RESEARCH | Article in Press

Expression diversity of cichlid MHC alleles

Seraina E. Bracamonte · Pascal I. Hablützel · Carlos Lozano-Martín · Marta Barluenga

Received: 22 March 2026 / Accepted: 15 September 2026
© The Author(s) 2026

Abstract

Background Diversification driven by natural selection is a major source of biodiversity. Parasites and pathogens impose important selective pressures on their hosts, contributing to high levels of polymorphism at immunologically relevant genes within and among populations. The antigen-recognizing class IIB genes of the major histocompatibility complex (MHC) are major targets of parasite-mediated selection and frequently show genetic differentiation and local adaptation among populations. Accumulating evidence suggests that, besides allele sequence, variation in MHC expression may contribute to functional differences among alleles and loci. Here, we studied allele-specific MHC IIB expression in wild cichlid populations and evaluated patterns of expression variation in relation to allele frequency, functional supertypes, and population structure.

Results Building on previous knowledge of MHC class IIB diversity associated with ecological divergence in cichlid fish, we examined MHC expression patterns, their relationship with allele frequency, and their variability among populations and species. In Midas cichlids, a few highly abundant alleles showed consistently low expression, while many less common alleles showed high expression and were restricted to particular host populations. MHC alleles belonging to the same functional supertypes showed similar expression patterns. Lowly expressed supertypes showed relatively consistent expression across populations, while highly expressed supertypes exhibited substantially greater among-population variation. In the African cichlid dataset, putative MHC class IIB loci also showed marked variation in expression. Interestingly, lowly expressed Midas cichlid alleles clustered phylogenetically with lowly expressed putative African loci, while most highly expressed variants formed separate clusters.

Conclusions Our findings reveal a strong relationship between MHC allele frequency and expression, with the most common alleles tending to show low expression and many less common alleles showing high expression. The phylogenetic association between lowly expressed Midas cichlid alleles and African putative loci suggests that low expression might be associated with distinct MHC lineages, although its functional significance remains unresolved. MHC expression also varied substantially among populations, particularly for highly expressed supertypes. Together, these results show that allele- and locus-specific expression provides an additional dimension of MHC diversity beyond sequence variation alone. Whether these expression patterns influence parasite recognition or are shaped by parasite-mediated selection remains to be tested using direct parasite and infection data.

Keywords Major histocompatibility complex · Gene expression · Cichlids · Host-parasite interactions

This Article in Press is shared early to give you faster access to new research. It is citable and carries a permanent DOI. The final edited version will replace it automatically.

Introduction

Host-parasite interactions are a major force driving evolution at different levels of biological organisation and in particular can lead to an excess of genetic diversity in some genomic regions [1–4]. Infection by parasites triggers the evolution of defence responses in the host that try to mitigate the costs of infection, which in turn induces parasites to adapt to the new conditions, culminating in an evolutionary arms race [5, 6]. These highly dynamic interactions generate and maintain genetic diversity among and within populations. A few genomic regions coding for genes of the host's immune system show evidence of parasite-mediated balancing selection and consequently maintain remarkably high levels of polymorphism [7, 8]. This is particularly true for the genes of the major histocompatibility complex (MHC). The MHC is at the core of the immune defence of vertebrates against parasites. MHC molecules present foreign antigens to T cells which initiates an adaptive immune response that is specific to the infecting agent [9, 10]. There are two classes of MHC molecules that are encoded by highly polymorphic gene families. MHC class I is expressed on most body cells and binds peptides from intracellular parasites and pathogens (e.g. viruses). MHC class II molecules are expressed mainly on specialized antigen-presenting cells, but their expression can be induced in other cell types as well. They present peptides loaded in endosomal compartments, typically derived from extracellular antigens or infectious agents located in intracellular vesicles, including metazoan parasites, bacteria, and extracellular viral particles [9, 10]. For simplicity we will further refer to these collectively as parasites.

MHC class I and class II are highly dynamic multigene families, whose evolution is characterized by birth-and-death processes. New genes arise through gene duplication, and some of these duplicated genes are retained in the genome, while others become pseudogenes, acquire new functions, or are eliminated [11, 12]. Positive selection acting on sites that interact directly with parasite-derived antigens promotes the accumulation of mutations across paralogous gene copies. This process results in extensive polymorphism and copy number variation among and within species and populations [13–15]. As a result, individuals, populations, and species each possess a distinctive repertoire of MHC class I and class II variants, enabling their immune systems to recognize antigens derived from a diverse array of parasites. These extreme levels of polymorphism are maintained by parasite-mediated balancing selection, with MHC allele pools adapted to the local parasite communities [8, 16, 17]. Differentiated local allele pools were inferred for a wide range of vertebrates [18–23] which tracks divergence of infecting parasite communities [24–27]. Additionally, specific alleles are associated with resistance or susceptibility to specific parasites [26, 28, 29]. Associations with parasite infections were also described for groups of functionally similar alleles, i.e. supertypes, rather than single alleles [30, 31].

The immune response to parasites may be modulated by the regulation of expression levels of particular MHC alleles [32]. Different studies have suggested that not all MHC alleles are expressed [33–36], and that expression can be tissue-specific [37–39]. However, less focus has been placed on the importance of regulating MHC gene expression for the response to parasites. Studies that focussed on the expression of specific MHC class IIB alleles or on overall expression found that alleles and genotypes differ in expression [38, 40], expression of alleles or genotypes increases in the presence of specific parasites [41], and increased expression may compensate for suboptimal individual MHC diversity [42]. Despite that, it remains unresolved whether MHC expression levels correlate with parasite infections. Identifying such associations requires higher, allele-specific resolution of MHC expression. Studies that thoroughly assess allele-specific expression are scarce, particularly for non-model organisms, and focuss predominantly on MHC class I. In birds, large variation in expression among MHC class I alleles were identified [43–45]. In these studies, a set of alleles showed consistently low expression, and they were attributed to the non-classical MHC, while classical MHC class I alleles were more polymorphic and showed pronounced variation in expression. For MHC class II, allele-specific high and low expression was identified in humans and mice with consequences for disease risk, although environmental factors may have a considerable effect on expression levels [32]. A study on newts inferring allele expression from whole transcriptome sequencing identified several highly expressed alleles and a strongly supported cluster of alleles with practically no

expression [35]. In salmonids and medaka, three clusters of MHC class IIB loci were identified: one highly polymorphic cluster with high expression across all tested tissues and features of classical MHC IIB and two clusters inferred to contain non-classical loci with low polymorphism and low, often tissue-specific expression [39, 46]. In this study, we used both allele-specific sequencing of the MHC IIB and transcriptome-wide expression data for a comparative exploration of expression variation in this gene family across cichlid fish.

Cichlid fish are one of the most species-rich vertebrate family [47, 48]. They inhabit tropical and subtropical freshwater systems where they have rapidly diversified as a result of natural and sexual selection [49–53]. Exposure to divergent parasite communities could be a mechanism contributing to cichlid diversification [54–57]. Cichlid fish have one of the largest MHC class II repertoires described so far [33, 58], which could be driven by parasite mediated selection. MHC allele pools exhibit considerable differentiation between populations and closely related species [54, 59, 60], and at the same time, allelic lineages are shared among independent radiations even across continents [61, 62]. Similar to other systems, MHC allele pool divergence in cichlid fish did not entirely correspond to divergence in parasite communities [60, 63], indicating that selection may act at different levels on the MHC, including expression.

In this study, we aimed to characterize patterns of MHC class II expression in cichlid fish. We generated MHC sequence data and identified MHC IIB alleles in the populations of the Neotropical Midas cichlid species complex (Amphilophus spp.), estimated their frequencies and expression levels, and assessed the relationship between frequency and expression. We then grouped functionally similar alleles into supertypes and examined their distribution and expression patterns across Midas cichlid populations. Finally, we used available genomic and transcriptomic data from African cichlids to quantitatively assess whether variation in expression among putative MHC class II loci showed patterns comparable to those observed among Midas cichlid alleles and supertypes. We also examined the phylogenetic relationship between highly and lowly expressed Midas cichlid alleles and putative African loci.

Material and methods

Sampling

We collected samples of Midas cichlid fish (340 in total) in Nicaragua in December in 2017 and 2018 from the two Great Nicaraguan lakes and five crater lakes (Fig. 1, Table 1). We caught fish with gill nets, and euthanized them with an overdose of Tricaine mesylate (MS-222) on cold water, following the procedures approved by the American Veterinary Medical Association Guidelines for Euthanasia of Animals: 2020 edition (available at https://www.avma.org/sites/default/files/2020-02/Guidelines-on-Euthanasia-2020.pdf). We stored muscle tissue in absolute ethanol at −20 °C and we collected the spleen in RNAlater Stabilization Solution (Ambion, USA) and stored it at −80 °C thereafter.

DNA and RNA extraction

We extracted DNA from fin, muscle, or spleen with the DNeasy Blood and Tissue kit (Qiagen, Germany) following the manufacturer's recommendations. We incubated a piece of tissue with 20 µl of Proteinase K in 180 µl of ATL buffer at 56 °C for approximately 3 h. We mixed the lysate with 200 µl of AL buffer and 200 µl of ethanol and transferred it to the spin column. We cleaned the DNA first with 500 µl of AW1 buffer and then with 500 µl of AW2 buffer and eluted it in 200 µl of DNase/RNAse-free water. We removed remnant RNA by incubating the DNA with 14 U of RNase A (Qiagen, Germany) at room temperature for 4 min and inactivated the RNase by freezing at −20 ºC. DNA was quantified with Qubit (Invitrogen, USA).

We extracted RNA from a fragment of 5 × 5 mm of spleen, which we homogenized with ceramic beads in 250 µl of TRI Reagent at 6800 rpm for 3 × 20 s in a Precellys Evolution homogenizer (Bertin Instruments, France). We

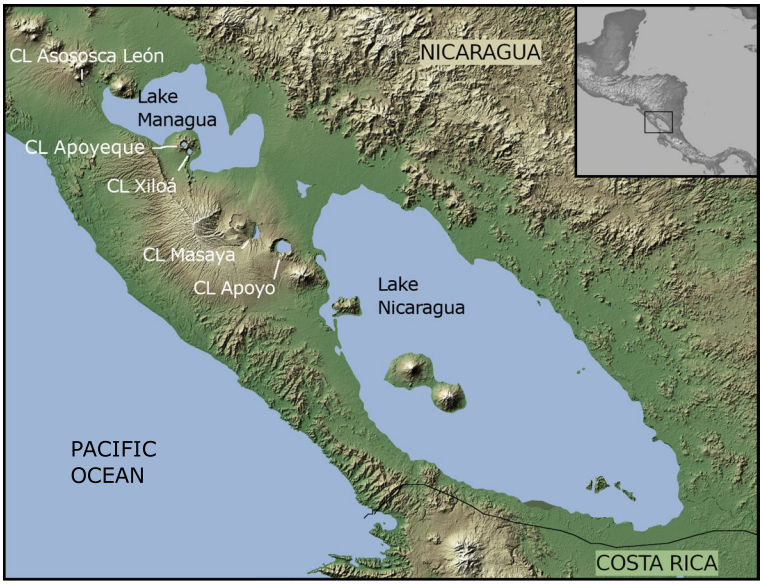


Fig. 1 Map of the great lakes region in Nicaragua (modified from NASA Photojournal PIA03364). Sampled lakes are indicated. CL = Crater Lake

Table 1 Number of individuals (n) of Midas cichlid included in the analyses in the different lake (L) and crater lakes (CL)

Lake	Population	n
L Nicaragua	*A. citrinellus* Nic	26
L Nicaragua	*A. labiatus* Nic	22
L Managua	*A. citrinellus* Man	50
L Managua	*A. labiatus* Man	12
CL Masaya	*A.* cf. *citrinellus* Mas	26
CL Masaya	*A.* cf. *labiatus* Mas	2
CL Apoyeque	*A.* cf. *citrinellus* Apo	12
CL Apoyeque	*A.* cf. *labiatus* Apo	6
CL Asososca León	*A. citrinellus* f. bent	18
CL Asososca León	*A. citrinellus* f. limn	20
CL Apoyo	*A. astorquii*	19
CL Apoyo	*A. chancho*	14
CL Apoyo	*A. globosus*	5
CL Apoyo	*A. zaliosus*	18
CL Xiloá	*A. amarillo*	19
CL Xiloá	*A. xiloaensis*	17
CL Xiloá	*A. sagittae*	35

incubated the homogenate for 5 min at room temperature, then added 50 µl of chloroform. We incubated the mixture for 2 min at room temperature and centrifuged it for 15 min at 12,500 rpm at 4 °C. We transferred the supernatant to a new tube, added 125 µl of isopropanol, and incubated it overnight at −80 °C. We transferred the liquid to a Direct-zol RNA Miniprep (Zymogen, USA) spin column and cleaned the RNA following the manufacturer's instructions. In short, we washed the spin column twice with 400 µl of RNA PreWash and once with 700 µl of RNA Wash Buffer. We eluted the RNA in 40 µl of DNase/RNase-free water and treated it with DNase I (Thermo Scientific, USA) following the manufacturer's recommendations. RNA quantity and quality were measured with a NanoDrop 1000 (Thermo Scientific, USA).

cDNA synthesis and qPCR testing

We synthesized cDNA with the RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific) following the manufacturer's instructions for gene-specific primers. We designed gene-specific reverse transcription primers

that targeted the MHC IIB (5'-AACATGGCAGGATG-3') and two housekeeping genes (ACTB 5'-TCCCTGTT GGCTTT-3', EF1A 5'-GTCGTTCTTGCTGT-3'). We mixed approximately 450 ng of DNase-treated RNA with 20 pmol of each reverse transcription primer, 4 µl of 5 × reaction buffer, 1 µl of RiboLock RNase Inhibitor, 2 µl of dNTPs, and 1 µl of RevertAid M-MuLV reverse transcriptase in a final volume of 20 µl and incubated it for 60 min at 42 °C, then for 5 min at 70 °C, and stored it at −80 °C.

We diluted the cDNA 1:20 and amplified MHC IIB, ACTB, and EF1A in a single PCR reaction. We reduced the amplification of the housekeeping genes with blocking reverse primers with a 3' amino linker. We determined the optimal blocking primer concentrations with qPCRs, gradually increasing the concentration of blocking reverse primers and decreasing the concentration of normal reverse primers, keeping the total concentration of reverse primers constant at 0.3 µM. We approximated the amount of PCR product with the height of peaks of the melting curves corresponding to each gene. We deemed the concentration of blocking primers that rendered peaks of similar heights as optimal. We prepared qPCR reactions as described below for regular PCRs but using the FastStart Universal SYBR Green Master (ROX). Cycling conditions were as follows: initial denaturation of 10 min at 95 °C and 40 cycle of 95 °C for 15 s, 50 °C for 30 s, and 72 °C for 1 min. This was followed by the melting analysis. We determined the optimal number of thermal cycles for regular PCRs from the amplification curves in order to avoid reaching the plateau phase.

MHC amplification and sequencing

MHC IIB was amplified from DNA with one forward and two reverse primers [60] and sequenced on a Illumina MiSeq by LGC Genomics (Germany). Twenty-one samples of the 340 were amplified and sequenced in duplicates. Briefly, PCRs were done in 20 µl containing 15 pmol of each primer and 2 × MyTaq PCR Mix with 1.5 U of polymerase. Cycling conditions were as follows: 2 min denaturation at 96 °C followed by 20 cycles of 96 °C for 15 s, 50 °C for 30 s, 70 °C for 1 min. Following primer removal, libraries were prepared using the MiSeq v3 chemistry. Paired-end sequencing resulted in a mean of 15,700 2 × 300 bp reads per sample.

We amplified MHC IIB and housekeeping genes from cDNA in a single reaction for each sample. We did PCRs in a reaction volume of 25 µl containing 12.5 µl 2 × FastStart PCR master mix (Roche, Germany), one forward and two reverse MHC IIB primers [60], one forward, one reverse, and one blocking reverse primer for each housekeeping gene, and 6.25 µl of diluted cDNA. Primer sequences and final concentrations are given in Additional file 1 table S1. Cycling conditions were the same as for qPCRs with only 23 cycles followed by a final elongation step of 7 min at 72 °C. We checked amplification success on an agarose gel. Adapter ligation and paired-end sequencing on an Illumina NextSeq 500/550 v2 were done at LGC Genomics (Germany) using 20 µl PCR product as starting material. This resulted in a mean of 636,677 2 × 150 bp reads per sample. We aimed for very high coverage to be able to detect alleles with very low expression.

Read processing

We demultiplexed reads from gDNA sequencing and trimmed adapters with bcl2fastq v2.20 [64] allowing for up to two mismatches in the barcode and discarding trimmed reads shorter than 20 base pairs (bp). We assembled the reads into alleles using the AmpliSAT pipeline [65]. We merged reads with AmpliMERGE and cleaned them with AmpliCLEAN, discarding reads with mean Phred scores < 30 and missing forward or reverse primers. We trimmed primers with a custom python script before calling alleles with AmpliSAS. To determine the best per amplicon frequency (PAF) filter, we ran AmpliSAS with values from 2 to 3.5 with incremental steps of 0.1. We compared the number of inferred alleles per sample between consecutive PAF values. The PAF value for which the number of alleles per sample differed least from the less stringent value and from the more stringent value was considered optimal. Additionally, we compared the agreement of samples that were PCR-amplified and sequenced in duplicates for each PAF. Both methods identified 2.7 as the best PAF for our data and all further analyses used the alleles called with this filter. Additional AmpliSAS parameters were as described in Bracamonte et al. [59].

Alleles could not be called for three samples due to the low number of reads that passed quality filtering. We renamed newly identified alleles in accordance with the naming system that we introduced in this previous study ("acXXX" where XXX follows consecutive numbering [59]). We removed an allele with a 1-bp deletion leading to a premature stop codon from further analyses. This procedure identified the composition of MHC alleles for each sample (i.e. the MHC profile) with all putative alleles present in the genome (expressed and non-expressed).

We demultiplexed and trimmed reads from cDNA sequencing in the same way as for gDNA sequencing (see above). We trimmed primers with a custom python script and merged them with BBMerge v38.90 [66] using the maxstrict setting and requiring a match between ratio mode and flat mode. We discarded reads shorter than 132 bp (expected length MHC −10 bp) and longer than 178 bp (expected length ACTB/EF1A+10 bp) using the script reformat.sh from BBMap v38.90. We mapped the surviving reads of each sample to the corresponding MHC profile, which included the sequences of the housekeeping genes ACTB and EF1A, with Bowtie2 v2.4.4 in end-to-end alignment mode [67]. Seed length for multiseed alignment was set to 32 bases with a seed interval of 1 base. Re-seeding was set to 3 allowing for up to 20 seed extension attempts. Maximum and minimum mismatch penalties were 12 and 6, respectively, and read gap opening and extension penalties were 8 and 5. We sorted SAM files by coordinate and converted them to BAM files with samtools v1.9 [68]. We removed reads with mapping quality < 8. We quantified reads per MHC allele and housekeeping gene for each sample separately with featureCounts [69] from the R package Rsubread v2.4.2 [70] using the MHC profile of the respective sample as reference. We could not map and quantify reads for one sample, because its MHC profile could not be inferred from DNA.

Quantification of African cichlid MHC expression

We used published data sets of transcriptome sequencing of the gills of adult individuals of several cichlid species from Lake Tanganyika ([71]; BioProject PRJNA552202; species, sex, and accession numbers are given in Additional file 1 table S2). We mapped the reads separately to the Oreochromis niloticus reference genome and to a curated list of representative MHC alleles of loci and homology groups defined in Hablützel et al. [61], hereafter referred to as putative loci, with HISAT2 v2.2.1 [72] and quantified the reads with HTSeq [73]. We combined these two sets of read counts for further analyses.

Data analysis

Phylogenetic reconstruction

We constructed a maximum likelihood phylogenetic tree of MHC IIB with all known Amphilophus alleles [33, 59] and the African alleles with IQ-TREE v2.0.6 [74]. We determined the best-fit nucleotide substitution model with ModelFinder [75] implemented in IQ-TREE, using separate partitions for each codon position. ModelFinder selected TVM+F+G4 for codon positions 1 and 2 and K3P+G4 for codon position 3, and the phylogeny was inferred using these partitions [76]. Node support was assessed using the ultrafast bootstrap method [77] with 10,000 replicates.

Analyses of MHC in the Midas cichlid

To group alleles into supertypes, we identified positively selected sites across all known Amphilophus MHC class IIB alleles using four different methods: FEL and FUBAR, available on the Datamonkey server [78, 79]), and two model comparisons implemented in CodeML (M2a vs. M1a and M8 vs. M7; [80]). We considered a site as positively selected if it was identified by at least two methods. We inferred the supertype of each allele as described in Bracamonte et al. [59].

Individuals for which either the MHC profile or expression estimates were unavailable were excluded from further analyses. All statistical analyses were performed in R v4.5.0 [81], unless otherwise indicated.

We first analysed read coverage for both housekeeping genes to identify samples with insufficient coverage or evidence of unsuccessful amplification or sequencing. For each sample, we calculated the proportion of housekeeping gene reads covering each gene and used boxplots to identify and remove outlier samples. We additionally excluded samples with coverage < 50 for either housekeeping gene. The analyses were robust to this threshold, as coverages thresholds ranging from 50 to 200 produced largely consistent results. We then normalized the expression of each MHC allele within each sample relative to the housekeeping genes. For this, we divided the number of reads of each MHC allele by the number of reads of each housekeeping gene and considered the mean of these normalizations as relative expression. We calculated frequency and mean expression of each allele and supertype within the Midas cichlid species. We mapped mean expression and supertype identity to the allele phylogeny with ggtree v4.0.5 [82].

Since the majority of alleles occurred in only very few individuals and were not shared among populations, we restricted analyses to the most common alleles (i.e. occurring in more than 5% of individuals). To test for differential expression among alleles, we fitted a generalized linear model (GLM) with a gamma hurdle distribution accounting for overdispersion. The common allele ac014 was not expressed in any individual and was thus removed from the GLM analysis. We used the package glmmTMB v1.1.14 [83] to fit the GLM and checked model assumptions with the package DHARMa v0.4.7 [84]. We included population as a random factor. We determined which alleles differed with pairwise post-hoc tests using the package emmeans v2.0.4 [85] and we accounted for multiple testing with the Benjamini–Hochberg correction. We further tested whether expression is dependent on the frequency of an allele using a Spearman correlation. For supertypes, we assessed the relationship between mean relative expression and frequency and mean relative expression and the number of alleles per supertype (i.e. supertype size) with Spearman correlations.

In order to estimate whether expression patterns and the frequency-expression relationship could be influenced by unequal sample sizes among alleles, we generated five resampled datasets with equal sample sizes across alleles allele by randomly selecting the same number of individuals for each allele. We then repeated the analyses described above, using a Tweedie distribution with overdispersion instead of a gamma hurdle distribution. We classified alleles as highly or lowly expressed based on clusters of similar expression, and this classification was subsequently extrapolated to rare alleles that were not included in the differential expression analysis.

For each individual, we determined the total number of alleles, the number of expressed alleles, and mean relative expression. These parameters were also calculated separately for highly and lowly expressed alleles. Relationships among these variables were examined using Spearman correlations. We then tested whether the number of supertypes and the number of expressed supertypes per individual differed among Midas cichlid populations by fitting a GLMs with a zero-truncated generalized Poisson distribution. We determined which populations differ in number of supertypes and number of expressed supertypes with pairwise tests. We further determined whether the frequency and mean expression of supertypes differed among populations. To test for differences in frequencies, we fitted a GLM with a beta distribution. We transformed frequencies with the formula $(y * (n-1) + 0.5)/n$, where y is the frequency and n is the number of samples for the respective lake, to ensure values were within the interval (0,1). The model included lake, supertype, and their interaction as fixed factors. To test for differences in expression, we fitted a GLM with a compound Poisson-gamma distribution that included population, supertype and their interaction as fixed factors. For both frequency and expression, we accounted for overdispersion of supertypes. We performed pairwise tests of lake or population differences by supertypes and supertype differences by lake or population.

African cichlid MHC

We standardized read counts by library size, estimated dispersion, and normalized counts with DESeq2 v1.36.0 [86]. We extracted normalized read counts of putative MHC IIB and MHC IIA loci and used them as proxy for

expression. We tested for differential expression among putative MHC IIB and MHC IIA loci by fitting GLMs with a Gamma distribution including species, locus and their interaction as fixed factors. We accounted for over-dispersion by species and for zero-inflation in the case of MHC IIA. We inferred pairwise expression differences between putative loci and between species. Since the interaction term was significant for MHC IIA, we did the pairwise tests by categories of the other factor.

Results

MHC expression patterns in the Midas cichlid

We found a total of 159 alleles in 321 Midas cichlid samples that clustered into 14 supertypes (four to 38 alleles per supertype) and two unassigned alleles that did not cluster in any supertype (Fig. 2, Additional file 1, Tables S2, S3). One allele (Amci-DXB*040101) was present in all individuals, and another (ac001) was present in all but three individuals, while 52 alleles were singletons, i.e. they occurred in a single individual (Additional file 1 Table S3). In total, 146 alleles were expressed in at least one individual, while 13 were not found expressed in any individual (Additional file 1 Table S3). For the non-expressed alleles, nine were singletons, three were rare (present in only a few individuals), while one was relatively abundant (allele ac014 present in 17.1% of all individuals; Additional file 2 Fig. S1). For the expressed alleles, 22 were only expressed in some of the individuals in which they occurred while the remaining 124 expressed alleles were always expressed (Additional file 1 Table S3). Mean relative expression (i.e., number of reads normalised by number of house-keeping gene reads) of the expressed alleles ranged from 4*10–5 to 24.2 (Additional file 1 Table S3).

Mean relative expression of supertypes (ST) ranged from 0.0002 for ST7 to 8.9 for ST3 (Additional file 1 Table S4). Supertypes ST1, ST2, and ST7 formed three well-defined clusters on the MHC phylogeny (Fig. 2) and had comparatively low nucleotide diversity (Additional file 1 Table S4). ST5, ST8 and ST9 also formed three well-defined clusters with low to moderate nucleotide diversity (Fig. 2, Additional file 1 Table S4). Despite that, ST7

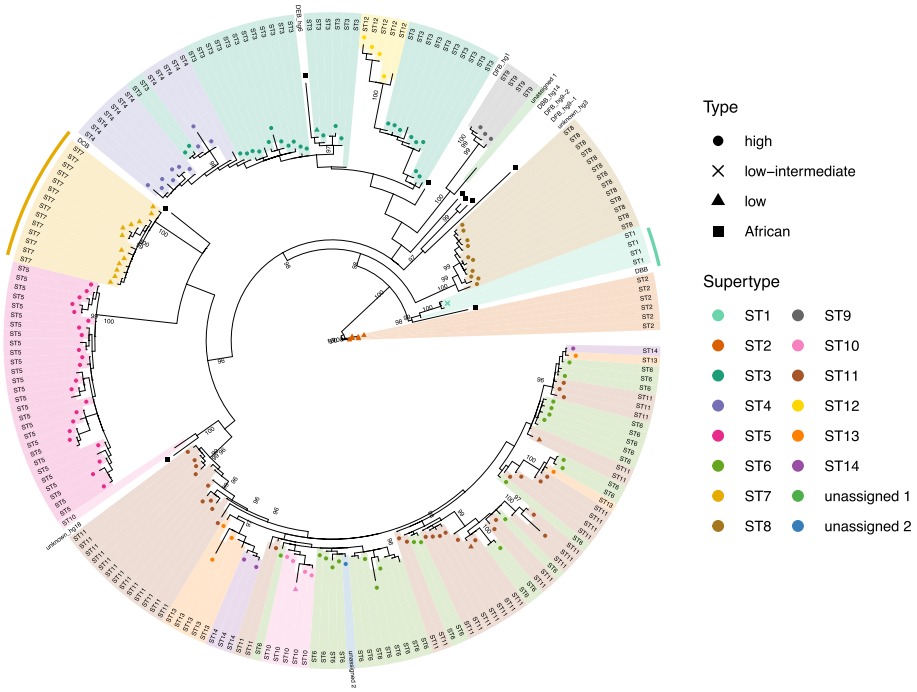


Fig. 2 Phylogeny of Midas cichlid and African cichlid MHC II*B* alleles. Symbols (type) indicate if alleles are highly expressed, intermediately expressed, lowly expressed, or of African origin. Colours refer to supertype for Midas alleles, African loci are shown in black

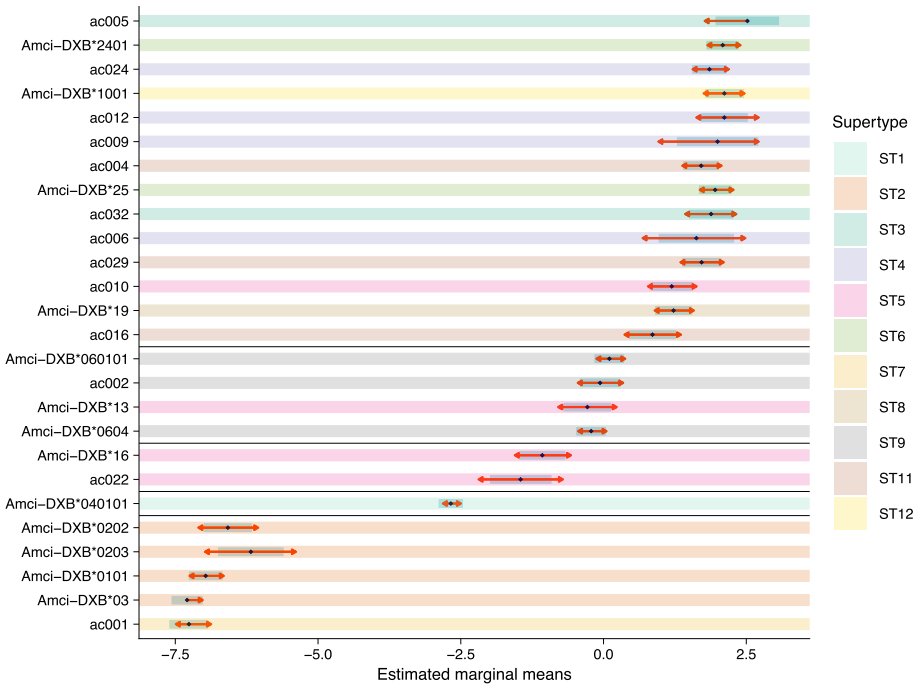


Fig. 3 Estimated marginal means of pairwise post-hoc tests of common Midas cichlid MHC II*B* alleles. Alleles are coloured by supertype, black lines separate groups of alleles with similar expression levels. Allele ac014 was excluded from the analysis, because it was not expressed in any individual

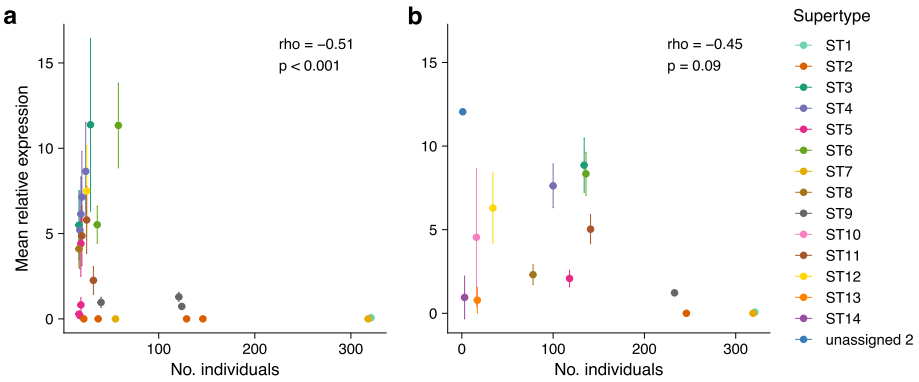


Fig. 4 Correlation between frequency and expression for the 27 common MHC II*B* alleles (**a**) and the supertypes (**b**) of the Midas cichlid. Spearman's rho and *p*-values are indicated

and ST8 had dN/dS ratios well above one, indicative of positive selection, while dN/dS of ST1, ST2 and ST9 was below one (Additional file 1 Table S4).

Frequency-expression relationship

Twenty-seven alleles occurred in more than 5% of all individuals and were therefore considered common alleles. Expression differed significantly among the common alleles ($\chi 2 = 18{,}127$, df = 25, $p < 0.001$). Pairwise comparisons identified three to five groups of alleles with similar expression levels (Fig. 3, Additional file 1 Table S5, Additional file 2 Fig. S2) and one non-expressed allele that was excluded from the analysis. Five common alleles were classified as lowly expressed and one allele, the ubiquitous allele of ST1, had low-intermediate expression. Despite their low expression, these alleles were among the most frequent, with two occurring in > 99% of individuals. Indeed, the four most frequent alleles belonged to the lowly expressed group (Fig. 4a). Among highly expressed alleles, the two most frequent showed comparatively low mean relative expression, while alleles with higher expression tended to occur at lower frequencies. Overall, expression of common alleles was negatively correlated with allele frequency (rho = −0.51, $p < 0.001$; Fig. 4a). This relationship was weaker at the supertype level (rho = −0.45, $p = 0.09$; Fig. 4b), although the three supertypes with the lowest mean relative expression (ST1:

Table 2 Descriptive diversity statistics per individual in the Midas cichlid. The ubiquitous allele with consistently low-intermediate expression is not included in the lowly expressed group

	Mean	SD	Median	Min	Max
No. alleles	6.9	1.5	7	3	12
No. highly expressed alleles	3.6	1.3	4	0	7
No. lowly expressed alleles	2.3	0.89	2	0	5
No. of expressed alleles	5.5	1.5	6	2	10
No. of expressed highly expressed alleles	3.5	1.3	4	0	7
No. of expressed lowly expressed alleles	0.97	0.88	1	0	4
Mean expression of alleles	2.2	2.0		0.007	16.5
Mean expression of highly expressed alleles	4.4	4.1		0	32.8
Mean expression of lowly expressed alleles	0.001	0.002		0	0.03
No. supertypes	5.9	1.1	6	3	8
No. expressed supertypes	4.9	1.2	5	2	8

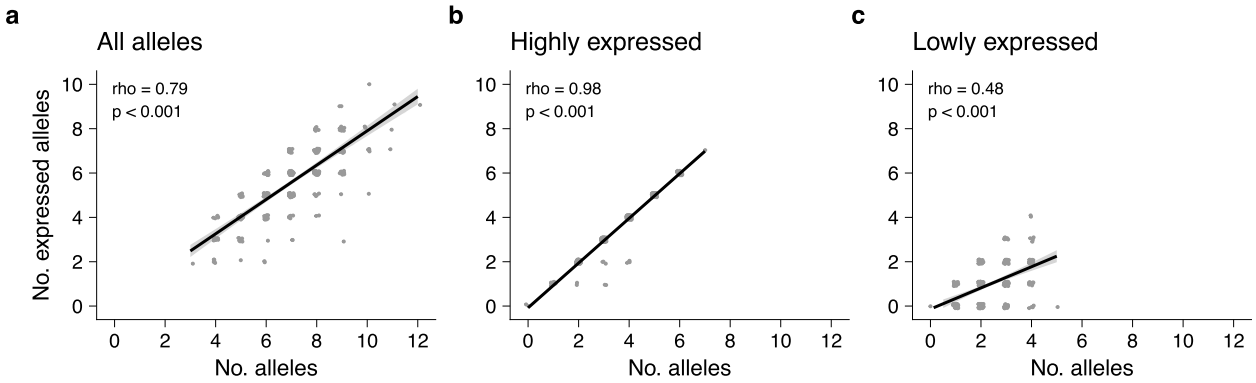


Fig. 5 Correlation between number of alleles and number of expressed alleles per individual for all alleles (**a**), only highly expressed alleles (**b**), and only lowly expressed alleles (**c**). The ubiquitous and consistently low-intermediate expression is not included in the lowly expressed group. Spearman's rho and p-values are given

0.067, ST2: 0.001, ST7: 0.0002) occurred in more than 75% of individuals. Applying the same 5% frequency threshold, ST10 (six alleles, five of which occur in individuals included in this analysis; mean relative expression = 4.6) and ST14 (four alleles, three of which occurred in individuals included in this analysis; mean relative expression = 0.95) were rare, while all other supertypes were common (Additional file 1 Table S4). Mean relative expression was not correlated with supertype size (rho = 0.19, p = 0.49).

Individuals carried a median of seven alleles, of which a median of six were expressed (Table 2). Across individuals, the number of alleles carried and the number expressed), were strongly correlated (rho = 0.79, $p < 0.001$; Fig. 5a). Individuals carried a median of four highly expressed alleles and two lowly expressed alleles (Table 2), while the ubiquitous allele with low-intermediate expression was expressed in all individuals. For both highly and lowly expressed alleles, the number of alleles carried was positively correlated with the number of expressed, although the relationship was considerably stronger for highly expressed alleles (highly expressed: rho = 0.98, $p < 0.001$, lowly expressed: rho = 0.48, $p < 0.001$; Fig. 5b, c). Only ten individuals did not express all of their highly expressed alleles, while 271 individuals did not express all of their lowly expressed alleles. Mean ± SD relative expression per individual was 2.2 ± 2.0 overall, 4.4 ± 4.1 for highly expressed alleles, and 0.001 ± 0.002 for lowly expressed alleles (Table 2). Individuals possessed a median of six supertypes, of which a median of five were expressed (Table 2). ST1 was present in every individual and ST7 was present in all but two. The proportion of individuals carrying only one allele of a given supertype ranged from 67% for ST2 to 100% for ST10, ST12 and ST14. Up to three alleles of the same supertype were observed, although infrequently, for ST2 (4.5% of individuals), ST3 (0.7%), ST7 (0.6%) and ST9 (0.4%).

Supertype frequency and expression diversity in the Midas cichlid populations

Midas cichlid populations differed significantly in the number of supertypes that an individual possessed (Table 3a). After accounting for multiple testing, 47 pairwise comparisons between populations showed significant differences (Additional file 1 Table S6). In general, populations from the large lakes possessed alleles from more supertypes than those from crater lakes, with the exception of Crater Lake (CL) Xiloá whose fish had similar numbers of supertypes to those from the large lakes. The deep benthic species A. xiloaensis from CL Xiloá had the highest number of supertypes. The deep benthic species A. chancho from CL Apoyo had similar number of supertypes to populations from the large lakes and higher numbers than other fish from CL Apoyo. On the other hand, A. labiatus from Lake Nicaragua had lower numbers than other populations from the large lakes and was more similar to those from the crater lakes. Populations also differed in the number of supertypes that an individual expressed (Table 3b), although less strongly than for the number of supertypes. In total, 36 pairwise comparisons were significant (Additional file 1 Table S6). As for the number of supertypes, fish from CL Xiloá and A. chancho expressed similar numbers of supertypes as fish from the large lakes and A. labiatus from Nicaragua expressed similar numbers of supertypes as fish from crater lakes.

Frequencies differed between supertypes and between fish from different lakes for a given supertype (Fig. 6a,c,e,g, Table 3c, Additional file 2, Fig. S4), although the differences between supertypes were much more pronounced (69% pairwise comparisons significant; Additional file 1 Table S7) than between lakes (18% pairwise comparisons significant). Supertypes could roughly be grouped into four groups based on similar frequencies and these differed only minimally among lakes (Additional file 1 Table S7). Supertypes ST1, and ST7 were both present in all or most fish in all lakes (Fig. 6a,e) while ST2 and ST9 were highly abundant in all lakes (Fig. 6g, Additional file 2 Fig. S4). Supertypes ST3, ST4, ST5, ST6, ST8 and ST11 were moderately abundant (Fig. 6c, Additional file 2 Fig. S4). They were present in all lakes, although not in all populations. ST12 and ST13 had low abundance and were not present in all lakes (Additional file 2 Fig. S4). ST10, and ST14 were rare and were thus not included in the analysis. The 'unassigned 2', a single allele, was only present in a single A. xiloaensis individual and was therefore also excluded. When comparing frequencies between lakes for each supertype, several supertypes were less frequent in CL Masaya and CL Apoyeque than in other lakes (Additional file 1 Table S7, Additional file 2 Fig. S4). CL Masaya was characterized by lower frequency of ST9, CL Apoyeque was characterized by lower frequencies of ST7 and ST9, and CL Asososca León was characterized by lower frequencies of ST5 and ST9 (Fig. 6e,g, Additional file 1 Table S7, Additional file 2 Fig. S4).

Mean supertype expression differed between supertypes and between populations (Fig. 6b,d,f,h, Table 3d, Additional file 2 Fig. S4). As for supertype frequency, differences between supertypes were more pronounced (52% pairwise comparisons significant; Additional file 1 Table S8) than differences between populations (11% pairwise comparisons significant; Additional file 1 Table S8). Contrary to frequencies, expression level differences were more gradual without clear distinctions, except for supertypes ST2, and ST7 for which expression was

Table 3 Results of generalized linear model analyses for Midas cichlid MHC II*B* supertypes

Factor	Chisq	df	*p*-value
a) Number of supertypes per individual			
Population	117.56	16	<0.001
b) Number of expressed supertypes per individual			
Population	76.32	16	<0.001
c) Supertype frequency			
Lake	88.14	6	<0.001
Supertype	2739.04	11	<0.001
Lake:Supertype	219.16	66	<0.001
d) Supertype mean expression			
Population	707.45	162	<0.001
Supertype	10,352.58	158	<0.001
Population:Supertype	402.88	165	<0.001

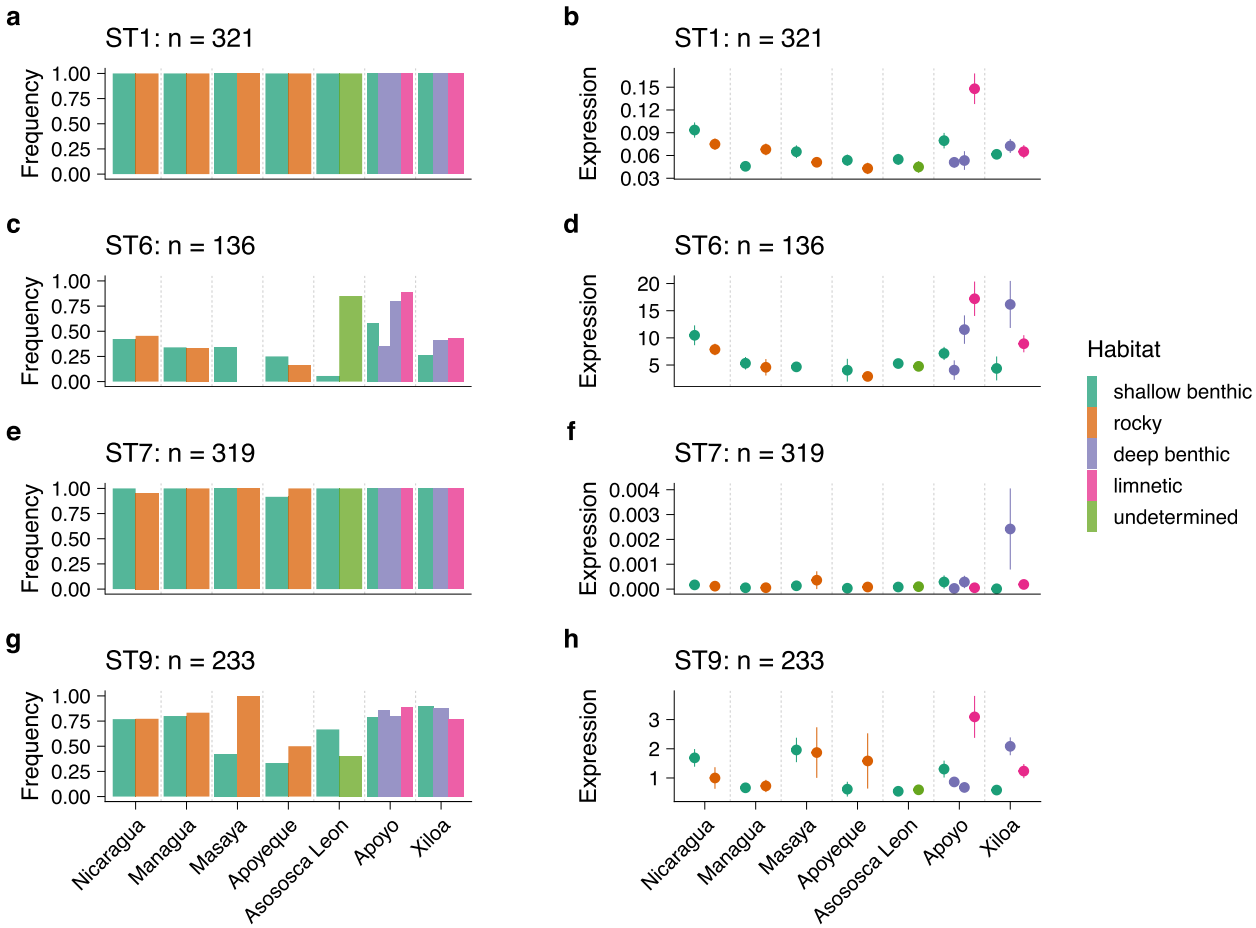


Fig. 6 Frequency and expression of four representative Midas cichlid supertypes across lakes and populations. Frequency (**a**, **c**, **e**, **g**) and expression levels (**b**, **d**, **f**, **h**) are shown for ST1, ST6, ST7 and ST9, respectively. Populations are coloured according to habitat type. The number of individuals carrying each supertype (n) is indicated in each panel. Expression values are shown as mean ± 95% CI

uniform with the exception of elevated expression in A. xiloaensis (Fig. 6f, Additional file 1 Table S8, Additional file 2 Fig. S4), and ST4, ST11 and ST13 that showed uniform expression. For all other supertypes, expression differed in some populations but not in others. When comparing expression of supertypes between populations, A. zaliosus from CL Apoyo and A. xiloaensis and A. sagittae from CL Xiloá showed higher expression than all other populations for several supertypes (Fig. 6, Additional file 1, Table S8, Additional file 2 Fig. S4). Furthermore, A. citrinellus from lake Nicaragua also showed higher expression for some supertypes (Fig. 6d, Additional file 1 Table S8, Additional file 2 Fig. S4). On the other hand, populations in CL Asososca León and CL Apoyeque showed lower expression for several supertypes (Fig. 6b,d,h, Additional file 1 Table S8, Additional file 2 Fig. S4).

Expression diversity in African cichlids

Similar to the Midas cichlid, we found variation in MHC IIB expression in African cichlids (Fig. 7a). Two putative loci (DBB & DCB) had low expression while the others had moderate to high expression. These two putative loci clustered with two Midas cichlid supertypes with exclusively lowly expressed alleles (DBB with ST1 and DCB with ST7), while most moderately to highly expressed putative loci clustered separately from Midas cichlid alleles (Fig. 2, Additional file 2 Fig. S1). Expression differed significantly among the putative loci, and these differences were consistent between species (Table 4a). Only the DEB and the DFB loci showed similar expression

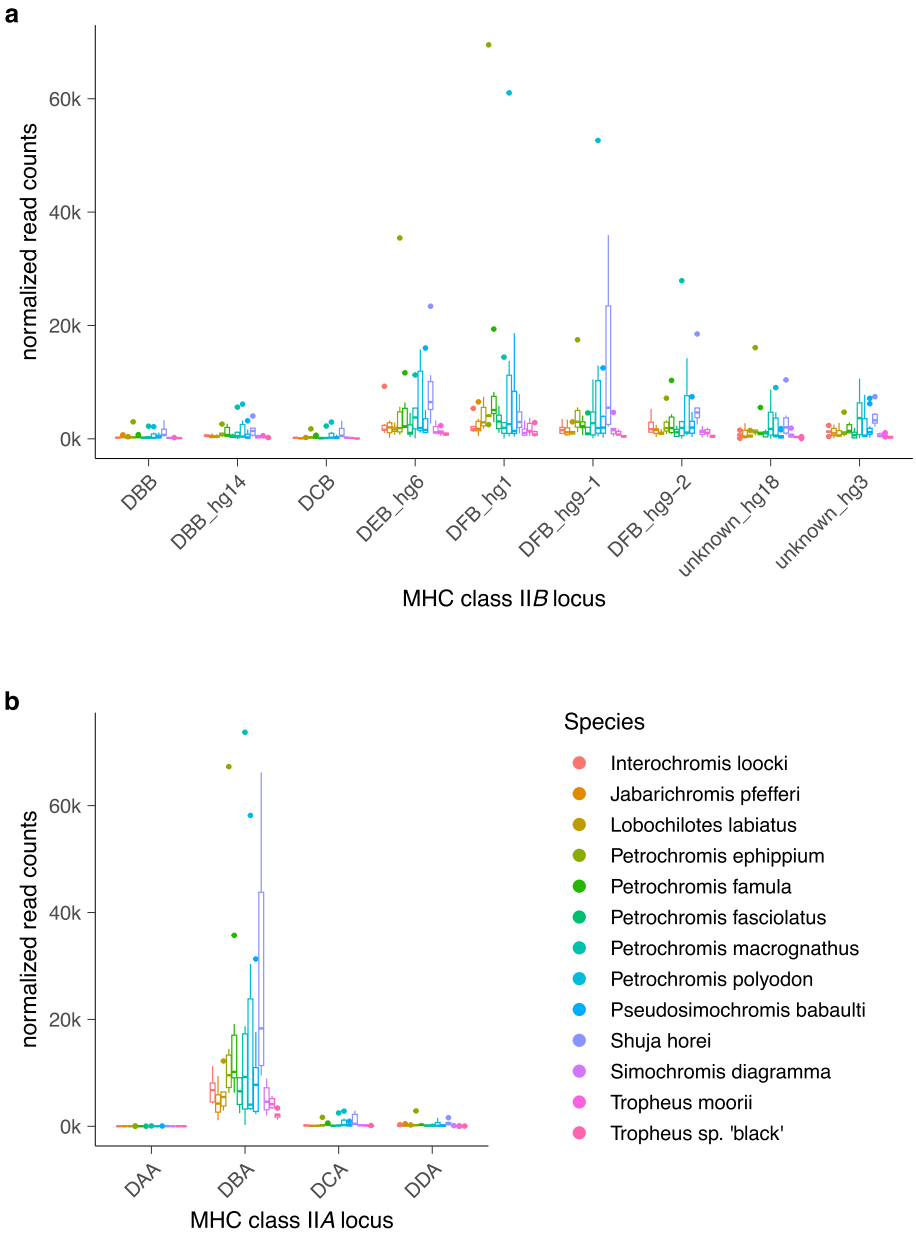


Fig. 7 Expression of putative African cichlid MHC II*B* (**a**) and II*A* (**b**) loci by species. Expression is approximated by normalized read counts

Table 4 Results of generalized linear model analyses for African MHCII loci

Factor	Chisq	df	*p*-value
a) MHC II*B*			
Species	783.494	12	<0.001
Locus	1473.649	8	<0.001
Species:Locus	111.157	96	0.138
b) MHC II*A*			
Species	417.505	12	<0.001
Locus	4804.484	3	<0.001
Species:Locus	60.228	36	0.007

levels, as did the two unknown hg loci (Additional file 1 Table S9). Most pairwise comparisons among species (64%) also indicated significantly different expression (Additional file 1 Table S9). Shuja horei had generally higher expression than other species, while the two Tropheus species showed lower expression than other species. Furthermore, P. macrognathus, P. polyodon and P. ephippium had similar levels of MHC IIB expression and J. pfefferi, I. loocki, L. labiatus, P. fasciolatus and S. diagramma also showed similar MHC IIB expression across loci. While expression varied considerably among putative MHC IIB loci only one putative MHC IIA locus showed high expression (Fig. 7b), although all four differed significantly from each other in all species analysed except for DCA vs DDA (Table 4b, Additional file 1 Table S10). Similar to MHC IIB, species differed in expression at MHC IIA loci, although less consistently than for MHC IIB (Table 4b, Additional file 1 Table S10). As for MHC IIB, S. horei had higher expression than most other species at all MHC IIA loci and Tropheus sp. 'black' had lowest. Across all MHC IIA loci, I. loocki, P. famula, P. macrognathus, P polyodon, P. ephippium and P. babaulti formed a cluster of similar expression and L. labiatus, P. fasciolatus, S. diagramma and T. moorii formed a second cluster (Additional file 1 Table S10).

Discussion

This study examines allele-specific MHC expression in the Neotropical Midas cichlid, a fish lineage with one of the largest MHC class II allele repertoires described to date. Consistent with previous findings [59], we found extensive MHC sequence diversity across Midas cichlid populations, including a few common and widespread alleles and a large proportion of less frequent alleles occurring in relatively few individuals. Expression varied markedly among alleles and was negatively associated with allele frequency: the most common alleles tended to show low expression, while many less common alleles showed substantially higher expression. A notable qualitative phylogenetic pattern was that lowly expressed Midas cichlid alleles clustered with lowly expressed putative MHC loci from African cichlids, while most highly expressed alleles and putative loci formed separate clusters in the African and Neotropical lineages.

Variation in expression among MHC alleles has previously been described in both model and non-model organisms for both MHC class I [43, 45, 87] and MHC class II [13, 35, 40]. In Midas cichlids, expression levels differed by several orders of magnitude among alleles. Moreover, MHC supertypes, which group alleles with putatively similar functional properties, tended to comprise alleles with similar mean expression levels. We also observed pronounced differences in expression among putative MHC class II loci within the African dataset. Together, these patterns suggest that expression level may represent an additional dimension of functional differentiation among MHC alleles and loci.

Allele-specific MHC expression has been shown to be modulated by genetic, epigenetic and environmental mechanisms [32, 88]. For MHC class I, previous studies have suggested an association between expression level, which was carried through from transcription to protein [87], and peptide-binding properties, with alleles differing in both the breadth of the peptides they bind and their level of cell-surface expression [89, 90]. Although MHC class I and MHC class II both present bound antigens to T cells, they differ substantially in antigen processing and presentation [91]. It therefore remains unknown whether a comparable relationship between expression level and peptide-binding properties occurs in MHC class II. If such a relationship exists, it could contribute to the expression patterns observed here, including similar expression levels among alleles within supertypes. However, our study did not measure antigen-binding properties or parasite infection. Therefore, whether the observed variation in MHC class II expression is related to differences in antigen recognition or contributes to parasite-mediated selection remains to be tested.

In the Midas cichlid, two clusters of lowly expressed alleles (ST1 and ST2) showed low sequence diversity and no evidence of positive selection, features that are more consistent with non-classical MHC class IIB alleles than with classical alleles [39, 46]. Indeed, a previous study placed these alleles closer to another putative non-classical allele or pseudogene than to the remaining MHC alleles [33]. A quantitatively similar phylogenetic

association was observed for the putative African locus DBB, which clustered tightly with ST1, and showed the second-lowest expression among the putative African loci. It was previously described to have low polymorphism and an amino acid conformation that may result in a distinct tertiary structure of the mature protein [61], all of which is consistent with non-classical MHC loci described in salmonids and suggested for other teleost fish, including the cichlid tilapia [46, 92].

A different pattern emerged for ST7. Although most ST7 alleles showed very low or undetectable expression and low genetic diversity, which could be compatible with a non-classical function [46], they also showed evidence of positive selection and clustered phylogenetically with putatively classical MHC alleles. The putative African locus DCB also showed low expression within the African dataset and was phylogenetically closely related to ST7. This combination of low expression with signatures more typical of classical MHC suggests that ST7 may represent a distinct mode of MHC regulation rather than simply a non-classical or degenerating lineage. One possibility is that these alleles retain classical antigen-presenting functions but are expressed only under particular conditions, in tissues not sampled in this study or at specific developmental stages [37–39].

In African cichlids, two MHC IIB lineages were identified without inter-lineage recombination: the DBB-DEB-DFB lineage and the DCB-DBB lineage [61]. Our phylogeny recovered a broadly similar grouping pattern. While the lowly expressed DCB forms a distinct lineage with the monophyletic, lowly expressed ST7 and the monophyletic sister ST5, the remaining alleles and putative loci, all belonging to the highly expressed groups, are either scattered throughout the phylogeny or form distinct cichlid lineage-specific clusters. The latter alleles and putative loci may thus have arisen by recent gene duplication [93] and be subject to concerted evolution [12]. Accordingly, substantial recombination was described for the DBB-DEB-DFB lineage [61]. Whether the contrasting patterns of expression and diversity observed for ST1, ST2 and ST7, and their phylogenetic associations with the putative African loci DBB and DCB, reflect shared evolutionary or functional properties remains unresolved. Addressing this question will require expression analyses across tissues and developmental stages, together with functional characterization of the encoded molecules.

Previous work in the Midas cichlid showed that many MHC alleles are geographically restricted, occurring only in individuals from a single lake or even in specific populations within lakes [59]. In our study, these less common alleles tended to show higher expression, while the most common alleles generally showed low expression, a pattern similar to that reported for MHC class I in house sparrows [45]. Additionally, supertypes composed of less common, highly expressed alleles showed greater variation in expression among populations. This again suggests that highly expressed alleles may be more responsive to environmental factors and may evolve under different selective or regulatory regimes than the more common, lowly expressed alleles. Whether this population-specific expression is related to local parasite communities or other environmental factors remains to be tested.

Interestingly, the highly expressed putative African cichlid loci, unlike the lowly expressed variants, clustered separately from the Midas cichlid alleles in the phylogeny, further suggesting distinct evolutionary trajectories for highly and lowly expressed MHC variants. Within the African cichlid dataset, we did not find evidence that expression of highly expressed putative loci varied more strongly among species than that of lowly expressed putative loci. However, such a pattern may have been obscured by the inclusion of all putative loci in the analysis without distinguishing between classical and putatively non-classical loci.

Technical and methodological considerations

We have taken care to minimize confounding effects in our study and tested the robustness of our analyses. Nonetheless, we cannot exclude the possibility that limitations intrinsic to the study designs influence our results. Expression estimates for Midas and African cichlids were obtained at different genomic levels, from different tissues, using different sequencing approaches and normalization procedures, precluding direct quantitative comparison of expression levels between the two datasets. Although a reference genome is available for the Midas cichlid, which allows MHC loci to be identified in the genome, the amplified exon 2 fragment is too short and

diverse to reliably assign alleles to loci. Similarly, allelic information cannot be reliably extracted from the available short-read transcriptome data for African cichlids. Nevertheless, two of the three least expressed Midas cichlid supertypes clustered with the two least expressed putative African loci. Given the methodological differences between datasets, we interpret this only as a qualitative phylogenetic association. Whether this association reflects conserved functional properties of these MHC lineages and requires direct functional investigation.

Co-amplification of multiple targets in a single PCR reaction introduces competition among targets for resources such as dNTPs and polymerase, potentially introducing an amplification bias in which more abundant targets and those with more specific primer binding are preferentially amplified, particularly in the later PCR stages when reagents become depleted [94, 95]. This, in turn, can bias quantification. We have taken precautions to reduce amplification bias. We used high concentrations of blocking primers to reduce amplification of the more abundant housekeeping genes to ensure sufficient amplification of the MHC target, which was extensively tested in quantitative PCRs, and we used a limited number of amplification cycles to avoid depletion of PCR reagents. Furthermore, our primers were designed to minimize dimerization which can further deplete reagents and reduce amplification efficiency of the targets. Finally, we removed samples from the analysis for which the ratio of reads between the two housekeeping genes suggested an amplification or sequencing bias. Hence, we do not expect that PCR amplification bias has substantially affected our analysis.

Conclusions

We found marked differences in expression among Midas cichlid MHC class IIB alleles that were associated with allele frequency. Moderate to high expression was characteristic of less common alleles, while low expression predominated among the most common alleles. Although many highly expressed alleles occurred at relatively low frequencies, the supertypes to which they belonged were widely distributed, suggesting that functionally similar alleles may be maintained despite allele turnover among individuals and populations. The lowly expressed putative loci of African cichlids clustered phylogenetically with lowly expressed Midas cichlid alleles, while this pattern was not observed for most highly expressed alleles and putative loci. This phylogenetic association suggests that low expression may be associated with distinct MHC lineages, although its functional significance remains unresolved. Overall, our results show that allele-specific and locus-specific expression could provide an important additional dimension to MHC sequence variation and should be considered when assessing the potential functional relevance of MHC alleles. In particular, allele frequency alone may not reflect functional importance. Whether the observed differences in allele frequency and expression are associated with parasite recognition or maintained by parasite-mediated selection remains to be tested.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1186/s12864-026-13370-x.

Acknowledgements We thank Mariana Leal-Cardín, Ana Santacruz and the local fishermen in Nicaragua for their help with sample acquisition. We thank Dr. Elena Arriero, Jesús García, and Rosa María Pérez from the Universidad Complutense de Madrid for their help with the qPCRs and determining the optimal concentrations of blocking primers. We acknowledge the support of the technical personnel of the Laboratorio de Sistemática Molecular y Genética de Poblaciones at MNCN for assistance with the laboratory procedures.

Authors' contributions MB and SEB conceived the idea and designed the research. PIH contributed with African data. SEB and CLM did the lab work and SEB, PIH, and CLM analysed the data. SEB wrote the first draft and all authors contributed to the final version. All authors read and approved the final manuscript.

Funding This work was supported by grant PID2020-115481 GB-I00 from the Spanish Ministry of Science and Innovation/Spanish State Research Agency (MCIN/AEI/10.13039/501100011033) to MB. SEB was supported by a Juan de la Cierva fellowship from MCIN/AEI/10.13039/501100011033 and NextGenerationEU/PRTR (FJC2020-042779-I) and by an Early PostDoc.Mobility Fellowship from the Swiss National Science Foundation (P2SKP3_191312). CLM was supported

by a predoctoral fellowship from MCIN/AEI/10.13039/501100011033 (FPI- PRE2018-085797).

Data availability The data for this study have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB110986 (https://www.ebi.ac.uk/ena/browser/view/PRJEB110986).

Declarations

Ethics approval and consent to participate Fish sampling was approved by the Ministerio del Ambiente y los Recursos Naturales of Nicaragua (MARENA).

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Schmid-Hempel P. Evolutionary parasitology: the integrated study of infections, immunology, ecology, and genetics. 1st ed. Oxford, United Kingdom: Oxford University Press; 2013. https://doi.org/10.1093/acprof:oso/9780199229482.001.0001.
2. Ebert D, Fields PD. Host-parasite co-evolution and its genomic signature. Nat Rev Genet. 2020;21:754–68. https://doi.org/10.1038/s41576-020-0269-1.
3. Summers K, McKeon S, Sellars J, Keusenkothen M, Morris J, Gloeckner D, et al. Parasitic exploitation as an engine of diversity. Biol Rev. 2003;78:639–75. https://doi.org/10.1017/s146479310300616x.
4. Betts A, Gray C, Zelek M, MacLean RC, King KC. High parasite diversity accelerates host adaptation and diversification. Science. 1979;360:907–11. https://doi.org/10.1126/science.aam9974.
5. Haldane JBS. Disease and evolution. Suppl La Ric Sci. 1949;19:68–76.
6. Woolhouse MEJ, Webster JP, Domingo E, Charlesworth B, Levin BR. Biological and biomedical implications of the co-evolution of pathogens and their hosts. Nat Genet. 2002;32:569–77. https://doi.org/10.1038/ng1202-569.
7. Dexter E, Fields PD, Ebert D. Uncovering the genomic basis of infection through co-genomic sequencing of hosts and parasites. Mol Biol Evol. 2023. https://doi.org/10.1093/molbev/msad145.
8. Eizaguirre C, Lenz TL. Major histocompatibility complex polymorphism: dynamics and consequences of parasite-mediated local adaptation in fishes. J Fish Biol. 2010;77:2023–47. https://doi.org/10.1111/j.1095-8649.2010.02819.x.
9. Murphy K, Weaver C. Janeway's immunobiology. 9th ed. New York, NY: Garland Science, Taylor & Francis Group, LLC; 2017.
10. Penn DJ, Ilmonen P. Major Histocompatibility Complex (MHC). Encyclopedia of Life Sciences, Wiley; 2005. https://doi.org/10.1038/npg.els.0003986.
11. Klein J, Sato A, O'hUigin C. Evolution by gene duplication in the major histocompatibility complex. Cytogenet Genome Res. 1998;80:123–7. https://doi.org/10.1159/000014967.
12. Nei M, Rooney AP. Concerted and birth-and-death evolution of multigene families. Annu Rev Genet. 2005;39:121–52. https://doi.org/10.1146/annurev.genet.39.073003.112240.
13. Efstratiou A, Gaigher A, Künzel S, Teles A, Lenz TL. Template-specific optimization of NGS genotyping pipelines reveals allele-specific variation in MHC gene expression. Mol Ecol Resour. 2024;24:e13935. https://doi.org/10.1111/1755-0998.13935.
14. Kelley J, Walter L, Trowsdale J. Comparative genomics of major histocompatibility complexes. Immunogenetics. 2005;56:683–95. https://doi.org/10.1007/s00251-004-0717-7.
15. O'Connor EA, Strandh M, Hasselquist D, Nilsson J-Å, Westerdahl H. The evolution of highly variable immunity genes across a passerine bird radiation. Mol Ecol. 2016;25:977–89. https://doi.org/10.1111/mec.13530.
16. Bernatchez L, Landry C. MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? J Evol Biol. 2003;16:363–77. https://doi.org/10.1046/j.1420-9101.2003.00531.x.

17. Spurgin LG, Richardson DS. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. Proc R Soc B Biol Sci. 2010;277:979–88. https://doi.org/10.1098/rspb.2009.2084.
18. Cammen K, Hoffman JI, Knapp LA, Harwood J, Amos W. Geographic variation of the major histocompatibility complex in Eastern Atlantic grey seals (*Halichoerus grypus*). Mol Ecol. 2011;20:740–52. https://doi.org/10.1111/j.1365-294X.2010.04975.x.
19. Kyle CJ, Rico Y, Castillo S, Srithayakumar V, Cullingham CI, White BN, et al. Spatial patterns of neutral and functional genetic variations reveal patterns of local adaptation in raccoon (*Procyon lotor*) populations exposed to raccoon rabies. Mol Ecol. 2014;23:2287–98. https://doi.org/10.1111/mec.12726.
20. Larson WA, Seeb JE, Dann TH, Schindler DE, Seeb LW. Signals of heterogeneous selection at an MHC locus in geographically proximate ecotypes of sockeye salmon. Mol Ecol. 2014;23:5448–61. https://doi.org/10.1111/mec.12949.
21. Miller KM, Kaukinen KH, Schulze AD. Expansion and contraction of major histocompatibility complex genes: a teleostean example. Immunogenetics. 2002;53:941–63. https://doi.org/10.1007/s00251-001-0398-4.
22. Minias P, Podlaszczuk P, Indykiewicz P, Ledwoń M, Nowakowski J, Chyb A, et al. Genetic variation at innate and adaptive immune genes – contrasting patterns of differentiation and local adaptation in a wild gull. Heredity (Edinb). 2023:1–10. https://doi.org/10.1038/s41437-023-00645-2.
23. Talarico L, Babik W, Marta S, Pietrocini V, Mattoccia M. MHC structuring and divergent allele advantage in a urodele amphibian: a hierarchical multi-scale approach. Heredity (Edinb). 2019;123:593–607. https://doi.org/10.1038/s41437-019-0221-3.
24. Eizaguirre C, Lenz TL, Kalbe M, Milinski M. Divergent selection on locally adapted major histocompatibility complex immune genes experimentally proven in the field. Ecol Lett. 2012;15:723–31. https://doi.org/10.1111/j.1461-0248.2012.01791.x.
25. Pavey SA, Sevellec M, Adam W, Normandeau E, Lamaze FC, Gagnaire P-A, et al. Nonparallelism in MHCIIβ diversity accompanies nonparallelism in pathogen infection of lake whitefish (*Coregonus clupeaformis*) species pairs as revealed by next-generation sequencing. Mol Ecol. 2013;22:3833–49. https://doi.org/10.1111/mec.12358.
26. Peng F, Ballare KM, Hollis Woodard S, Haan S, Bolnick DI. What evolutionary processes maintain MHC IIβ diversity within and among populations of stickleback? Mol Ecol. 2021;30:1659–71. https://doi.org/10.1111/mec.15840.
27. Stutz WE, Bolnick DI. Natural selection on MHC IIβ in parapatric lake and stream stickleback: balancing, divergent, both or neither? Mol Ecol. 2017;26:4772–86. https://doi.org/10.1111/mec.14158.
28. Bonneaud C, Pérez-Tris J, Federici P, Chastel O, Sorci G. Major histocompatibiliy alleles associated with local resistance to malaria in a passerine. Evolution (N Y). 2006;60:383–9. https://doi.org/10.1111/j.0014-3820.2006.tb01114.x.
29. Eizaguirre C, Lenz TL, Kalbe M, Milinski M. Rapid and adaptive evolution of MHC genes under parasite selection in experimental vertebrate populations. Nat Commun. 2012;3:621. https://doi.org/10.1038/ncomms1632.
30. Schmid DW, Meyer M, Wilhelm K, Tilley T, Link-Hessing T, Fleischer R, et al. MHC class II genes mediate susceptibility and resistance to coronavirus infections in bats. Mol Ecol. 2023;00:1–14. https://doi.org/10.1111/mec.16983.
31. Sepil I, Lachish S, Hinks AE, Sheldon BC. Mhc supertypes confer both qualitative and quantitative resistance to avian malaria infections in a wild bird population. Proc R Soc B Biol Sci. 2013;280:20130134. https://doi.org/10.1098/rspb.2013.0134.
32. Handunnetthi L, Ramagopalan SV, Ebers GC, Knight JC. Regulation of major histocompatibility complex class II gene expression, genetic variation and disease. Genes Immun. 2010;11:99–112. https://doi.org/10.1038/gene.2009.83.
33. Hofmann MJ, Bracamonte SE, Eizaguirre C, Barluenga M. Molecular characterization of MHC class IIB genes of sympatric Neotropical cichlids. BMC Genet. 2017;18:15. https://doi.org/10.1186/s12863-017-0474-x.
34. Bracamonte SE, Baltazar-Soares M, Eizaguirre C. Characterization of MHC class II genes in the critically endangered European eel (*Anguilla anguilla*). Conserv Genet Resour. 2015;7:859–70. https://doi.org/10.1007/s12686-015-0501-z.
35. Gaczorek TS, Marszałek M, Dudek K, Arntzen JW, Wielstra B, Babik W. Interspecific introgression of MHC genes in *Triturus* newts: evidence from multiple contact zones. Mol Ecol. 2023;32:867–80. https://doi.org/10.1111/mec.16804.
36. Sin YW, Dugdale HL, Newman C, Macdonald DW, Burke T. MHC class II genes in the European badger (*Meles meles*): characterization, patterns of variation, and transcription analysis. Immunogenetics. 2012;64:313–27. https://doi.org/10.1007/s00251-011-0578-9.
37. Gerdol M, Lucente D, Buonocore F, Poerio E, Scapigliati G, Mattiucci S, et al. Molecular and structural characterization of MHC class II β genes reveals high diversity in the cold-adapted icefish *Chionodraco hamatus*. Sci Rep. 2019;9:5523. https://doi.org/10.1038/s41598-019-42003-5.
38. Harstad H, Lukacs MF, Bakke HG, Grimholt U. Multiple expressed MHC class II loci in salmonids; details of one non-classical region in Atlantic salmon (*Salmo salar*). BMC Genomics. 2008;9:193. https://doi.org/10.1186/1471-2164-9-193.
39. Dijkstra JM, Grimholt U, Leong J, Koop BF, Hashimoto K. Comprehensive analysis of MHC class II genes in teleost fish genomes reveals dispensability of the peptide-loading DM system in a large part of vertebrates. BMC Evol Biol. 2013;13:260. https://doi.org/10.1186/1471-2148-13-260.

40. Schwensow N, Castro-Prieto A, Wachter B, Sommer S. Immunological MHC supertypes and allelic expression: how low is the functional MHC diversity in free-ranging Namibian cheetahs? Conserv Genet. 2019;20:65–80. https://doi.org/10.1007/s10592-019-01143-x.
41. Axtner J, Sommer S. The functional importance of sequence versus expression variability of MHC alleles in parasite resistance. Genetica. 2012;140:407–20. https://doi.org/10.1007/s10709-012-9689-y.
42. Wegner KM, Kalbe M, Rauch G, Kurtz J, Schaschl H, Reusch TBH. Genetic variation in MHC class II expression and interactions with MHC sequence polymorphism in three-spined sticklebacks. Mol Ecol. 2006;15:1153–64. https://doi.org/10.1111/j.1365-294X.2006.02855.x.
43. Drews A, Strandh M, Råberg L, Westerdahl H. Expression and phylogenetic analyses reveal paralogous lineages of putatively classical and non-classical MHC-I genes in three sparrow species (*Passer*). BMC Evol Biol. 2017;17:152. https://doi.org/10.1186/s12862-017-0970-7.
44. Drews A, Westerdahl H. Not all birds have a single dominantly expressed MHC-I gene: transcription suggests that siskins have many highly expressed MHC-I genes. Sci Rep. 2019;9:19506. https://doi.org/10.1038/s41598-019-55800-9.
45. Watson H, Drews A, Skogsmyr KH, Neto JM, Roved J, Westerdahl H. Only rare classical MHC-I alleles are highly expressed in the European house sparrow. Proc R Soc B Biol Sci. 2024. https://doi.org/10.1098/rspb.2023.2857.
46. Grimholt U. MHC and evolution in teleosts. Biology (Basel). 2016;5:6. https://doi.org/10.3390/biology5010006.
47. Salzburger W. Understanding explosive diversification through cichlid fish genomics. Nat Rev Genet. 2018;19:705–17. https://doi.org/10.1038/s41576-018-0043-9.
48. Seehausen O, Wagner CE. Speciation in freshwater fishes. Annu Rev Ecol Evol Syst. 2014;45:621–51. https://doi.org/10.1146/annurev-ecolsys-120213-091818.
49. Barluenga M, Meyer A. Phylogeography, colonization and population history of the Midas cichlid species complex (*Amphilophus* spp.) in the Nicaraguan crater lakes. BMC Evol Biol. 2010;10:326. https://doi.org/10.1186/1471-2148-10-326.
50. Kautt AF, Kratochwil CF, Nater A, Machado-Schiaffino G, Olave M, Henning F, et al. Contrasting signatures of genomic divergence during sympatric speciation. Nature. 2020;588:106–11. https://doi.org/10.1038/s41586-020-2845-0.
51. Matschiner M, Böhne A, Ronco F, Salzburger W. The genomic timeline of cichlid fish diversification across continents. Nat Commun. 2020;11:5895. https://doi.org/10.1038/s41467-020-17827-9.
52. Ronco F, Matschiner M, Böhne A, Boila A, Büscher HH, El Taher A, et al. Drivers and dynamics of a massive adaptive radiation in cichlid fishes. Nature. 2021;589:76–81. https://doi.org/10.1038/s41586-020-2930-4.
53. Salzburger W. The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. Mol Ecol. 2009;18:169–85. https://doi.org/10.1111/j.1365-294X.2008.03981.x.
54. Blais J, Rico C, van Oosterhout C, Cable J, Turner GF, Bernatchez L. MHC adaptive divergence between closely related and sympatric African cichlids. PLoS ONE. 2007;2:e734. https://doi.org/10.1371/journal.pone.0000734.
55. Hablützel PI, Vanhove MPM, Deschepper P, Grégoir AF, Roose AK, Volckaert FAM, et al. Parasite escape through trophic specialization in a species flock. J Evol Biol. 2017;30:1437–45. https://doi.org/10.1111/jeb.13111.
56. Karvonen A, Wagner CE, Selz OM, Seehausen O. Divergent parasite infections in sympatric cichlid species in Lake Victoria. J Evol Biol. 2018;31:1313–29. https://doi.org/10.1111/jeb.13304.
57. Maan ME, Van Rooijen AMC, Van Alphen JJM, Seehausen O. Parasite-mediated sexual selection and species divergence in Lake Victoria cichlid fish. Biol J Lin Soc. 2008;94:53–60. https://doi.org/10.1111/j.1095-8312.2008.00989.x.
58. Málaga-Trillo E, Zaleska-Rutczynska Z, McAndrew B, Vincek V, Figueroa F, Sültmann H, et al. Linkage relationships and haplotype polymorphism among cichlid Mhc class II B loci. Genetics. 1998;149:1527–37. https://doi.org/10.1093/genetics/149.3.1527.
59. Bracamonte SE, Hofmann MJ, Lozano-Martín C, Eizaguirre C, Barluenga M. Divergent and non-parallel evolution of MHC IIB in the Neotropical Midas cichlid species complex. BMC Ecol Evol. 2022;22:41. https://doi.org/10.1186/s12862-022-01997-9.
60. Hablützel PI, Grégoir AF, Vanhove MPM, Volckaert FAM, Raeymaekers JAM. Weak link between dispersal and parasite community differentiation or immunogenetic divergence in two sympatric cichlid fishes. Mol Ecol. 2016;25:5451–66. https://doi.org/10.1111/mec.13833.
61. Hablützel PI, Volckaert FAM, Hellemans B, Raeymaekers JAM. Differential modes of MHC class IIB gene evolution in cichlid fishes. Immunogenetics. 2013;65:795–809. https://doi.org/10.1007/s00251-013-0725-6.
62. Lozano-Martín C, Bracamonte SE, Barluenga M. Evolution of MHC IIB diversity across cichlid fish radiations. Genome Biol Evol. 2023;15. https://doi.org/10.1093/gbe/evad110.
63. Meyer BS, Hablützel PI, Roose AK, Hofmann MJ, Salzburger W, Raeymaekers JAM. An exploration of the links between parasites, trophic ecology, morphology, and immunogenetics in the Lake Tanganyika cichlid radiation. Hydrobiologia. 2019;832:215–33. https://doi.org/10.1007/s10750-018-3798-2.
64. Illumina. bcl2fastq2 Conversion Software. 2017.
65. Sebastian A, Herdegen M, Migalska M, Radwan J. Amplisas: a web server for multilocus genotyping using next-generation amplicon sequencing data. Mol Ecol Resour. 2016;16:498–510. https://doi.org/10.1111/1755-0998.12453.

66. Bushnell B, Rood J, Singer E. BBMerge – accurate paired shotgun read merging via overlap. PLoS ONE. 2017;12:e0185056. https://doi.org/10.1371/journal.pone.0185056.
67. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9. https://doi.org/10.1038/nmeth.1923.
68. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9. https://doi.org/10.1093/bioinformatics/btp352.
69. Liao Y, Smyth GK, Shi W. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30:923–30. https://doi.org/10.1093/bioinformatics/btt656.
70. Liao Y, Smyth GK, Shi W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. Nucleic Acids Res. 2019;47:e47. https://doi.org/10.1093/nar/gkz114.
71. El Taher A, Böhne A, Boileau N, Ronco F, Indermaur A, Widmer L, et al. Gene expression dynamics during rapid organismal diversification in African cichlid fishes. Nat Ecol Evol. 2020;5:243–50. https://doi.org/10.1038/s41559-020-01354-3.
72. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37:907–15. https://doi.org/10.1038/s41587-019-0201-4.
73. Putri GH, Anders S, Pyl PT, Pimanda JE, Zanini F. Analysing high-throughput sequencing data in Python with HTSeq 2.0. Bioinformatics. 2022;38:2943–5. https://doi.org/10.1093/bioinformatics/btac166.
74. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37:1530–4. https://doi.org/10.1093/molbev/msaa015.
75. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. Modelfinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14:587–9. https://doi.org/10.1038/nmeth.4285.
76. Chernomor O, von Haeseler A, Minh BQ. Terrace aware data structure for phylogenomic inference from supermatrices. Syst Biol. 2016;65:997–1008. https://doi.org/10.1093/sysbio/syw037.
77. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol. 2018;35:518–22. https://doi.org/10.1093/molbev/msx281.
78. Pond SLK, Poon AFY, Velazquez R, Weaver S, Hepler NL, Murrell B, et al. HyPhy 2.5—a customizable platform for evolutionary hypothesis testing using phylogenies. Mol Biol Evol. 2020;37:295–9. https://doi.org/10.1093/molbev/msz197.
79. Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Pond SLK. Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. Mol Biol Evol. 2018;35:773–7. https://doi.org/10.1093/molbev/msx335.
80. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24:1586–91. https://doi.org/10.1093/molbev/msm088.
81. R Core Team. R: A language and environment for statistical computing. 2025.
82. Yu G, Smith DK, Zhu H, Guan Y, Lam TT. Ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol. 2017;8:28–36. https://doi.org/10.1111/2041-210X.12628.
83. Brooks M, Kristensen K, Benthem K, Magnusson A, Berg C, Nielsen A, et al. GlmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. R J. 2017;9:378. https://doi.org/10.32614/RJ-2017-066.
84. Hartig F. DHARMa: Residual diagnostics for hierarchical (multi-level / mixed) regression models. CRAN: Contributed Packages 2016. https://doi.org/10.32614/CRAN.package.DHARMa.
85. Lenth R V., Piaskowski J. emmeans: Estimated marginal means, aka least-squares means. CRAN: Contributed Packages 2017. https://doi.org/10.32614/CRAN.package.emmeans.
86. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550. https://doi.org/10.1186/s13059-014-0550-8.
87. Wallny H-J, Avila D, Hunt LG, Powell TJ, Riegert P, Salomonsen J, et al. Peptide motifs of the single dominantly expressed class I molecule explain the striking MHC-determined response to Rous sarcoma virus in chickens. Proc Natl Acad Sci U S A. 2006;103:1434–9. https://doi.org/10.1073/pnas.0507386103.
88. St. Pierre CL, Macias-Velasco JF, Wayhart JP, Yin L, Semenkovich CF, Lawson HA. Genetic, epigenetic, and environmental mechanisms govern allele-specific gene expression. Genome Res. 2022;32:1042–57. https://doi.org/10.1101/gr.276193.121.
89. Chappell PE, Meziane EK, Harrison M, Magiera Ł, Hermann C, Mears L, et al. Expression levels of MHC class I molecules are inversely correlated with promiscuity of peptide binding. Elife. 2015. https://doi.org/10.7554/eLife.05345.
90. Kaufman J. Generalists and specialists: a new view of how MHC class I molecules fight infectious pathogens. Trends Immunol. 2018;39:367–79. https://doi.org/10.1016/j.it.2018.01.001.
91. Morris A, Hewitt C, Young S. The major histocompatibility complex: its genes and their roles in antigen presentation. Mol Aspects Med. 1994;15:377–503. https://doi.org/10.1016/0098-2997(94)90041-8.

92. Dijkstra JM, Grimholt U. Major histocompatibility complex (MHC) fragment numbers alone - in Atlantic cod and in general - do not represent functional variability. F1000Res. 2018;7:963. https://doi.org/10.12688/f1000research.15386.1.
93. Sato A, Dongak R, Hao L, Shintani S, Sato T. Organization of MHC class II A and B genes in the tilapiine fish *Oreochromis*. Immunogenetics. 2012;64:679–90. https://doi.org/10.1007/s00251-012-0618-0.
94. Edwards MC, Gibbs RA. Multiplex PCR: advantages, development, and applications. Genome Res. (n.d.);3:S65–75.
95. Markoulatos P, Siafakas N, Moncany M. Multiplex polymerase chain reaction: a practical approach. J Clin Lab Anal. 2002;16:47–51. https://doi.org/10.1002/jcla.2058.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Seraina E. Bracamonte[1,2] · Pascal I. Hablützel[3,4] · Carlos Lozano-Martín[1] · Marta Barluenga[1]

✉ Seraina E. Bracamonte
seraina.bracamonte@ebc.uu.se

✉ Marta Barluenga
marta.barluenga@mncn.csic.es

Pascal I. Hablützel
pascal.hablutzel@vliz.be

Carlos Lozano-Martín
clmvg2612@gmail.com

[1] Museo Nacional de Ciencias Naturales, CSIC, José Gutiérrez Abascal 2, Madrid 28006, Spain

[2] Department of Ecology and Genetics, Uppsala University, Norbyvägen 18D, Uppsala 75236, Sweden

[3] Flanders Marine Institute, Jacobsenstraat 1, Ostend 8400, Belgium

[4] Vrije Universiteit Brussel, Pleinlaan 2, Brussels 1050, Belgium