Divergent and non-parallel evolution of MHC IIB in the Neotropical Midas cichlid species complex

Seraina E. Bracamonte[1], Melinda J. Hofmann[1], Carlos Lozano-Martín[1], Christophe Eizaguirre[2] and Marta Barluenga[1*]

Abstract

Background: Ecological diversification is the result of divergent natural selection by contrasting habitat characteristics that favours the evolution of distinct phenotypes. This process can happen in sympatry and in allopatry. Habitat-specific parasite communities have the potential to drive diversification among host populations by imposing selective pressures on their host's immune system. In particular, the hyperdiverse genes of the major histocompatibility complex (MHC) are implicated in parasite-mediated host divergence. Here, we studied the extent of divergence at MHC, and discuss how it may have contributed to the Nicaraguan Midas cichlid species complex diversification, one of the most convincing examples of rapid sympatric parallel speciation.

Results: We genotyped the MHC IIB for individuals from six sympatric Midas cichlid assemblages, each containing species that have adapted to exploit similar habitats. We recovered large allelic and functional diversity within the species complex. While most alleles were rare, functional groups of alleles (supertypes) were common, suggesting that they are key to survival and that they were maintained during colonization and subsequent radiations. We identified lake-specific and habitat-specific signatures for both allelic and functional diversity, but no clear pattern of parallel divergence among ecomorphologically similar phenotypes.

Conclusions: Colonization and demographic effects of the fish could have contributed to MHC evolution in the Midas cichlid in conjunction with habitat-specific selective pressures, such as parasites associated to alternative preys or environmental features. Additional ecological data will help evaluating the role of host–parasite interactions in the Midas cichlid radiations and aid in elucidating the potential role of non-parallel features differentiating crater lake species assemblages.

Keywords: Major histocompatibility complex, Ecological divergence, Adaptive radiation, *Amphilophus* species complex

Background

Diversification into ecologically and morphologically distinct forms upon colonization of novel habitats is the result of natural selection, and this can lead to convergent evolution of similar forms in similar but geographically separated habitats [1–4]. Both species interactions and abiotic environmental properties of the habitat contribute to this process [5]. Parasites are ubiquitous biotic agents [6], that are recognized as powerful drivers of diversification [7–10]. Experiments on antagonistic coevolution of hosts and parasites evidenced rapid evolutionary changes and divergence of host populations [11–13].

*Correspondence: marta.barluenga@mncn.csic.es
[1] Museo Nacional de Ciencias Naturales, CSIC, José Gutiérrez Abascal 2, 28006 Madrid, Spain
Full list of author information is available at the end of the article

Bracamonte *et al. BMC Ecology and Evolution*   (2022) 22:41

Page 2 of 17

Parasites interact directly with the host immune system, and therefore, variability in parasite communities leaves detectable signatures in the immune response and its underlying genetic basis [14]. The immunogenes of the major histocompatibility complex (MHC) play a crucial role in the adaptive immune response of jawed vertebrates [15, 16]. They are hyperdiverse and tend to segregate among populations [17–19]. Thus, MHC genes are excellent candidates for studying the link between parasites, immunogenetic adaptation, and host diversification [7, 20, 21]. MHC genes encode cell-surface glycoproteins that bind peptide fragments derived from parasites and present them to T-cells, thereby activating the adaptive immune response [16]. Broadly, intracellular parasites (e.g. viruses) are recognized by MHC class I, while MHC class II recognizes extracellular parasites (e.g. bacteria or helminths). MHC class I molecules are heterodimers consisting of two polypeptide chains, Iα and a β2-microglobulin. MHC class II molecules are also heterodimers, with two homologous chains IIα and IIβ, encoded by MHC IIA and MHC IIB genes, respectively. MHC genes are the most polymorphic vertebrate genes known to date [16, 22], with variation concentrated in the antigen-binding site, the region determining specificity of the molecule [23, 24]. For MHC class II, polymorphism is mostly contained by exon 2 of MHC IIB. Experimental evidence and theoretical models suggest that the high levels of polymorphism may be the result of balancing selection mediated by parasites [25, 26]. Furthermore, there is increasing evidence that MHC also interacts with an individual's microbiota [27, 28].

Spatial variation in selection on the MHC can lead to a patchwork of immunogenetic divergence and local adaptation among populations occupying different habitats [29]. Divergence may then be reinforced by MHC-assortative mate choice to increase resistance and attain immunogenetic optimality of offspring [17, 20, 30, 31]. Similar parasite and microbial communities within habitat types associated to alternative preys or substrates are expected to lead to similarities at MHC and parallel divergence among habitat types, although empirical evidence is inconclusive. Some studies in the freshwater fish model three-spined stickleback indeed found repeated parallel divergence at MHC among ecotypes [17, 32, 33]; however, other studies in sticklebacks and African cichlids identified population-specific MHC pools and associated divergence [18, 34]. On the other hand, divergent ecotypes can retain similar MHC pools among very contrasting habitats, as shown in the livebearing freshwater fish *Poecilia mexicana* [35]. Hence, predicting the evolutionary outcome of processes on MHC allele segregation is difficult.

Cichlid fish are textbook examples of adaptive radiations, forming the most species-rich and phenotypically diverse clade of teleosts [36]. Ecologically and morphologically similar species are frequently found within and among lakes, and even between continents [3, 37, 38]. An increasing number of studies investigates the importance of parasites in host diversification [39–43] and investigates their ecological interaction ([43], Santacruz et al., [44]). Cichlids possess an extremely large number of MHC class II genes, up to 13 loci per haplotype [45], and show high allelic diversity [46–48], a feature that could facilitate immune specialization [13], enable habitat and trophic adaptation [32], and influence assortative mating [49], which may ultimately contribute to speciation.

The Neotropical Midas cichlid, *Amphilophus* spp., has recently colonized several isolated volcanic crater lakes, independently and asynchronously, from source populations in the tectonic great Nicaraguan lakes, Managua and Nicaragua [50–52] (Fig. 1). Colonization was followed by ecological divergence and sympatric speciation producing closely related species assemblages within each lake [51–54]. Convergent phenotypes can be found along different ecological axes among lakes (e.g. benthic *vs* limnetic in crater lakes (CL) Apoyo and Xiloá; [55, 56], or the rocky (exploited by thick-lipped fish) vs sandy (exploited by normal-lipped fish) substrates in the great lakes; [57, 58]) that are more closely related to species within their respective lakes that to similar phenotypes in other lakes [51, 54]. These eco-morphotypes have been described as distinct species in some lakes [59, 60] while in others, until more work is done, they are still considered to form single polymorphic populations [54].

Here we investigate the contribution of MHC to repeated and parallel divergence of the Midas cichlid in six sympatric assemblages inhabiting isolated crater lakes. First, we characterized the sequence diversity of MHC class IIB alleles in the Midas cichlid species complex across Nicaraguan lakes using high-throughput sequencing. We identified sites under positive selection in the MHC alleles, inferred their phylogenetic relationship, and assigned supertypes of functionally similar alleles. Second, we identified divergence at MHC IIB within the species complex. We tested for differences in sequence diversity among populations, lakes, and habitats. We identified the distribution of alleles among fish in different lakes, populations or species within lakes, and fish in shared habitats across lakes, and tested for divergence in allele and supertype pools. We then identified alleles and supertypes that contributed most to this divergence.

Fig. 1 Map of the Nicaraguan great lakes region showing sampled water bodies (modified from NASA Photojournal PIA03364)

Results

Molecular sequence analyses

We analysed a total of 287 individuals from six sympatric assemblages, each containing species that have adapted to exploiting benthic-limnetic or sandy-rocky habitats (Table 1). A total of 4,158,917 paired-end reads passed the initial quality control, with a mean of 14,441 reads per sample (range: 1713–61,269). These were clustered into 152 unique MHC class II exon 2 alleles, 130 new and 22 corresponding to previously identified alleles [47]. Alleles identified in this study were named ac001 to ac130. Two alleles were excluded from further analyses: the putative non-classical allele Amci-DXB*000101 (see [47]), and one allele with premature stop codons due to a 1 bp deletion. One allele had a 3 bp deletion, and since this did not result in premature stop codons, it was kept in the analyses (Additional file 1: Fig. S1). The resulting 150 alleles translated into 146 unique amino acid sequences. The fragment length was 142 bp of which 97 sites were variable. In the translated amino acid sequences, 35 of 47 sites were variable (Additional file 1: Fig. S2). The nucleotide p-distance among all sequences (π) $\pm$ SE was 0.20 ± 0.018, the amino acid p-distance (aa p-dist) was 0.33 ± 0.042, the dN was 0.24 ± 0.042, and the dS was 0.21 ± 0.053.

A codon-based Z-test of selection did not provide evidence for gene-wide positive selection on alleles over evolutionary times ($Z = 0.8$, $p = 0.2$). However, codon models allowing for site-specific or branch-site-specific positive selection provided a better fit for the data than models without positive selection (Table 2). Selection models implemented in CodeML identified 13 positively selected sites (PSS), MEME identified 16 PSS, and analyses with MrBayes identified 10 PSS. Eight sites under positive selection were identified by all methods

Bracamonte *et al. BMC Ecology and Evolution*   (2022) 22:41

Page 4 of 17

Table 1 Midas cichlid populations used in this study with their characteristic morphology and preferred habitat

Lake	Lake type	Population	Morphotype	Habitat	n
Nicaragua	Tectonic	*A. citrinellus*	High-bodied	Shallow benthic	24
		A. labiatus	Thick-lipped	Rocky	25
Managua	Tectonic	*A. citrinellus*	High-bodied	Shallow benthic	20
		A. labiatus	Thick-lipped	Rocky	4
Masaya	Crater	*A. cf. citrinellus*	High-bodied	Shallow benthic	27
		A. cf. labiatus	Thick-lipped	Rocky	18
Asososca León	Crater	*A. citrinellus* f. benthic	High-bodied	Shallow benthic	23
		A. citrinellus f. limnetic	Long-bodied	Limnetic	20
Apoyo	Crater	*A. astorquii*	High-bodied	Shallow benthic	20
		A. chancho	High-bodied	Deep benthic	21
		A. zaliosus	Long-bodied	Limnetic	21
Xiloá	Crater	*A. amarillo*	High-bodied	Shallow benthic	22
		A. xiloaensis	High-bodied	Deep benthic	11
		A. sagittae	Long-bodied	Limnetic	31
Total					287

n, number of sampled individuals per population

Table 2 Positive site-specific selection identified with CodeML models, BUSTED and MEME, and MrBayes

Model	logL	Parameter estimates	PSS
M1a[a]	− 3068.35	$p_0 = 0.606$, $p_1 = 0.394$ $\omega_0 = 0.128$, $\omega_1 = 1$	NA
M2a[a]	− 2985.41	$p_0 = 0.379$, $p_1 = 0.366$, $p_2 = 0.255$ $\omega_0 = 0.073$, $\omega_1 = 1$, $\omega_2 = 4.138$	1, 3, 10, 12, 13, 22, 28, 35, 40, 42, 44, 45, 46
M7[a]	− 3052.97	$p = 0.171$, $q = 0.184$	NA
M8[a]	− 2984.37	$p_0 = 0.749$, $p_1 = 0.251$, $p = 0.194$ $q = 0.219$, $\omega = 3.767$	1, 3, 10, 12, 13, 22, 28, 35, 40, 42, 44, 45, 46
Constrained[b]	− 2885.7	$p_1 = 0.369$, $p_2 = 0$, $p_3 = 0.631$ $\omega_1 = 0$, $\omega_2 = 0.54$, $\omega_3 = 1$	NA
Unconstrained[b]/MEME	− 2805.1	$p_1 = 0.351$, $p_2 = 0.543$, $p_3 = 0.106$ $\omega_1 = 0.37$, $\omega_2 = 0.37$, $\omega_3 = 43.67$	1, 3, 5, 12, 13, 16, 21, 22, 35, 36, 39, 42, 44, 45, 46, 47
MrBayes		$p_- = 0.356$, $p_N = 0.359$, $p_+ = 0.284$ $\omega_- = 0.082$, $\omega_N = 1$, $\omega_+ = 4.577$	1, 3, 13, 22, 28, 35, 36, 42, 45, 46

logL, log-likelihood value; PSS, positively selected sites; NA, not allowed; PSS indicated for the unconstrained model are identified with MEME. Model parameters are: M1a and M2a: $p_0 =$ proportion of sites with $0 < \omega_0 < 1$, $p_1 =$ proportion of sites with $\omega_1 = 1$, $p_2 =$ proportion of sites with $\omega_2 > 1$; M7 and M8: p, q = β distribution parameters, $p_0 =$ proportion of sites with ω within the β distribution, $p_1 =$ proportion of sites with ω > 1; constrained and unconstrained: p_1, $p_2 =$ proportion of sites with $0 \le \omega_1 \le \omega_2 \le 1$, $p_3 =$ proportion of sites with $\omega_3 = 1$ (constrained) or $\omega_3 > 1$ (unconstrained); MrBayes: $p_- =$ proportion of sites with $0 \le \omega_- < 1$, $p_N =$ proportion of sites with $\omega_N = 1$, $p_+ =$ proportion of sites with $\omega_+ > 1$

[a] CodeML models

[b] BUSTED models

and an additional four sites were identified by two methods (Additional file 1: Fig. S2). Reducing alleles to PSS resulted in 115 unique amino acid sequences which clustered in 13 supertypes of putatively functionally similar alleles. The number of alleles per supertype varied from 4 to 29 (Fig. 2).

A split network grouped alleles into seven major clusters, though they were only weakly delimited (Fig. 2). Bayesian phylogenetic inference also indicated that alleles fall into seven clusters with posterior probabilities > 80 (Additional file 1: Fig. S3). All major clusters but one were composed of one or two supertypes, which were largely monophyletic. The last cluster consisted of three paraphyletic supertypes.

Population divergence analyses

The mean ± SE number of alleles per individual was 7.5 ± 0.1 (range: 4–14), with a minimum of 6.2 ± 0.35 for *A. zaliosus* in CL Apoyo and a maximum of 8.9 ± 0.30 for *A. citrinellus* in L Nicaragua (Table 3,
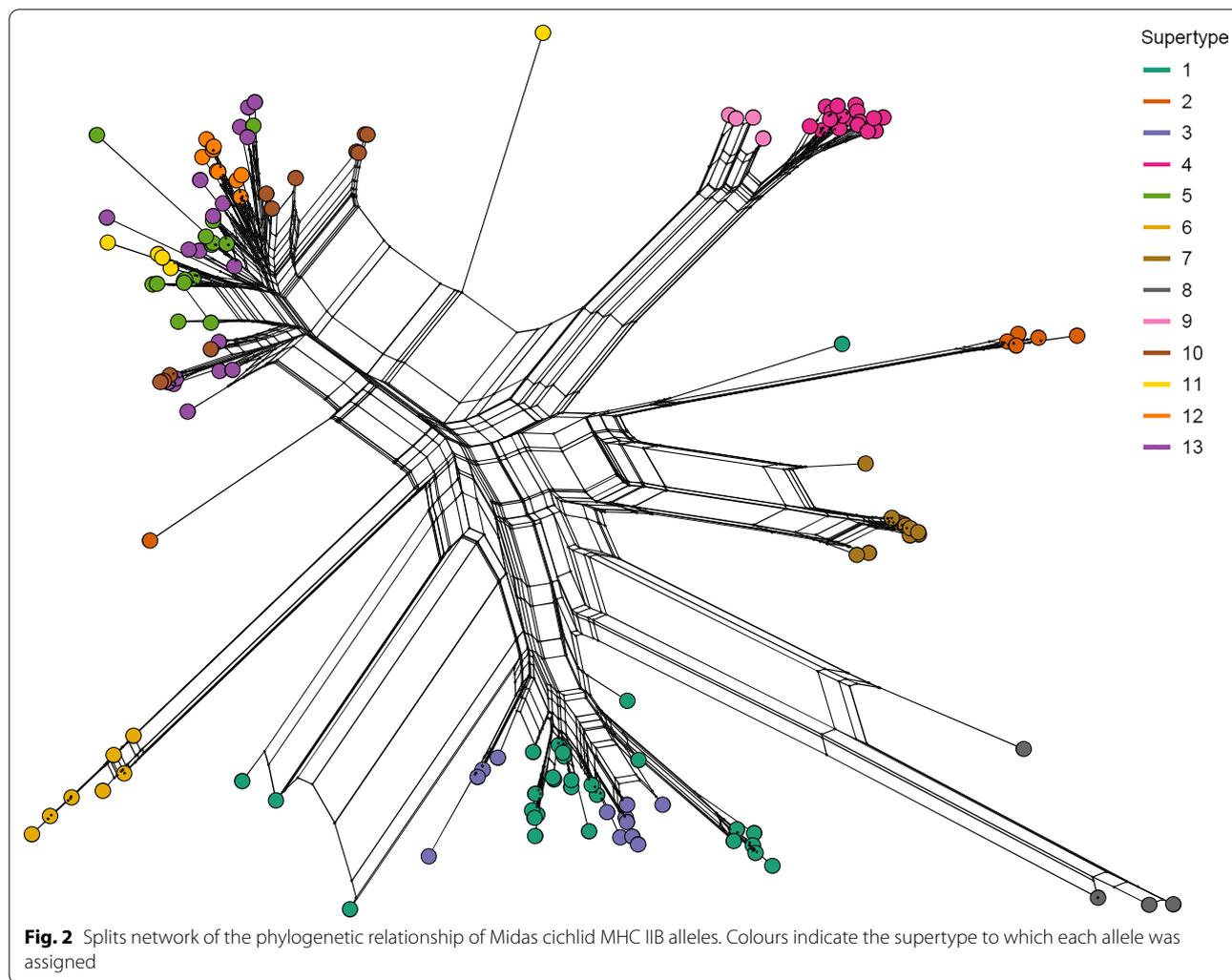
Fig. 2 Splits network of the phylogenetic relationship of Midas cichlid MHC IIB alleles. Colours indicate the supertype to which each allele was assigned

Fig. 3A). The mean ± SE number of supertypes per individual was 5.7 ± 0.07. It was lowest in *A. citrinellus* f. limnetic in CL Asososca León and highest in *A. citrinellus* in L Nicaragua and L Managua (Table 3, Fig. 3B). We additionally calculated four diversity indices for each individual. Mean ± SE π, aa p-dist, dN, and dS per individual were 0.254 ± 0.001, 0.392 ± 0.001, 0.299 ± 0.001, and 0.411 ± 0.003, respectively. Π, aa p-dist, and dN were lowest in *A. citrinellus* f. benthic in CL Asososca León and highest in *A. zaliosus* in CL Apoyo and *A. sagittae* in CL Xiloá (Table 3, Fig. 3C, D). DS was lowest in *A. labiatus* in L Nicaragua and highest in *A. cf. labiatus* in CL Masaya (Table 3, Fig. 3E). Aa p-dist and dN were highly correlated (Pearson's R = 0.94, p-value < 0.001), hence only dN was analysed further. Within-individual sequence diversity indices except dS differed significantly among populations (Table 4). The number of alleles and the number of
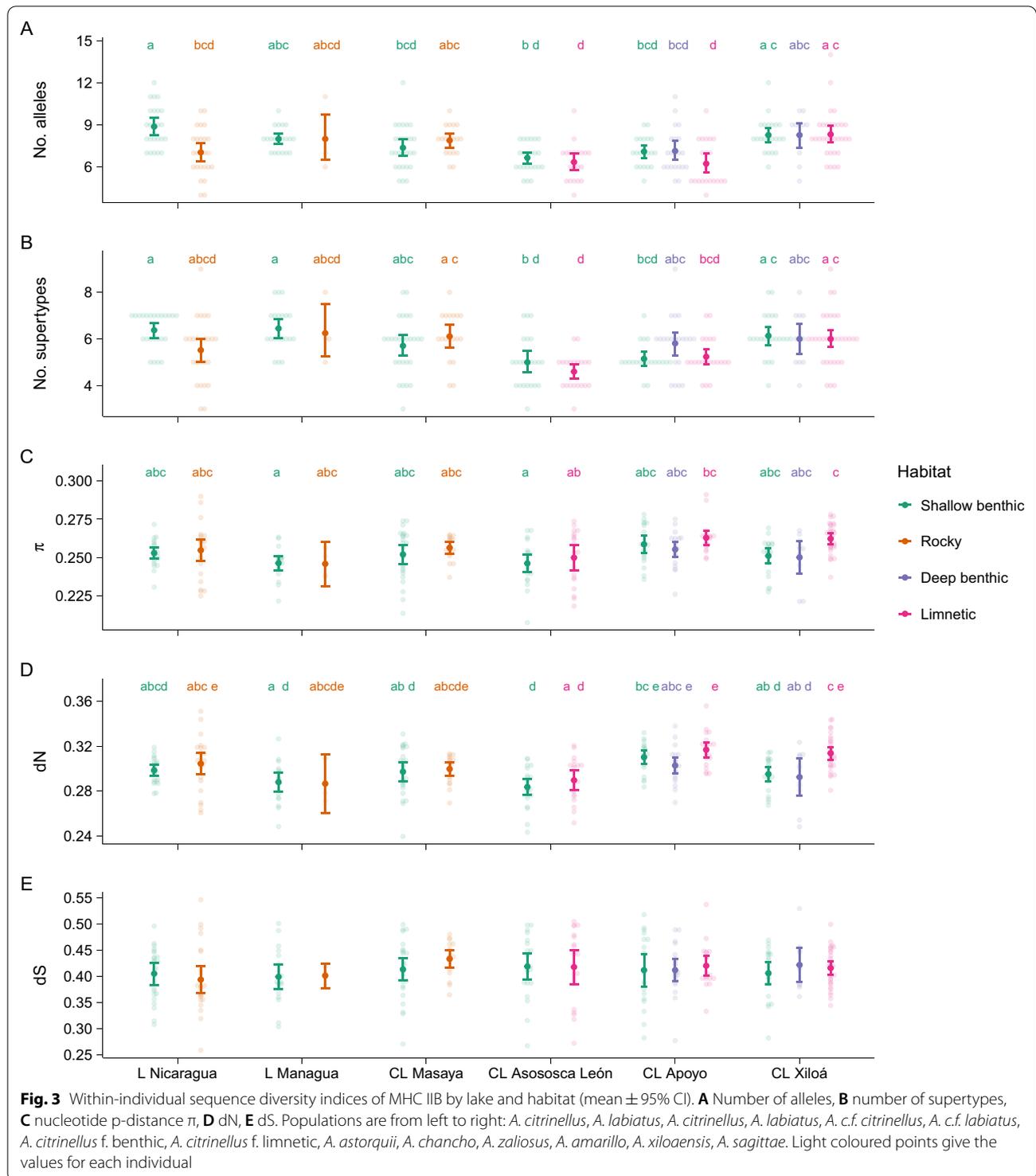
supertypes per individual differed significantly among lakes but not among habitats. However, there was a significant interaction between lake and habitat. Within-individual π and dN differed significantly among lakes and habitats. There was no significant difference in dS among lakes or habitats (Table 4).

The total number of alleles per population ranged from 22 in *A. zaliosus* in CL Apoyo to 55 in *A. citrinellus* in L Managua. The number of private alleles per population ranged from 0 in *A. cf. labiatus* in CL Masaya to 19 in *A. citrinellus* in L Managua. At least 9 supertypes were present in each population (Table 3). The total number of alleles per lake ranged from 27 in CL Asososca León to 68 in CL Xiloá, and the number of private alleles ranged from 3 in CL Masaya to 22 in L Managua. At least 10 supertypes were present in the Midas cichlid community of each lake (Table 3). The number of alleles, the number of supertypes, and the

Bracamonte *et al. BMC Ecology and Evolution*   (2022) 22:41

Page 6 of 17

Table 3 Diversity indices for populations, lakes, and averaged across all samples of the Midas cichlid populations

	n	No. alleles	No. priv. alleles	No. supertypes	Mean no. alleles	Mean no. supertypes	π	aa p-dist	dN	dS
L Nicaragua	49	66	14	13	7.9±0.26	5.9±0.17	0.254±0.002	0.397±0.003	0.299±0.003	0.396±0.008
A. citrinellus	24	50	8	13	8.9±0.30	6.4±0.16	0.253±0.002	0.395±0.003	0.295±0.003	0.400±0.011
A. labiatus	25	38	4	11	7.0±0.33	5.5±0.28	0.255±0.003	0.400±0.005	0.303±0.005	0.392±0.013
L Managua	24	66	22	13	8.0±0.22	6.4±0.20	0.246±0.002	0.382±0.004	0.288±0.003	0.399±0.009
A. citrinellus	20	55	19	13	8.0±0.19	6.4±0.21	0.246±0.002	0.382±0.004	0.288±0.004	0.399±0.011
A. labiatus	4	23	3	10	8.0±1.08	6.2±0.63	0.246±0.005	0.384±0.009	0.287±0.008	0.401±0.007
CL Masaya	45	46	3	12	7.6±0.21	5.9±0.18	0.254±0.002	0.392±0.003	0.298±0.003	0.421±0.007
A. cf. citrinellus	27	40	0	12	7.4±0.30	5.7±0.24	0.252±0.003	0.391±0.004	0.297±0.004	0.413±0.010
A. cf. labiatus	18	43	1	12	7.9±0.25	6.1±0.25	0.256±0.002	0.394±0.003	0.300±0.003	0.433±0.008
CL Asososca León	43	27	7	10	6.5±0.18	4.8±0.15	0.248±0.002	0.377±0.003	0.286±0.003	0.418±0.010
A. citrinellus f. benthic	23	23	2	9	6.7±0.21	5.0±0.23	0.246±0.003	0.374±0.004	0.284±0.004	0.419±0.012
A. citrinellus f. limnetic	20	24	2	10	6.3±0.30	4.6±0.17	0.250±0.004	0.381±0.004	0.290±0.004	0.418±0.016
CL Apoyo	62	47	16	11	6.8±0.19	5.4±0.12	0.259±0.001	0.400±0.002	0.310±0.002	0.414±0.007
A. astorquii	20	36	9	10	7.1±0.24	5.2±0.15	0.259±0.003	0.399±0.004	0.310±0.003	0.412±0.015
A. chancho	21	28	1	10	7.1±0.35	5.8±0.26	0.255±0.002	0.396±0.003	0.303±0.003	0.412±0.010
A. zaliosus	21	22	2	10	6.2±0.35	5.2±0.18	0.263±0.002	0.404±0.003	0.317±0.003	0.420±0.009
CL Xiloá	64	68	17	13	8.3±0.19	6.0±0.13	0.256±0.002	0.396±0.002	0.304±0.002	0.413±0.005
A. amarillo	22	46	7	13	8.3±0.26	6.1±0.20	0.251±0.002	0.389±0.003	0.295±0.003	0.406±0.010
A. xiloaensis	11	32	4	11	8.3±0.49	6.0±0.36	0.250±0.005	0.386±0.007	0.292±0.007	0.422±0.014
A. sagittae	31	31	3	12	8.3±0.30	6.0±0.21	0.262±0.002	0.404±0.003	0.314±0.003	0.416±0.007
Total	287	150		13	7.5±0.10	5.7±0.07	0.254±0.001	0.392±0.001	0.299±0.001	0.411±0.003

n, number of individuals; no. alleles, total number of different alleles; no. priv. alleles, number of private alleles; no. supertypes, total number of different supertypes; mean no. alleles, mean ± SE number of alleles per individual; mean no. supertypes, mean ± SE number of supertypes per individual; π, mean ± SE nucleotide p-distance; aa p-dist, mean ± SE amino acid p-distance; dN, mean ± SE dN; dS, mean ± SE dS

Fig. 3 Within-individual sequence diversity indices of MHC IIB by lake and habitat (mean ± 95% CI). **A** Number of alleles, **B** number of supertypes, **C** nucleotide p-distance π, **D** dN, **E** dS. Populations are from left to right: *A. citrinellus, A. labiatus, A. citrinellus, A. labiatus, A. c.f. citrinellus, A. c.f. labiatus, A. citrinellus* f. benthic, *A. citrinellus* f. limnetic, *A. astorquii, A. chancho, A. zaliosus, A. amarillo, A. xiloaensis, A. sagittae*. Light coloured points give the values for each individual

number of private alleles were all independent of sample size both at the level of populations and lakes.

Twenty-nine alleles out of 150 occurred in at least 5% of individuals (Fig. 4A), and were considered as common alleles. Forty-six were singleton alleles. One allele, Amci-DXB*040101, was present in all individuals, and another allele, ac001, was present in all but one. Common alleles varied greatly in frequency among populations

Table 4 Model statistics for sequence diversity indices, allele pools, and supertype pools across Midas cichlid populations

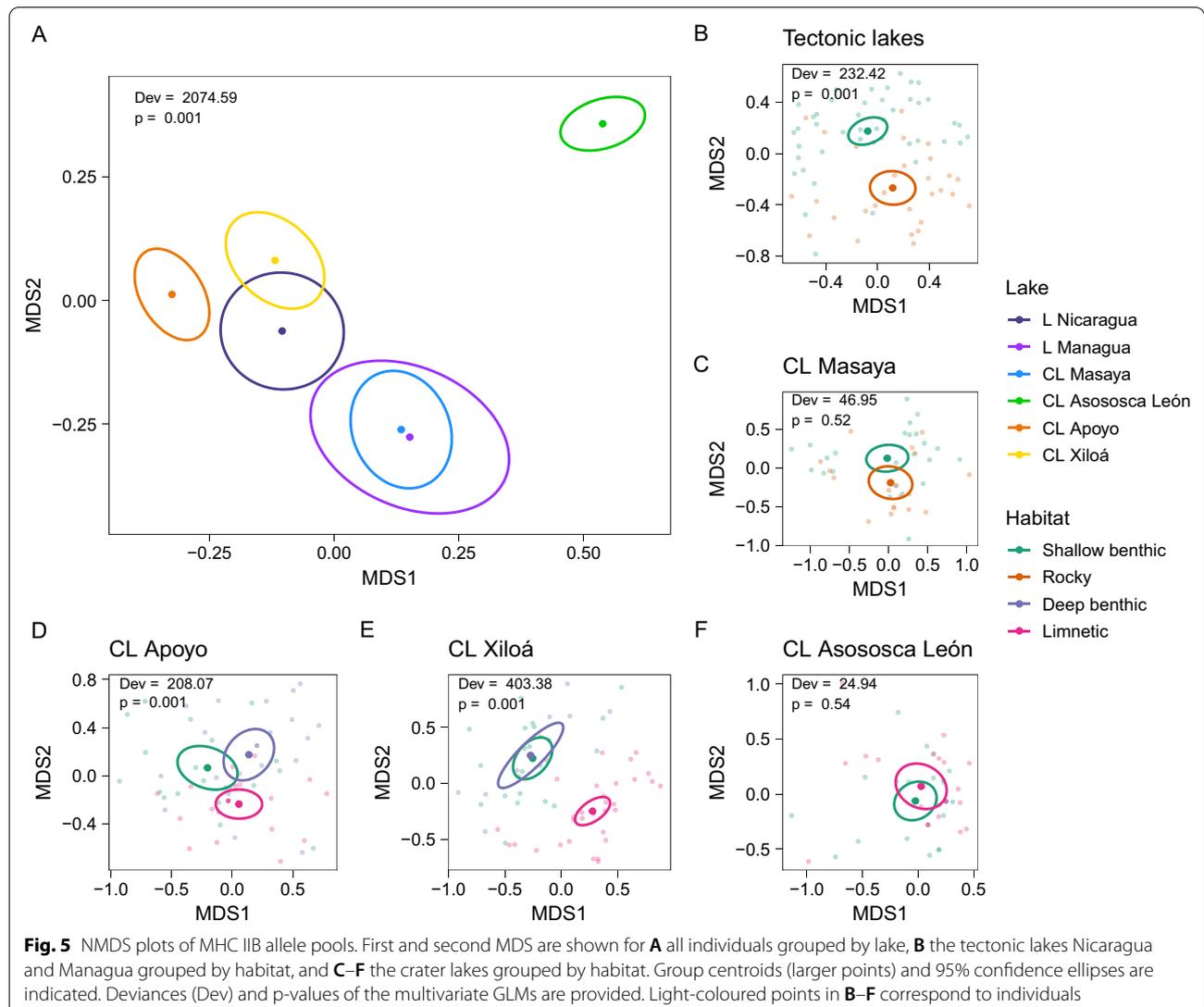| Predictor	df	No. alleles		No. supertypes		π			dN			dS			Allele pools		Supertype pools	
		Chisq	p-value	Chisq	p-value	Sum Sq	F	p-value	Sum Sq	F	p-value	Sum Sq	F	p-value	Dev	p-value	Dev	p-value
Population	13	89.338	**<0.001**	74.892	**<0.001**	0.008	3.877	**<0.001**	0.03	7.306	**<0.001**	0.03	0.769	0.693	1858.01	**0.001**	472.26	**0.001**
Residuals	273					0.043			0.08			0.736						
Lake	5	63.275	**<0.001**	54.005	**<0.001**	0.005	5.768	**<0.001**	0.019	12.909	**<0.001**	0.015	1.139	0.340	2074.59	**0.001**	334.51	**0.001**
Habitat	3	7.847	**0.049**	4.67	0.198	0.003	5.506	**0.001**	0.008	8.761	**<0.001**	0.001	0.181	0.909	694.91	**0.001**	45.20	**0.024**
Lake × Habitat	5	19.052	**0.002**	11.852	**0.037**	0.001	0.633	0.674	0.001	0.941	0.46	0.008	0.570	0.723	244.51	**0.001**	92.56	**0.001**
Residuals	273					0.043			0.080			0.736						

Significant *p*-values are highlighted in bold
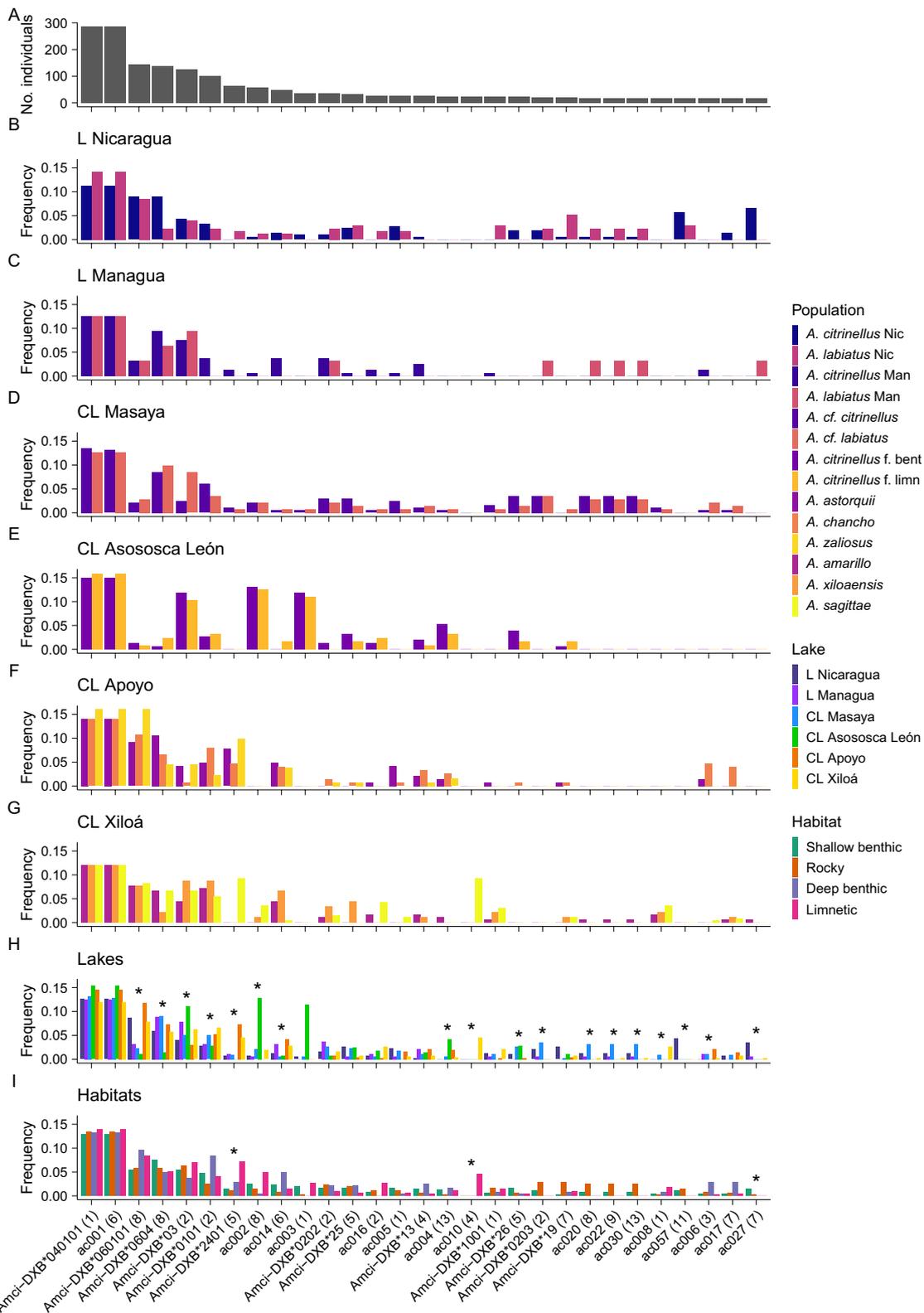
Fig. 4 Distribution of MHC IIB alleles present in > 5% of individuals. **A** Number of individuals carrying an allele in the entire data set, **B**–**G** frequency of the alleles within populations split by lake, **H** frequency of alleles within lakes and **I** habitats. Alleles labelled with asterisks (*) contributed to allele pool differences among groups. The supertype that an allele was assigned to is given in parenthesis. Bent = benthic, limn = limnetic

Bracamonte *et al. BMC Ecology and Evolution*   (2022) 22:41

Page 10 of 17

(Fig. 4B–G) and cichlid communities of each lake (Fig. 4H). Five alleles were present in all populations and 11 were present in all lake communities. Each supertype was present in at least 26 individuals (ca. 10%). Supertype 1, consisting of 29 alleles, was found in all individuals, and supertype 6, with only 7 alleles, was found in all but one. Supertype 2 (6 alleles) and supertype 8 (4 alleles) were found in 220 (77%) and 255 (89%) individuals, respectively (Additional file 1: Fig. S4A). One supertype was absent in fish from CL Masaya (12), two in fish from CL Apoyo (9, 11), and three were absent in fish from CL Asososca León (10, 11, 12; Additional file 1: Fig. S4C).

Allele pools differed among populations (Table 4). This was due to 26 alleles that differed in frequency, 20 of them being common alleles (Fig. 4B–G). The frequency of all supertypes also differed among populations (Table 4, Additional file 1: Fig. S4B). Allele and supertype pools

further differed among fish in different lakes, and fish in different habitats across lakes, and the interaction term was also significant (Table 4). Thirty-three alleles (18 being common alleles; Fig. 4H) and 8 supertypes (Additional file 1: Fig. S4C) differed in abundance in fish from different lakes, while 4 alleles (3 of them common; Fig. 4I) and 1 supertype (Additional file 1: Fig. S4D) showed different abundances among fish living in different habitats. Alleles and supertypes that differentiated fish from different habitats also contributed to differentiation of populations among lakes. Pairwise comparisons indicated that, averaged over habitats, fish in each lake had distinct allele and supertype pools that differed from those of fish in all other lakes (Fig. 5A, Additional file 1: Fig. S5). Similarly, allele and supertype pools of all fish exploiting a specific habitat were distinct when averaged over lakes. A codon usage analysis at PSS indicated that MHC alleles between

Fig. 5 NMDS plots of MHC IIB allele pools. First and second MDS are shown for **A** all individuals grouped by lake, **B** the tectonic lakes Nicaragua and Managua grouped by habitat, and **C**–**F** the crater lakes grouped by habitat. Group centroids (larger points) and 95% confidence ellipses are indicated. Deviances (Dev) and p-values of the multivariate GLMs are provided. Light-coloured points in **B**–**F** correspond to individuals

populations in a specific habitat were more similar than expected with shared ancestry (91.6–96.1% observed vs 86.1–90.0% expected identical codons, all p-values < 0.001) and they were highly unlikely to have arisen by convergent evolution (60.3–65.3% expected identical codons, all p-values < 0.001). Within lakes, populations exploiting different habitats differed in their respective allele pools in the large lakes (Fig. 5B) and in crater lakes Apoyo (Fig. 5D) and Xiloá (Fig. 5E), but not in crater lakes Masaya (Fig. 5C) and Asososca León (Fig. 5F).

Discussion

In this study we investigated the potential contribution of MHC class IIB variation to diversification in the Neotropical crater lake Midas cichlid radiation, one of the most convincing cases of sympatric parallel speciation [51, 52, 61]. We report extensive allelic and also putative functional diversity within and among Midas cichlid populations. Fish from different lakes showed divergent MHC IIB allele pools, as did fish exploiting different habitats within lakes, although no clear parallelism was found. High MHC IIB diversity in the Midas cichlid in particular [47], and in cichlids in general [45], may facilitate the propensity to speciate in this group, as it may facilitate differential local responses and niche specialization upon colonizing new habitats.

The large allelic diversity detected in the Midas cichlid resulted in most alleles being rare, and each occurring only in few individuals. Less than 20% of the alleles were detected in at least 5% of the individuals. Of the 29 most common alleles, eleven were present in fish from all lakes, and only five were recovered from all populations. Since antigen specificity of MHC alleles is predominantly determined by the antigen-binding sites rather than the full coding sequence [62], we clustered alleles into functional supertypes of putatively similar specificity [63]. The 150 alleles recovered in the Midas cichlid converged into 13 supertypes. All supertypes were fairly common across populations, and in each population at least nine were recovered. This suggests that maintaining diverse functionality is relevant for coping with parasite infections, but which variant provides the function may be secondary. This is in line with the hypothesis that balancing selection acts to maintain functional supertypes with rapid turn-over of alleles within supertypes due to arms race dynamics which is supported by a simulation study [64]. Balancing selection on supertypes rather than on alleles can therefore explain why populations and species tend to be markedly differentiated at MHC alleles despite the commonly observed pattern of lineage and supertype sharing. Codon usage indicated that the pattern of lineage sharing was indeed due to shared ancestry and is not an artefact of convergent evolution on sites involved

in antigen binding. The importance of functional rather than sequence diversity has also been proposed for other fish species [64, 65], amphibians [66], birds [67], and mammals [68].

We detected lake-specific signatures in the Midas cichlid MHC IIB diversity, suggesting that either phylogenetic ancestry and demographic history have shaped immunogenetic diversity, or within lake characteristics are driving divergent selection among them. Midas cichlid species and populations are more closely related within lakes than among lakes [51, 54, 69], and MHC IIB evolution may at least partially be governed by neutral processes due to colonization patterns. Evidence for neutral evolution at MHC has been described in different systems and at different scales [70, 71]. MHC allele composition of fish in CL Masaya, although differentiated, resembled more that of fish in the great lakes than in any other crater lake. This pattern is also observed for neutral genetic variation and morphological divergence and suggests more recent connections or faunal exchange [61]. On the other extreme CL Asososca León harbours the most genetically differentiated population of all lakes [51, 54], and the most distinct MHC IIB signature was observed for fish of this lake. Also, fish from CL Apoyo and from CL Xiloá each show unique MHC IIB signatures, in line with clear genome-wide divergence [51, 54]. Similarly, populations in the isolated crater lakes Asososca León and Apoyo, that suffered from marked founder effects and/or bottlenecks during their colonization history [51], have eroded MHC diversity. Notably, three widespread supertypes that are present in individuals of all other lakes, are missing in fish from CL Asososca León and two are missing in fish from CL Apoyo. Also, the total number of alleles and functional supertypes per population and the number of alleles and supertypes per individual were lowest within these two lakes. Indeed, population bottlenecks were repeatedly shown to cause severe reductions of MHC diversity that may exceed those at neutrally evolving loci [72, 73]. However, we did not observe a reduction in other sequence diversity parameters, suggesting that to some degree selection has maintained MHC diversity.

Divergent MHC IIB signatures between fish inhabiting different lakes may also be a result of lake features affecting entire species flocks. Abiotic parameters, such as salinity or nutrient levels differ considerably among the Nicaraguan lakes [74, 75]. These can affect parasite communities and may consequently select for different sets of MHC alleles and supertypes among fish inhabiting different lakes. Furthermore, the isolated crater lakes Asososca León and Apoyo harbour an impoverished fish fauna [51, 75] and fewer parasite taxa (Santacruz et al., [44]) compared to the more connected crater lakes Masaya and Xiloá. This difference in fish and parasite diversity is

paralleled by differences in MHC IIB diversity among fish inhabiting these lakes. Reduced MHC diversity in populations exposed to less diverse parasite communities is well documented in the well-studied three-spined stickleback [32, 71], and this may also apply to Midas cichlid populations.

Ecomorphologically similar Midas cichlid species assemblages have evolved among lakes providing similar habitat structures [51, 61, 76], despite the unique properties of each lake. Accordingly, we found an MHC IIB signature linked to habitat type, although it was less pronounced than the lake-specific signature. This pattern is the result of divergence from an ancestral allele pool, rather than convergence, consistent with the colonization history and the young age of the Midas cichlid radiations. Notably, limnetic fish had higher genetic diversity per individual than their benthic counterparts when averaged across lakes. Furthermore, allele pools in the deeper crater lakes Apoyo and Xiloá were most divergent between benthic and limnetic species. This is not fully consistent with the phylogenetic relationships in these two lakes which provide evidence that the limnetic species is most distantly related to the benthic species in CL Apoyo but not in CL Xiloá [54, 76]. Allele pools also diverged between *A. citrinellus* and *A. labiatus* that inhabit sandy and rocky substrates in the great lakes, respectively. On the other hand, MHC IIB allele pools were not differentiated between morphotypes in crater lakes Masaya and Asososca León. In these lakes ecomorphotypes are not considered different species, and rather form single polymorphic populations within each lake [61]. Divergence in habitat use goes along with divergence in a number of characteristics that have the potential to alter selection on MHC genes. Parasite communities often differ among habitats, lake and stream, benthic and limnetic, as reported in several freshwater fish species (sticklebacks, whitefish or African cichlids; [18, 40, 77, 78]). Also, different feeding preferences were shown to influence exposure to divergent trophically transmitted parasites in freshwater fish [79–82]. The Midas cichlid species may therefore encounter contrasting parasite communities along the different habitats and prey items they have specialized on. In African cichlids, feeding strategies were shown to affect both MHC allele pool and parasite community compositions [42]. Furthermore, the gut microbiota of the Midas cichlid may have diverged among the benthic-limnetic axis and with trophic ecology within lakes, although with limited parallelism among lakes ([83, 84] but see [85]). Commensal and symbiotic microbial diversity was found to be associated with MHC diversity across the vertebrate clade [27, 28, 86, 87], hence it is conceivable that the microbiota may also exert selective pressure on the Midas cichlid MHC.

Despite the divergence of MHC IIB among habitats, distinct sets of alleles and supertypes were involved in the divergence of populations within each lake. This suggests that differences in selection pressure among habitats may not follow a fully parallel pattern among lakes. An absence of parallelism in MHC allele pools was also reported for whitefish, in which it was associated with a corresponding non-parallelism in microbial pathogens [34]. Information on parasite and microbial infection patterns in the Midas cichlid can elucidate the extent to which such intimate species interactions shape divergence, assortative mating, and speciation in this system.

Conclusions

Our results provide evidence for substantial immunogenetic divergence among allopatric populations and between habitats among sympatric populations in the Midas cichlid species complex. Despite indication of habitat-specific signatures on MHC IIB, we do not recover clear patterns of parallel divergent selection in ecomorphologically similar species. This suggests that colonization and demographic processes may have shaped MHC evolution along with natural selection.

Methods

Sampling

Samples of the Midas cichlid radiation were collected from the two great Nicaraguan lakes, Nicaragua and Managua, and four crater lakes, Apoyo, Asososca León, Masaya, and Xiloá, in December 2009 and 2010. We collected approximately 20 individuals per species or described morphotype inhabiting clearly distinct habitats within each of the six lakes (Fig. 1, Table 1). Fish were caught with gill nets and anaesthetised with Tricaine mesylate (MS-222). Fish were photographed for species and morphotype identification and euthanised on ice. Fin clips were preserved in 100% ethanol for DNA analysis. All methods were carried out in accordance with current Spanish and European Union laws (ECC/566/2015 and 2010/63/UE, respectively) and with the permission of the Ministerio del Ambiente y los Recursos Naturales Nicaragua (MARENA; permit number 001-012012). The study is reported in accordance to ARRIVE guidelines.

DNA extraction and sequencing

DNA was extracted with the DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer's recommendations with RNase digestion. DNA concentrations were measured with a NanoDrop 1000 (Thermo Fisher Scientific, Bonn, Germany), standardized to 20 ng/μl, and re-quantified with Qubit (Thermo Fisher Scientific).

For PCR amplification of MHC class IIB exon 2, the forward primer AcMHCIIBF5 (5′ CCACKGAGCTG AASGACATSGAG 3′) was used that discriminates against the putatively non-classical alleles [47]. Since the 3′-end of exon 2 was not conserved enough for a single primer to recover all previously characterized *Amphilophus* alleles, two reverse primers were newly designed (AcMHCIIBR10, 5′ GCAGTAYNTCYCCYTCTGAG 3′ and AcMHCIIBR11 5′ GCAGWMTSTCTCCTTYKCAG 3′). These new primers are designed to recover a 142 bp fragment of exon 2 of all but two rare alleles of the known *Amphilophus* alleles.

PCR amplification, library preparation, and amplicon sequencing were done at LGC Genomics (Berlin, Germany). In short, PCR was done in a volume of 20 µl using the MHC-specific primers, tailed with 20 bp derived from Illumina adapters. The mix consisted of 1× MyTaq buffer containing 1.5 U MyTaq polymerase (Bioline, Luckenwalde, Germany), 15 pmol of forward and reverse primers, 2 µl of BioStabII PCR Enhancer (Sigma-Aldrich, Taufkirchen, Germany), and 10–80 ng of DNA. The cycling parameters consisted of 2 min denaturation at 96 °C followed by 20 cycles of 96 °C for 15 s, 50 °C for 30 s, 70 °C for 60 s. Primers were removed with ExoI digestion (New England Biolabs, Ipswich, USA). For library preparation, a second round of PCRs was done with dual index combinations of full-length Illumina adaptors. PCR products were pooled and purified using 1 vol AgencourtXP beads (Agilent) and used in an emulsion PCR with standard Illumina primers using buffers and enzymes from the emPCR Kit (Roche 454) and the oil-surfactant-mixture from the Micellula DNA Emulsion & Purification Kit (EURx). The emulsion was broken and DNA purified following the instructions of the Micellula DNA Emulsion & Purification Kit. A final size selection was done on an LMP Agarose gel. PCR products were sequenced on an Illumina MiSeq V3 resulting in approximately 5 M paired-end reads of 2 × 300 bp.

Data processing and allele calling

Libraries were demultiplexed with bcl2fastq v1.8.4 (Illumina Inc.), allowing up to two mismatches per barcode. Paired reads with incomplete or conflicting barcodes were discarded, as were reads shorter than 100 bases and those containing more than one N. Remaining reads were trimmed at the 3′ end to achieve an average Phred score ≥ 20 over a window of 10 bases. For primer clipping, up to three mismatches were allowed per primer and primer pair. Sequences that did not contain forward and reverse primers were discarded. Primer-clipped forward and reverse read pairs were merged with BBmerge v34.48 [88].

Alleles were called using the command line version of AmpliSAS [89]. Reads were clustered with default parameters for Illumina technology. Per amplicon frequencies (PAF) of clusters, sorted in descending order, were inspected for a distinct drop in frequency which would indicate an optimal PAF [90]. In most samples, this drop occurred between 1 and 2%. PAF (min_amplicon_seq_frequency) was then set to 1.5% which gave a consistent genotype for the sample that was amplified and sequenced in duplicate. The maximum number of expected alleles per individual was set to 50. *Amphilophus* alleles identified by Hofmann et al. [47] were supplied as reference. Alleles newly identified in this study were numbered continuously starting with ac001.

This approach led to somewhat lower numbers of alleles per individual than previously described for *Amphilophus* [47]. Alleles may differ outside the sequenced fragment, and the more stringent filtering approach in this study may collapse alleles differing in only few bases. Previously identified alleles that were grouped together in the current study mostly differed only in synonymous substitutions or < 2 amino acids in the entire exon 2. Since our main focus lies on detecting signatures of divergence, this represents a conservative approach. Also, such high sequence similarity likely represents functional similarity [63] and will only have a minor impact on supertype inference.

Molecular sequence analyses

Alleles were aligned and variable sites were identified for nucleotide and translated amino acid sequences. Amino acid variability was visualized with a sequence motive logo using the R package ggseqlogo v0.1 [91]. Overall nucleotide p-distance (π), overall amino acid p-distance (aa p-dist), and the overall number of non-synonymous substitutions per non-synonymous site (dN) and synonymous substitutions per synonymous site (dS) were calculated in MEGA X [92]. A gamma distribution with a site-rate parameter of 0.39 was assumed, as indicated by the best-fit substitution model determined with MEGA X. For dN and dS, the Nei–Gojobori method with Jukes–Cantor correction was used. Variance was estimated with 9999 bootstrap replicates.

The phylogenetic relationship among alleles was inferred with a split network constructed in SplitsTree5 v5.0.0 alpha [93] using the Jukes–Cantor method to calculate the distance matrix. The phylogenetic relationship was further determined with Bayesian inference implemented in MrBayes v3.2.7a [94] using codon substitution models. The ω parameter was set to the Ny98 model which allows ω to vary among codons and identifies positively selected sites. Four independent runs with eight chains each with a

heating parameter of 0.1 and two swaps per cycle were run for 5×10^6 generations and sampled every 1000 generations. The first 25% were discarded as burn-in. Default values were used for the remaining parameters. Convergence among runs was examined with the R package RWTY v1.0.2 [95]. Majority consensus trees were visualized in R [96].

We estimated historical positive selection over evolutionary timescales on MHC alleles based on ratios of non-synonymous to synonymous substitutions among all alleles. Analyses were conducted on all sites. Gene-wide positive selection on alleles, i.e. along all codons (sites) of all alleles, was estimated with a codon-based Z-test using the Nei–Gojobori distance model with Jukes–Cantor correction and partial deletion in MEGA X. Variance estimation was done with 9999 bootstrap replicates. Gene-wide episodic positive selection was estimated for the full phylogeny with the branch-site unrestricted statistical test for episodic diversification (BUSTED; [97]) available on the datamonkey server [98]. BUSTED tests whether at least one site on at least one branch of the allele phylogeny has evolved under positive selection. Site-specific positive selection was estimated with CodeML implemented in PAML v4.9j [99], the mixed effects model of evolution (MEME; [100]) on the datamonkey server, and the Ny98 model in MrBayes (see above). For CodeML, positive selection was identified by comparing models M1a (nearly neutral, $0 < \omega_0 < 1$, $\omega_1 = 1$) and M2a (positive selection, $0 < \omega_0 < 1$, $\omega_1 = 1$, $\omega_2 > 1$) and models M7 (10 ω classes following a β distribution) and M8 ($\beta + \omega > 1$) using codon frequencies estimated from a F3 × 4 nucleotide frequency model. One ω ratio was used for all branches. The codon of amino acid site 41 was excluded due to a deletion in one allele. Likelihood ratio test were used to compare the respective models and PSS were identified with the Bayes empirical Bayes method. The majority consensus tree from MrBayes was used as starting tree. Sites identified by at least two methods were considered to be under positive selection.

Alleles were clustered into functional supertypes following Sepil et al. [101]. For this, 5 z-scores describing the physicochemical properties of amino acids [102] were assigned to each position of unique sequences of concatenated PSS. The resulting matrix was used for discriminant analysis of principle components (DAPC) implemented in the R package adegenet [103, 104]. The optimal number of clusters was identified by K-means clustering using the find.cluster function with the "goodfit" selection criterion. Automatic cross-validation (xvalDapc function) was used for the DAPC. Alleles were assigned to supertypes based on consensus of 5×10 clusterings.

Population divergence analyses

The total number of alleles, private alleles, and super-types were identified for each population and for each lake. The number of alleles and supertypes per individual was counted. Four additional within-individual sequence diversity indices were calculated in MEGA X [92] using the same parameters as for estimating overall sequence diversity (see above): π, aa p-dist, dN, and dS. Statistical analyses were done in R [96]. Within-individual sequence diversity indices were compared among populations and among lakes, habitats, and their interaction with generalized linear models (GLM). For comparing the number of alleles and the number of supertypes, a quasipoisson distribution with a log link function was used and significance was inferred with X^2 tests. For comparisons involving all other diversity indices, linear models (LM) were used. Post hoc pairwise comparisons using Tukey's HSD and custom comparisons were done with the emmeans package v1.4.7 [105].

Alleles present in at least 5% of individuals and singleton alleles were identified. For alleles present in >5% of individuals, frequencies per population and per lake were calculated. To assess whether the allelic composition and the supertype composition of individuals differ among populations and among lakes and habitats, multivariate GLM were performed using the function manyglm of the mvabund package v4.1.3 [106]. Likelihood-ratio-tests were used to evaluate the significance of model terms. Alleles contributing to the difference between groups were identified from the models by reporting univariate statistics with adjusted p-values. Differences in allelic composition were assessed on the full data set. Pairwise multivariate GLM were done to identify which lakes and habitats differ from each other. Benjamini–Hochberg adjustment of p-values was used to account for multiple testing. GLMs were used to assess if populations within tectonic lakes and within each crater lake segregate by habitat based on allelic composition. For lakes with >2 habitats, this was followed by pairwise multivariate GLM. Segregation of allelic composition was visualized with nonmetric multidimensional scaling (NMDS) based on Jaccard dissimilarities using the vegan package v2.5-6 [107]. Six dimensions were chosen which resulted in a stress close to 0.1. A minimum of 30 and a maximum of 75 random starts were performed, each with a maximum of 1000 iterations. NMDS were calculated for all lakes combined and separately for the tectonic lakes and each crater lake. Four to five dimensions were chosen for single lake analyses.

For populations exploiting the same habitat in different lakes, a codon usage analysis was done with custom scripts provided by Lenz et al. [108] to infer whether co-ancestry or convergent evolution is the most likely scenario for

Bracamonte *et al. BMC Ecology and Evolution*   (2022) 22:41

Page 15 of 17

sharing of similar alleles. In brief, for each population pair the number of identical amino acids at PSS was counted between all allele pairs, recording whether they were coded by the same codon. The number of identical codons was compared to a theoretical distribution of expected identical codons under the convergent evolution scenario. This distribution was obtained with 1000 Monte Carlo simulations based on observed codon frequencies of the full sequence in both populations and the observed number of identical amino acids at PSS. Similarly, a theoretical distribution of expected identical codons was obtained for the co-ancestry scenario. Proportion tests were used to estimate significance for each scenario.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12862-022-01997-9.

Additional file 1: Figure S1. Alignment of 142 bp of exon 2 of 150 *Amphilophus* MHC class IIB alleles. Conserved sites are shaded in grey. **Figure S2.** Amino acid variability for 47 positions of 150 aligned Midas cichlid MHC class IIB alleles. **Figure S3.** Phylogenetic relationship of *Amphilophus* MHC IIB alleles. **Figure S4.** Distribution of MHC IIB supertypes. **Figure S5.** NMDS plot for all individuals grouped by lake showing the variation among individuals.

Authors' contributions
MB conceived the study. MJH and SEB carried out the experimental work with the supervision of CE. SEB and CLM analysed the data. SEB wrote the first draft and all authors contributed to subsequent drafts. All authors read and approved the final manuscript.

Availability of data and materials
Raw reads were deposited in the NCBI Sequence Read Archive (BioProject PRJNA736256). Allele sequences are available from GenBank (Accession numbers: MZ368756-MZ368841, MZ368843-MZ368885).

Declarations

Ethics approval and consent to participate
The study protocol was approved by The Ministry of Natural Resources (MARENA) of Nicaragua (No. 001-012015). Methods to euthanise fish were carried out in strict accordance with current Spanish and European Union laws (ECC/566/2015 and 2016/63/UE, respectively), and approved by the CEEA-MNCN. Voucher individuals were deposited in the collections of the Museo Nacional de Ciencias Naturales CSIC in Spain.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
¹Museo Nacional de Ciencias Naturales, CSIC, José Gutiérrez Abascal 2, 28006 Madrid, Spain. ²School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK.

References
1. Blom MPK, Horner P, Moritz C. Convergence across a continent: adaptive diversification in a recent radiation of Australian lizards. Proc R Soc B Biol Sci. 2016;283:20160181.
2. Magalhaes IS, Whiting JR, D'Agostino D, Hohenlohe PA, Mahmud M, Bell MA, et al. Intercontinental genomic parallelism in multiple three-spined stickleback adaptive radiations. Nat Ecol Evol. 2020;193:1–11.
3. Muschick M, Indermaur A, Salzburger W. Convergent evolution within an adaptive radiation of cichlid fishes. Curr Biol. 2012;22:2362–8.
4. Ravinet M, Westram A, Johannesson K, Butlin R, André C, Panova M. Shared and nonshared genomic divergence in parallel ecotypes of *Littorina saxatilis* at a local scale. Mol Ecol. 2016;25:287–305.
5. Nosil P. Ecological speciation. New York: Oxford University Press Inc.; 2012.
6. Windsor DA. Most of the species on earth are parasites. Int J Parasitol. 1998;28:1939–41.
7. Eizaguirre C, Lenz TL, Traulsen A, Milinski M. Speciation accelerated and stabilized by pleiotropic major histocompatibility complex immunogenes. Ecol Lett. 2009;12:5–12.
8. Karvonen A, Seehausen O. The role of parasitism in adaptive radiations—when might parasites promote and when might they constrain ecological speciation? Int J Ecol. 2012. https://doi.org/10.1155/2012/280169.
9. 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.
10. Yoder JB, Nuismer SL. When does coevolution promote diversification? Am Nat. 2010;176:802–17.
11. Betts A, Gray C, Zelek M, MacLean RC, King KC. High parasite diversity accelerates host adaptation and diversification. Science. 2018;360:907–11.
12. Buckling A, Rainey PB. Antagonistic coevolution between a bacterium and a bacteriophage. Proc R Soc Lond B Biol Sci. 2002;269:931–6.
13. 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.
14. Lazzaro BP, Little TJ. Immunity in a variable world. Philos Trans R Soc B Biol Sci. 2009;364:15–26.
15. Klein J. Natural history of the major histocompatibility complex. 99th ed. New York: Wiley; 1986.
16. Murphy K, Weaver C. Janeway's immunobiology. 9th ed. New York: Garland Science, Taylor & Francis Group, LLC; 2017.
17. Eizaguirre C, Lenz TL, Sommerfeld RD, Harrod C, Kalbe M, Milinski M. Parasite diversity, patterns of MHC II variation and olfactory based mate choice in diverging three-spined stickleback ecotypes. Evol Ecol. 2011;25:605–22.
18. Stutz WE, Bolnick DI. Natural selection on MHC II beta in parapatric lake and stream stickleback: balancing, divergent, both or neither? Mol Ecol. 2017;26:4772–86.
19. 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. 2019;123:593–607.

Bracamonte *et al. BMC Ecology and Evolution*   (2022) 22:41

Page 16 of 17

20. 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.

21. Malmstrøm M, Matschiner M, Tørresen OK, Star B, Snipen LG, Hansen TF, et al. Evolution of the immune system influences speciation rates in teleost fishes. Nat Genet. 2016;48:1204–10.

22. 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.

23. Radwan J, Babik W, Kaufman J, Lenz TL, Winternitz J. Advances in the evolutionary understanding of MHC polymorphism. Trends Genet. 2020;36:298–311.

24. Reche PA, Reinherz EL. Sequence variability analysis of human class I and class II MHC molecules: functional and structural correlates of amino acid polymorphisms. J Mol Biol. 2003;331:623–41.

25. Fijarczyk A, Dudek K, Niedzicka M, Babik W. Balancing selection and introgression of newt immune-response genes. Proc R Soc B Biol Sci. 2018;285:20180819.

26. Spurgin LG, Richardson DS. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. Proc R Soc B-Biol Sci. 2010;277:979–88.

27. Bolnick DI, Snowberg LK, Caporaso JG, Lauber C, Knight R, Stutz WE. Major histocompatibility complex class IIb polymorphism influences gut microbiota composition and diversity. Mol Ecol. 2014;23:4831–45.

28. Kubinak JL, Stephens WZ, Soto R, Petersen C, Chiaro T, Gogokhia L, et al. MHC variation sculpts individualized microbial communities that control susceptibility to enteric infection. Nat Commun. 2015;6:13.

29. Hedrick PW. Pathogen resistance and genetic variation at MHC loci. Evolution. 2002;56:1902–8.

30. Gahr CL, Boehm T, Milinski M. Female assortative mate choice functionally validates synthesized male odours of evolving stickleback river–lake ecotypes. Biol Lett. 2018;14:20180730.

31. Milinski M. The major histocompatibility complex, sexual selection, and mate choice. Annu Rev Ecol Evol Syst. 2006;37:159–86.

32. 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.

33. Natsopoulou ME, Palsson S, Olafsdottir GA. Parasites and parallel divergence of the number of individual MHC alleles between sympatric three-spined stickleback *Gasterosteus aculeatus* morphs in Iceland. J Fish Biol. 2012;81:1696–714.

34. Pavey SA, Sevellec M, Adam W, Normandeau E, Lamaze FC, Gagnaire PA, et al. Nonparallelism in MHCII 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.

35. Tobler M, Plath M, Riesch R, Schlupp I, Grasse A, Munimanda GK, et al. Selection from parasites favours immunogenetic diversity but not divergence among locally adapted host populations. J Evol Biol. 2014;27:960–74.

36. Salzburger W. Understanding explosive diversification through cichlid fish genomics. Nat Rev Genet. 2018;19:705–17.

37. Colombo M, Diepeveen ET, Muschick M, Santos ME, Indermaur A, Boileau N, et al. The ecological and genetic basis of convergent thick-lipped phenotypes in cichlid fishes. Mol Ecol. 2013;22:670–84.

38. Kocher TD, Conroy JA, McKaye KR, Stauffer JR. Similar morphologies of cichlid fish in lakes Tanganyika and Malawi are due to convergence. Mol Phylogenet Evol. 1993;2:158–65.

39. 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;2:12.

40. Gobbin TP, Vanhove MPM, Pariselle A, Groothuis TGG, Maan ME, Seehausen O. Temporally consistent species differences in parasite infection but no evidence for rapid parasite-mediated speciation in Lake Victoria cichlid fish. J Evol Biol. 2020;33:556–75.

41. 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.

42. 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.

43. Vanhove MPM, Hablützel PI, Pariselle A, Simkova A, Huyse T, Raeymaekers JAM. Cichlids: a host of opportunities for evolutionary parasitology. Trends Parasitol. 2016;32:820–32.

44. Santacruz A, Barluenga M, Pérez-Ponce de León G. The macroparasite fauna of cichlid fish from Nicaraguan lakes, a modelsystem for understanding host–parasite diversification and speciation. Sci Rep. 2022;12:3944

45. Malaga-Trillo E, Zaleska-Rutczynska Z, McAndrew B, Vincek V, Figueroa F, Sultmann H, et al. Linkage relationships and haplotype polymorphism among cichlid Mhc class II B loci. Genetics. 1998;149:1527–37.

46. Hablützel PI, Volckaert FAM, Hellemans B, Raeymaekers JAM. Differential modes of MHC class IIB gene evolution in cichlid fishes. Immunogenetics. 2013;65:795–809.

47. Hofmann MJ, Bracamonte SE, Eizaguirre C, Barluenga M. Molecular characterization of MHC class IIB genes of sympatric Neotropical cichlids. BMC Genet. 2017;18:15.

48. Klein D, Ono H, Ohuigin C, Vincek V, Goldschmidt T, Klein J. Extensive MHC variability in cichlid fishes of Lake Malawi. Nature. 1993;364:330–4.

49. Andreou D, Eizaguirre C, Boehme T, Milinski M. Mate choice in sticklebacks reveals that immunogenes can drive ecological speciation. Behav Ecol. 2017;28:953–61.

50. Barluenga M, Meyer A. The Midas cichlid species complex: incipient sympatric speciation in Nicaraguan cichlid fishes? Mol Ecol. 2004;13:2061–76.

51. 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;10:20.

52. 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;8:15.

53. Barluenga M, Stolting KN, Salzburger W, Muschick M, Meyer A. Sympatric speciation in Nicaraguan crater lake cichlid fish. Nature. 2006;439:719–23.

54. 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.

55. Kautt AF, Machado-Schiaffino G, Torres-Dowdall J, Meyer A. Incipient sympatric speciation in Midas cichlid fish from the youngest and one of the smallest crater lakes in Nicaragua due to differential use of the benthic and limnetic habitats? Ecol Evol. 2016;6:5342–57.

56. Kusche H, Recknagel H, Elmer KR, Meyer A. Crater lake cichlids individually specialize along the benthic- limnetic axis. Ecol Evol. 2014;4:1127–39.

57. Machado-Schiaffino G, Kautt AF, Torres-Dowdall J, Baumgarten L, Henning F, Meyer A. Incipient speciation driven by hypertrophied lips in Midas cichlid fishes? Mol Ecol. 2017;26:2348–62.

58. Sowersby W, Cerca J, Wong BBM, Lehtonen TK, Chapple DG, Leal-Cardín M, et al. Pervasive admixture and the spread of a large-lipped form in a cichlid fish radiation. Mol Ecol. 2021;30:5551–71.

59. Stauffer JR Jr, McKaye KR. Descriptions of three new species of cichlid fishes (Teleostei: Cichlidae) from Lake Xiloá, Nicaragua. Cuad Investig UCA. 2002;12:1–18.

60. Stauffer JR Jr, McCrary JK, Black KE. Three new species of cichlid fishes (Teleostei: Cichlidae) from Lake Apoyo, Nicaragua. Proc Biol Soc Wash. 2008;121:117–29.

61. Kautt AF, Machado-Schiaffino G, Meyer A. Lessons from a natural experiment: allopatric morphological divergence and sympatric diversification in the Midas cichlid species complex are largely influenced by ecology in a deterministic way. Evol Lett. 2018;2:323–40.

62. Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL, et al. 3-dimensional structure of the human class-II histocompatibility antigen HLA-DR1. Nature. 1993;364:33–9.

63. Doytchinova IA, Flower DR. In silico identification of supertypes for class II MHCs. J Immunol. 2005;174:7085–95.

64. Lighten J, Papadopulos AST, Mohammed RS, Ward BJ, Paterson GI, Baillie L, et al. Evolutionary genetics of immunological supertypes reveals two faces of the Red Queen. Nat Commun. 2017;8:1294.

Bracamonte *et al. BMC Ecology and Evolution*   (2022) 22:41

Page 17 of 17

65. Smallbone W, Ellison A, Poulton S, van Oosterhout C, Cable J. Depletion of MHC supertype during domestication can compromise immuno-competence. Mol Ecol. 2021;30:736–46.

66. Lillie M, Grueber CE, Sutton JT, Howitt R, Bishop PJ, Gleeson D, et al. Selection on MHC class II supertypes in the New Zealand endemic Hochstetter's frog. BMC Evol Biol. 2015;15:63.

67. 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;280:8.

68. 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.

69. Geiger MF, McCrary JK, Schliewen UK. Not a simple case—a first comprehensive phylogenetic hypothesis for the Midas cichlid complex in Nicaragua (Teleostei: Cichlidae: *Amphilophus*). Mol Phylogenet Evol. 2010;56:1011–24.

70. Landry C, Bernatchez L. Comparative analysis of population structure across environments and geographical scales at major histocompatibil-ity complex and microsatellite loci in Atlantic salmon (*Salmo salar*). Mol Ecol. 2001;10:2525–39.

71. Peng F, Ballare KM, Woodard SH, den Haan S, Bolnick DI. What evolu-tionary processes maintain MHCIIβ diversity within and among popula-tions of stickleback? Mol Ecol. 2021;30:1659–71.

72. Eimes JA, Bollmer JL, Whittingham LA, Johnson JA, Oosterhout CV, Dunn PO. Rapid loss of MHC class II variation in a bottlenecked popula-tion is explained by drift and loss of copy number variation. J Evol Biol. 2011;24:1847–56.

73. Sutton JT, Nakagawa S, Robertson BC, Jamieson IG. Disentangling the roles of natural selection and genetic drift in shaping variation at MHC immunity genes. Mol Ecol. 2011;20:4408–20.

74. Barlow G, Baylis J, Roberts D. Chemical analyses of some crater lakes in relation to adjacent Lake Nicaragua. Investig Ichthyofauna Nicar Lakes. 1976;4:17–20.

75. Waid RM, Raesly RL, McKaye KR, McCrary JK. Zoogeografía íctica de lagunas cratéricas de Nicaragua. Encuentro. 1999;51:65–80.

76. Elmer KR, Fan S, Kusche H, Spreitzer ML, Kautt AF, Franchini P, et al. Paral-lel evolution of Nicaraguan crater lake cichlid fishes via non-parallel routes. Nat Commun. 2014;5:5168.

77. Karvonen A, Lundsgaard-Hansen B, Jokela J, Seehausen O. Differentia-tion in parasitism among ecotypes of whitefish segregating along depth gradients. Oikos. 2013;122:122–8.

78. 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.

79. 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.

80. Knudsen R, Amundsen PA, Klemetsen A. Inter- and intra-morph pat-terns in helminth communities of sympatric whitefish morphs. J Fish Biol. 2003;62:847–59.

81. Siwertsson A, Refsnes B, Frainer A, Amundsen P-A, Knudsen R. Diver-gence and parallelism of parasite infections in Arctic charr morphs from deep and shallow lake habitats. Hydrobiologia. 2016;783:131–43.

82. Stutz WE, Lau OL, Bolnick DI. Contrasting patterns of phenotype-dependent parasitism within and among populations of threespine stickleback. Am Nat. 2014;183:810–25.

83. Franchini P, Fruciano C, Frickey T, Jones JC, Meyer A. The gut microbial community of Midas cichlid fish in repeatedly evolved limnetic-benthic species pairs. PLoS ONE. 2014;9:7.

84. Härer A, Torres-Dowdall J, Rometsch SJ, Yohannes E, Machado-Schiaf-fino G, Meyer A. Parallel and non-parallel changes of the gut microbiota during trophic diversification in repeated young adaptive radiations of sympatric cichlid fish. Microbiome. 2020;8:149.

85. Baldo L, Riera JL, Salzburger W, Barluenga M. Phylogeography and eco-logical niche shape the cichlid fish gut microbiota in Central American and African lakes. Front Microbiol. 2019;10:2372.

86. Hernández-Gómez O, Briggler JT, Williams RN. Influence of immunoge-netics, sex and body condition on the cutaneous microbial communi-ties of two giant salamanders. Mol Ecol. 2018;27:1915–29.

87. Leclaire S, Strandh M, Dell'Ariccia G, Gabirot M, Westerdahl H, Bona-donna F. Plumage microbiota covaries with the major histocompatibil-ity complex in blue petrels. Mol Ecol. 2019;28:833–46.

88. Bushnell B, Rood J, Singer E. BBMerge—Accurate paired shotgun read merging via overlap. PLoS ONE. 2017;12:e0185056.

89. 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.

90. Biedrzycka A, Sebastian A, Migalska M, Westerdahl H, Radwan J. Testing genotyping strategies for ultra-deep sequencing of a co-amplifying gene family: MHC class I in a passerine bird. Mol Ecol Resour. 2017;17:642–55.

91. Wagih O. ggseqlogo: a versatile R package for drawing sequence logos. Bioinformatics. 2017;33:3645–7.

92. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evo-lutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35:1547–9.

93. Huson DH, Bryant D. Application of phylogenetic networks in evolu-tionary studies. Mol Biol Evol. 2006;23:254–67.

94. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, et al. MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012;61:539–42.

95. Warren DL, Geneva AJ, Lanfear R. RWTY (R We There Yet): an R package for examining convergence of Bayesian phylogenetic analyses. Mol Biol Evol. 2017;34:1016–20.

96. R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2020. http://www.R-project.org

97. Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward A, et al. Gene-wide identification of episodic selection. Mol Biol Evol. 2015;32:1365–71.

98. Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Kosakovsky Pond SL. Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. Mol Biol Evol. 2018;35:773–7.

99. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24:1586–91.

100. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Pond SLK. Detecting individual sites subject to episodic diversifying selection. PLoS Genet. 2012;8:e1002764.

101. Sepil I, Moghadam HK, Huchard E, Sheldon BC. Characterization and 454 pyrosequencing of Major Histocompatibility Complex class I genes in the great tit reveal complexity in a passerine system. BMC Evol Biol. 2012;12:68.

102. Sandberg M, Eriksson L, Jonsson J, Sjöström M, Wold S. New chemi-cal descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids. J Med Chem. 1998;41:2481–91.

103. Jombart T. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics. 2008;24:1403–5.

104. Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. 2010;11:94.

105. Lenth R. emmeans: Estimated Marginal Means, aka Least-Squares Means. 2020. https://CRAN.R-project.org/package=emmeans

106. Wang Y, Naumann U, Wright ST, Warton DI. mvabund—an R package for model-based analysis of multivariate abundance data. Methods Ecol Evol. 2012;3:471–4.

107. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. Vegan: community ecology package. 2019. https://CRAN.R-project.org/package=vegan

108. Lenz TL, Eizaguirre C, Kalbe M, Milinski M. Evaluating patterns of conver-gent evolution and trans-species polymorphism at MHC immunogenes in two sympatric stickleback species. Evolution. 2013;67:2400–12.

Publisher's Note