

RESEARCH ARTICLE

Open Access



Molecular characterization of MHC class IIB genes of sympatric Neotropical cichlids

Melinda J. Hofmann¹, Seraina E. Bracamonte^{2,3}, Christophe Eizaguirre^{2,4} and Marta Barluenga^{1*}

Abstract

Background: The Major Histocompatibility Complex (MHC) is a key component of the adaptive immune system of all vertebrates and consists of the most polymorphic genes known to date. Due to this complexity, however, MHC remains to be characterized in many species including any Neotropical cichlid fish. Neotropical crater lake cichlids are ideal models to study evolutionary processes as they display one of the most convincing examples of sympatric and repeated parallel radiation events within and among isolated crater lakes.

Results: Here, we characterized the genes of MHC class IIB chain of the Midas cichlid species complex (*Amphilophus* cf. *citrinellus*) including fish from five lakes in Nicaragua. We designed 19 new specific primers anchored in a stepwise fashion in order to detect all alleles present. We obtained 866 genomic DNA (gDNA) sequences from thirteen individuals and 756 additional sequences from complementary DNA (cDNA) of seven of those individuals. We identified 69 distinct alleles with up to 25 alleles per individual. We also found considerable intron length variation and mismatches of alleles detected in cDNA and gDNA suggesting that some loci have undergone pseudogenization. Lastly, we created a model of protein structure homology for each allele and identified their key structural components.

Conclusions: Overall, the Midas cichlid has one of the most diverse repertoires of MHC class IIB genes known, which could serve as a powerful tool to elucidate the process of divergent radiations, colonization and speciation in sympatry.

Keywords: Major Histocompatibility Complex, Sympatric, Neotropical, Midas cichlid fish, *Amphilophus*

Background

The Major Histocompatibility Complex (MHC) is a key component of the adaptive immune system of all jawed vertebrates [1, 2]. The function of the MHC molecules is to present short self and non-self peptides often derived from parasites and pathogens for recognition by T-lymphocytes [3]. This sets off the cascade of targeted immune defenses against those specific parasites and pathogens. MHC also plays a role in establishing a memory to rapidly eliminate those agents in case of future encounters [3]. MHC molecules are encoded by the most polymorphic genes in all jawed vertebrates, and most species have different number of loci that are co-dominantly expressed (e.g. [4, 5]). There are two classical antigen presenting MHC molecules, MHC class I, that is expressed on all nucleated

cells and elicits a response against intracellular parasites, and MHC class II, that is only expressed on antigen-presenting cells (macrophages, B-cells and dendritic cells), which actively engulf and process inter-cellular parasites [3]. Here, we focus on MHC class II which is composed of two chains, α and β , which together form the peptide-binding groove [4]. The peptide-binding region of the β chain is the most polymorphic and hence the most studied region of the MHC. In general, the MHC IIB region is divided into 5 to 6 exons with varying intron lengths depending on species and haplotypes [5–7].

The highly polymorphic multigene nature of MHC causes some technical difficulties when trying to simultaneously detect all alleles, particularly those that are rare in the target population. Cloning and Sanger sequencing have associated PCR-based errors and PCR amplification biases [8–10], making accurate amplification a laborious and costly process. Next generation sequencing technologies have, to some extent, facilitated population level

* Correspondence: marta.barluenga@mncn.csic.es

¹Museo Nacional de Ciencias Naturales, CSIC, José Gutiérrezz Abascal, 2, 28006 Madrid, Spain

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



studies of MHC, although those new techniques tend to overestimate allelic diversity [11, 12]. Overcoming those challenges allows the use of MHC as a powerful tool to study biodiversity [13, 14], disease dynamics [15], evolutionary processes [16, 17], and even to estimate the number of founders of a population [18].

Cichlid fish are excellent model systems to study evolutionary processes since they demonstrate some of the most extreme examples of explosive adaptive radiations (e.g. [19–22]). They are some of the most species-rich families of freshwater fishes worldwide, and their hotspots of diversification are the great lakes of East Africa. They are also present in Central and South America [23, 24]. Particularly, the Neotropical Midas cichlid species complex (*Amphilophus spp.*) is a valuable model system for the study of recent speciation [25–27]. This group not only comprises one of the most compelling examples of sympatric speciation [28], but also recent independent colonization events and in situ rapid diversification [29], which makes it an excellent natural experiment of adaptation and incipient speciation [25].

Many studies have attributed cichlid's rapid speciation events to various factors, including phenotypic plasticity [30], reproductive behavior and local adaptation [31, 32], and even genomic processes [33–35]. It has been suggested that the mechanism of adaptive speciation in general, and in sympatry in particular, may result from a pleiotropic role of the MHC in co-evolutionary dynamics of local host-parasites and odor-mediated mate choice ultimately leading to reproductive isolation [14, 36, 37]. Here, we characterized the β chain of the MHC class II in the Neotropical Midas cichlid species complex to establish the baseline for evaluating the role of parasites and immune system in sympatric speciation.

A striking characteristic of MHC polymorphism is the occurrence of similar alleles in related species, known as trans-species polymorphism (TSP) [38]. This similarity might have arisen by convergence [39, 40], although a more commonly accepted idea is that this polymorphism is maintained, mostly by balancing selection, beyond the species formation [41]. This polymorphism transmitted through several speciation phases can be a useful tool to study speciation itself [41]. TSP has been found to occur across related species of reptiles [42], mammals [43], amphibians [44], birds [45, 46], and fish [47]. In this study we also characterize events of TSP.

There is some knowledge about MHC diversity patterns of cichlid fish [37, 48–51], but this comes exclusively from African species. Some studies have focused on the diversity of MHC class I in cichlids from Lake Victoria, finding many common alleles across species [52]. Other studies have found high diversity of MHC class IIB alleles in different species of Lake Malawi cichlids [48, 53]. A population genetic analysis on MHC class IIB of Lake Malawi

cichlids even suggested that adaptive divergence at this locus could be linked to speciation in cichlids [36].

Old and New World cichlids have been geographically separated for a very long time [54–56], therefore MHC evolution in Neotropical cichlids is likely to have followed its own evolutionary trajectory. Therefore, MHC has to be characterized *de novo* in the Midas cichlid in order to understand its role in their adaptation and speciation. We first sequenced exons 1–6 of the MHC IIB and described intron and exon conformation as well as most intron length variability. Then we used both genomic (gDNA) and expressed transcripts (mRNA) to characterize the allelic diversity existing in exon 2 – that which encodes for the peptide binding groove. We then tested for various modes of evolution of the MHC and modeled the tertiary structure of each detected allele to identify the structural components of the MHC molecules.

Methods

Sampling, DNA and RNA isolation

Sampling of Midas cichlid fish took place in several Nicaraguan lakes (Fig. 1). Adult fish were captured using gill nets (collection permit number 001-012012), anesthetized with MS 222 following standard procedures and euthanized on ice before processing. Fin tissues of 13 randomly selected individuals were preserved in 100% ethanol at 4 °C. Those 13 samples represent a good portion of the diversity of this species complex (Fig. 1, Additional file 1: Table S1). Additional spleen tissue samples of 7 of these individuals was preserved in RNAlater® (Qiagen, Hilden, Germany) and stored at -80 °C.

Total genomic DNA (gDNA) was extracted using DNeasy spin columns for Blood and Tissue Kit® (Qiagen, Hilden, Germany) according to the manufacturer's protocol, with the addition of RNase. DNA was quantified using Nanodrop 1000 (ThermoFisher Scientific, Bonn) and standardized to a concentration of 20 ng / μ l. RNA was extracted with Invitrap Spin tissue RNA mini kit® (Berlin, Germany) and the reverse transcription performed with the QuantiTect® Reverse Transcription kit (Qiagen, Hilden, Germany).

Primers design

Firstly, to characterize the allelic diversity in the exon 2 of the MHC IIB gene in the Midas cichlid, we retrieved MHC sequences from GenBank choosing different sequences from related fish species with well-characterized MHC genes. We aligned the sequences (see Additional file 1: Table S2) and designed a reverse primer (MHC_Rev3 "GATCTGTTTGGGGTAGAAGTCG") located in the most conserved region in the middle of exon 3. We did a PCR by pairing this reverse primer with two published forward primers designed for sticklebacks located in conserved upstream regions of exon 2



Fig. 1 Map of the Pacific coast of Nicaragua (Central America) with the great lakes and several crater lakes where samples were collected (source <http://photojournal.jpl.nasa.gov/catalog/PIA03364>; wikimedia.org). Samples belonged to the four species *Amphilophus citrinellus*, *A. labiatus*, *A. amarillo* and *A. xiloaensis*

(SatoGa11_mod1 [57], GAIEx2startF [58]). Using the resulting sequences we designed 14 additional primers in a stepwise manner. We designed new primers in adjacent regions of the sequence, considered sets of new amplicons, and designed additional primers on the new sequences maximizing allele amplification (Table 1). We paired all forward and reverse primers. Additionally, we specifically designed 4 primers (AcMHCIIF3, AcMHCIIF4, AcMHCIIF5, and AcMHCIIF6) to discriminate a group of particularly abundant alleles that were not very variable (later referred to as Group I), which could hinder the detection of rarer alleles.

PCR Amplification and gel extraction

PCR amplifications were performed following the recommendations of Lenz and Becker [10] in order to reduce PCR artifacts, common in the amplification of multigene families. We used Taq Polymerase with no proof-reading capacity, extended elongation times, excess of primers to avoid incomplete amplicons acting as heteroduplex primers, and duplicated reactions. However, we did not reduce PCR cycles or do a reconditioning PCR since under those recommendations our bands were too weak for cloning. Each amplification was done in two independent reactions, each consisting of 2 μ l 10x Dreamtaq Buffer, 1 μ l dNTP's (10 mM), 2 μ l of each primer (5 pmol / μ l), 0.2 μ l Taq Polymerase (Dreamtaq[®]) and 2 μ l of template DNA in a total volume of 20 μ l. The thermal profile started with an initial denaturation step of 95 °C for 3 min, followed by 30 cycles of denaturation at 94 °C for 30 s, annealing at specific temperature for each primer

pair (ranging from 44 to 64 °C, Table 1) for 1 min, elongation at 72 °C for 1 min, with a final elongation at 72 °C for 10 min. The PCR reactions for each individual and primer pair were then pooled and loaded into a 2% agarose gel and run at 40 V for 4 h. Gels were then stained with Ethidium bromide to visualize bands. The bands of interest were cut and extracted with NucleoSpin Gel and PCR Clean-up[®] (MACHEREY-NAGEL GmbH & Co. KG) for further cloning.

Cloning

PCR amplicons were cloned with the Qiagen PCR cloning Kit[®] (Qiagen, Hilden, Germany). Cloning followed the protocol described in Bracamonte et al. [59]. For clone screening, 1 μ l of the denatured clones was used as template for a PCR with the universal M13 primers: M13_Funi (5'ACGACGTTGTAAAACGACGGCCAG 3'), and M13_RP15 (5'TTCACACAGGAAACAGCTATGACC 3'). The reaction had a final volume of 10 μ l, included 1 μ l 10x Dreamtaq Buffer, 0.5 μ l dNTPs (10 mM), 1 μ l of each primer (5 pmol / μ l), 0.1 μ l Taq Polymerase (Dreamtaq[®]), and ran using the following thermal profile: initial denaturing step at 95 °C for 1 min, followed by 25 cycles of denaturing at 96 °C for 10 s, annealing at 50 °C for 10 s, elongation at 72 °C for 1 min, with a final elongation at 72 °C for 7 min. Two μ l of this PCR product were then loaded in a 1% agarose gel and run for 30 min at 90 V, and ultimately stained with Ethidium bromide to visualize bands. We sequenced the clones that were positive for the amplicon. We sequenced between 16 and 48 clones per amplification in order to detect rare alleles.

Table 1 Primer combinations used for amplification of MHC IIB

Forward	Primer Reverse Primer	TA	Fragment
SatoGa11_mod1 AACTCCACKGAKCTGAAGRAC	MHC Rev3 GATCTGTTGGGGTAGAAGTCG	59 °C	E2 - E3
GAllEx2startF GTCITTAACCTCCACGGAGCTGAAGG	MHC Rev3 GATCTGTTGGGGTAGAAGTCG	55 °C	E2 - E3
GAllEx2startF GTCITTAACCTCCACGGAGCTGAAGG	AcMHCII BR1 GGRGTGAAGTCTGACTRATGG	59 °C	E2 - E3
AaMHCII BE1F1b ATGGCTYCATCCTTYMTC	MHC Rev3 GATCTGTTGGGGTAGAAGTCG	56 °C	E1 - E3
AaMHCII BE1F1b ATGGCTYCATCCTTYMTC	AcMHCII BR4 ACCCAGGATCAGTCCTGAGG	59 °C	E1 - E5
AcMHCII BF2 (excludes group I) TTAACTCCACKGAGCTGAASGACA	AcMHCII BR3 (excludes gop I) GAYGATGAAYCATAACTCACCTGAT	64 °C	E2 - E2/I2
AcMHCII BF3 (only group I) TCAGGTGAGTYATGDITTCATC	AcMHCII BR4 ACCCAGGATCAGTCCTGAGG	46 °C	E2/I2 - E5
AcMHCII BF3 (only group I) TCAGGTGAGTYATGDITTCATC	AcMHCII BR5 TTCCTCTGTAGTAGATGAATCC	44 °C	E2/I2 - R5
AcMHCII BF4 (excludes goup I) TCAGGTGAGTCTGTTTCTGTG	AcMHCII BR4 ACCCAGGATCAGTCCTGAGG	54 °C	E2/I2 - E5
AcMHCII BF5 (excludes group I) CCACKGAGCTGAASGACATSGAG	MHC Rev3 GATCTGTTGGGGTAGAAGTCG	55 °C	E2 - E3
AcMHCII BF5 (excludes group I) CCACKGAGCTGAASGACATSGAG	AcMHCII BR4 ACCCAGGATCAGTCCTGAGG	59 °C	E2 - E5
AcMHCII BF7 CGAGTACGTTTCGATCTTTGTATTGC	AcMHCII BR9 DCTGATTTAGTCAGAGCAGTCT	50 °C	E2 - E2
AcMHCII BF8 CGAGTWCATCARVICTTACTAYTWC	AcMHCII BR9 DCTGATTTAGTCAGAGCAGTCT	55 °C	E2 - E2
AcMHCII BF9 GAAACCTGTTACAGCAGTCCCTC	AcMHCII BR8 CATGTGCTACATGCAACATATCA	NA	I2 - I2
AcMHCII BF6 CCACTGAGCTGAASGACATSGAG	MHC Rev3 GATCTGTTGGGGTAGAAGTCG	55 °C	E2 - E3

Fragment indicates the region of the gene amplified: 'E' exon, 'I' intron, '/' exon-intron boundary. The primers AcMHCII BF9 and AcMHCII BR8 were used only to sequence through clones with long intron 2

Sequencing

Cycle sequencing was done using the Big Dye Terminator v3.1 using the Cycle Sequencing Kit (Life Technologies) following the manufacturer's protocol scaled to 10 µl total reaction volume per sample. The product was then purified using 50 µl BigDye X Terminator™ Purification Kit® mix (Life technologies, Thermo Fisher Scientific Inc). Sequencing took place on an ABI 3730 Genetic Analyzer (Life Technologies). Even though MHC sequence variants may stem from different loci and therefore may be paralogous, we will refer to them as alleles.

Identifying and naming alleles

The taxonomic status of the species within the Midas cichlid complex is under considerable debate. Although

some species have been described recently within this species complex, due to their very recent history (<50000 years) we name our alleles under the generic name "Amci" for *Amphilophus cf. citrinellus*. Sequences were aligned using CODON-CODE ALIGNER® (Codon-code Corporation 2002–2009) and analyzed with BIOEDIT v1 [60]. To avoid sequence artefacts, generally alleles were only considered true alleles if they were amplified in at least two independent PCR and cloning events (see Results for exceptions on this). We manually inspected the alignment and removed all sequences that were identified as chimeras by being partially identical to one allele group, and partially identical to another allele group and only existing in one PCR product. Sequences identified as chimeras were removed from all

further analyses. A consensus was created from all identical sequences of different lengths and this was hereafter considered an allele. All alleles were aligned with well annotated published sequences [58] and checked for stop codons within exons to further rule out the presence of pseudogenes. Since there is no complete genome for this species, identifying the family and locus that each allele belonged to is not currently possible. Alleles were named following the allele nomenclature guidelines for MHC established by Klein et al. [61]. We looked for tandem repeats in all alleles using Tandem Repeat Finder v4.09 [62].

We did a blast search of all alleles on the *Amphilophus citrinellus* draft genome shotgun sequencing project (BioProject PRJEB6974) [63] to confirm the exon structure of the MHC IIB alleles we found and to confirm intron length variability.

Alleles were assigned to groups according to estimates of evolutionary divergence between sequences. Analyses were conducted using the Maximum Composite Likelihood model [64] in MEGA v5.2 [65]. Codon positions included were 1st, 2nd, 3rd, and noncoding. All positions containing gaps and missing data were eliminated. The final dataset had a total of 201 positions. We validated the resulting groups with a randomization analysis in R [66]. The mean pairwise distances within each group was compared to a null distribution generated by a random selection of equivalent alleles repeated 999 times to calculate the p-value. The results of this analysis were contrasted with the phylogenetic reconstruction of the allele relationships.

Phylogenetic and statistical analyses

In order to determine the phylogenetic relationships between alleles, we used exon 2 and 3 to construct a phylogenetic tree using Bayesian inference with BEAST v2.0 [67]. We found the most appropriate substitution model (HKY + G) and partitioning scheme according to the Bayesian Information Criterion implemented in PARTITIONFINDER [68]. We specified the parameters in BEAUti v2.0 [67] for the BEAST-run, and the MCMCs were run for 10^9 generations sampling every 100,000 trees. We used a strict clock model and a Yule speciation prior. We inspected the traces for convergence with TRACER 1.5 (<http://tree.bio.ed.ac.uk/software/tracer/>) and checked that they were higher than 200. The 10,000 resulting trees were summarized with TREEANNOTATOR v2.1.2 (<http://beast.bio.ed.ac.uk/treeannotator>) applying a 10% burn-in. We depicted the phylogeny with the corresponding posterior values of each node with FIGTREE v1.4.2 (<http://tree.bio.ed.ac.uk/software/figtree/>).

To further analyze the relationship between the alleles we constructed a neighbor-net network with SPLITSTREE4 [69]. We calibrated the network by calculating the best

fit model using MEGA and estimated the probability of invariable sites.

To evaluate TSP we gathered previously published sequences from several fish species. We selected 11 sequences of MHC IIB of a well-studied African cichlid, Nile tilapia (*Oreochromis niloticus*), from Sato et al. [70], and 9 sequences from a search at NCBI's Genbank database with BLASTx from other cichlids and two other fish families: Sciaenidae and Sebastidae which had alleles closest to cichlid ones (Additional file 1: Table S3). We constructed a phylogeny with all the alleles (69 from the Midas cichlid and 20 from other species) using the same methodology and parameters as described above.

Tests of selection

In order to elucidate on the evolutionary history of MHC IIB of Midas cichlid, to determine if different selective pressures have shaped the different domains of the molecule, and if selection has acted differently on the different groups of alleles, we performed selection tests, both by domain and on the entire sequence, as implemented in MEGA. We acknowledge the limitations of these tests given the fact that we cannot allocate alleles to specific loci. Selection tests were performed on the groups of alleles established by sequence divergence. We tested separately the alleles of Group I since they appear to follow a different evolutionary pathway. We calculated rates of substitution (d_N and d_S), and tested for overall positive or purifying selection with a Z-test of selection for each domain separately. For alleles in Group I we applied the Pamilo-Bianchi-Li method with Kimura 2-parameter correction, and for the rest of alleles we applied the Nei-Gojobori method with Jukes-Cantor correction, in accordance with their corresponding best fit substitution model. Significance levels were estimated with 10,000 bootstrap replicates. We used the Nei-Gojobori method for calculating the "absolute" number of synonymous and non-synonymous sites since the Pamilo-Bianchi-Li method is not available for this test. We do not have the full sequence coverage for all alleles, and therefore we used a pairwise method that compares two sequences at a time and then averages over all possible comparisons. With this method we could compare all 69 allele despite some not having complete sequences.

To further evaluate the mode of evolution of each group of alleles, to identify possible peptide binding sites, and to identify potentially non-classical MHC alleles, we used site-specific tests of selection focusing on exons 2 and 3 that are the exons for which we have full sequence coverage for most alleles. We tested site-specific positive selection within each of the four groups of alleles, and across all alleles excluding Group I, due to its likely different evolutionary history. Site-specific selection was tested using CodeML in PAML v4.7 [71], assuming

different ω parameters among codons with no a priori knowledge of which class of selection (neutral, purifying, or positive) a given codon belongs to. We estimated parameters under five different codon substitution models: Beta models M7 (no positive selection), M8 (positive selection), and M8a (no positive selection and $\omega = 1$), and models M1a (nearly neutral) and M2a (positive selection) [71]. A likelihood ratio test (LRT) was performed to compare the fit of the models with and without selection. Statistical significance was determined by comparing twice the difference of log-likelihood scores ($2\Delta\ln L$) to the χ^2 distribution with degrees equal to the difference in the number of parameters between the models to be compared.

Protein tertiary structure homology models

To characterize the tertiary structure of each allele, and determine if they have the proper characteristics to allow them to fold into a potentially functional MHC molecule we built protein homology models using the Swiss-Model workspace v8.05 [72]. This method identifies structural templates from a protein data bank, aligns the target sequence to a template structure, builds a 3D model, and evaluates the quality of the model. The parameters considered were: *overall model quality* (QMEAN4) in which less negative values indicate more reliable homology models, and *global model quality estimation* (GMQE) in which the higher numbers indicate the more reliable models. We also located the four conserved cysteine residues that are essential for the stability of the β_1 and β_2 domains of MHC IIB, and analyzed the structure of the N- and C-terminal areas of the protein.

Results

Cloning, sequencing and MHC organization

For a total of 13 individuals we obtained 867 gDNA and 756 cDNA sequences, all belonging to the MHC IIB. These sequences revealed 69 distinct alleles (GenBank accession numbers: KY039442-KY039474 and KY354964-KY355011). Three of these alleles (DXB*05, DXB*1002 and DXB*13) were only represented by one single sequence while one was only represented by 2 sequences (DXB*0402), and would remain to be confirmed by an additional PCR amplification. We found an average of 12.5 (SD = 6.1) alleles per individual and a maximum of 25 (Table 2 and Fig. 2). We identified six exons and five introns, but we did not sequence all exons for all alleles (Additional file 1: Table S4). We characterized the gene from exon 1 through exon 6 focusing efforts on exon 2 where we expect most variability (Fig. 2). For fifteen alleles we amplified exon 1 and found that it has 55 bp, and includes the start codon "ATG" that codes for a methionine residue at the N-terminus. Exon 1 mainly codes for the signal peptide

formed by 16–20 amino acids. Exon 2 is the most polymorphic of all exons with up to ten amino acids per site (Additional file 1: Figure S1), and consists of 273 bp, of which we sequenced at least 231 bp for all 69 alleles. Exon 2 encodes the beta sheets and alpha helices of the β_1 domain of the mature MHC class II molecule. For most alleles we amplified intron 2 which revealed extreme length variation ranging from 200 to 2500 bp (Fig. 3). Alleles with long intron 2 contained three tandem repeats, a 42mer repeated 6.4 times, a 21mer repeated 12.6–18.6 times, and a 127mer repeated twice which contributed to their length. For exon 3 we amplified from 99 bp to 214 bp of most alleles, exon 3 is 214 bp long and is enriched in the amino acid proline, aiding in the formation of beta turns that are essential to the structure of the β_2 domain. We obtained partial sequences of intron 3, for two alleles, and a complete sequence for one allele that was 405 bp long. We obtained sequences of exon 4 for ten alleles that were 69 bp. For nine alleles we sequenced 64 bp of exon 5 which is expected to have 114 bp. Two of them (DXB*060101 and DXB*07) have additional 12 bp on the 5' end of exon 5. Introns 4 and 5 were not sequenced for any allele because of their known low variability [2]. We generated only partial sequences of exon 6, which we cannot assign to individual alleles as they do not span any allele specific polymorphisms due to the low variability of this region. The summary of the length in bp of each sequenced intron and exon is shown in Additional file 1: Table S4 with the reference of the sequence of *Oreochromis niloticus*.

To further corroborate our results, we did a BLAST search on the *A. citrinellus* shotgun genome project (BioProject PRJEB6974) and recovered three complete MHC IIB genes from contig 1595 position 54549–67334 (CCOE01001596.1), contig 1079 position 91043–98760 (CCOE01001080.1), and unplaced scaffold 125 position 28791–32030 (CCOE01001892.1), as well as several incomplete MHC IIB sequences. The three complete sequences contained all 6 exons, with the same length we describe. This data confirms that the length of exon 5 is 114 bp (of which we only retrieved partial sequences), but only contained one of the two variants we found (the short variant). Intron length variability was also represented in the genome sequences. We did not find in this data all intron variants we found in our sequences (e.g. we found intron 2 with 702 bp and 1.5 kb, but not with 248 bp), but found new intron length variants (e.g. intron 2 of contig 1595 is 10 kb). This data must be taken with caution since the shotgun sequences have not been confirmed with PCR and sequencing standard protocols employed when working with MHC. Nevertheless, this data provides valuable corroboration about intron lengths that is otherwise unavailable at this time.

Table 2 List of all of MHC IIb alleles found in the Midas cichlid (Continued)

AncI-DXB*0405	3	26	1	16	20	66
AncI-DXB*0406		1		1		2
AncI-DXB*05	1					1
AncI-DXB*060101		1	43	31	2	136
AncI-DXB*060102			1	1	42	2
AncI-DXB*0602			1	6		7
AncI-DXB*0603		5			3	8
AncI-DXB*0604	1			1	16	18
AncI-DXB*07			2			2
AncI-DXB*08			1	40		41
AncI-DXB*09		50			9	59
AncI-DXB*1001				1	25	26
AncI-DXB*1002				1		1
AncI-DXB*110101	31	42				73
AncI-DXB*110102		2				2
AncI-DXB*1102		8				8
AncI-DXB*12				19		20
AncI-DXB*13			1			1
AncI-DXB*14					38	38
AncI-DXB*15						16
AncI-DXB*16	4			34		38
AncI-DXB*17	4	18				22
AncI-DXB*18			17	44	23	134
AncI-DXB*19					9	9
AncI-DXB*20					2	2
AncI-DXB*21				2		2
AncI-DXB*2201	2	23				25
AncI-DXB*2202		2				2
AncI-DXB*23		2				2
AncI-DXB*2401		13				13
AncI-DXB*25			1		3	4
AncI-DXB*26						6
AncI-DXB*27	2					2
Total sequences	118	29	144	20	30	144
Total Alleles	17	8	18	10	9	20
			2	14	11	8
			18	80	69	109
			2	159	137	2
			103	2	2	1623
			9	14	8	69

Alleles found in 13 Midas cichlid samples (listed on the top) showing number of copies recovered per allele amplified in gDNA and cDNA in each individual. All alleles except DXB*05, DXB*1002, DXB*0402, and DXB*13 were confirmed with at least two independent PCR reactions. At the bottom we summarize total number of sequences and total number of alleles obtained

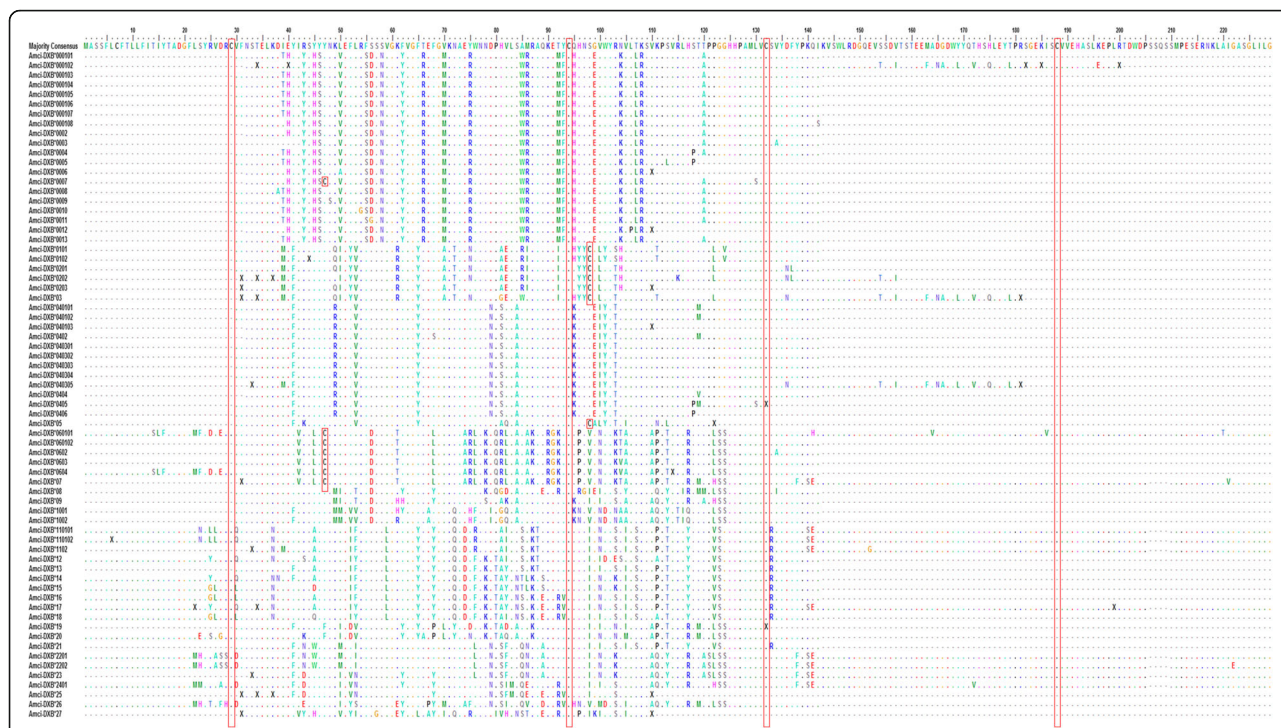


Fig. 2 Alignment of amino acid sequences of MHC IIB. A majority consensus sequence was made for comparison. ‘-’ represent identical amino acids, ‘-’ represent gaps or introns that were not sequenced. Cysteine residues are outlined in red boxes

Although in general alleles are defined as alternate sequences of a single locus, since there is no complete reference genome for this species, there is no information on how many MHC loci to expect, or how variable they may be. It must be noted that our alleles cannot be assigned to specific loci at this stage.

Grouping of alleles and Phylogenetic analyses

We constructed a phylogeny with all 69 alleles recovered from exons 2 and 3 (Additional file 1: Figure S2). We also

constructed a neighbor-net network (Fig. 4). According to estimates of evolutionary divergence between sequences, alleles grouped into five major groups (Additional file 1: Figure S3). When plotting these groups into the network one of these groups was not consistent and we left them as ungrouped (14 alleles). The remaining alleles clustered in four groups (I, II, III, IV). The results of the randomization analysis validating the groups are shown in Additional file 1: Table S5. Group I grouped 20 alleles, which were only found in gDNA. These alleles showed

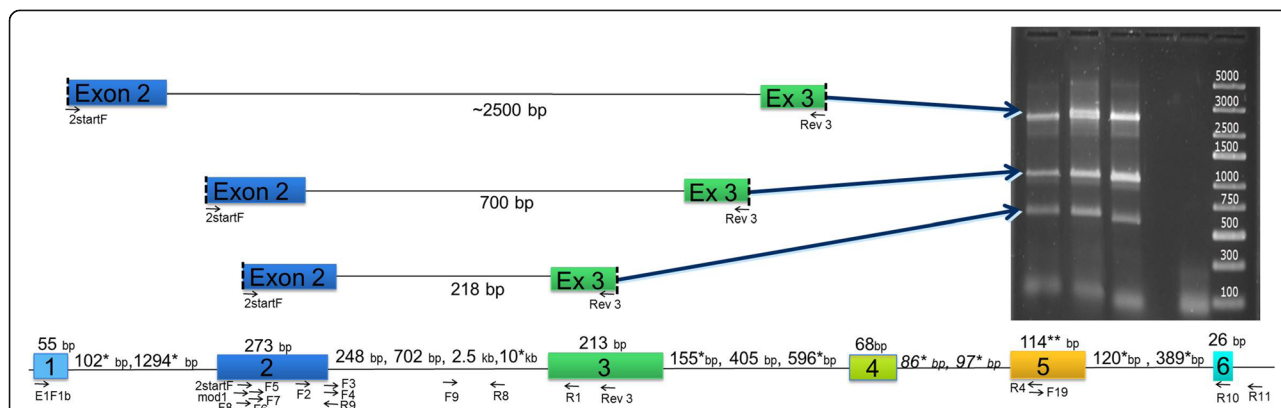


Fig. 3 Representation of intron-exon organization of MHC IIB sequences showing length of fragments and position of primers. The gel shows variation in the length of intron 2 (showing exon 2 / intron 2 / exon 3) for three individuals. * indicates lengths observed in sequences of the *Amphilophus citrinellus* draft genome shotgun sequencing project and ** indicate the exon that we found to be 12 bp longer in allele DXB*060101 and DXB*07

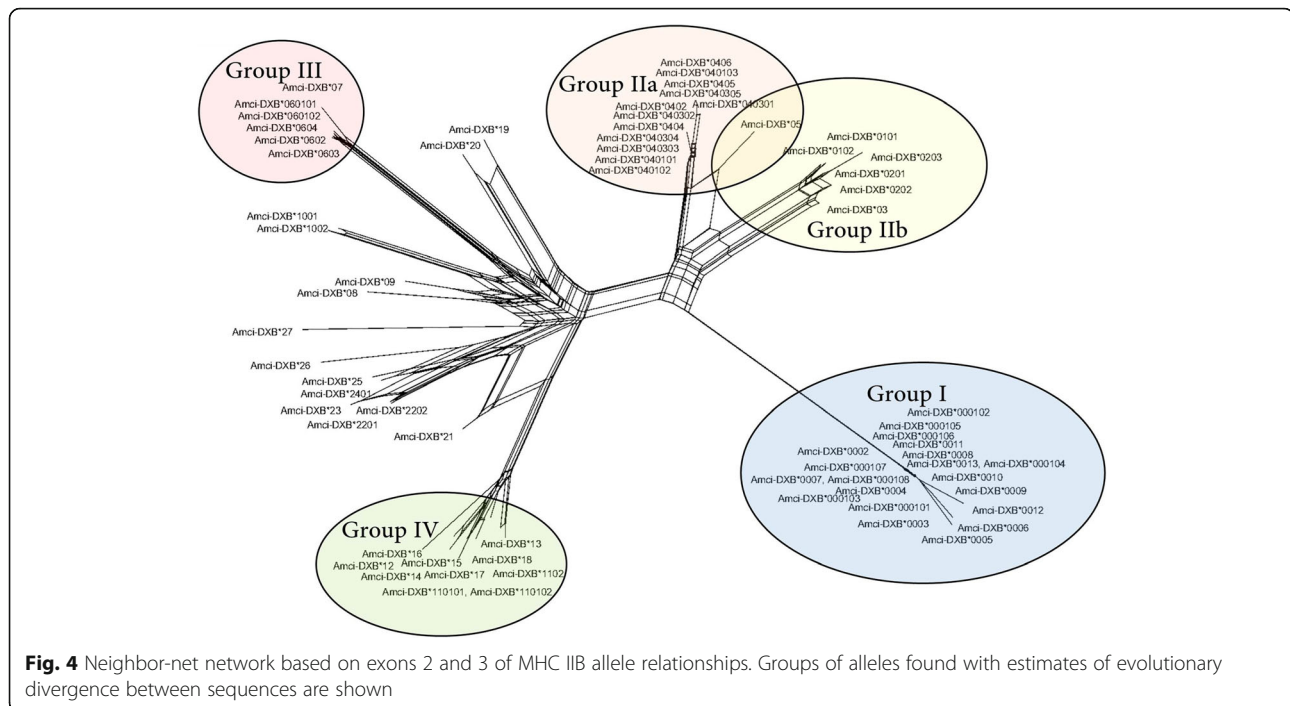


Fig. 4 Neighbor-net network based on exons 2 and 3 of MHC IIB allele relationships. Groups of alleles found with estimates of evolutionary divergence between sequences are shown

very low polymorphism and might represent non-classical MHC genes or MHC pseudogenes (see selection analyses section). Group II was further divided into two subgroups, II-a and II-b. Group II-a assembled 13 alleles, 9 found in gDNA, 1 found in cDNA, and 3 found in both cDNA and gDNA. Group II-b had 6 alleles, 5 found in gDNA and 1 found in both gDNA and cDNA. Group III assembled 6 alleles, 5 found in cDNA and 1 found in both gDNA and cDNA. Ten alleles were included in Group IV, 7 found in cDNA, and 3 in both gDNA and cDNA. From the 14 ungrouped alleles, 12 were found in cDNA, 1 in gDNA, and 1 in both cDNA and gDNA.

Taken together the Midas cichlid MHC IIB alleles belonged to two types, those alleles not found expressed in the tissue examined in this study (alleles in Group I), and those found in both gDNA and cDNA (alleles in the rest of the groups).

Trans-species polymorphism

We reconstructed the phylogenetic relationships of exon 2 MHC IIB of all alleles we retrieved in the Midas cichlid, and 20 alleles of several other fish species of cichlids and non-cichlids (*Sebastes caurinus*, *Sebastes maliger*, *Müichthys miiuy*, *Pundamilia nyererei*, *Haplochromis sp. 'rockribensis'*, *Haplochromis xenognathus*, and *Oreochromis niloticus* (see Additional file 1: Table S3)). All of these alleles clustered within the Midas cichlid alleles providing evidence for trans-species polymorphism. Nile tilapia (*Oreochromis niloticus*), the species for which there exists most MHC information, had alleles distributed throughout

the tree (Fig. 5). One of these alleles clustered with our Group I (DJB accession # AB677258.1). Several alleles from different species clustered with alleles in Group IV, and also with the ungrouped alleles.

Test of selection

The analyses of site-specific selection revealed that alleles in groups II, III and IV had sites that fitted best the model with positive selection (Group II, $p = 0.02$; groups III and IV, $p < 0.001$), while alleles in Group I did not ($p = 0.10$). Alleles in Group I did not have any synonymous substitutions, therefore d_N / d_S ratios could not be calculated. We tested all alleles including those in Group I, and found 19 positively selected sites, 13 of which were inferred at 99% level (Table 3). When we excluded alleles in Group I, 19 and 20 sites were inferred to be positively selected for models M2a and M8 respectively, 15 and 16 of them were inferred at 99% level. However, only 4 sites overlapped with those in the previous analysis. For the groups II, III and IV, we found between 1 and 9 sites to be under positive selection (Table 3). However, these sites were not shared between groups whether or not Group I was included, indicating that the groups might have followed different evolutionary trajectories.

The tests of overall selection among all alleles showed that the entire sequence (LP, β_1 and β_2 domains) of all alleles is under purifying selection ($p = 0.021$), and this pattern remains when the alleles of Group I are excluded ($p = 0.005$) (Table 4). Nine alleles were excluded from the analysis of the β_2 domain because we did not obtain sequences of this

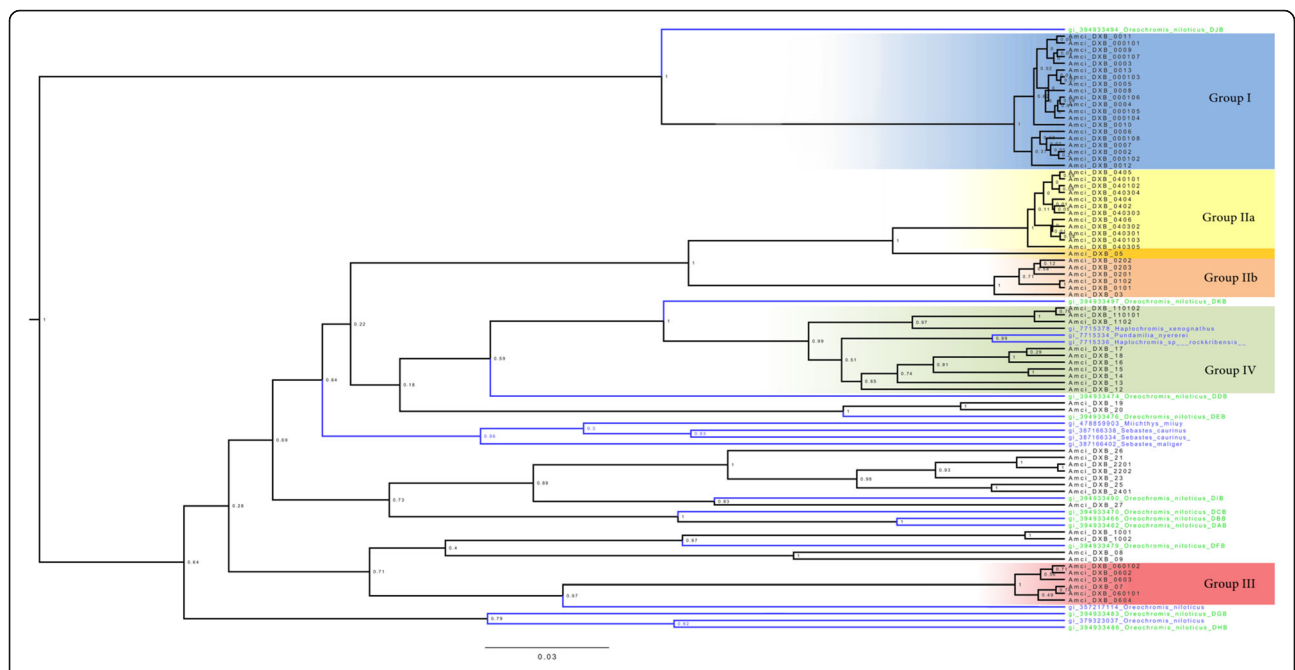


Fig. 5 Phylogenetic tree showing trans-species polymorphism of exons 2 and 3 of all MHC IIB alleles found in the Midas cichlid (in black) and DAB alleles from other species (in blue), with posterior probabilities. Alleles are grouped according to how they cluster in Fig. 4

Table 3 Summary of likelihood ratio tests for site-specific positive selection of MHC IIB genes comparing groups of alleles

Allele Group	Models compared	LRT ($2\Delta l$)	p-value	Estimate for $\omega > 1$	Proportion of PSS	PSS
All groups	M1a versus M2a	81.738836	1.09E-06	3.71226	0.34292	1, 2, 7, 9, 10, 32, 33, 37, 38, 39, 41, 42, 43, 46, 57, 58, 60, 61, 62
All groups	M7 versus M8	84.941740	6.61E-07	3.55150	0.35599	1, 2, 7, 9, 10, 32, 33, 37, 38, 39, 41, 42, 43, 46, 57, 58, 60, 61, 62
All groups	M8a versus M8	73.238718	6.12E-07			
All except I	M1a versus M2a	83.327546	1.16E-06	3.89788	0.32156	8, 9, 16, 17, 32, 39, 40, 44, 45, 46, 48, 49, 50, 53, 64, 65, 67, 68, 69
All except I	M7 versus M8	80.631206	6.74E-07	3.72628	0.33118	8, 9, 14, 16, 17, 32, 39, 40, 44, 45, 46, 48, 49, 50, 53, 64, 65, 67, 68, 69
All except I	M8a versus M8	74.046658	8.90E-07			
Group I only	M1a versus M2a	4.594664	0.10			
Group I only	M7 versus M8	4.594238	0.10			
Group I only	M8a versus M8	4.594000	0.10			
Group II only	M1a versus M2a	7.702834	0.021	3.16267	0.48401	49, 65
Group II only	M7 versus M8	7.706184	0.021	3.16268	0.48401	13, 49, 65
Group II only	M8a versus M8	7.702818	0.021			
Group III only	M1a versus M2a	14.605000	0.0007	31.27408	0.03609	73
Group III only	M7 versus M8	14.663852	0.0007	31.37037	0.03603	73
Group III only	M8a versus M8	14.604918	0.0007			
Group IV only	M1a versus M2a	20.610518	0.00003	7.27232	0.15698	11, 48, 55, 67, 69, 106
Group IV only	M7 versus M8	20.981230	0.00003	7.22462	0.15872	11, 42, 48, 55, 66, 67, 69, 77, 106

The LRT models compared by the likelihood ratio test with five codon substitution models: Beta models M7 (no positive selection), M8 (positive selection), and M8a (no positive selection and $\omega = 1$), and models M1a (nearly neutral) and M2a (positive selection). The $\Delta l = 2(lb - la)$, $\omega = d_n / d_s$, positively selected sites (PSS) are inferred by empirical Bayesian posterior probabilities. PSS in bold are inferred at 99%

Table 4 Tests of overall selection and selection by domain

Domain	Alleles	length (AAs)	dN	dN Number	dS	dS Number	dN/dS	P- value (purifying)	P- value (positive)
LP, β_1 & β_2	Group I-IV and ungrouped (N69)	228	0.221 (0.024)	44.488 (4.998)	0.327 (0.053)	18.525 (2.234)	0.676	0.021	1.000
β_1	Group I-IV and ungrouped (N69)	91	0.254 (0.032)	37.703 (3.988)	0.301 (0.059)	12.807 (1.679)	0.844	0.219	1.000
β_2	Group I-IV and ungrouped (N60)	93	0.129 (0.029)	8.502 (2.154)	0.459 (0.101)	7.238 (1.585)	0.282	<0.001	1.000
β_1 & β_2	Group I (N20)	113	0.008 (0.003)	1.616 (0.521)	0.000 (0.000)	0.000 (0.000)	-	-	-
LP, β_1 & β_2	Group II-IV and ungrouped (N49)	229	0.193 (0.022)	41.637 (4.845)	0.306 (0.044)	18.101 (2.142)	0.633	0.005	1.000
β_1	Group II-IV and ungrouped (N49)	91	0.216 (0.028)	33.987 (3.776)	0.270 (0.049)	11.568 (1.579)	0.800	0.131	1.000
β_2	Group II-IV and ungrouped (N40)	94	0.138 (0.034)	9.378 (2.293)	0.463 (0.114)	8.030 (1.679)	0.298	0.001	1.000

Tests of overall selection on the entire length of sequences, and on the different domains were performed. The leading peptide LP, and the two domains β_1 and β_2 , comprise the entire known sequence, despite not all alleles have full coverage of both domains and the LP. The amino acid lengths for each test are given as the number of sites that were included in the test. Three alleles (0005, 0006 and 0012) from Group I, and six alleles (0203, 040103, 0406, 05, 25, 27) from Groups II-IV were excluded for the test of selection on the β_2 domain, as they lack sequence information for this domain. d_N , number of non-synonymous substitution per non-synonymous site; d_S , number of synonymous substitutions per synonymous site; standard error (SE). Significant p-values (at 99% confidence level) are shown bold

domain for those alleles. The β_2 domain shows purifying selection for all alleles ($p < 0.001$), as well as when excluding Group I ($p = 0.001$). However, neither the β_1 domain, nor the entire sequence showed signs of overall positive selection ($p = 1.00$) excluding Group I ($p = 0.001$) (Table 4).

Protein structure homology models

We built protein homology models for all alleles to characterize their tertiary structure, and to determine if they can fold into a potentially functional MHC molecule. The QMEAN4 ranged between -1.72 and -4.32, indicating generally good quality of the proposed models (see Additional file 1: Table S6 for details). GMQE score ranged between 0.61 and 0.75, indicating an overall good quality for most models (Additional file 1: Table S6).

We located the cysteine residues in the 3D structure of alleles to see if they were in the correct position to make structural disulfide bonds. All alleles had an unpaired cysteine at position 7 of the leader peptide, and the four expected cysteine residues at positions 29, 94, and 132, 188, which pair covalently to form disulfide bonds that increase the stability of the β_1 and β_2 subunits respectively (Fig. 6a). All alleles in Group III had an additional unpaired cysteine at position 47 (Fig. 6b), and all alleles in Group II-b had an additional cysteine at position 98. Alleles of Group I showed no notable anomalies in their 3D structure, but allele DXB*0007 presented an unpaired cysteine at position 47 similarly to alleles of Group III. The N-terminal area of MHC IIB protein included an alpha helical region and a beta sheet of four strands in antiparallel orientation. It also showed that the C-terminal area mainly has a beta-fold structure and is characterized by an immunoglobulin-like beta-sandwich made of two anti-parallel sheets. Interestingly, our work revealed that all 3-D models were similar among all groups of alleles, with the exception of the unpaired cysteine positions.

Discussion

We found a total of 69 alleles of MHC IIB exons 2 and 3 in 13 Midas cichlid individuals. This represents very high allelic diversity in this species despite the small sample size tested. Individual Midas cichlids harbor a large number of alleles, with a maximum of 25 per individual, and an average of 12.5 (SD = 6.1). Together, this implies that the Midas cichlid has at least 13 distinct MHC IIB loci, although this may be an underestimation as we cannot detect allele sharing between loci, and we may have insufficient sequences for some alleles. In other Old World cichlid species, up to 17 polymorphic loci have been found, and between 1–13 alleles per individual [51]. Hence, our characterization of Neotropical cichlids revealed comparable structure and diversity between Old and New World cichlids.

Large variation in number of MHC loci, total number of alleles at the population level, and heterozygosity at the individual level, have been reported among fish taxa. For example, pipefish (*Syngnathus typhle*) and Atlantic cod (*Gadus morhua*) have lost all MHC class II loci [73, 74], and appear to compensate for it by larger diversification of MHC class I [75]. Seahorse (*Hippocampus abdominalis*) has a single MHC IIB locus, and a maximum of two alleles per individual [76]. Spotted halibut (*Verasper variegatus*) and Japanese flounder (*Paralichthys olivaceus*) have up to five alleles per individual [77, 78], and guppy (*Poecilia reticulata*) and eel (*Anguilla anguilla*) have up to six alleles [59, 79], which suggests that they all have at least 3 distinct loci. Stickleback (*Gasterosteus aculeatus*), a species with thoroughly studied MHC, has between 3 and 6 MHC class IIB loci depending on the population of origin [57, 80]. However, cichlid MHC IIB alleles are more numerous than those of any other fish that has been studied. This may have contributed to, or be a result of their great diversification.

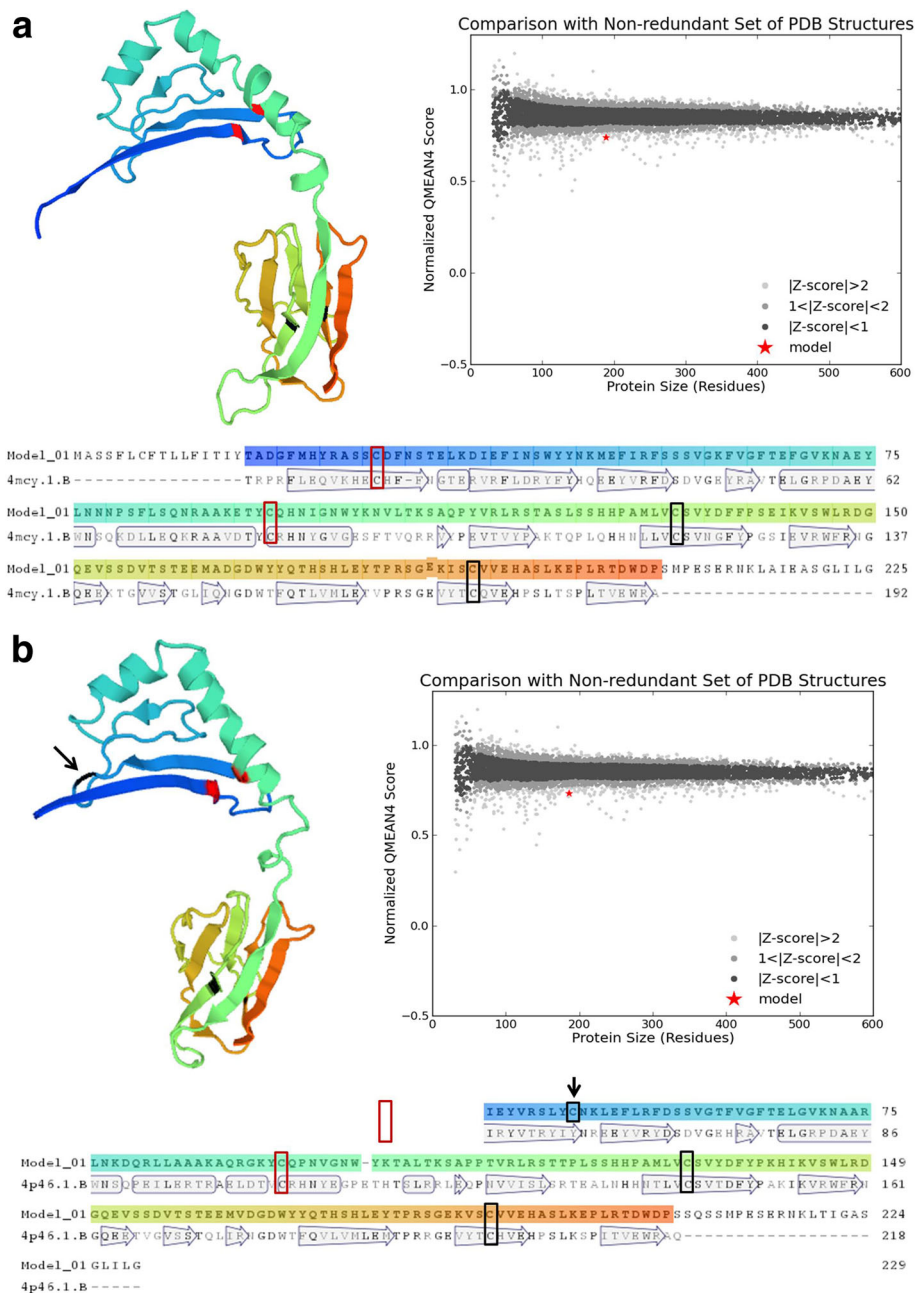


Fig. 6 Models of tertiary structure of MHC IIB sequences, where red boxes represent cysteine amino acids forming the disulfide bond in the β_1 domain, and black boxes represent cysteine residuals that form the disulfide bonds of the β_2 domain. **a** Model of allele DXB*2202; **b** Model of allele DXB*060101, with an unpaired cysteine at position 47 (indicated with a black arrow). The graphs show how each model (red star) compares to a non-redundant set of Protein data bank (PDB) structures, indicating the quality of the model compared to molecules of the same size as a value equivalent to a z-score test

Because MHC genes are directly responding to local parasite pressure [81, 82] they may encode for a magic trait contributing to local adaptation and ultimately to ecological speciation [14, 83]. Determining the contemporary MHC diversity is however challenging due to its multigene nature and fast turnover, often affecting conserved regions that are typically used for setting up primers for amplification [84].

Neotropical cichlids are model species for speciation in sympatry [28], but their MHC genes had not been characterized to date. The nearest relatives with well characterized MHC class II genes are African cichlids [50, 70, 85], from which they split 93 MY ago [54]. We therefore carried out this study to establish reference sequences in order to obtain specific primers that would

amplify a comprehensive diversity of MHC IIB genes for future population-based studies.

Like most other fish, the sequences we obtained of the Midas cichlid MHC IIB genes are composed of 6 exons [5, 85]. We did not obtain the sequence of all exons for all alleles, but for those we have the complete sequence there are always 6 exons. Published sequences of these genes in other cichlids show the same intron-exon organization [5, 85]. This is the most standard structure in fish, even though species with 5 exons and functional MHC have also been described (e.g., sea bream *Chrysophrys major* [86] or Japanese flounder *Paralichthys olivaceus* [87]). Variation in exon number is due to either exon fusion [88], or exon splitting [5, 89].

We did however find considerable variation in the length of intron 2, which ranged from 239 bp to 2.5 kb in the sequences we obtained, and 10 kb in the genome sequences. This intron length variation is likely the reason why we were unable to amplify some alleles that were obtained from cDNA when using the same primers spanning introns in gDNA. Length variation in intron 2 has been reported in several species including other cichlids [51, 90, 91]. The tandem repeats found in intron 2 contribute to the increased length of the long intron. Reusch and Langefors [7] reported a 10-mer repeat in intron 2 of three-spined sticklebacks, responsible for important changes in sequence length, demonstrating that this mode of intron evolution can happen in other fish species. The genome sequences also revealed length variation in all other introns, most notably in intron 3 that varies between 155 to 5 kb.

Within the 69 alleles found in the Midas cichlid we distinguished different groups. One group of alleles (Group I) resembled a pattern of non-classical MHC IIB genes [92]. These alleles showed low variability, are apparently not expressed, and none of their positions seemed to have evolved under positive selection according to our selection analyses. A pattern of low polymorphism is typical of non-classical MHC IIB genes, since their primary function is to assist in loading the antigenic peptides onto classic MHC class II molecules [93]. Because non-classical MHC molecules do not have to bind to a wide array of antigen peptides, the sequence between the different alleles do not follow the typical “patchwork” pattern of classic MHC II alleles, especially in the peptide binding region [93]. Non-classical MHC IIB genes similar to those in Group I of the Midas cichlid have also been described in Atlantic Salmon (*Salmo salar*) [94]. All other groups of alleles and the ungrouped alleles showed classic MHC class II gene patterns. They all displayed sites subjected to strong positive selection, suggesting that they might have evolved under strong parasite-mediated balancing selection [95, 96]. However, the results of the selection tests have to be taken with caution since we cannot allocate

alleles to specific loci and alleles within groups could potentially belong to different evolutionary lineages. This might also have an influence on the overall selection tests for which all groups were combined. It might explain some of the discrepancies between strong positive selection in site-specific tests and the absence of positive selection in the overall tests. Antagonistic coevolution between hosts and parasites is recognized as a powerful force capable of driving rapid evolutionary changes, which might significantly contribute to biodiversity (e.g., [91]). In fish, MHC frequency shifts of resistance alleles have been observed as a response to local parasite-mediated selection [83]. Combined with MHC-based mate choice reported in almost all jawed vertebrates [82], host-parasite interaction through MHC genes has been suggested to contribute to speciation, even in sympatry [37, 97–99].

The phylogenetic tree of all 69 alleles plus 20 of other fish species, displayed a pattern that strongly supports trans-species polymorphism. Some alleles of the Midas cichlid seem to be more closely related to alleles of other species than to other alleles of the same individual. For example, alleles of Group I are more closely linked to Nile tilapia DJB allele (DJB accession # AB677258.1) than to any other Midas MHC IIB allele. Similarly allele DXB*27 of the Midas cichlid is closely related to allele DIB of Nile tilapia, indicating homology and hence TSP. TSP is a common pattern of the MHC and has been observed in many taxonomic groups (reptiles [42], amphibians [44], mammals [43], birds [45, 46], and fish [39, 47]). TSP is evident throughout the phylogenetic tree, and seemed to be most common with alleles of Group IV and ungrouped alleles.

The tertiary structure models showed that similar to the Nile tilapia [100], the Midas cichlid MHC IIB sequence has all the necessary features for the molecule to be functional, including two pairs of cysteine residues. The biological function of unpaired cysteine residues in the MHC molecules remains unknown. It has however been suggested that they could play a role in the formation of exosomal dimers [101]. We found two groups of alleles with extra unpaired cysteines (groups II-b and III), but nothing noteworthy was found in the structure of these alleles. Future studies focusing on the tertiary structure of MHC molecules should focus on determining the function of unpaired cysteines, to further reveal their contribution to immunity specifically, and species' evolution in general.

Despite considerable sequencing effort we were not able to find all alleles in both cDNA and gDNA. Alleles in Group I were never detected in cDNA, which makes us think they might be putative non-classical or even pseudogenes. On the other hand, we had difficulties in amplifying alleles from Group III in gDNA while they were readily obtained in cDNA. We only succeeded in

amplifying these alleles in gDNA by using primers that excluded intron 2. Intron 2 is therefore likely causing sequencing difficulties due to particularly long sequences or rich GC content. Another explanation for these difficulties might be alternative splicing, which is known to occur in MHC [102]. Indeed, in salamanders over 20% of the transcripts can be alternatively spliced, with variation in different organs, see Bulut et al. [103]. As the alleles discovered here seem to be functional and variable, and they may be contributing to the dynamic response of MHC to parasite selection [99].

Conclusions

Taken altogether, MHC IIB genes in the Midas cichlid showed enormous richness in allele diversity and copy number. This diversity is larger than that described in most other fish, and is only comparable to that found in other cichlids. Our findings promise great potential in studying the processes of evolution and speciation in this model system and should be further studied at the ecotype, population and species levels to elucidate the role that parasites may play in sympatric speciation.

Additional file

Additional file 1: Figure S1. Aminoacid variability of all obtained MHC IIB sequences. **Figure S2.** Phylogenetic inference tree of MHC IIB alleles. **Figure S3.** Estimate of evolutionary divergence between sequences that support allele groupings. **Table S1.** List of samples used for this study, Species, Lake and ID numbers are given. **Table S2.** List of sequences from GeneBank used to design primers MHC-Rev_3. **Table S3.** Sequences of different species used to evaluate trans-species polymorphism. **Table S4.** Length in base pairs for the exons and Introns sequenced for each allele and *O. niloticus* sequence used as reference. **Table S5.** Mean pairwise distances of randomization analysis on the groups of alleles. **Table S6.** 3D Protein homology models for all alleles with the global model quality estimation, the overall model quality scores, and the summary of estimated Z-scores. (DOCX 8379 kb)

Abbreviations

cDNA: Complementary DNA; gDNA: Genomic DNA; MHC: Major Histocompatibility Complex; TSP: Trans-species polymorphism

Acknowledgements

We thank M Peláez Aller, and L Benítez Rico for help in the lab, and O Roth, P Hablützel and J Calatayud for helpful discussions, as well as three anonymous reviewers for their valuable comments and suggestions.

Funding

Ministerio de Economía y Competitividad del Gobierno de España, Programa de Formación de Personal Investigador FPI BES-2011-047645 to MJH, Programa Estatal de Fomento de la Investigación Científica y Técnica de Excelencia Proyecto CGL 2010-16103 to MB. This project was further enabled through two German Science Foundation grants to CE (DFG, EI841/4-1 and EI841/6-1) both part of the SPP 1399 priority programme on “host-parasite interactions”.

Availability of data and materials

All Data is publically available, sequences can be downloaded from GenBank accession numbers KY039442-KY039474 and KY354964-KY355011.

Authors' contributions

MJH and MB conceived the study. MB and CE supervised the study. MJH carried out the experimental work supported by SB. MJH and MB wrote the manuscript. All authors approved the final version of the manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

The Ministry of Natural Resources (MARENA) in Nicaragua provided permission for collection permits (No. 001-012012).

Author details

¹Museo Nacional de Ciencias Naturales, CSIC, José Gutiérrez Abascal, 2, 28006 Madrid, Spain. ²GEOMAR Helmholtz Centre for Ocean Research Kiel, Duesternbrooker weg 20, 24105 Kiel, Germany. ³Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Müggelseedamm 310, 12587 Berlin, Germany. ⁴Queen Mary University of London, School of Biological and Chemical Sciences, Mile End Road, London E1 4NS, UK.

Received: 30 August 2016 Accepted: 13 January 2017

Published online: 15 February 2017

References

- Pastoret P-P, Griebel P, Bazin H, Govaerts A, editors. Handbook of Vertebrate Immunology. San Diego, California 92101-4495. USA: Academic; 1998.
- Dixon B, Van Erp SHM, Rodrigues PNS, Egberts E, Stet RJM. Fish major histocompatibility complex genes: an expansion. *Dev Comp Immunol.* 1995;19:109–33.
- Janeway CA, Travers P, Walport M. Immunobiology: the immune system in health and disease. 2005.
- Madden DR. The three-dimensional structure of peptide-mhc complexes. *Annu Rev Immunol.* 1995;13:587–622.
- Ono H, O'huigin C, Vinček V, Klein J. Exon-intron organization of fish major histocompatibility complex class II B genes. *Immunogenetics.* 1993;38:223–34.
- Schwaiger FW, Weyers E, Epplen C, Brün J, Ruff G, Crawford A, et al. The paradox of MHC-DRB exon/intron evolution: alpha-helix and beta-sheet encoding regions diverge while hypervariable intronic simple repeats coevolve with beta-sheet codons. *J Mol Evol.* 1993;37:260–72.
- Reusch TBH, Langefors A. Inter- and intralocus recombination drive MHC class IIB gene diversification in a teleost, the three-spined stickleback *Gasterosteus aculeatus*. *J Mol Evol.* 2005;61:531–41.
- Cummings SM, McMullan M, Joyce DA, van Oosterhout C. Solutions for PCR, cloning and sequencing errors in population genetic analysis. *Conserv. Genetics.* 2010;11:1095–7.
- Burri R, Promerová M, Goebel J, Fumagalli L. PCR-based isolation of multigene families: lessons from the avian MHC class IIB. *Mol Ecol.* 2014;14:778–88.
- Lenz TL, Becker S. Simple approach to reduce PCR artefact formation leads to reliable genotyping of MHC and other highly polymorphic loci - implications for evolutionary analysis. *Gene.* 2008;427:117–23.
- Babik W. Methods for MHC genotyping in non-model vertebrates. *Mol Ecol.* 2010;10:237–51.
- Lighten J, Van Oosterhout C, Bentzen P. Critical review of NGS analyses for de novo genotyping multigene families. *Mol Ecol.* 2014;23:3957–72.
- Sommer S. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Front Zool.* 2005;18:1–18.
- Eizaguirre C, Lenz TL, Traulsen A, Milinski M. Speciation accelerated and stabilized by pleiotropic major histocompatibility complex immunogenes. *Ecol Lett.* 2009;12:5–12.
- Trowsdale J. The MHC, disease and selection. *Immunol. Lett.* Elsevier B.V.; 2011;137:1–8.
- Klein J, Sato A, Nikolaidis N. MHC, TSP, and the origin of species: from immunogenetics to evolutionary genetics. *Annu Rev Genet.* 2007;41:281–304.
- 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.

18. Vincek V, O'Huigin C, Satta Y, Takahata N, Boag PT, Grant PR, et al. How large was the founding population of Darwin's finches? *Proc R Soc London, Ser B*. 1997;264:111–8.
19. Kocher TD. Adaptive evolution and explosive speciation: the cichlid fish model. *Nat Rev Genet*. 2004;5:288–98.
20. Seehausen O. African cichlid fish: a model system in adaptive radiation research. *Proc R Soc B*. 2006;273:1987–98.
21. Turner GF. Adaptive radiation of cichlid fish. *Curr Biol*. 2007;17:R827–31.
22. Salzburger W. The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. *Mol Ecol*. 2009;18:169–85.
23. Kullander SO. Cichlidae. Pp. 605–654. In: Reis RE, Kullander SO, Ferraris Jr CJ, editors. Check list of the freshwater fishes of South and Central America. Porto Alegre: Edipucrs; 2003. p. 729.
24. Barlow GW, Munsey JW. The red devil-Midas-arrow cichlid species complex in Nicaragua. *Pap Biol Sci*. 1976;1:157–369.
25. Elmer KR, Lehtonen TK, Kautt AF, Harrod C, Meyer A. Rapid sympatric ecological differentiation of crater lake cichlid fishes within historic times. *BMC Biol*. 2010;8:60.
26. Gavrillets S, Vose A, Barluenga M, Salzburger W, Meyer A. Case studies and mathematical models of ecological speciation. 1. Cichlids in a crater lake. *Mol. Ecol*. 2007;16:2893–909.
27. Barluenga M, Meyer A. The Midas cichlid species complex: incipient sympatric speciation in Nicaraguan cichlid fishes? *Mol Ecol*. 2004;13:2061–76.
28. Barluenga M, Stölting KN, Salzburger W, Muschick M, Meyer A. Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature*. 2006;439:719–23.
29. 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.
30. Muschick M, Barluenga M, Salzburger W, Meyer A. Adaptive phenotypic plasticity in the Midas cichlid fish pharyngeal jaw and its relevance in adaptive radiation. *BMC Evol Biol*. 2011;11:116.
31. Thibert-Plante X, Gavrillets S. Evolution of mate choice and the so-called magic traits in ecological speciation. *Ecol Lett*. 2013;16:1004–13.
32. Kornfield I, Smith PF. African Cichlid Fish: Model Systems for Evolutionary Biology. *Annu Rev Ecol Syst*. 2000;31:163–82.
33. Elmer KR, Fan S, Gunter HM, Jones JC, Boekhoff S, Kuraku S, et al. Rapid evolution and selection inferred from the transcriptomes of sympatric crater lake cichlid fishes. *Mol Ecol*. 2010;19:197–211.
34. Kautt AF, Elmer KR, Meyer A. Genomic signatures of divergent selection and speciation patterns in a "natural experiment", the young parallel radiations of Nicaraguan crater lake cichlid fishes. *Mol Ecol*. 2012;21:4770–86.
35. Santos EM, Braasch I, Boileau N, Meyer BS, Sauter L, Böhne A, et al. The evolution of cichlid fish egg-spots is linked with a cis-regulatory change. *Nat Commun*. 2014;5:5149.
36. Summers K, McKeon S, Sellars J, Keusenkothen M, Morris J, Gloeckner D, et al. Parasitic exploitation as an engine of diversity. *Biol Rev Camb Philos Soc*. 2003;78:639–75.
37. 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.
38. Klein J. Generation of diversity at MHC loci: Implications for T-cell receptor repertoires. In: Fougereau M, Dausset J, editors. *Immunol*. 1980;80: 239–53.
39. Lenz TL, Eizaguirre C, Kalbe M, Milinski M. Evaluating patterns of convergent evolution and trans-species polymorphism at MHC immunogenes in two sympatric stickleback species. *Evolution*. 2013;67:2400–12.
40. Xu S, Chen B, Zhou K, Yang G. High similarity at three MHC loci between the baiji and finless porpoise: trans-species or convergent evolution? *Mol Phylogenet Evol*. 2008;47:36–44.
41. Klein J, Sato A, Nagl S, O'Huigin C. Molecular trans-species polymorphism. *Annu Rev Ecol Syst*. 1998;29:1–21.
42. Stiebens VVA, Merino SE, Chain FJJ, Eizaguirre C. Evolution of immunogenes in the endangered loggerhead sea turtle (*Caretta caretta*) revealed by 454 amplicon sequencing. *BMC Evol Biol*. 2013;13:95.
43. Bryja J, Galan M, Charbonnel N, Cosson JF. Duplication, balancing selection and trans-species evolution explain the high levels of polymorphism of the DQA MHC class II gene in voles (Arvicolinae). *Immunogenetics*. 2006;58:191–202.
44. Kiemnec-Tyburczy KM, Richmond JQ, Savage AE, Zamudio KR. Selection, trans-species polymorphism, and locus identification of major histocompatibility complex class IIβ alleles of New World ranid frogs. *Immunogenetics*. 2010;62:741–51.
45. Eimes JA, Townsend AK, Sepil I, Nishiumi I, Satta Y. Patterns of evolution of MHC class II genes of crows (*Corvus*) suggest trans-species polymorphism. *PeerJ*. 2015;3:e853.
46. Kikkawa EF, Tsuda TT, Sumiyama D, Naruse TK, Fukuda M, Kurita M, et al. Trans-species polymorphism of the Mhc class II DRB-like gene in banded penguins (genus *Spheniscus*). *Immunogenetics*. 2009;61:341–52.
47. Ottová E, Simková A, Martin J-F, de Bellocq JG, Gelnar M, Allienne J-F, et al. Evolution and trans-species polymorphism of MHC class IIβ genes in cyprinid fish. *Fish Shellfish Immunol*. 2005;18:199–222.
48. Klein D, Ono H, O'Huigin C, Vincek V, Goldschids T, Klein J. Extensive MHC variability in cichlids of Lake Malawi. *Nature*. 1993;364:330–4.
49. Murray BW, Shintani S, Sülmann H, Klein J. Major histocompatibility complex class II A genes in cichlid fishes: identification, expression, linkage relationships, and haplotype variation. *Immunogenetics*. 2000;51:576–86.
50. Figueroa F, Mayer WE, Sülmann H, O'Huigin C, Tichy H, Satta Y, et al. Mhc class II B gene evolution in East African cichlid fishes. *Immunogenetics*. 2000;51:556–75.
51. Málaga-Trillo E, Zaleska-Rutczynska Z, McAndrew B, Vincek V, Figueroa F, Sülmann H, et al. Linkage relationships and haplotype polymorphism among cichlid Mhc class II B loci. *Genetics*. 1998;149:1527–37.
52. Sato A, Klein D, Sülmann H, Figueroa F, O'Huigin C, Klein J. Class I mhc genes of cichlid fishes: identification, expression, and polymorphism. *Immunogenetics*. 1997;46:63–72.
53. Ono H, O'Huigin C, Tichy H, Klein J. Major-histocompatibility-complex variation in two species of cichlid fishes from Lake Malawi. *Mol Biol Evol*. 1993;10:1060–72.
54. Albert JS, Reis RE. Historical Biogeography of Neotropical Freshwater Fishes. Albert JS, Reis RE, editors. Berkeley Los Angeles London: University of California Press; 2011.
55. Friedman M, Keck BP, Dornburg A, Eytan RI, Martin CH, Hulseley CD, et al. Molecular and fossil evidence place the origin of cichlid fishes long after Gondwanan rifting. *Proc R Soc London, Ser B*. 2013;280:20131733.
56. Matschiner M, Musilová Z, Barth JMI, Starostová Z, Salzburger W, Steel M, et al. Bayesian Phylogenetic Estimation of Clade Ages Supports Trans-Atlantic Dispersal of Cichlid Fishes. *Syst Biol*. 2016. syw076. doi:10.1093/sysbio/syw076.
57. Sato A, Figueroa F, O'Huigin C, Steck N, Klein J. Cloning of major histocompatibility complex (Mhc) genes from threespine stickleback, *Gasterosteus aculeatus*. *Mol Mar Biol Biotechnol*. 1998;7:221–31.
58. Lenz TL, Eizaguirre C, Becker S, Reusch TBH. RSCA genotyping of MHC for high-throughput evolutionary studies in the model organism three-spined stickleback *Gasterosteus aculeatus*. *BMC Evol Biol*. 2009;9:57.
59. 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.
60. Hall T. BioEdit: a user-friendly biological sequence alignment editor and 975 analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser*. 1999;41: 95–8.
61. Klein J, Bontrop RE, Dawkins RL, Erlich HA, Gyllensten UB, Heise ER, et al. Nomenclature for the major histocompatibility complex of different species: a proposal. *Immunogenetics*. 1990;31:217–9.
62. Benson G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res*. 1999;27:573–80.
63. Elmer KR, Fan S, Kusche H, Luise Spreitzer M, Kautt AF, Franchini P, et al. Parallel evolution of Nicaraguan crater lake cichlid fishes via non-parallel routes. *Nat Commun*. 2014;5:5168.
64. Tamura K, Nei M, Kumar S. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc Natl Acad Sci U S A*. 2004; 101:11030–5.
65. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 2011;28:2731–9.
66. Bunn A, Korpela M. Time Series Analysis in dPLR. 2016. p. 4.
67. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, et al. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Comput Biol*. 2014;10:1–6.
68. Lanfear R, Calcott B, Kainer D, Mayer C, Stamatakis A. Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Evol Biol*. 2014;14:82.
69. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol*. 2006;23:254–67.
70. Sato A, Dongak R, Hao L, Shintani S. Organization of Mhc class II A and B genes in the tilapia fish *Oreochromis*. *Immunogenetics*. 2012;64:679–90.

71. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 2007;24:1586–91.
72. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. *Bioinformatics.* 2006;22:195–201.
73. Haase D, Roth O, Kalbe M, Schmiedeskamp G, Scharsack JP, Rosenstiel P, et al. Absence of major histocompatibility complex class II mediated immunity in pipefish, *Syngnathus typhle*: evidence from deep transcriptome sequencing. *Biol Lett.* 2013;9:20130044.
74. Star B, Nederbragt AJ, Jentoft S, Grimholt U, Malmstrøm M, Gregers TF, et al. The genome sequence of Atlantic cod reveals a unique immune system. *Nature.* 2011;477:207–10.
75. Star B, Jentoft S. Why does the immune system of Atlantic cod lack MHC II? *BioEssays.* 2012;34:648–51.
76. Bahr A, Wilson AB. The impact of sex-role reversal on the diversity of the major histocompatibility complex: insights from the seahorse (*Hippocampus abdominalis*). *BMC Evol Biol.* 2011;11:121.
77. Xu T, Chen S, Zhang YX. MHC class II alpha gene polymorphism and its association with resistance/susceptibility to *Vibrio anguillarum* in Japanese flounder (*Paralichthys olivaceus*). *Dev Comp Immunol.* 2010;34:1042–50.
78. Li H, Jiang L, Han J, Su H, Yang Q, He C. Major histocompatibility complex class IIA and IIB genes of the spotted halibut *Verasper variegatus*: Genomic structure, molecular polymorphism, and expression analysis. *Fish Physiol Biochem.* 2011;37:767–80.
79. Lighten J, Van Oosterhout C, Paterson IG, McMullan M, Bentzen P, van Oosterhout C. Ultra-deep Illumina sequencing accurately identifies MHC class IIb alleles and provides evidence for copy number variation in the guppy (*Poecilia reticulata*). *Mol Ecol.* 2014;14:753–67.
80. Lenz TL, Eizaguirre C, Scharsack JP, Kalbe M, Milinski M. Disentangling the role of MHC-dependent “good genes” and “compatible genes” in mate-choice decisions of three-spined sticklebacks *Gasterosteus aculeatus* under semi-natural conditions. *J Fish Biol.* 2009;75:2122–42.
81. Spurgin LG, Richardson DS. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proc R Soc B Biol Sci.* 2010;277:979–88.
82. Milinski M. Arms races, ornaments and fragrant genes: The dilemma of mate choice in fishes. *Neurosci Biobehav Rev.* 2014;46:567–72.
83. 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.
84. Wegner KM. Massive parallel MHC genotyping: titanium that shines. *Mol Ecol.* 2009;18:1818–20.
85. Pang J, Gao F, Lu M, Ye X, Zhu H, Ke X. Major histocompatibility complex class IIA and IIB genes of Nile tilapia *Oreochromis niloticus*: genomic structure, molecular polymorphism and expression patterns. *Fish Shellfish Immunol.* 2013;34:486–96.
86. Chen SL, Zhang YX, Xu MY, Ji XS, Yu GC, Dong CF. Molecular polymorphism and expression analysis of MHC class II B gene from red sea bream (*Chrysophrys major*). *Dev Comp Immunol.* 2006;30:407–18.
87. Zhang YX, Chen SL, Liu YG, Sha ZX, Liu ZJ. Major histocompatibility complex class IIb allele polymorphism and its association with resistance/susceptibility to *Vibrio anguillarum* in Japanese flounder (*Paralichthys olivaceus*). *Mar Biotechnol.* 2006;8:600–10.
88. Dijkstra JM, Katagiri T, Hosomichi K, Yanagiya K, Inoko H, Ototake M, et al. A third broad lineage of major histocompatibility complex (MHC) class I in teleost fish; MHC class II linkage and processed genes. *Immunogenetics.* 2007;59:305–21.
89. Figueroa F, Ono H, Tichy H, O’Huigin C, Klein J. Evidence for Insertion of a New Intron into an Mhc Gene of Perch-Like Fish. *Proc R Soc B Biol Sci.* 1995;259:325–30.
90. Xu T, Sun Y, Shi G, Cheng Y, Wang R. Characterization of the Major Histocompatibility Complex Class II Genes in Miiuy Croaker. *Liu Z, editor. PLoS One.* 2011;6:e23823.
91. Jiang J, Li C, Zhang Q, Wang X. Locus number estimation of MHC class II B in stone flounder and Japanese flounder. *Int J Mol Sci.* 2015;16:6000–17.
92. Grimholt U. MHC and Evolution in Teleosts. *Bilology.* 2016;5:6.
93. Kropshofer H, Hämmerling GJ, Vogt AB. The impact of the non-classical MHC proteins HLA-DM and HLA-DO on loading of MHC class II molecules. *Immunol Rev.* 1999;172:267–78.
94. 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.
95. Xu T, Liu J, Sun Y, Zhu Z, Liu T. Characterization of 40 full-length MHC class IIA functional alleles in miiuy croaker: Polymorphism and positive selection. *Dev Comp Immunol.* 2016;55:138–43.
96. Takahata N, Satta Y, Klein J. Polymorphism and balancing selection at major histocompatibility complex. *Genetics.* 1992;130TSK92:925–38.
97. Buckling A, Rainey PB. The role of parasites in sympatric and allopatric host diversification. *Nature.* 2002;420:496–9.
98. Eizaguirre C, Yeates SE, Lenz TL, Kalbe M, Milinski M. MHC-based mate choice combines good genes and maintenance of MHC polymorphism. *Mol Ecol.* 2009;18:3316–29.
99. 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.
100. Zhou F, Dong Z, Fu Y, Li T, Zeng Y, Ji X, et al. Molecular cloning, genomic structure, polymorphism and expression analysis of major histocompatibility complex class II B gene of Nile tilapia (*Oreochromis niloticus*). *Aquaculture.* 2013;372-375:149–57.
101. Lynch S, Santos SG, Campbell EC, Nimmo AMS, Botting C, Prescott A, et al. Novel MHC class I structures on exosomes. *J Immunol.* 2009;183:1884–91.
102. Laurens V, Chapusot C, del Rosario OM, Bentrari F, Padros MR, Tournefier A. Axolotl MHC class II beta chain: predominance of one allele and alternative splicing of the beta1 domain. *Eur J Immunol.* 2001;31:506–15.
103. Bulut Z, McCormick CR, Bos DH, DeWoody AJ. Polymorphism of alternative splicing of major histocompatibility complex transcripts in wild tiger salamanders. *J Mol Evol.* 2008;67:68–75.

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

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

Submit your manuscript at
www.biomedcentral.com/submit

