields more alleles, akin to a species accumulation curve in ecology (Gotelli & Colwell, 2001). In contrast to this straightforward sample size effect, the relationship between number of alleles found and number of genes assayed is more challenging to interpret, as researchers often do not know whether the number of MHC-IIA genes being assayed is representative of the number of MHC-IIA genes actually present in the genome. One possible interpretation of the positive relationship

between number of alleles found and number of genes assayed is a sampling artefact: studies find more alleles if they assay a greater proportion of the (known or unknown number of) existing genes. This explanation is valid only if the number of MHC-IIA genes assayed in a given study does not tightly correspond to the true number of IIA genes in the focal species' genome. An alternative explanation is that the true number of IIA alleles and the true number of IIA genes in a species might covary biologically. For example, evolutionary forces favouring MHC diversification in a species or lineage could lead to gene duplication and subsequent divergence, producing both more genes and more unique alleles over time.

The second main conclusion is that MHC-IIA genes within a species usually showed strong signatures of positive selection, rather than evolving solely via purifying selection and drift. Our CodeML analyses found that, in over 70% of species, MHC-IIA alleles exhibited a significantly better fit to an evolutionary model with positively selected sites than to a null model that allows only purifying selection and drift. Furthermore, a median of 19% of sites in exon 2 were identified as being under positive selection, and the estimated value of dN/dS at these positively selected sites was high, with a median value of 5.4.

As with allelic polymorphism, there was substantial variation among species in the strength of support for positive selection on MHC-IIA. However, the predictors of that interspecific variation were not the same as for IIA polymorphism levels. The signal strength of positive selection on MHC-IIA in a species showed no effect of sample size or number of genes assayed, but was stronger in fish than in mammals, birds, or amphibians. The lack of effect of sample size and number of genes might mean that even a relatively small sample of a species' alleles captures enough of the phylogenetic diversity of MHC sequences to yield a meaningful test for positive selection.

The detection of stronger evidence of positive selection on MHC-IIA in fish than in other groups has several candidate interpretations. One centres on aquaculture. Eight of the 13 IIA data sets from fish were from hatchery or aquaculture settings, where the high density of a single fish species and the spatial overlap of adults and juveniles can set the stage for more pronounced disease transmission (Krkošek et al., 2006; Pulkkinen et al., 2010). If enough time elapses, the resulting pathogen-mediated selection on MHC (Cao et al., 2018; Yang et al., 2016) might not only alter allele frequencies of standing genetic variation, but also lead to signals of positive selection. There might also be direct artificial selection for parasite-resistant traits in aquaculture species (Kube et al., 2012), given that the proportion of farmed or domesticated species in our data set was somewhat higher in fish (8 of 13 species) than in mammals (9 of 31 species). Another explanation for greater positive selection in fish could be the different genome structure of teleosts, in which the major histocompatibility genes are not packed into a dense complex (Grimholt, 2016; Stet et al., 2003). In particular, the lack of linkage between class I and class II genes in fish can decouple how those genes evolve, which

might create more scope for MHC-IIA to respond to pathogen-mediated selection even in the presence of strong selection on class I from different pathogens.

The third main conclusion from our analyses is that although MHC-IIA generally exhibits polymorphism and a history of positive selection, both phenomena are more pronounced in the population-matched IIB data sets that we examined. In the subset of 27 species with appropriate data, studies of MHC-IIB reported an average of three more alleles than studies of MHC-IIA (10.6 vs. 7.9) when controlling for potential effects of sample size, number of genes assayed, and taxonomic group. Likewise, in the subset of 20 species amenable to CodeML analysis for both types of genes, MHC-IIB showed stronger evidence for positive selection than MHC-IIA when controlling for those same covariates. Given the tight interaction of alpha and beta subunits in the MHC heterodimer (Figure 1), and the key role of the MHC in immune defence and mate choice, our findings nevertheless underscore the importance of examining both IIB and IIA sequences in such studies. As genomic organization of the MHC becomes better understood in a wider range of species, it should become possible to test whether polymorphism and diversifying selection in MHC-IIA are more evident in A-B gene pairs that experience low recombination rates, as predicted by the idea that high A-B recombination constrains one subunit to a "best average fit" with the alleles of the other, more polymorphic, subunit (Kaufman, 1999).

## 4.2 | Calls for future studies

Class IIA of the MHC clearly warrants more attention. Our analysis of MHC-IIA from 50 species of vertebrates provides variable but important evidence for (i) allelic polymorphism and (ii) a history of positive selection. This evidence is in line with theoretical expectations, given the contribution of IIA to the peptide binding groove, but is strongly at odds with the widespread lack of IIA data in studies of the functional importance of class II diversity. This leads us to make three broad calls for future studies.

First, we encourage researchers to characterize allelic polymorphism and selection history in MHC-IIA in additional species. Given that 44 of the 50 MHC-IIA data sets in our meta-analysis came from mammals and bony fish, future studies should give increased attention to other taxa, for example, birds, other reptiles, amphibians, and cartilaginous fish.

Second, we need to better ascertain the number, location, and orthology of MHC-IIA genes as part of studying MHC genomics in a wide range of taxa. The known variation in MHC genomics – even within major vertebrate groups, such as between songbirds and chicken *Gallus domesticus* (O'Connor et al., 2019) or between Atlantic salmon *Salmo salar* and Atlantic cod *Gadus morhua* (Grimholt, 2016) – emphasizes the shortcomings in what is known about most other species. As can be true of MHC-IIB, work on the evolution of MHC-IIA alleles is hampered by this lack of fundamental genomic knowledge. For example, if a study cannot assign MHC-IIA alleles to

specific MHC-IIA genes, then it becomes impossible to answer questions about allele frequencies and mechanisms of balancing selection (Phillips et al., 2018) or about lineages of alleles over deep time (Goebel et al., 2017). Although such work has progressed much further in mammals, where characterization and annotation of MHC-IIA lineages is far advanced (e.g., Amills et al., 2008), we found only four mammalian species with sufficient data for measuring MHC-IIA polymorphism in all three genes (DPA, DQA, DRA), and only two of those species had complete data for assessing positive selection. Our analysis therefore could not break down MHC-IIA variability to gene/lineage level, not even in mammals. More broadly, additional comparative data on MHC genomics are required to understand not only orthology but also whether MHC-IIA resembles MHC-IIB in phenomena such as gene expansion (He et al., 2021) and copy number variation (Minias et al., 2019; Talarico et al., 2022).

To tackle these genomic questions, MHC research should take advantage of improvements in long-read sequencing technologies, as several recent studies have done (Cheng et al., 2021; Fuselli et al., 2018; He et al., 2021). PacBio and NanoPore sequencing now can generate reads from 10s to 100s of kb (Amarasinghe et al., 2020; Hon et al., 2020), circumventing the problems inherent in making MHC assemblies from short sequencing reads (He et al., 2021). A wider understanding of MHC gene identity, lineage history, and genomic organization would greatly enhance our ability to explore patterns of polymorphism and selection in MHC-IIA, and should lead to a clearer picture of important phenomena such as linkage, gene duplication, and copy number variation.

Third, an important area for future studies is to expand our understanding of class II heterodimer formation. Molecular characterization of MHC-IIA genes is an important step, but ultimately a functional perspective on class II MHC will require an improved understanding of the joint action of the alpha and beta subunits of the protein. This will be most easily achievable in species with only a single classical MHC-IIA and MHC-IIB gene (e.g., Atlantic salmon) (Grimholt, 2016), but will be more complicated in other taxa. In species with a single MHC-IIA and multiple MHC-IIB genes (e.g., chicken), the alpha subunit may be evolutionarily constrained by the need to work effectively with different beta subunits – that is, a best average fit (Salomonsen et al., 2003). In other species, there are multiple MHC-IIA and multiple MHC-IIB genes (e.g., Leach's storm-petrel, *Hydrobates leucorhous*), raising the question of whether specific A-B gene combinations always pair in a heterodimer and thus might coevolve as a team (Rand et al., 2019). Outside of a few primate and rodent models (Askew & Harding, 2002; Braunstein & Germain, 1987; Temme et al., 2019), we currently lack data to investigate these questions.

## 4.3 | Outlook

Our meta-analysis shows that – despite their historical underrepresentation in immunogenetic studies – MHC-IIA genes are nearly as polymorphic as MHC-IIB and show clear signals of positive

selection. Undoubtedly, there are some species where MHC-IIA diversity is quite limited and where a strong focus on MHC-IIB is probably justified, but researchers can no longer simply assume that this is the case in less well-studied species. We therefore suggest that it is conceptually important to incorporate MHC-IIA data into studies of pathogen-mediated selection, mate choice, and other aspects of ecology and evolution. The class II peptide binding groove is encoded by two types of genes, and it is time to consider them both to gain an integrative understanding of their functional importance.

#### AUTHOR CONTRIBUTIONS

Donald C. Dearborn designed research, performed research, analysed data, and wrote the manuscript. Sophie Warren performed research and contributed to writing the manuscript. Frank Hailer designed research and wrote the manuscript.

#### ACKNOWLEDGEMENTS

Funding was provided to DCD by Institutional Development Awards (P20GM0103423) from the National Institute of General Medical Sciences of the National Institutes of Health. We acknowledge support from the National Evolutionary Synthesis Center (NESCent) to DCD and FH, NSF no. EF-0905606, and from Cardiff University School of Biosciences to FH. We are grateful for assistance from Pete Schlax and Bridget Tweedie, and to Rob J. Thomas for discussions and advice. We thank the editor and anonymous reviewers for their helpful comments and suggestions.

#### CONFLICT OF INTEREST

The authors declare no conflict of interest.

#### DATA AVAILABILITY STATEMENT

The DNA sequence data were retrieved from published studies as listed in Appendix 1. Our single-species alignments, measures of polymorphism and positive selection, associated covariates, and R scripts for the GLMs and GLMMs are available on DataDryad [data set] (Dearborn et al., 2022) at <https://doi.org/10.5061/dryad.fbg79cnx0>.

#### BENEFIT-SHARING STATEMENT

Benefits generated: Benefits from this research accrue from the sharing of our data and R scripts on public databases as described above.

#### ORCID

Donald C. Dearborn  <https://orcid.org/0000-0003-2464-9368>

Frank Hailer  <https://orcid.org/0000-0002-2340-1726>

#### REFERENCES

- Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., & Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome Biology*, 21(1), Article 30. <https://doi.org/10.1186/s13059-020-1935-5>
- Amills, M., Ramírez, O., Tomàs, A., Obexer-Ruff, G., & Vidal, O. (2008). Positive selection on mammalian MHC-DQ genes revisited from a multispecies perspective. *Genes and Immunity*, 9(8), 651–658. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-57249094355&doi=10.1038%2fgene.2008.62&partnerID=40&md5=b82ebafa1b24c01b096d937466126508>
- Arbanasić, H., Galov, A., Ambriović-Ristov, A., Grizelj, J., Arsenos, G., Marković, B., Dovenski, T., Vince, S., & Čurik, I. (2013). Extensive polymorphism of the major histocompatibility complex DRA gene in Balkan donkeys: Perspectives on selection and genealogy. *Animal Genetics*, 44(6), 711–716. <https://doi.org/10.1111/age.12054>
- Askew, D., & Harding, C. V. (2002). Differences in antigen processing with haplotype-mismatched MHC class II heterodimers: AαdAβ heterodimers participate in early endosomal processing. *European Journal of Immunology*, 32(10), 2726–2736. [https://doi.org/10.1002/1521-4141\(200210\)32:10<2726::AID-IMMU2726>3.0.CO;2-8](https://doi.org/10.1002/1521-4141(200210)32:10<2726::AID-IMMU2726>3.0.CO;2-8)
- Averdam, A., Kuschal, C., Otto, N., Westphal, N., Roos, C., Reinhardt, R., & Walter, L. (2011). Sequence analysis of the grey mouse lemur (*Microcebus murinus*) MHC class II DQ and DR region. *Immunogenetics*, 63(2), 85–93. <https://doi.org/10.1007/s00251-010-0487-3>
- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84943645306&doi=10.18637%2fjss.v067.i01&partnerID=40&md5=4d9bb7b70f27f3c5078ee1a34ff386d>
- Bernatchez, L., & Landry, C. (2003). MHC studies in nonmodel vertebrates: What have we learned about natural selection in 15 years? *Journal of Evolutionary Biology*, 16(3), 363–377. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-0037884727&partnerID=40&md5=8867ccd1178aa917a84f06699a7f2976>
- Blancher, A., Aarnink, A., Tanaka, K., Ota, M., Inoko, H., Yamanaka, H., Nakagawa, H., Apoil, P. A., & Shiina, T. (2012). Study of cynomolgus monkey (*Macaca fascicularis*) Mhc DRB gene polymorphism in four populations. *Immunogenetics*, 64(8), 605–614. <https://doi.org/10.1007/s00251-012-0613-5>
- Blancher, A., Aarnink, A., Yamada, Y., Tanaka, K., Yamanaka, H., & Shiina, T. (2014). Study of MHC class II region polymorphism in the Filipino cynomolgus macaque population. *Immunogenetics*, 66(4), 219–230. <https://doi.org/10.1007/s00251-014-0764-7>
- Bontrop, R. E., Otting, N., De Groot, N. G., & Doxiadis, G. G. M. (1999). Major histocompatibility complex class II polymorphisms in primates. *Immunological Reviews*, 167, 339–350. <https://doi.org/10.1111/j.1600-065X.1999.tb01403.x>
- Bowen, L., Aldridge, B. M., Gulland, F., Woo, J., Van Bonn, W., Delong, R., Stott, J. L., & Johnson, M. L. (2002). Molecular characterization of expressed DQA and DQB genes in the California sea lion (*Zalophus californianus*). *Immunogenetics*, 54(5), 332–347. <https://doi.org/10.1007/s00251-002-0472-6>
- Bowen, L., Aldridge, B. M., Miles, A. K., & Stott, J. L. (2006). Expressed MHC class II genes in sea otters (*Enhydra lutris*) from geographically disparate populations. *Tissue Antigens*, 67(5), 402–408. <https://doi.org/10.1111/j.1399-0039.2006.00559.x>
- Bracamonte, S. E., Baltazar-Soares, M., & Eizaguirre, C. (2015). Characterization of MHC class II genes in the critically endangered European eel (*Anguilla Anguilla*). *Conservation Genetics Resources*, 7(4), 859–870. <https://doi.org/10.1007/s12686-015-0501-z>
- Braunstein, N. S., & Germain, R. N. (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(9), 2921–2925. <https://doi.org/10.1073/pnas.84.9.2921>
- Brown, J. H., Jardetzky, T. S., Gorga, J. C., Stern, L. J., Urban, R. G., Strominger, J. L., & Wiley, D. C. (1993). Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature*, 364(6432), 33–39. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-0027217789&partnerID=40&md5=c53cb97487f38d67ac6aef0a07326db8>



- Brown, J. J., Thomson, W., Clegg, P., Eyre, S., Kennedy, L. J., Matthews, J., Carter, S., & Ollier, W. E. R. (2004). Polymorphisms of the equine major histocompatibility complex II DRA locus. *Tissue Antigens*, 64(2), 173–179. <https://doi.org/10.1111/j.1399-0039.2004.00269.x>
- Bryja, J., Galan, M., Charbonnel, N., & Cosson, J. F. (2006). Duplication, balancing selection and trans-species evolution explain the high levels of polymorphism of the DQA MHC class II gene in voles (Arvicolinae). *Immunogenetics*, 58(2–3), 191–202. <https://doi.org/10.1007/s00251-006-0085-6>
- Burri, R., Promerová, M., Goebel, J., & Fumagalli, L. (2014). PCR-based isolation of multigene families: Lessons from the avian MHC class IIB. *Molecular Ecology Resources*, 14(4), 778–788. <http://www.scopus.com/inward/record.url?eid=2-s2.0-84902304355&partnerID=40&md5=4dbfd6ce4bb1ad938678c38046baf44f>
- Canal, D., Alcaide, M., Anmarkrud, J. A., & Potti, J. (2010). Towards the simplification of MHC typing protocols: Targeting classical MHC class II genes in a passerine, the pied flycatcher *Ficedula hypoleuca*. *BMC Research Notes*, 3, Article 236. <http://www.scopus.com/inward/record.url?eid=2-s2.0-77957033334&partnerID=40&md5=0a67f9b89facfee788ca2aee6a989d30>
- Cao, Z., Wang, L., Xiang, Y., Liu, X., Tu, Z., Sun, Y., & Zhou, Y. (2018). MHC class IIa polymorphism and its association with resistance/susceptibility to *Vibrio harveyi* in golden pompano (*Trachinotus ovatus*). *Fish and Shellfish Immunology*, 80, 302–310. <https://doi.org/10.1016/j.fsi.2018.06.020>
- Chen, Y., Liu, Y., Song, M., Lai, J., Sun, J., & Gong, Q. (2020). Molecular polymorphism and expression of MHC I  $\alpha$ , II  $\alpha$ , II  $\beta$  and II invariant chain in the critically endangered Dabry's sturgeon (*Acipenser dabryanus*). *Developmental and Comparative Immunology*, 103, Article 103494. <https://doi.org/10.1016/j.dci.2019.103494>
- Chen, Y. Y., Zhang, Y. Y., Zhang, H. M., Ge, Y. F., Wan, Q. H., & Fang, S. G. (2010). Natural selection coupled with intragenic recombination shapes diversity patterns in the major histocompatibility complex class II genes of the giant panda. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 314 B(3), 208–223. <https://doi.org/10.1002/jez.b.21327>
- Cheng, Y., Grueber, C., Hogg, C. J., & Belov, K. (2021). Improved high-throughput MHC typing for non-model species using long-read sequencing. *Molecular Ecology Resources*, 22, 862–876. <https://doi.org/10.1111/1755-0998.13511>
- Cuesta, A., Ángeles Esteban, M., & Meseguer, J. (2006). Cloning, distribution and up-regulation of the teleost fish MHC class II alpha suggests a role for granulocytes as antigen-presenting cells. *Molecular Immunology*, 43(8), 1275–1285. <https://doi.org/10.1016/j.molimm.2005.07.004>
- Dearborn, D. C., Warren, S., & Hailer, F. (2022). *Polymorphism and selection in major histocompatibility complex (MHC) class IIA*. Dryad, Dataset. <https://doi.org/10.5061/dryad.fbg79cnx0>
- Deng, Q., Zhang, H., Xiang, R., Zhang, Z., Ling, F., Zhuo, M., Du, H., & Wang, X. (2013). Identification of Mamu-DPA1, Mamu-DQA1, and Mamu-DRA alleles in a cohort of Chinese rhesus macaques. *Immunogenetics*, 65(12), 901–904. <https://doi.org/10.1007/s00251-013-0736-3>
- Dudek, K., Gaczorek, T. S., Zieliński, P., & Babik, W. (2019). Massive introgression of major histocompatibility complex (MHC) genes in newt hybrid zones. *Molecular Ecology*, 28(21), 4798–4810. <https://doi.org/10.1111/mec.15254>
- Fuselli, S., Baptista, R. P., Panziera, A., Magi, A., Guglielmi, S., Tonin, R., Benazzo, A., Bauzer, L. G., Mazzoni, C. J., & Bertorelle, G. (2018). A new hybrid approach for MHC genotyping: high-throughput NGS and long read MinION nanopore sequencing, with application to the non-model vertebrate Alpine chamois (*Rupicapra rupicapra*). *Heredity*, 121(4), 293–303. <https://doi.org/10.1038/s41437-018-0070-5>
- Gao, F., Chen, C., Arab, D. A., Du, Z., He, Y., & Ho, S. Y. W. (2019). EasyCodeML: A visual tool for analysis of selection using CodeML. *Ecology and Evolution*, 9(7), 3891–3898. <https://doi.org/10.1002/ece3.5015>
- Ge, F., Memon, S., Xi, D., Li, S., Liu, X., Li, G., Wang, L., Leng, J., Khan, S., & Deng, W. (2016). Cloning and characterization of MHC-DQA1 and MHC-DQA2 molecules from yak (*Bos grunniens*). *Archives Animal Breeding*, 59(3), 395–400. <https://doi.org/10.5194/aab-59-395-2016>
- Godwin, U. B., Antao, A., Wilson, M. R., Chinchar, V. G., Miller, N. W., Clem, L. W., & McConnell, T. J. (1997). MHC class II B genes in the channel catfish (*Ictalurus punctatus*). *Developmental and Comparative Immunology*, 21(1), 13–23. [https://doi.org/10.1016/S0145-305X\(97\)00003-7](https://doi.org/10.1016/S0145-305X(97)00003-7)
- Godwin, U. B., Flores, M., Quiniou, S., Wilson, M. R., Miller, N. W., Clem, L. W., & McConnell, T. J. (2000). MHC class II A genes in the channel catfish (*Ictalurus punctatus*). *Developmental and Comparative Immunology*, 24(6–7), 609–622. [https://doi.org/10.1016/S0145-305X\(00\)00005-7](https://doi.org/10.1016/S0145-305X(00)00005-7)
- Goebel, J., Promerová, M., Bonadonna, F., McCoy, K. D., Serbielle, C., Strandh, M., Yannic, G., Burri, R., & Fumagalli, L. (2017). 100 million years of multigene family evolution: Origin and evolution of the avian MHC class IIB. *BMC Genomics*, 18(1), Article 460. <https://doi.org/10.1186/s12864-017-3839-7>
- Gómez, D., Conejeros, P., Marshall, S. H., & Consuegra, S. (2010). MHC evolution in three salmonid species: A comparison between class II alpha and beta genes. *Immunogenetics*, 62(8), 531–542. <https://doi.org/10.1007/s00251-010-0456-x>
- Gotelli, N. J., & Colwell, R. K. (2001). Quantifying biodiversity: Procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, 4(4), 379–391. <https://doi.org/10.1046/j.1461-0248.2001.00230.x>
- Grimholt, U. (2016). MHC and evolution in teleosts. *Biology*, 5(1), Article 6. <https://doi.org/10.3390/biology5010006>
- Gyllensten, U. B., & Erlich, H. A. (1989). Ancient roots for polymorphism at the HLA-DQ $\alpha$  locus in primates. *Proceedings of the National Academy of Sciences of the United States of America*, 86(24), 9986–9990. <https://doi.org/10.1073/pnas.86.24.9986>
- He, K., Minias, P., & Dunn, P. O. (2021). Long-read genome assemblies reveal extraordinary variation in the number and structure of MHC loci in birds. *Genome Biology and Evolution*, 13(2), 1–13. <https://doi.org/10.1093/gbe/evaa270>
- Hess, C. M., & Edwards, S. V. (2002). The evolution of the major histocompatibility complex in birds. *Bioscience*, 52(5), 423–431. <http://www.scopus.com/inward/record.url?eid=2-s2.0-0036241218&partnerID=40&md5=a392848a518cd39b2912823e4086fc60>
- Höglund, J. (2009). *Evolutionary conservation genetics*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199214211.001.0001>
- Hon, T., Mars, K., Young, G., Tsai, Y. C., Karalius, J. W., Landolin, J. M., Maurer, N., Kudrna, D., Hardigan, M. A., Steiner, C. C., Knapp, S. J., Ware, D., Shapiro, B., Peluso, P., & Rank, D. R. (2020). Highly accurate long-read HiFi sequencing data for five complex genomes [Data paper]. *Scientific Data*, 7(1), Article 399. <https://doi.org/10.1038/s41597-020-00743-4>
- Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, 50(3), 346–363. <https://doi.org/10.1002/bimj.200810425>
- Janova, E., Matiasovic, J., Vahala, J., Vodicka, R., Van Dyk, E., & Horin, P. (2009). Polymorphism and selection in the major histocompatibility complex DRA and DQA genes in the family equidae. *Immunogenetics*, 61(7), 513–527. <https://doi.org/10.1007/s00251-009-0380-0>
- Jeffery, K. J. M., & Bangham, C. R. M. (2000). Do infectious diseases drive MHC diversity? *Microbes and Infection*, 2(11), 1335–1341. [https://doi.org/10.1016/S1286-4579\(00\)01287-9](https://doi.org/10.1016/S1286-4579(00)01287-9)
- Kamath, P. L., & Getz, W. M. (2011). Adaptive molecular evolution of the Major Histocompatibility Complex genes, DRA and DQA, in the genus *Equus*. *BMC Evolutionary Biology*, 11(1), Article 128. <https://doi.org/10.1186/1471-2148-11-128>

- Kamiya, T., O'Dwyer, K., Westerdahl, H., Senior, A., & Nakagawa, S. (2014). A quantitative review of MHC-based mating preference: The role of diversity and dissimilarity. *Molecular Ecology*, 23(21), 5151–5163. <https://doi.org/10.1111/mec.12934>
- Karl, J. A., Heimbruch, K. E., Vriezen, C. E., Mironczuk, C. J., Dudley, D. M., Wiseman, R. W., & O'Connor, D. H. (2014). Survey of major histocompatibility complex class II diversity in pig-tailed macaques. *Immunogenetics*, 66(11), 613–623. <https://doi.org/10.1007/s00251-014-0797-y>
- Kaufman, J. (1999). Co-evolving genes in MHC haplotypes: The 'rule' for nonmammalian vertebrates? *Immunogenetics*, 50(3–4), 228–236. <https://doi.org/10.1007/s002510050597>
- Kennedy, L. J., Angles, J. M., Barnes, A., Carmichael, L. E., Radford, A. D., Ollier, W. E. R., & Happ, G. M. (2007). DLA-DRB1, DQA1, and DQB1 alleles and haplotypes in North American gray wolves [Conference Paper]. *Journal of Heredity*, 98(5), 491–499. <https://doi.org/10.1093/jhered/esm051>
- Koutsogiannouli, E. A., Moutou, K. A., Sarafidou, T., Stamatis, C., Spyrou, V., & Mamuris, Z. (2009). Major histocompatibility complex variation at class II DQA locus in the brown hare (*Lepus europaeus*). *Molecular Ecology*, 18(22), 4631–4649. <https://doi.org/10.1111/j.1365-294X.2009.04394.x>
- Krkošek, M., Lewis, M. A., Morton, A., Frazer, L. N., & Volpe, J. P. (2006). Epizootics of wild fish induced by farm fish. *Proceedings of the National Academy of Sciences of the United States of America*, 103(42), 15506–15510. <https://doi.org/10.1073/pnas.0603525103>
- Kube, P. D., Taylor, R. S., & Elliott, N. G. (2012). Genetic variation in parasite resistance of Atlantic salmon to amoebic gill disease over multiple infections. *Aquaculture*, 364–365, 165–172. <https://doi.org/10.1016/j.aquaculture.2012.08.026>
- Kuhn, M., Weston, S., Wing, J., Forester, J., & Thaler, T. (2021). *contrast: A collection of contrast methods*. R package version 0.24.1. <https://CRAN.R-project.org/package=contrast>
- Kumánovics, A., Takada, T., & Fischer Lindahl, K. (2003). Genomic organization of the mammalian MHC. *Annual Review of Immunology*, 21, 629–657.
- Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, 35(6), 1547–1549. <https://doi.org/10.1093/molbev/msy096>
- Lenth, R. V. (2022). *emmeans: Estimated marginal means, aka least-squares means*. R package version 1.7.2 <https://CRAN.R-project.org/package=emmeans>
- Li, W., Sun, W. X., Hu, J. F., Hong, D. W., & Chen, S. L. (2014). Molecular characterization, polymorphism and expression analysis of swamp eel major histocompatibility complex class II B gene, after infection by *Aeromonas Hydrophilia*. *Journal of Animal and Plant Sciences*, 24(2), 481–491 <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84899438780&partnerID=40&md5=0ca5596f407e68f2a8f0f38d1e3e0fc4>
- Li, W., Sun, W. X., Meng, L., & Hong, D. W. (2014). Molecular cloning, genomic structure, polymorphism and expression analysis of major histocompatibility complex class IIA gene of swamp eel *Monopterus albus*. *Biologia (Poland)*, 69(2), 236–246. <https://doi.org/10.2478/s11756-013-0307-y>
- Lighten, J., Papadopulos, A. S. T., Mohammed, R. S., Ward, B. J., Paterson, G. I., Baillie, L., Bradbury, I. R., Hendry, A. P., Bentzen, P., & Van Oosterhout, C. (2017). Evolutionary genetics of immunological supertypes reveals two faces of the Red Queen. *Nature Communications*, 8(1), Article 1294. <https://doi.org/10.1038/s41467-017-01183-2>
- Lillie, M., Cui, J., Shine, R., & Belov, K. (2016). Molecular characterization of MHC class II in the Australian invasive cane toad reveals multiple splice variants. *Immunogenetics*, 68(6–7), 449–460. <https://doi.org/10.1007/s00251-016-0919-9>
- Liu, J., Sun, Y., & Xu, T. (2016). Identification of 48 full-length MHC-DAB functional alleles in miyu croaker and evidence for positive selection. *Fish and Shellfish Immunology*, 54, 544–550. <https://doi.org/10.1016/j.fsi.2016.05.008>
- Liu, L. X., Zhao, S. G., Lu, H. N., Yang, Q. L., Huang, X. Y., & Gun, S. B. (2015). Association between polymorphisms of the swine MHC-DQA gene and diarrhoea in three Chinese native piglets. *International Journal of Immunogenetics*, 42(3), 208–216. <https://doi.org/10.1111/iji.12186>
- Maccari, G., Robinson, J., Ballingall, K., Guethlein, L. A., Grimholt, U., Kaufman, J., Ho, C. S., De Groot, N. G., Flicek, P., Bontrop, R. E., Hammond, J. A., & Marsh, S. G. E. (2017). IPD-MHC 2.0: An improved inter-species database for the study of the major histocompatibility complex. *Nucleic Acids Research*, 45(D1), D860–D864. <https://doi.org/10.1093/nar/gkw1050>
- Magalhães, V., Abrantes, J., Munõz-Pajares, A. J., & Esteves, P. J. (2015). Genetic diversity comparison of the DQA gene in European rabbit (*Oryctolagus cuniculus*) populations. *Immunogenetics*, 67(10), 579–590. <https://doi.org/10.1007/s00251-015-0866-x>
- Marsden, C. D., Mable, B. K., Woodroffe, R., Rasmussen, G. S. A., Cleaveland, S., McNutt, J. W., Emmanuel, M., Thomas, R., & Kennedy, L. J. (2009). Highly endangered african wild dogs (*Lycan pictus*) lack variation at the major histocompatibility complex. *Journal of Heredity*, 100, S54–S65. <https://doi.org/10.1093/jhered/esp031>
- Memon, S., Wang, L., Li, G., Liu, X., Deng, W., & Xi, D. (2018). Isolation and characterization of the major histocompatibility complex DQA1 and DQA2 genes in gayal (*Bos frontalis*). *Journal of Genetics*, 97(1), 121–126. <https://doi.org/10.1007/s12041-018-0882-3>
- Milinski, M. (2006). The major histocompatibility complex, sexual selection, and mate choice. *Annual Review of Ecology, Evolution, and Systematics*, 37, 159–186.
- Minias, P., Pikus, E., Whittingham, L. A., & Dunn, P. O. (2018). A global analysis of selection at the avian MHC. *Evolution*, 72(6), 1278–1293. <https://doi.org/10.1111/evo.13490>
- Minias, P., Pikus, E., Whittingham, L. A., & Dunn, P. O. (2019). Evolution of copy number at the MHC varies across the avian tree of life. *Genome Biology and Evolution*, 11(1), 17–28. <https://doi.org/10.1093/gbe/evy253>
- Murphy, K., & Weaver, C. (2017). *Janeway's immunobiology* (Ninth ed.). Garland Science.
- Nei, M., & Rooney, A. P. (2005). Concerted and birth-and-death evolution of multigene families. *Annual Review of Genetics*, 39, 121–152. <http://www.scopus.com/inward/record.url?eid=2-s2.0-27944477670&partnerID=40&md5=9ba6b71cf48fa82a88349ffcd b085c79>
- O'Connor, E. A., Westerdahl, H., Burri, R., & Edwards, S. V. (2019). Avian MHC evolution in the era of genomics: Phase 1.0. *Cell*, 8(10), 1–21. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85077538189&doi=10.3390%2fcells8101152&partnerID=40&md5=e7c535155f1eab6bb7b1db2b031bc64a>
- Otting, N., de Groot, N. G., & Bontrop, R. E. (2019). Limited MHC class II gene polymorphism in the West African chimpanzee is distributed maximally by haplotype diversity. *Immunogenetics*, 71(1), 13–23. <https://doi.org/10.1007/s00251-018-1080-4>
- Pang, J. C., Gao, F. Y., Lu, M. X., Ye, X., Zhu, H. P., & Ke, X. L. (2013). Major histocompatibility complex class IIA and IIB genes of Nile tilapia *Oreochromis niloticus*: Genomic structure, molecular polymorphism and expression patterns. *Fish and Shellfish Immunology*, 34(2), 486–496. <https://doi.org/10.1016/j.fsi.2012.11.048>
- Penn, D. J., Damjanovich, K., & Potts, W. K. (2002). MHC heterozygosity confers a selective advantage against multiple-strain infections. *Proceedings of the National Academy of Sciences of the United States of America*, 99(17), 11260–11264 <Go to ISI>://000177606900057.



- Phillips, K. P., Cable, J., Mohammed, R. S., Herdegen-Radwan, M., Raubic, J., Przesmycka, K. J., Van Oosterhout, C., & Radwan, J. (2018). Immunogenetic novelty confers a selective advantage in host–pathogen coevolution. *Proceedings of the National Academy of Sciences of the United States of America*, 115(7), 1552–1557. <https://doi.org/10.1073/pnas.1708597115>
- Piertney, S. B., & Oliver, M. K. (2006). The evolutionary ecology of the major histocompatibility complex. *Heredity*, 96(1), 7–21 <Go to ISI>://000234053400004.
- Plasil, M., Mohandesan, E., Fitak, R. R., Musilova, P., Kubickova, S., Burger, P. A., & Horin, P. (2016). The major histocompatibility complex in Old World camelids and low polymorphism of its class II genes. *BMC Genomics*, 17(1), Article 167. <https://doi.org/10.1186/s12864-016-2500-1>
- Pulkkinen, K., Suomalainen, L. R., Read, A. F., Ebert, D., Rintamäki, P., & Valtonen, E. T. (2010). Intensive fish farming and the evolution of pathogen virulence: The case of columnaris disease in Finland. *Proceedings of the Royal Society B: Biological Sciences*, 277(1681), 593–600. <https://doi.org/10.1098/rspb.2009.1659>
- R Core Team. (2021). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing.
- Radwan, J., Babik, W., Kaufman, J., Lenz, T. L., & Winternitz, J. (2020). Advances in the evolutionary understanding of MHC polymorphism. *Trends in Genetics*, 36(4), 298–311. <https://doi.org/10.1016/j.tig.2020.01.008>
- Rand, L. M., Woodward, C., May, R., Ackerman, R. A., Tweedie, B., Zicarelli, T. B., & Dearborn, D. C. (2019). Divergence between genes but limited allelic polymorphism in two MHC class II A genes in Leach's storm-petrels *Oceanodroma leucorhoa*. *Immunogenetics*, 71(8–9), 561–573. <https://doi.org/10.1007/s00251-019-01130-z>
- Ren, L., Yang, Z., Wang, T., Sun, Y., Guo, Y., Zhang, Z., Fei, J., Bao, Y., Qin, T., Wang, J., Huang, Y., Hu, X., Zhao, Y., & Li, N. (2011). Characterization of the MHC class II  $\alpha$ -chain gene in ducks. *Immunogenetics*, 63(10), 667–678. <https://doi.org/10.1007/s00251-011-0545-5>
- Salomonsen, J., Marston, D., Avila, D., Bumstead, N., Johansson, B., Juul-Madsen, H., Olesen, G. D., Riegert, P., Skjødtt, K., Vainio, O., Wiles, M. V., & Kaufman, J. (2003). The properties of the single chicken MHC classical class II  $\alpha$  chain (B-LA) gene indicate an ancient origin for the DR/E-like isotype of class II molecules. *Immunogenetics*, 55(9), 605–614. <https://doi.org/10.1007/s00251-003-0620-7>
- Scott, P. C., Gogolin-Ewens, K. J., Adams, T. E., & Brandon, M. R. (1991). Nucleotide sequence, polymorphism, and evolution of ovine MHC class II DQA genes. *Immunogenetics*, 34(2), 69–79. <https://doi.org/10.1007/BF00211419>
- Seddon, J. M., & Baverstock, P. R. (1999). Variation on islands: Major histocompatibility complex (Mhc) polymorphism in populations of the Australian bush rat. *Molecular Ecology*, 8(12), 2071–2079. <https://doi.org/10.1046/j.1365-294X.1999.00822.x>
- Sena, L., Schneider, M. P. C., Brenig, B., Honeycutt, R. L., Womack, J. E., & Skow, L. C. (2003). Polymorphisms in MHC-DRA and -DRB alleles of water buffalo (*Bubalus bubalis*) reveal different features from cattle DR alleles. *Animal Genetics*, 34(1), 1–10. <https://doi.org/10.1046/j.1365-2052.2003.00920.x>
- Sepil, I., Lachish, S., Hinks, A. E., & Sheldon, B. C. (2013). Mhc supertypes confer both qualitative and quantitative resistance to avian malaria infections in a wild bird population. *Proceedings of the Royal Society B: Biological Sciences*, 280(1759), 1–8. <http://www.scopus.com/inward/record.url?eid=2-s2.0-84875288964&partnerID=40&md5=d657aea428f915faa95a7a622a730108>
- Shen, T., He, X., Lei, M., Wang, J., Li, X., & Li, J. (2014). Cloning and structure of a histocompatibility class IIA gene (Lelo-DAA) in Chinese longsnout catfish (*Leiocassis longirostris*). *Genes and Genomics*, 36(6), 745–753. <https://doi.org/10.1007/s13258-014-0208-7>
- Shen, T., Xu, S., Yang, M., Pang, S., & Yang, G. (2011). Molecular cloning, expression pattern, and 3D structural analysis of the MHC class IIB gene in the Chinese longsnout catfish (*Leiocassis longirostris*). *Veterinary Immunology and Immunopathology*, 141(1–2), 33–45. <https://doi.org/10.1016/j.vetimm.2011.02.002>
- Sin, Y. W., Dugdale, H. L., Newman, C., MacDonald, D. W., & Burke, T. (2012). MHC class II genes in the European badger (*Meles meles*): Characterization, patterns of variation, and transcription analysis. *Immunogenetics*, 64(4), 313–327. <https://doi.org/10.1007/s00251-011-0578-9>
- Snibson, K. J., Maddox, J. F., Fabb, S. A., & Brandon, M. R. (1998). Allelic variation of ovine MHC class II DQA1 and DQA2 genes. *Animal Genetics*, 29(5), 356–362. <https://doi.org/10.1046/j.1365-2052.1998.295351.x>
- Spurgin, L. G., & Richardson, D. S. (2010). How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings of the Royal Society B: Biological Sciences*, 277(1684), 979–988 <http://www.scopus.com/inward/record.url?eid=2-s2.0-77950816077&partnerID=40&md5=1d96a68c8ade5976e6150e6e4bb19c64>
- Spurgin, L. G., Van Oosterhout, C., Illera, J. C., Bridgett, S., Gharbi, K., Emerson, B. C., & Richardson, D. S. (2011). Gene conversion rapidly generates major histocompatibility complex diversity in recently founded bird populations. *Molecular Ecology*, 20(24), 5213–5225. <https://doi.org/10.1111/j.1365-294X.2011.05367.x>
- Stefanović, M., Čirović, D., Bogdanović, N., Knauer, F., Heltai, M., Szabó, L., Lanszki, J., Zhelev, C. D., Schaschl, H., & Suchentrunk, F. (2021). Positive selection on the MHC class II DLA-DQA1 gene in golden jackals (*Canis aureus*) from their recent expansion range in Europe and its effect on their body mass index. *BMC Ecology and Evolution*, 21(1), Article 122. <https://doi.org/10.1186/s12862-021-01856-z>
- Stet, R. J. M., Kruiswijk, C. P., & Dixon, B. (2003). Major histocompatibility lineages and immune gene function in teleost fishes: The road not taken. *Critical Reviews in Immunology*, 23(5–6), 441–471 <http://www.scopus.com/inward/record.url?eid=2-s2.0-1242314225&partnerID=40&md5=73faef5db3897d62df8f06dbd260aab7>
- Subramaniam, R., White, S. N., Herrmann-Hoesing, L. M., & Srikumaran, S. (2012). MHC class II DR allelic diversity in bighorn sheep. *Gene*, 506(1), 217–222. <https://doi.org/10.1016/j.gene.2012.06.048>
- Talarico, L., Bryjová, A., Čížková, D., Douda, K., & Reichard, M. (2022). Individual copy number variation and extensive diversity between major MHC-DAB1 allelic lineages in the European bitterling. *Immunogenetics*, 74(5), 497–505. <https://doi.org/10.1007/s00251-021-01251-4>
- Taniguchi, Y., Matsumoto, K., Matsuda, H., Yamada, T., Sugiyama, T., Homma, K., Kaneko, Y., Yamagishi, S., & Iwaisaki, H. (2014). Structure and polymorphism of the major histocompatibility complex class II region in the Japanese crested ibis, *Nipponia nippon*. *Plos One*, 9(9), Article e108506. <https://doi.org/10.1371/journal.pone.0108506>
- Temme, S., Temme, N., & Koch, N. (2019). Assembly, intracellular transport, and release of MHC class II peptide receptors. *Methods in Molecular Biology*, 1988, 297–314. [https://doi.org/10.1007/978-1-4939-9450-2\\_22](https://doi.org/10.1007/978-1-4939-9450-2_22)
- Trowsdale, J., & Knight, J. C. (2013). Major histocompatibility complex genomics and human disease. *Annual Review of Genomics and Human Genetics*, 14, 301–323.
- Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (Fourth ed.). Springer.
- Wagner, J. L., Burnett, R. C., DeRose, S. A., & Storb, R. (1996). Molecular analysis and polymorphism of the DLA-DQA gene. *Tissue Antigens*, 48(3), 199–204. <https://doi.org/10.1111/j.1399-0039.1996.tb02629.x>
- Wegner, K. M., Kalbe, M., Kurtz, J., Reusch, T. B. H., & Milinski, M. (2003). Parasite selection for immunogenetic optimality. *Science*, 301(5638), 1343 <http://www.scopus.com/inward/record.url?eid=2-s2.0-0042190374&partnerID=40&md5=ba0ab1bbfd442a960a92648432b6a2c0>
- Winternitz, J. C., Minchey, S. G., Garamszegi, L. Z., Huang, S., Stephens, P. R., & Altizer, S. (2013). Sexual selection explains more functional variation in the mammalian major histocompatibility complex than

- parasitism. *Proceedings of the Royal Society B: Biological Sciences*, 280(1769), 20131605 <http://www.scopus.com/inward/record.url?eid=2-s2.0-84897358088&partnerID=40&md5=9d59ab09f83abb9975249717d440882e>
- Xu, T., Liu, J., Sun, Y., Zhu, Z., & Liu, T. (2016). Characterization of 40 full-length MHC class IIA functional alleles in miyu croaker: Polymorphism and positive selection. *Developmental and Comparative Immunology*, 55, 138–143. <https://doi.org/10.1016/j.dci.2015.10.019>
- Xu, T. J., Chen, S. L., Ji, X. S., & Sha, Z. X. (2009). Molecular cloning, genomic structure, polymorphism and expression analysis of major histocompatibility complex class IIA and IIB genes of half-smooth tongue sole (*Cynoglossus semilaevis*). *Fish and Shellfish Immunology*, 27(2), 192–201. <https://doi.org/10.1016/j.fsi.2009.04.009>
- Xu, T. J., Chen, S. L., & Zhang, Y. X. (2010). MHC class II $\alpha$  gene polymorphism and its association with resistance/susceptibility to *Vibrio anguillarum* in Japanese flounder (*Paralichthys olivaceus*). *Developmental and Comparative Immunology*, 34(10), 1042–1050. <https://doi.org/10.1016/j.dci.2010.05.008>
- Yang, M., Wei, J., Li, P., Wei, S., Huang, Y., & Qin, Q. (2016). MHC class II $\alpha$  polymorphisms and their association with resistance/susceptibility to Singapore grouper iridovirus (SGIV) in orange-spotted grouper, *Epinephelus coioides*. *Aquaculture*, 462, 10–16. <https://doi.org/10.1016/j.aquaculture.2016.04.026>
- Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, 24(8), 1586–1591. <https://doi.org/10.1093/molbev/msm088>
- Yang, Z., Wong, W. S. W., & Nielsen, R. (2005). Bayes empirical Bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution*, 22(4), 1107–1118. <https://doi.org/10.1093/molbev/msi097>
- Yao, G., Zhu, Y., Wan, Q. H., & Fang, S. G. (2015). Major histocompatibility complex class II genetic variation in forest musk deer (*Moschus berezovskii*) in China. *Animal Genetics*, 46(5), 535–543. <https://doi.org/10.1111/age.12336>
- Yu, S., Ao, J., & Chen, X. (2010). Molecular characterization and expression analysis of MHC class II  $\alpha$  and  $\beta$  genes in large yellow croaker (*Pseudosciaena crocea*). *Molecular Biology Reports*, 37(3), 1295–1307. <https://doi.org/10.1007/s11033-009-9504-8>
- Zhu, R., Chen, Z. Y., Wang, J., Yuan, J. D., Liao, X. Y., Gui, J. F., & Zhang, Q. Y. (2014). Extensive diversification of MHC in Chinese giant salamanders *Andrias davidianus* (Anda-MHC) reveals novel splice variants. *Developmental and Comparative Immunology*, 42(2), 311–322. <https://doi.org/10.1016/j.dci.2013.10.001>

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Dearborn, D. C., Warren, S., & Hailer, F. (2022). Meta-analysis of major histocompatibility complex (MHC) class IIA reveals polymorphism and positive selection in many vertebrate species. *Molecular Ecology*, 31, 6390–6406. <https://doi.org/10.1111/mec.16726>

## APPENDIX 1

Key variables for 50 species with MHC-IIA data that met our inclusion criteria. See Data Availability statement for .csv file of complete MHC-IIA and MHC-IIB data set

Class and order	Family	Common Name	Latin Name	MHC-IIA No. of Animals <sup>a</sup>	MHC-IIA No. of Genes <sup>b</sup>	MHC-IIA No. of Alleles <sup>c</sup>	IIA MHC-IIA MB Log Likelihood <sup>d</sup>	MHC-IIA Source	MHC-IIB Source, if any
Mammalia: Primates	Cercopithecoidea	Rhesus Macaque	<i>Macaca mulatta</i>	30	3	30	23.2	Deng et al. (2013)	N/A
		Cynomolgus Macaque	<i>Macaca fascicularis</i>	117	3	26	18.5	Blancher et al. (2014)	Blancher et al. (2012, 2014)
Mammalia: Primates	Hominoidea	Pigtailed Macaque	<i>Macaca nemestrina</i>	32	3	31	23.3	Karl et al. (2014)	Karl et al. (2014)
		Chimpanzee	<i>Pan troglodytes</i>	67	3	13	15.6	Otting et al. (2019)	N/A
Mammalia: Primates	Cheirogaleoidea	Grey Mouse Lemur	<i>Microcebus murinus</i>	10	2	16	41.5	Averdam et al. (2011)	N/A
		European Rabbit	<i>Oryctolagus cuniculus</i>	40	1	30	91.4	Magalhães et al. (2015)	N/A
Mammalia: Rodentia	Leporidae	European Brown Hare	<i>Lepus europaeus</i>	674	1	37	60.7	Koutsogiannouli et al. (2009)	N/A
		Common Vole	<i>Microtus arvalis</i>	16	2	8	30.9	Bryja et al. (2006)	N/A
Mammalia: Rodentia	Cricetidae	Bank Vole	<i>Clethrionomys glareolus</i>	16	2	7	25	Bryja et al. (2006)	N/A
		Water Vole	<i>Arvicola terrestris</i>	96	2	7	9.5	Bryja et al. (2006)	N/A
Mammalia: Rodentia	Muridae	Australian Bush Rat	<i>Rattus fuscipes greyii</i>	305	1	36	73.1	Seddon and Baverstock (1999)	N/A
		Bighorn Sheep	<i>Ovis canadensis</i>	24	1	3	0	Subramaniam et al. (2012)	Subramaniam et al. (2012)
Mammalia: Artiodactyla	Bovidae	Domestic Sheep	<i>Ovis aries</i>	126	2	16	13.2	Sribson et al. (1998)	N/A
		Gayal	<i>Bos frontalis</i>	3	2	2		Memon et al. (2018)	N/A
Mammalia: Artiodactyla	Bovidae	Water Buffalo	<i>Bubalus bubalis</i>	75	1	2		Sena et al. (2003)	Sena et al. (2003)
		Yak	<i>Bos grunniens</i>	6	2	2		Ge et al. (2016)	N/A
Mammalia: Artiodactyla	Moschidae	Forest Musk Deer	<i>Moschus berezovskii</i>	71	3	12	1.3	Yao et al. (2015)	Yao et al. (2015)
		Domestic Pig	<i>Sus scrofa</i>	471	1	8	7.3	Liu et al. (2015)	N/A
Mammalia: Artiodactyla	Suidae	Wild Camel	<i>Camelus ferus</i>	13	2	5	0.4	Plasil et al. (2016)	Plasil et al. (2016)
		Bactrian Camel	<i>Camelus bactrianus</i>	38	2	6	0.3	Plasil et al. (2016)	Plasil et al. (2016)
Mammalia: Artiodactyla	Camelidae	Dromedary	<i>Camelus dromedarius</i>	28	2	4	1.5	Plasil et al. (2016)	Plasil et al. (2016)
		Plains Zebra	<i>Equus quagga</i>	71	1	9	0	Kamath and Getz (2011)	N/A
Mammalia: Perissodactyla	Equidae	Balkan Donkey	<i>Equus asinus africanus</i>	248	1	7	0.1	Arbanasic et al. (2013)	N/A
		Domestic Horse	<i>Equus caballus</i>	539	1	4	0	Brown et al. (2004)	N/A
Mammalia: Carnivora	Mustelidae	European Badger	<i>Meles meles</i>	7	2	4	6.1	Sin et al. (2012)	Sin et al. (2012)
		Sea Otter	<i>Enhydra lutris</i>	3	1	1		Bowen et al. (2006)	Bowen et al. (2006)
Mammalia: Carnivora	Otariidae	California Sea Lion	<i>Zalophus californianus</i>	19	2	2		Bowen et al. (2002)	Bowen et al. (2002)
		Giant Panda	<i>Ailuropoda melanoleuca</i>	121	3	12	32.8	Chen et al. (2010)	Chen et al. (2010)
Mammalia: Carnivora	Canidae	African Wild Dog	<i>Lycan pictus</i>	234	2	2		Marsden et al. (2009)	N/A
		Golden Jackal	<i>Canis aureus</i>	164	1	3	19.8	Stefanović et al. (2021)	N/A
Mammalia: Carnivora	Canidae	Grey Wolf	<i>Canis lupus</i>	175	1	11	22.8	Kennedy et al. (2007)	Kennedy et al. (2007)

## APPENDIX 1 (Continued)

Class and order	Family	Common Name	Latin Name	MHC-IIA No. of Animals <sup>a</sup>	MHC-IIA No. of Genes <sup>b</sup>	MHC-IIA No. of Alleles <sup>c</sup>	IIA MHC-IIA MB Log Likelihood <sup>d</sup>	MHC-IIA Source	MHC-IIIB Source, if any
Aves: Galliformes	Phasianidae	Chicken	<i>Gallus gallus</i>	21	1	1		Salomonsen et al. (2003)	N/A
Aves: Anseriformes	Anatidae	Duck	<i>Anas platyrhynchos</i>	9	1	4	0	Ren et al. (2011)	N/A
Aves: Procellariiformes	Hydrobatidae	Leach's Storm-petrel	<i>Hydrobates leucorhous</i>	30	2	4	12.7	Rand et al. (2019)	Rand et al. (2019)
Aves: Ciconiiformes	Threskiornithidae	Japanese Crested Ibis	<i>Nipponia nippon</i>	5	3	2		Taniguchi et al. (2014)	Taniguchi et al. (2014)
Amphibia: Urodela	Cryptobranchidae	Chinese Giant Salamander	<i>Andrias davidianus</i>	8	2	15	3.7	Zhu et al. (2014)	Zhu et al. (2014)
Amphibia: Anura	Bufoiidae	Australian Cane Toad	<i>Rhinella marina</i>	5	4	7	3.8	Lillie et al. (2016)	Lillie et al. (2016)
Actinopterygii: Acipenseriformes	Acipenseridae	Dabry's Sturgeon	<i>Acipenser dabryanus</i>	7	4	17	77.5	Chen et al. (2020)	Chen et al. (2020)
Actinopterygii: Anguilliformes	Anguillidae	European Eel	<i>Anguilla anguilla</i>	4	2	14	84.7	Bracamonte et al. (2015)	Bracamonte et al. (2015)
Actinopterygii: Siluriformes	Bagridae	Chinese Longsnout Catfish	<i>Leiocassis longirostris</i>	4	1	3	35.2	Shen et al. (2014)	Shen et al. (2011)
Actinopterygii: Salmoniformes	Ictaluridae	Channel Catfish	<i>Ictalurus punctatus</i>	2	2	5	17.2	Godwin et al. (2000)	Godwin et al. (1997)
Actinopterygii: Cichliformes	Salmonidae	Atlantic Salmon	<i>Salmo salar</i>	27	1	9	55.9	Gómez et al. (2010)	Gómez et al. (2010)
Actinopterygii: Perciformes	Cichlidae	Nile Tilapia	<i>Oreochromis niloticus</i>	6	2	11	54.2	Pang et al. (2013)	Pang et al. (2013)
Actinopterygii: Pleuronectiformes	Sciaenidae	Large Yellow Croaker	<i>Pseudosciaena crocea</i>	2	1	3	4.3	Yu et al. (2010)	Yu et al. (2010)
		Miluy Croaker	<i>Milichthys miluy</i>	26	4	27	190.9	Xu et al. (2016)	Liu et al. (2016)
	Sparidae	Gilthead Seabream	<i>Sparus aurata</i>	2	3	9	87.1	Cuesta et al. (2006)	N/A
	Carangidae	Golden Pompano	<i>Trachinotus ovatus</i>	80	2	29	231.5	Cao et al. (2018)	N/A
	Cynoglossidae	Half-smooth Tongue Sole	<i>Cynoglossus semilaevis</i>	4	2	8	30.3	Xu et al. (2009)	Xu et al. (2009)
Actinopterygii: Synbranchiformes	Paralichthyidae	Japanese Flounder	<i>Paralichthys olivaceus</i>	212	2	55	256.9	Xu et al. (2010)	N/A
	Synbranchidae	Swamp Eel	<i>Monopterus albus</i>	5	3	5	8.7	Li, Sun, Meng, and Hong (2014)	Li, Sun, Hu, et al. (2014)

<sup>a</sup>Sample size.<sup>b</sup>Number of MHC-IIA genes assayed in the study.<sup>c</sup>Number of unique MHC-IIA exon 2 nucleotide sequences.<sup>d</sup>Evidence for positive selection on a species' MHC-IIA alleles.