8-30-2022

Polymorphism & Parasites: Structure, Diversity and Selection of the MHCII Genes in a Weakly Electric Fish, Brachyhypopomus Occidentalis

Bruna L. Silva
University of Massachusetts Boston

POLYMORPHISM & PARASITES:

STRUCTURE, DIVERSITY AND SELECTION OF THE MHCII GENES IN A WEAKLY ELECTRIC FISH, *BRACHYHYPOPOMUS OCCIDENTALIS*

A Thesis Presented

by

BRUNA L. SILVA

Submitted to the Office of Graduate Studies,

University of Massachusetts Boston,

In partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

AUGUST 2022

Biology Program

POLYMORPHISM & PARASITES: STRUCTURE, DIVERSITY AND SELECTION OF THE MHCII GENES IN A WEAKLY ELECTRIC FISH, *BRACHYHYPOPOMUS OCCIDENTALIS*

A Thesis Presented
by
Bruna Silva

Approved as to style and content by:

_____
Luis De León, Assistant Professor
Chairperson of Committee

_____
Mayra Cadorin Vidal, Assistant Professor
Member

_____
Brook Moyers, Assistant Professor
Member

_____
Linda Huang, Program Director
Biology Program

_____
Rachel Skvirsky, Chair
Biology Department

ABSTRACT

POLYMORPHISM & PARASITES: STRUCTURE, DIVERSITY AND SELECTION OF THE MHCII GENES IN A WEAKLY ELECTRIC FISH, *BRACHYHYPOPOMUS OCCIDENTALIS*

August 2022

Bruna L. Silva, B.A., University of New Hampshire
M.S., University of Massachusetts Boston

Directed by Dr. Luis De León

The major histocompatibility complex (MHC) class II molecules play a key role in inducing an immune response, by presenting foreign peptides to T-lymphocytes. They are considered one of the most polymorphic genes in the vertebrate genome and diversity has been associated with species diversification mediated by parasite, viral and bacterial infections. While MHC genes are well documented in teleost fish, none thus far have been described in the Gymnotiform order – a highly diverse group of neotropical electric fishes. Using a combination of a recently annotated genome and whole genome resequencing data, I identified and characterized both the classical MHCII DAB, DAA and non-classical DBB genes in populations of the electric fish *Brachyhypopomus occidentalis*. I found highly polymorphic sites within the classical MHCII genes, and significant genetic divergence between populations widespread throughout the isthmus of Panama. To explore whether drift or selection is driving genetic diversity in MHCII genes, I compared variation between

the MHCIIβ and a neutral mitochondrial gene (COI). Geographic distance between sites was only correlated with variation in the COI gene, suggesting that selective pressures could be driving diversification in the MHCIIβ. Further analysis showed that prevalence of parasites within drainages was associated with rate of non-synonymous mutations in the MHCIIβ, which highlights the potential role parasites play in driving genetic diversity. Overall, my results highlight the highly polymorphic nature of the MHCII genes in electric fishes, and provides evidence for selection as a driver of variation at this locus. Although further analyses are needed, these findings contribute to a better understanding of what drives diversification in Neotropical electric fishes.

ACKNOWLEDGEMENTS

I couldn’t have done this without the support and expertise of Dr. Luis De León and Dr. Carlos Arias. Nor without my incredible lab-mates, Elaine, Mickayla, and Daniel for all the advice and emotional support. I want to acknowledge my wonderful boyfriend and friends who have kept me grounded throughout this process. Lastly, my cat Lincoln, for the constant companionship while I worked. Furthermore, this project would never have happened without the generous funding from the Secretaría Nacional de Ciencia, Tecnología e Innovación (Panama), UMass Boston, and the McNair program.

TABLE OF CONTENTS

CHAPTER Page

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Terminology

Major Histocompatibility Complex Structure

The major histocompatibility complex (MHC) is a set of genes which encode for molecules vital in the immune response of vertebrates (Altmann and Trowsdale 1989). Two major classifications for MHC molecules exist, Class I and II. Class I molecules are found on all nucleated cells and function by binding to viral peptides and presenting them to CD8+ T-cells, which in turn exterminates the infected cell (Bjorkman and Parham 1990). Class II genes encode for molecules only located on specialized antigen presenting cells (macrophages, dendric and B-cells) and bind to extracellular foreign peptides (Alfonso & Karlsson, 2000; Hughes & Nei, 1989). MHCII molecules work with CD4+ T-cells to coordinate an appropriate immune response (Podivinský 1966; Engelhard 1994) and play a key role in the adaptive immunity of an individual. The Class II molecule is comprised of an alpha and beta chain non-covalently associated to form a heterodimer (Figueroa et al. 1997; Amadou et al. 1999; Doherty and Zinkernagel 1975). Each chain is composed of two extracellular domains ($\alpha1/\alpha2$ and $\beta1/\beta2$ respectively), and the connecting peptide (CP), transmembrane (TM), and cytoplasmic tail (CY) regions (Fig 1).

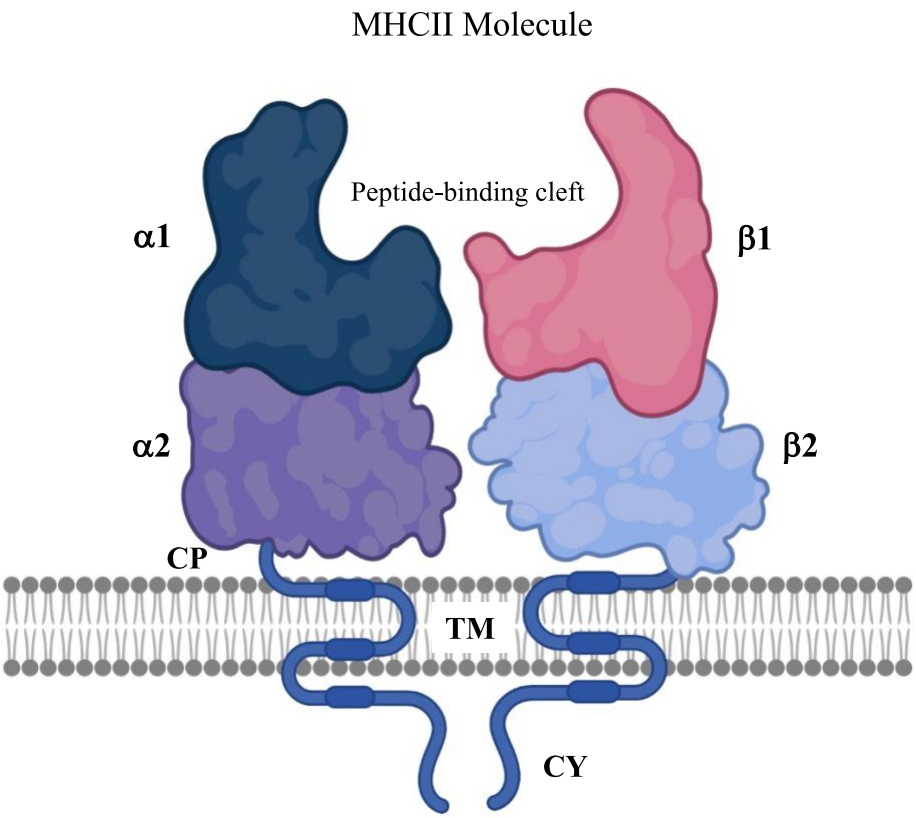

Fig 1. Structure of the MHCII α and β chains. The extracellular domains are composed of the α1/α2 and β1/β2, while the connecting peptide (CP), transmembrane (TM), and cytoplasmic tail make up the remainder of the molecule.

Classical and Non-Classical

MHC genes can be further sorted into Classical and non-Classical categories. Classical MHC genes have been the focus of most studies; attention has especially been given to the first domain of the classical α and β gene. The α1 and β1 domains interact to form an extracellular groove made of a β-sheet and two α-helix walls that directly bind with foreign peptides (Alfonso & Karlsson, 2000; Hughes & Nei, 1989). Overall, classical MHC genes are highly polymorphic, and diversity is thought to be maintained through

balancing selection (Doherty & Zinkernagel, 1975; Klein & O'Huigin, 1994; Takahata & Nei, 1990; Wegner, 2004). Non-Classical MHC genes can influence the peptide binding process, but they lack the ability to directly bind with antigens and are therefore often less diverse than classical genes (Alfonso and Karlsson 2000). Unlike in most vertebrates, classical and non-classical genes are often unlinked in teleost (Sato et al. 2000) and there is some uncertainty to the function of non-classical molecules (Alfonso and Karlsson 2000). A single classical MHCII lineage exists in teleost (DA-) and two non-classical lineages (DB- and DE-) (Onott et al., 1992; Sambrook et al., 2005).

Parasites

Although number of parasite species are grossly undercounted, it is estimated that parasites are one of the most abundant life forms to exist (Dobson et al. 2008). Parasites influence their hosts in a variety of negative ways, and due to their typically short lifespan they can multiply and adapt quickly to host defense systems (Summers et al. 2003; Gandon and Nuismer 2009; Start and Gilbert 2016). Some ectoparasites exist in freshwater fish, the most common being ciliated protozoans (Lin, Clark, and Dickerson 1996), which can multiply quickly and spread rapidly throughout a population. However, most parasites found on fish are internal and typically found within the gills, eyes, or gut of the host (Reed et al. 2009; Dobson et al. 2008). Trematodes, either Digeneans or Monogeneans, are one of the most abundant parasites found in fish (Reed et al. 2009; Dobson et al. 2008; Zargar et al. 2012). Monogeneans are oviparous and have a direct lifecycle, species are usually generalists and can infect a variety of hosts (Reed et al. 2009). Although Digeneans can look like Monogeneans, they are viviparous and have complex life cycles involving

multiple hosts (Timofeeva, Gerasev, and Gibson 1997; Rehulková, Benovics, and Šimková 2020; Reed et al. 2009). Many Digeneans are specialists and require specific hosts at different life stages (Reed et al. 2009; Scholz, Aguirre-Macedo, and Salgado-Maldonado 2001; Pinto and de Melo 2012).

Gymnotiform

Gymnotiform is an order of neotropical electric fish consisting of more than 200 identified species (W. G. Crampton 1996; Albert 2001). Primarily found in freshwater rivers of South and Lower Central America (Albert and Crampton 2006; Alda et al. 2013; Aguilar et al. 2019), fish in this order are slender and lack a dorsal or caudal fin, however they use an elongated anal fin to propel forward and backward through the water (Youngerman, Flammang, and Lauder 2014). A distinct feature of this group is their ability to produce species specific electric organ discharges (EODs) (Albert 2001; Hagedorn 2008; Fugère, Ortega, and Krahe 2011). Species often have small eyes and rely on continuous electric pulses to communicate with conspecifics, and navigate murky waters (Hagedorn 2008; Fugère, Ortega, and Krahe 2011). The closely related order of Siluriform (Catfish) is a common predator of Gymnotiformes and convergently evolved electroreceptors to detect electric fields (Gallant et al. 2014; W. G. R. Crampton 2019). Certain species within Gymnotiform have adapted to this threat by changing the frequency of their EODs, thus remaining undetected by predators (Stoddard 1999). In an extreme example, *Electrophorus sp.* is capable of producing electric discharges strong enough to stun or kill prey (Traeger et al. 2015; W. G. Crampton 1996).

Brachyhypopomus occidentalis

Of the few Gymnotiform species to now colonize Central America, *Brachyhypopomus occidentalis* is one of the most widely distributed throughout the isthmus of Panama. They are typically small sized fish (6 - 45cm) that spend most of their time hiding under rocks or leaf litter (Albert and Crampton 2006). Multiple colonization events has led to isolated populations of *B. occidentalis* and increasing genetic divergence (Picq et al. 2014). Previous research in this system has highlighted vast genetic and EOD differences (Picq et al. 2016) among populations, making this a particularly good system to explore the drivers of genetic diversification. Differences in electric pulses are associated with geographic distance, and likely due to genetic drift (Picq et al. 2016). However, drift and selection are not mutually exclusive and selective pressures could still be promoting other variations.

Research Objectives

Gymnotiformes are not well represented in evolutionary studies, and little is known about what drives diversification within this order. The objective of this study is to help fill in this gap of knowledge by providing novel research on the genetic diversity at the MHC loci of a neotropical electric fish. I aim to identify the MHCII genes of *B. occidentalis*, characterize their organization, as well as explore variations seen across populations in Panama. Lastly, I explore potential associations between parasite communities and MHC diversity.

CHAPTER 1. ORGANIZATION AND POLYMORPHISM OF MHCII GENES OF A WEAKLY ELECTRIC NEOTROPICAL FISH, *BRACHYHYPOPOMUS OCCIDENTALIS*

Abstract

The major histocompatibility complex (MHC) class II molecules play a key role in inducing an immune response, by presenting foreign peptides to T lymphocytes. They represent one of most polymorphic genes in the vertebrate genome and have associated with diversification withing lineages. MHC genes are well documented in teleost, but have been largely overlooked in order Gymnotiform, a highly diverse group of neotropical electric fishes. We used a combination of a recently annotated de novo genome assembly and whole genome resequencing data to identify and characterize both classical MHCII DAB, DAA and non-classical DBB genes in populations of the electric fish *Brachyhypopomus occidentalis*. We also examined patterns of selection acting on these genes, and their phylogenetic relationships with other fish taxa. In total, we detected 17 *Broc*DAA and 19 *Broc*DAB unique sequences from 10 *B. occidentalis* individuals. The extracellular domains, encoded by exon 2, showed remarkable levels of diversity; however, there were no statistical differences in variation between the peptide binding residues (PBR) and non-peptide binding sites. Phylogenetic analysis using either the putative amino acid sequence or the exon 2 of each gene showed evidence of trans-species polymorphisms

(TSP) in the DAB but not in the DAA genes amongst Gymnotiformes. While we found the classical MHCII genes of *B. occidentalis* to be highly polymorphic, the non-classical DBB gene showed no variation and all individuals were homozygote for the same allele. Overall, our results describe the MHCII genes in a previously unreported clade of fish and, provides novel molecular tools to further explore the drivers of diversification in electric fishes.

Introduction

The Major Histocompatibility Complex (MHC) is one of the most polymorphic set of genes in the vertebrate genome (Malmstrøm et al. 2016; Hedrick 2002). MHC molecules are present on virtually all cell surfaces and function by binding with foreign peptides and presenting them to nearby T-cells, thus inducing an immune response (Altmann and Trowsdale 1989). The complex of genes that encode these molecules can be categorized into Class I, II, and III based on their structure and function. Class I molecules are located on all nucleated cells, and their function is to present viral antigens to nearby CD8+ cytotoxic T cells, which in turn terminate the infected cell (Bjorkman and Parham 1990). Class II molecules are heterodimers comprised of two non-covalently associated α and β chains. Each chain is composed of two extracellular domains ($\alpha1/\alpha2$, and $\beta1/\beta2$, respectively) and the connecting peptide (CP), transmembrane I, cytoplasmic tail (CY) regions. Class II molecules are only present on specialized antigen presenting cells (B lymphocytes, macrophages, dendritic cells) and work with CD4+ helper T cells to coordinate an appropriate response for a range of extracellular pathogens, including bacterial (Lohm et al. 2002) and parasitic infections (Podivinský 1966; Engelhard 1994).

Class III genes encode for a group of proteins involved in immune response but are not involved in antigen presentation (Milner and Campbell 2001).

MHC molecules can be further categorized into classical and non-classical subgroups based on key features and expression patterns. Non-classical MHC genes are less polymorphic and, although they can influence the peptide presentation process, they often lack direct binding capacity to t-cells (Alfonso and Karlsson 2000). Classical MHC genes are characterized by a high degree of polymorphism, especially in the peptide binding region (PBR) of the α1 and β1 domain (Hughes and Nei 1989; Carroll, Penn, and Potts 2002). The first domain of each molecule is mostly encoded by exon 2 of the 4 −5 exon gene. Historically, studies have focused exclusively on this region since the MHCII α and β chains form an extracellular groove made of a β-sheet floor and two α-helix walls that directly bind with foreign peptides for antigen presentation (Hughes and Nei 1989; Alfonso and Karlsson 2000).

Higher levels of diversity at these loci have been associated with lower parasitic (J. Klein and O'Huigin 1994; K Mathias Wegner 2004; Nevo and Beiles 1992), viral (Thursz et al. 1995; Lekstrom-Himes et al. 1999), fungal (Savage and Zamudio 2011) and bacterial (Lohm et al. 2002) infections. The high allelic diversity found at these loci is thought to be maintained by balancing selection acting through a combination of non-mutually exclusive mechanisms, namely: negative frequency selection (Takahata and Nei 1990) and heterozygote advantage (Doherty and Zinkernagel 1975; J. Klein and O'Huigin 1994). Their central role in adaptive immunity and genetic variability have made the MHC genes a preferred system to study the origin and maintenance of genetic diversity in natural

populations. Indeed, Class I and II genes have been well documented in many vertebrates, including mammals (Alves de Sá et al. 2019; Horton et al. 2004; Amadou et al. 1999), birds (Zhang et al. 2019; Wang et al. 2012; Pink et al. 1977), reptiles (K. M. Reed and Settlage 2021; Glaberman, Moreno, and Caccone 2009), and teleost fish. Unlike most other vertebrates, fish MHC genes are unlinked (Sato et al. 2000), and their organization can vary greatly between species. For instance, salmon (*Salmo salar*) have one functional MHC II locus (R. J. Stet et al. 2002), while some cichlids have over ten (Málaga-Trillo et al. 1998). In an extreme example, the Atlantic cod (*Gadus morhua*) have lost all MHC class II genes and instead have an above-average number of class I genes (Malmstrøm et al. 2016; Persson et al. 1999). Since the first MHC gene was isolated in carp (Hashimoto, Nakanishi, and Kurosawa 1990), several other fish lineages have been studied, including zebrafish (Sambrook, Figueroa, and Beck 2005; Onott et al. 1992), rainbow trout (Stet et al., 2001), sticklebacks (Christophe Eizaguirre et al. 2011; K. M. Wegner et al. 2006), salmon (Grimholt et al., 2000; Stet et al., 2002), cichlids (Murray et al. 2000), halibut (Li et al. 2011), yellow croaker (Yu, Ao, and Chen 2010a), and seabream (Cuesta, Ángeles Esteban, and Meseguer 2006). However, little is known about the organization or characteristics of MHC genes in electric fishes of the order Gymnotiformes. Thus, the goal of this study is to characterize, for the first time, the MHC Class II genes in the neotropical gymnotiform *Brachyhypopomus occidentalis.*

Gymnotiformes are freshwater fish primarily endemic to Central and South America, where over 200 species have been identified (Albert and Crampton 2006). A unique feature of this clade is the ability to produce species – specific electric organ discharges (EODs) that are used for navigation and communication with conspecifics (Hagedorn 2008).

Although this clade originated in South America, six species and five genera have been described in Lower Central America (Albert and Crampton 2006; Alda et al. 2013). Found in rivers throughout Panama and southern Costa Rica, *B. occidentalis* (bluntnose knifefish) exhibit substantial genetic (Aguilar et al., 2018; Picq et al., 2014) and EOD (Picq et al., 2016) interpopulation variation. Differences between populations have been shaped by historical events, such as multiple waves of colonization and genetic drift (Aguilar et al., 2018; Martin & Bermingham, 1998; Picq et al., 2014, Picq et al., 2016). Additionally, the increasing availability of genomic sequencing data is likely to facilitate the study of the drivers shaping genetic diversity in electric fishes, including variation at the MHC loci. In particular, the annotated genome of *B. occidentalis* (Arias et al., 2021) provides an opportunity to study the diversity of the MHC genes in a species-rich genus of the gymnotiform clade. Here, we take advantage of the available genomic sequencing data, as well as sequence data from two populations of *B. occidentalis*, to characterize variation in a non-classical (DBB) and two classical (DAA and DAB) MHCII genes. Our goals for this study were to: **(1) characterize the genomic structure and organization of the MHCII genes of *B. occidentalis*, (2) test for selection on the classical MHC loci, by comparing inter- and intra-genetic substitution rates**, and **(3) assess phylogenetic relationships between *B. occidentalis* and other teleost fishes based on available MHC genes**.

Materials and Methods

Fish Sampling

Genomic DNA for *B. occidentalis* were obtained from ten fish collected at two rivers, Tugamantí (n=5; Lat: 09.2232°, Long: -078.8887°) and Tapagrilla (n = 5; Lat: 09.1833°, Long: -079.20°) belonging to the Bayano drainage (weight = 5.56g ± 3.68; length = 12.67cm ± 4.23) in eastern Panama. Fish were located using electrodes connected to a rod and plugged into a handheld amplifier (RadioShack, Fort Wort Texas). Since *B. occidentalis* emit continuous electric pulses, it is possible to hear their distinct buzzing through the amplifier when the electrodes are in water near their presence. Once fish were detected, we used a dip net for capture, and euthanasia was done by submerging individuals in eugenol (C10H12O2) for 60 seconds as described in Davis et al., (2015). Samples were stored in -20°C until DNA extraction from muscle tissue using protocol for Dneasy Blood & Tissue Kit (Qiagen). This project was approved by the Institutional Animal Care and Use Committee at the University of Massachusetts Boston (IACUC 2019121), and collecting permits were authorized by Panama's Ministry of the Environment (MiAmbiente; permit number: SE/AB-1-19).

DNA Sequencing

To characterize the MCH II genes, we used the recently assembled and annotated genome of *B. occidentalis* (Arias et al., 2021) as a reference, as well as whole genome resequencing data from the ten *B. occidentalis* obtained from the Bayano drainage. The annotated reference genome is 540.3 Mb in size and includes 23,935 functionally annotated

genes, out of a total 34,347 predicted protein coding genes (Arias et al., 2021). Completeness of the reference genome was estimated by searching for 3,640 Benchmarking Universal Single-Copy Orthologs (BUSCOs); of which 3,414 (93.8%) were detected, 27 (.7%) were fragmented, and 99 (5.5%) orthologs were missing (Arias et al., 2021). For the resequencing protocol, we obtained DNA of ten *B.occidentalis* individuals from frozen preserved tissues using Dneasy Blood & Tissue Kit (Qiagen, Valencia, CA). Libraries were prepared following the protocol for KAPA Library Preparation kits (Kapa Biosystems, Wilmington, MA). Samples were sent for next generation sequencing on an IlluminaX Ten platform (Novogene, Sacramento, CA) with minimum of 20x coverage. Sequences were quality trimmed using Trimmomatic 0.40 (Bolger et al., 2014), before following the GTAK pipeline for variant calling (Broad Institute, Cambridge, MA).

MHC bioinformatics

Genes in the reference genome were predicted through the GeMoMa Pipeline (v1.6.4, Keilwagen et al., 2018). The MHC Class II was annotated based on the transcriptome data of the closely related electric eel, *Electrophorus electricus* (Traeger et al. 2015; Gallant et al. 2014). Predicted MHCII sequences were isolated and extracted from the ten re-sequenced *B. occidentalis* individuals using BLAST function within Geneious (v1.2). Basic amino acid analysis was performed on EMBOSS Pepstats (Rice, Longden, and Bleasby 2000). To further confirm the identity of the region, we aligned all sequences against other teleost MHC sequences, using the Clustal Omega program (Thompson et al. 1994) as implemented in Geneious (v1.2). Teleost MHC sequences were obtained by using the BLAST function in the NCBI database to search for genes with high similarity to *B.*

occidentalis sequences (Accession numbers in Supplemental Materials). In total, DAA and DAB sequences from six teleost species were obtained (*Electrophorus electricus, Ictalurus punctatus, Salmo salar, Oncorhynchus mykiss, Danio rerio,* and *Cyprinus carpio;* Table S1), in addition to the homologous HLA genes found in *Homo sapiens.* The non-classical MHCII DBB sequences were obtained from the available data on five teleost species (*Electrophorus electricus, Danio rerio, Lates calcarifer, Chionodraco hamatus*, and *Oreochromis niloticus;* Table S1). Putative peptide binding regions (PBR), conserved residues, and potential N-linked glycosylation sites were identified based on similarity to other sequences used in previous studies (Jeon, Won, and Suk 2019; Cosson and Bonifacino 1992; Yu, Ao, and Chen 2010b; Amadou et al. 1999).

Tests of Selection

To test for selection on MHC II genes, we performed several maximum likelihood (ML) tests, including FEL (Fixed Effects Likelihood), MEME (Mixed Effects Model of Evolution), FUBAR (Fast, Unconstrained Bayesian AppRoximation), SLAC (Single Likelihood Ancestor Counting), and GERD (Genetic Algorithm for Recombination Detection). Tests for selection were all run in the HyPhy package implemented in DataMonkey (Kosakovsky Pond and Frost 2005b). We used SLAC to calculate synonymous (S) and non-synonymous (N) substitutions in the exon 2 of *Broc*DAA and *Broc*DAB. Non-synonymous mutation rates (dN) were compared with synonymous mutations (dS) using a Welch's *t*-test to detect inter- and intra-genetic differences in substitution rates. Specifically, we tested two hypotheses; whether dN/dS was higher in the

exon 2 of each ($\alpha1 > \alpha - \alpha1$ & $\beta1 > \beta - \beta1$), and whether the beta gene showed higher dN/dS than the alpha ($\beta > \alpha$ & $\beta1 > \alpha1$).

To further test whether sites are under positive or purifying selection pressures, MEME analysis was used to calculate the mutation rate (ω) of the protein coding sequence of each gene (Murrell et al., 2012). Contrary to the other methods, this model allows for mixed selection pressures by using a single dS value (α) but two separate dN values ($\beta+$ and $\beta-$). Two ω rates are calculated for each site, and a ML approach is used to calculate the probability that the site evolved under diversifying ($\omega > 1$) or purifying ($\omega < 1$) selection. Only sites with a probability greater than 95% chance ($p < 0.05$) of occurring were considered to be under selection. Given that selection tests assume that no recombination event occurred at the sites being considered, we used GERD to test for recombination prior to selection (Pond et al. 2006).

Phylogenetic Analysis

Phylogenetic analyses were performed by using either the complete protein sequences of the DAA and DAB genes or the predicted sequences of the extracellular $\alpha1$ and $\beta1$ domains. Neighbor – joining (NJ) trees (Saitou and Nei, 1987) were constructed with Geneious Tree Builder (v1.2) using 100,000 bootstrap replicates, and only bootstrap values over 50% were retained. Genetic distances were calculated using Jukes-Cantor distance measures in Geneious (v1.2). To confirm evolutionary clustering, Maximum Likelihood trees were also constructed using the MEGA 11.0 program (Tamura et al., 2021). A total of 39 homologous genes were obtained from the NCBI database (Accession

numbers in, table S1) and included sequences from 19 teleost species (*Electrophorus electricus, Ictalurus punctatus, Salmo salar, Oncorhynchus mykiss, Danio rerio, Cyprinus carpio, Ctenopharyngodon idella, Oncorhynchus tshawytscha, Oncorhynchus nerka, Paralichthys olivaceus, Cynoglossus semilaevis, Sparus aurata, Pagrus major, Miichthys miiuy, Dicentrarchus labrax, Plecoglossus altivelis, Osmerus mordax, Megalobrama amblycephala,* and *Leiocassis longirostris*), and 2 mammalian species (*Homo sapiens,* and *Mus musculus*), which we used as outgroups.

Results

*Structure and Polymorphisms of Broc*DAA, *Broc*DAB, *Broc*DBB

In total five MHCII genes were detected in the annotated reference genome, however only two fell under the classical MHCII DA- lineage. A non-classical beta gene was also detected, belonging to the DB- group of MHC genes. The remaining two of the predicted class II genes were detected in *Electrophorus electricus*, but they were not found in any other teleost lineage. Low evidence scores (< 500) and poor alignments with other species suggest this was likely due to duplication events leading to a pseudogene. Our study focused on the genes with high evidence scores (> 500) and included the two classical MHCII DAA and DAB genes, as well as the non-classical DBB.

We obtained 20 complete MHC IIα and IIβ gene sequences of *B. occidentalis* collected from the two rivers, Tugamantí and Tapagrilla (2n = 20), in addition to the reference genome sequence of a fish caught in the Juan Grande River (Arias et al., 2021, n = 1). In total, we found 17 unique *Broc*DAA sequences and 19 unique *Broc*DAB sequences or alleles from mostly heterozygote individuals. Genetic sequences were named

*Broc*DAA*0101 to *Broc*DAA*0801 and *Broc*DAB*0101 to *Broc*DAB*1102 in accordance with nomenclature used by Shum et al. (2001). Only five alleles were detected in both rivers, including *Broc*DAB*01, *Broc*DAB*03, *Broc*DAA*01, *Broc*DAA*02 and *Broc*DAA*03. In contrast, the non-classical gene showed astonishingly low variation; all 20 *Broc*DBB sequences were identical, and all individuals were homozygote for the *Broc*DBB*0101 allele.

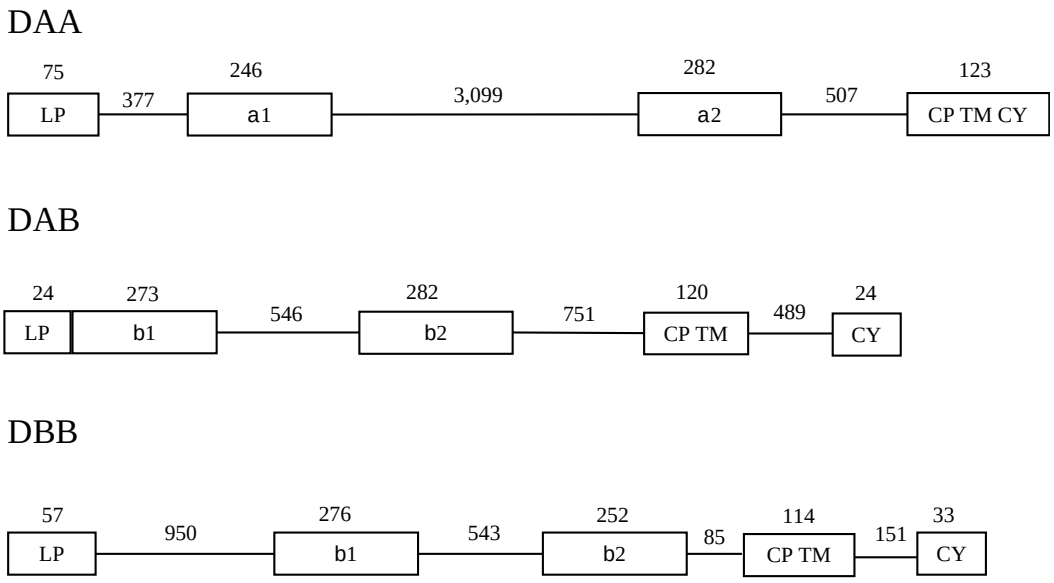

Fig 1.1 Exon/intron organization of *Broc*DAA, *Broc*DAB, and *Broc*DAA in *Brachyhypopomus occidentalis*. Boxes indicate relative position of exons, lines represent introns, and numbers show base pair (bp) length

The MHCII alpha chain in *B.occidentalis* is a 4,798 bp gene made of four exons (Fig 1) that translate to 241 residues and a predicted molecular weight of 26.4kDa. Exon 1 makes up the untranslated leading peptide (LP) chain and the first codon of α1, which is separated from exon 2 by a relatively short intron. High rates of variability are expected in

the α1 domain, which is mostly encoded by exon 2, and is responsible for directly binding with foreign peptides. Similar to previous teleost studies (Yu et al., 2010), we detected the 83 residues at this site to be highly polymorphic. Although it encompasses only 34.4% of the entire gene, 55.2% of all polymorphisms were found in this region (Table 1). Exon 3 encodes for another extracellular domain, the α2, and a rather large intron (3,099 bp) separates exons 2 and 3. Likely due to replication events, this intron is lacking in the predicted MHC IIα gene of *E. electricus* but was detected in all *B. occidentalis* individuals. Lastly, exon 4 encodes for the transmembrane domain and cytoplasmic tail.

The classical MHC IIβ chain in *B.occidentalis* is a 2,509bp gene with 5 exons (Fig 1) that encode a polypeptide of 240 amino acids (aa), and a predicted molecular weight of 27.4kDa. Similar to *Broc*DAA, the first exon in *Broc*DAB encodes for a short leading peptide and the beginning of the β1 domain. The majority of the highly polymorphic β1 domain is encoded for by exon 2, and composed of 91 residues, 23.1% of which were variable (Table 1). The remaining exons encode the rest of the polypeptide chain; exon 3 encodes for the β2 domain, while exon 4 encodes for the hydrophobic transmembrane region and exon 5 encodes the short cytoplasmic tail (Fig 1). The non-classical MHCIIβ chain followed a similar structure to its classical counterpart, except that an intron was detected separating the leading peptide and β1 of the *Broc*DBB gene. Overall, the DBB gene is 2,607bp with a predicted molecular weight of 27.8kDa; and a total of 5 exons (Fig 1) are predicted to form a 243aa chain.

		β1 /α1	α2/β2 CP TM CY
DAA	**(a)** Total Sites	83	158
	(b) Variable Sites	16 (19.3%)	13 (8.2%)
	(c) Total Variability	55.2%	44.8%
DAB	**(a)** Total Sites	91	149
	(b) Variable Sites	21 (23.1%)	7 (4.7%)
	(c) Total Variability	75.0%	25.0%

Table 1.1 Variability in the first domain (α1/β1) compared to all other regions (α2/β2 and CP/TM/CY) of the Classical MHCII genes of *B. occidentalis*. Letters indicate (a) number of total sites per region, (b) number of variable sites within each region (and percent of region that shows variability), and (c) percentage of total variability that falls within each region.

The classical DA- alpha and beta chains form a heterodimer, the first domain of each (α1 and β1 respectively) form an extra-membrane cleft in which foreign peptides are bound to and presented to CD4+ T-cells (Engelhard 1994). We found that of the 28 variable sites detected in the DAB gene, 75.0% occurred in the β1 domain (Table 1). In comparison, the β2/CP/TM/CY domains encompass 62.0% of the gene but only possess 25.0% of the variability detected. Of the seven variable sites that fell outside of the β1, all were detected within the extracellular β2; in contrast, the trans- and inner- membrane regions (CP/TM/CY) were uniform across all individuals.

Sequence Comparisons

The translated amino acid sequences of all unique *Broc*DAA and *Broc*DAB alleles were aligned with sequences from other species (Figs. 2 and 3), including *E. electricus* (*Elel*DA-), *Danio rerio* (*Dare*DA-), *Salmo salar* (*Sasa*DA-), *Cyprinus carpio* (*Cyca*DA-),

Ictalurus punctatus (*Icpu*DA-), and the human HLA-DR genes (Table S1). Similarities of *Broc*DAA with other vertebrate sequences ranged from 67.9% amino acid identity (*Elel*DAA) to 32.1% (HLA-DRA). Comparisons of *Broc*DAB showed a similar pattern, albeit slightly more uniformity between teleost species, with overall identity ranging from 75.8% (*Elel*DAB) to 28.4% (HLA-DRB). The predicted non-classical *Broc*DBB gene was best aligned to other teleost sequences of the DB- lineage; namely, *E. electricus, D. rerio, Lates calcarifer, Chionodraco hamatus, Oreochromis niloticus* (Table S1). Similarities between sequences ranged from 72.1% (*Elel*DBB) to 39.7% (*Orni*DBB). Pairwise identity scores are comparable to other MHCII studies in fish (Murray et al., 2000; Yu et al., 2010; Stet et al., 2002; Gerdol et al., 2019), adding confidence in the characterization of these regions in *B. occidentalis.* Furthermore, several conserved residues were detected across all taxa.

The α and β chain each contained four cysteine sites well conserved among all teleosts. The cysteines of the α1 (C13 and C66) and α2 (C108 and C164) are predicted to form an intra-domain disulfide bridge, based on mammalian studies (Dick et al. 2002). The same predicted disulfide bridge was detected in the β1 (C9 and C75) and β2 (C113 and C169) of *Broc*DAB, and the β1 (C11 and C74) and β2 (C104 and C160) of *Broc*DBB. All *B. occidentalis* sequences shared the same potential N-linked glycosylation site in the α2 domain (NLT, position 127-130). This site was identical to *S. salar*, but differed from *E. electricus* by one mutation (NVT, Fig 3). A potential N- linked glycosylation site in the β1 domain (NST, position 37 – 39) was identical in all *B. occidentalis* as well as in the DAB sequences of *E. electricus, S. salar* and *I. punctatus* (Fig 4). While all sequences had some

number of substitutions in the aligned protein sequences, the member proximal domains of the CP/TM/CY regions in all genes showed little change and high uniformity across species.

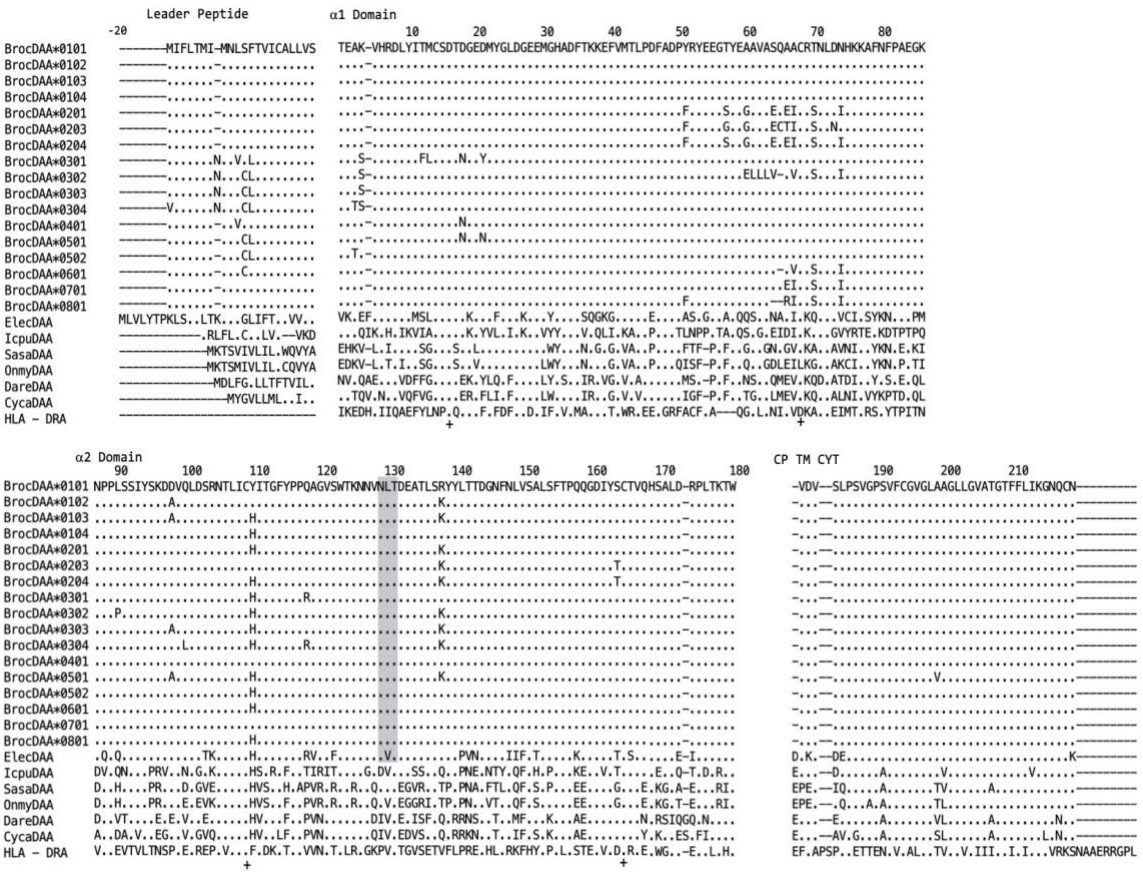

Fig 1.2. Translated amino acid sequence alignment of MHCII DAA genes. Sequences of *B. occidentalis* are aligned with several other species shown in Table S1. Conserved cysteines sites are displayed with a plus (+), predicted peptide binding sites are shown with a hash (#), and predicted N-glycosylation site is shaded in gray. Domains are labeled above the start of their respective regions; identical sites are represented by a dot (.) while introduced gaps are shown with dashes (-).

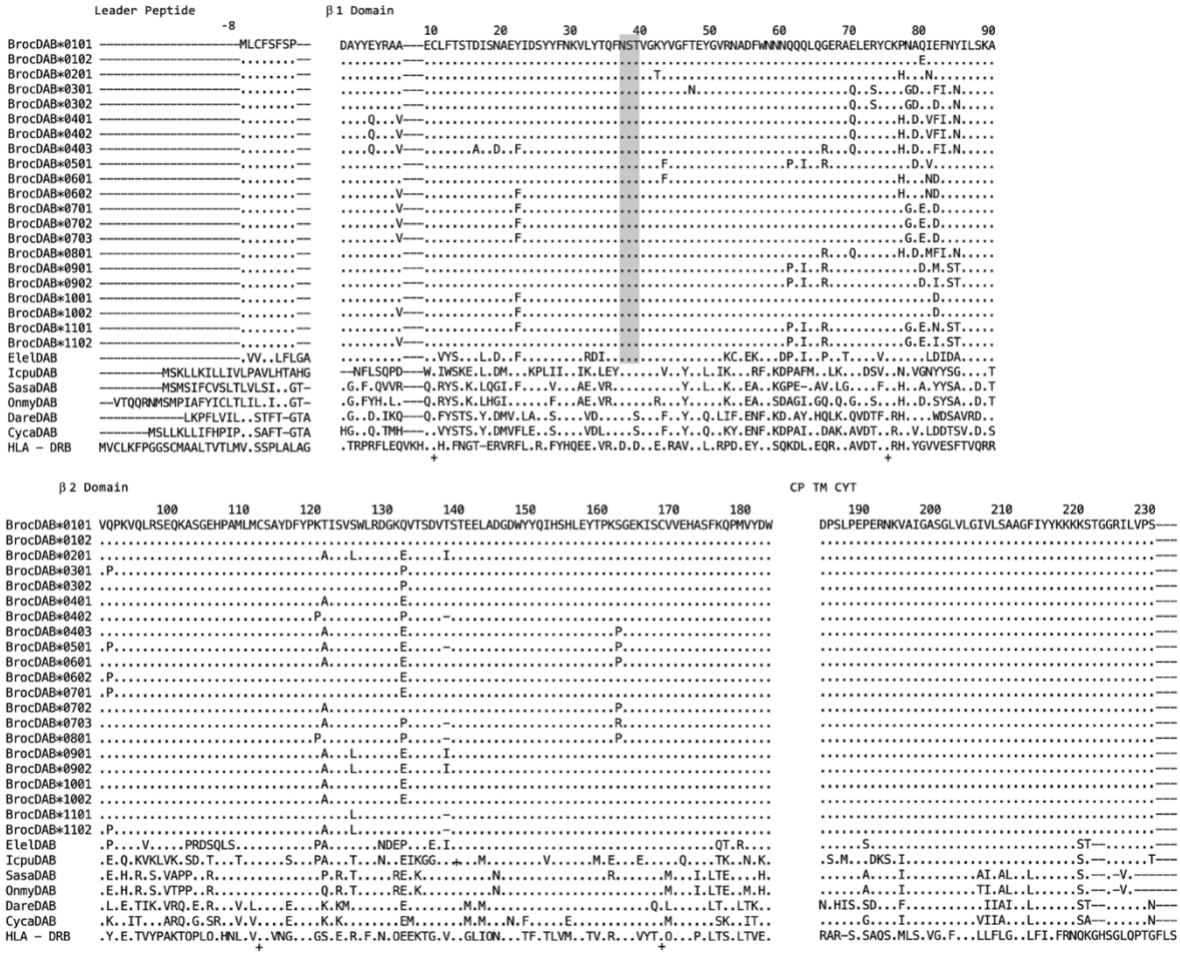

Fig 1.3. Translated amino acid sequence alignment of MHCII DAB genes. Sequences of *B. occidentalis* are aligned with several other species shown in Table S1. Conserved cysteines sites are displayed with a plus (+), predicted peptide binding sites are shown with a hash (#), and predicted N-glycosylation site is shaded in gray. Domains are labeled above the start of their respective regions; identical sites are represented by a dot (.) while introduced gaps are shown with dashes (-).

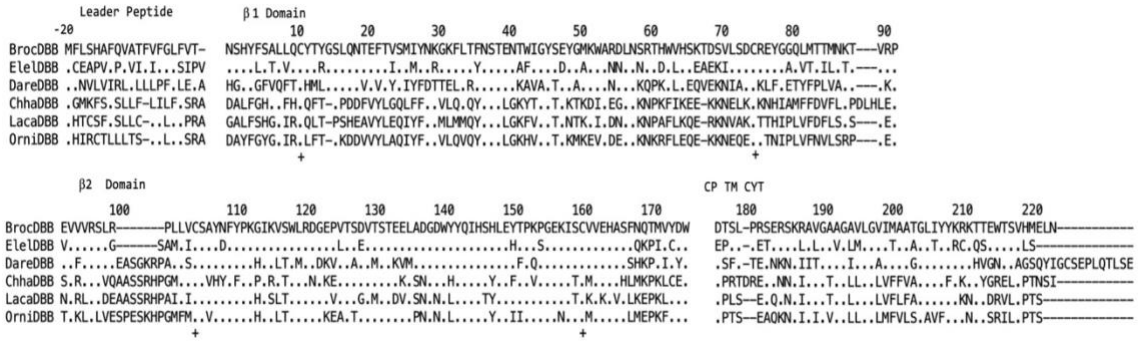

Fig 1.4. Translated amino acid sequence alignment of MHCII DBB genes. The solely unique *B. occidentalis* sequence is aligned with several other teleost species shown in Table S1. Conserved cysteines sites are displayed with a plus (+). Domains are labeled above the start of their respective regions; identical sites are represented by a dot (.) while introduced gaps are shown with dashes (-).

Selection Analysis

Comparisons of average dN/dS rates between the putative peptide binding regions (PBR) of *Broc*DAA and *Broc*DAB yielded no significant difference. Furthermore, overall rates of mutation between the entire coding regions of both genes were similar ($\beta > \alpha$, $p = 0.56$). However, there was a significant difference in the substitution rates of the exon 2 compared to other exons within each gene (Fig 5). As expected, the extracellular domain, primarily responsible for recognizing and binding with pathogens, had higher rates of non-synonymous mutations than the trans- and inner- membrane regions ($\alpha 1 > \alpha - \alpha 1$, $p = 0.012$; $\beta 1 > \beta - \beta 1$, $p = 0.0001$).

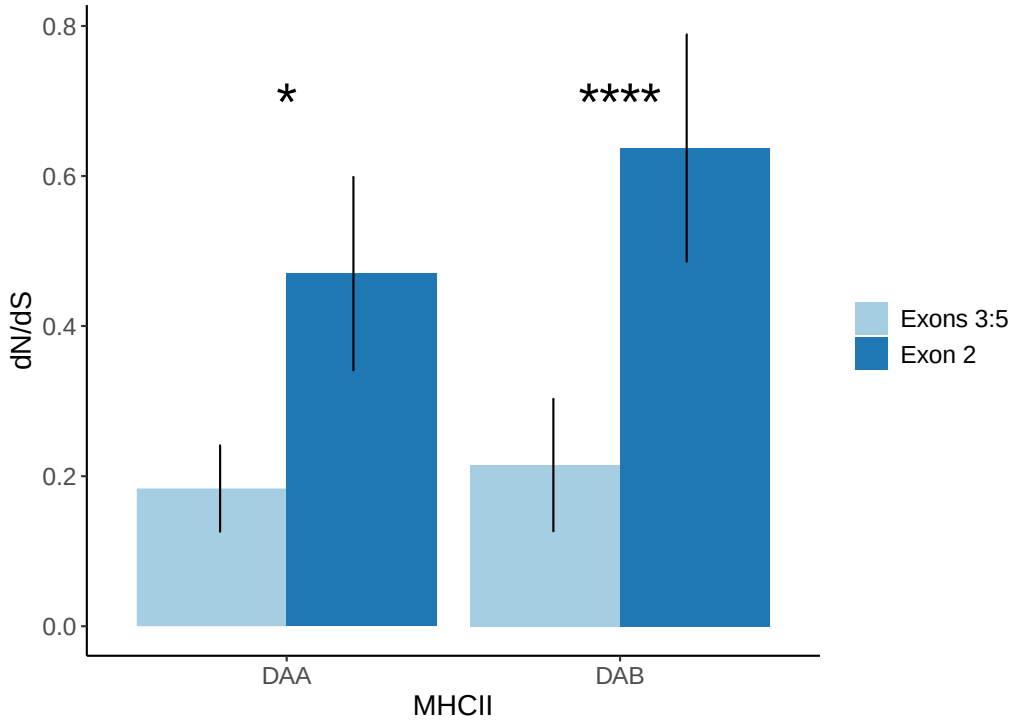

Fig 1.5. Mutation rates of the classical MHCII in *B.occidentalis*. Comparisons show average non-synonymous to synonymous substitution rate (dN/dS) in the exon 2 and in exons 3 through 5 of *Broc*DAA and *Broc*DAB. Stars indicate significant p-value (* < 0.05, ** < 0.0001), there was no significant difference in mutation rates between genes (p = 0.56).**

The mixed effects model of evolution (MEME) showed that both classical genes have sites that seem to be under diversifying selection. Specifically, five sites in the *Broc*DAA gene, and eight sites in the *Broc*DAB showed evidence of diversifying selection ($p < 0.05$). All sites predicted to be under selection in the alpha chain occur in the $\alpha 1$ domain. Eight variable sites were predicted to be under diversifying selection in *Broc*DAB: four sites in the $\beta 1$ domain and four in the $\beta 2$ (Table 2). Previous research suggests that different selection pressures may act on different regions of the MHC gene; while the extra-cellular domains may be under diversifying selection, purifying selecting may be acting on the trans- and inner- membrane regions (Ottová et al., 2005). Our analysis showed no

evidence of purifying selection in the MHCII of *B. occidentalis*, and none of the sites in the trans- or inner- membrane regions were predicted be under selection.

	Site #	p-value	Region
DAA	26	0.04	Non-PBR
	34	0.04	PBR
	83	0.01	PBR
	87	0.01	Non-PBR
	88	0.05	PBR
DAB	17	0.02	Non-PBR
	72	0.00	PBR
	87	0.00	PBR
	91	0.01	PBR
	130	0.00	Non-PBR
	131	0.03	Non-PBR
	142	0.03	Non-PBR
	148	0.01	Non-PBR

Table 1.2. Sites under positive (diversifying) selection in the PBR and non-PBR of *Broc*DAA and *Broc*DAB. Only sites where $\omega > 1$ and $p < .05$ were considered under positive selection.

Phylogenetic Analysis

To evaluate evolutionary relationships, NJ and ML phylogenetic trees were constructed using either the putative amino acid chain of exon 2 or the complete predicted chain of the DAB and DAA genes. Trees constructed using the DAB sequences were highly uniform, and showed reciprocal monophyly for most clades, with minor changes in the clustering of the complex perciform clade (Supplementary Figure S1). Overall similar topology was obtained when using the DAA sequences, although there were some discrepancies when using exclusively the α1. Specifically, the Gymnotiformes formed a monophyletic group when using the complete amino acid sequence, but the position of *E. electricus* remained unresolved when solely using the α1 sequences. The Siluriformes, represented by *I. punctatus* (*Icpu*DAA), is expected to cluster closely with the

Gymnotiformes as seen in the phylogenetic analysis using the complete amino acid sequence (Fig. S2b). However, *I. punctatus* clusters closely with the Osmeriformes when using only the α1 domain (Fig. S2a).

Discussion

The major histocompatibility complex plays a key role in the adaptive immune response of vertebrates. Classical MHCII genes are highly polymorphic genes and have been extensively studied due to their association with pathogen resistance (Lekstrom-Himes et al., 1999; Thursz et al., 1995; Engelhard, 1994; Podivinský, 1966; Wegner, 2004; Lohm et al., 2002), and ability to promote diversification (Malmstrøm et al. 2016). We take advantage of the reference genome of *Brachyhypopomus occidentalis* (Arias et al. 2021) and whole genome resequencing data to identify predicted MHCII genes. To our knowledge, this is the first study to characterize these immunologically vital genes in the electric fishes of the order Gymnotiform. Overall, the MHC loci in *B. occidentalis* showed extraordinary diversity, especially in the extracellular exon 2 of the *Broc*DAB and *Broc*DAA genes.

MHC Organization in B. occidentalis

Five MHCII genes were predicted in the annotated reference genome of *B. occidentalis* (Arias et al. 2021), including the DAA, DAB, and DBB genes. Premature stop codons and poor sequence alignments suggest that the remaining two genes predicted are most likely non-functioning pseudogenes only found in the electric fishes, *B. occidentalis* and *Electrophorus electricus* (Traeger et al. 2015; Gallant et al. 2014). MHC pseudogenes

are common in other organisms, for example the majority of the 224 classical human MHC genes are thought to be pseudogenes (Beck et al. 1999). That is, only approximately two in every five genes are predicted to influence the immune function and play a role in disease resistance. We found a similar ratio in *B.occidentalis*, with three of the five genes predicted to be functional MHC genes. Our findings of one functional α chain, and two β chain loci are typical for MHCII genes, given that an α chain can bind to multiple β chains (Alfonso & Karlsson, 2000). Notably, both pseudogenes predicted in *B.occidentalis* were also detected in the most closely related taxon, *E. electricus*, but not in other fish species. This suggests the two pseudogenes likely evolved via duplication events in a common ancestor of electric fishes. Future genetic studies in other gymnotiforms and the closely related siluriform are needed to better understand the timing of this duplication event. We also detected another likely duplication event that occurred more recently in *B. occidentalis*; a large repetitive intron was detected in the IIα gene of all *B. occidentalis* individuals, but such intron was not found in *E. electricus.* Thus, our results suggest that duplication events are playing an important role in shaping the structure of MHC genes in electric fishes; yet more genetic data in other gymnotiform clades are needed to unveil the functional consequence of these duplication events.

MHC Variation

Alignments of the putative amino acid sequences with homologous genes in other teleost species revealed several highly conserved structural features. We identified four cysteine residues in each of the *BrocDAA, BrocDAB,* and *BrocDBB* sequences that were shared among all teleost sequences. Previous studies also found two pairs of cysteine

residues that bind together to form two structurally important bridges among the α1/α2 and β1/β2, respectively (Beck et al., 1999; Grimholt et al., 2000; Li et al., 2011). The potential N- glycosylation site detected in the β1 (NST, position 37-39) was identical amongst the Gymnotiformes, Siluriformes, and Salmoniformes. There was less uniformity at this site in the α2 (NLT, position 127 – 129), with only *Salmo salar* sharing the same sequence as *BrocDAA*. Interestingly, the potential N-glycosylation site differed within the Salmoniformes and Gymnotiformes by a single mutation (V, position 128).

Variability amongst sequences were particularly high in the extra-membrane peptide binding regions of both genes (Table 1). In contrast, the transmembrane region and inner-membrane cytoplasmic tail showed little to no variability among species. All vertebrate sequences shared the same GxxxGxxGxxx motif in the transmembrane region of the α chain. This is not surprising, given that the G residues in this motif play a key role in the interaction between the α and β chains when forming the α-β dimer (Cosson & Bonifacino, 1992). The leading peptide domain of each sequence was also highly polymorphic, and it showed the largest number of differences between the two species of Gymnotiformes included in the analyses. Although not much is known about the purpose of the leading peptide, it is an untranslated region and variation seen here likely has little functional importance.

Evidence of Selection

Previous MHC studies have focused on exon 2 for its remarkable polymorphism, and its above-average number of non-synonymous mutations. Several sites within the exon

2 are predicted to evolve under strong selective pressures as they directly interact and bind with foreign peptides (Engelhard 1994; Sambrook, Figueroa, and Beck 2005; Beck et al. 1999). However, contrary to other teleosts (Gerdol et al. 2019; Yu, Ao, and Chen 2010b), we did not find significant differences in dN between the PBR of *Broc*DAA and *Broc*DAB. While exon 2 of *Broc*DAA and *Broc*DAB harbored the majority of polymorphisms observed, the peptide-binding region (PBR) was no more likely than the non-PBR to be under positive selection. This could be due to our sampling being limited to a single watershed. Including data from other populations representing separate drainages under potentially different selective regimes could reveal MHC sites under positive selection. Differing selection pressures may also act on different regions of the MHC gene; while the PBR may be under diversifying selection, purifying selection may be acting on the non-PBR (Ottová et al., 2005). Our analysis showed no evidence of purifying selection in the MHCII of *B. occidentalis*, perhaps due to recent bottleneck events.

Phylogenetic Analysis

Our phylogenetic analyses showed a consistent clustering across fish taxa, regardless of sequence used (Supplementary Figures S1 and S2). Furthermore, *B. occidentalis* alleles grouped as expected and phylogenetic clustering of all sequences confirmed that *Broc*DAA and *Broc*DAB were in the classical DA- lineage, while *Broc*DBB belonged to the non-classical DB- group. The only slight deviation from this pattern was within the Neighbor-Joining tree based solely on the $\alpha 1$ domain of BrocDAA. Incongruities could be due to reduced genetic information when only using the first domain.

Phylogenetic and sequence alignment results could also suggest the presence of trans-species polymorphism (TSP), or allelic lineages that persist across species (Jan Klein et al. 1998). The existence of TSP indicates that balancing selection may be acting to maintain advantageous alleles among populations, and is associated with persistent selective pressures (Jan Klein, Sato, and Nikolaidis 2007). The presence of TSP was detected between species of mice and rats, with a rare mutation observed across lineages and likely predating speciation (Figueroa, Gúnther, and Klein 1988). While TSP appears to be present among the *Broc*DAB alleles, the *Broc*DAA showed less uniformity with closely related species such as *E. electricus*. The non-classical *Broc*DBB and *Elel*DBB also overlapped substantially, and a deletion across six sites of the β2 was only detected within the DBB genes of the Gymnotiformes. While we can speculate on the existence of TSP, more genetic information on diverse species within the order Gymnotiform is needed to make any conclusions.

Conclusion

Our results suggest that the classical MHCII genes are highly variable within *B. occidentalis* and between two Gymnotiform species. Although there was little variation within the member proximal domains; the extracellular domain was highly polymorphic. In contrast, the non-classical MHCIIβ was identical in all *B. occidentalis* and coincided with *Electrophorus electricus* but varied across all teleosts. Overall, more research focused on the non-classical MHC genes of teleosts is needed to evaluate if the extreme homozygosity seen in *Broc*DBB is within normal range. While fish belonging to the order Gymnotiform can account for a significant portion of biomass found in rivers throughout

Lower Central and South America (W. G. Crampton 1996), little research has been conducted on the drivers of diversification within this lineage. Our study provides novel insight into the MHCII variation of electric fishes and is a necessary precursor for future studies.

Supplementary Figures and Tables:

Name	Accession Number	Common Name	Scientific Name	Order
ElelDAB	XM_026996251	Electric eel	*Electrophorus electricus*	Gymnotiformes
IcpuDAB	U77597	Channel catfish	*Ictalurus punctatus*	Silruriformes
SasaDAB	AJ439067	Atlantic salmon	*Salmo salar*	Salmoniformes
OnmyDAB	CBX11173	Rainbow trout	*Oncorhynchus mykiss*	Salmoniformes
DareDAB	BC124461	Zebrafish	*Danio rerio*	Cypriniformes
CycaDAB	HQ380378	Common carp	*Cyprinus carpio*	Cypriniformes
HLA-DRB	X02902	Human	*Homo Sapien*	Primates
ElelDAA	XM_026996499	Electric eel	*Electrophorus electricus*	Gymnotiformes
IcpuDAA	AF103002	Channel catfish	*Ictalurus punctatus*	Siluriformes
SasaDAA	BT150000	Atlantic Salmon	*Salmo salar*	Salmoniformes
OnmyDAA	AJ251432	Rainbow trout	*Oncorhynchus mykiss*	Salmoniformes
DareDAA	L19446	Zebrafish	*Danio rerio*	Cypriniformes
CycaDAA	X95432	Common carp	*Cyprinus carpio*	Cypriniformes
HLA-DRA	V00523	Human	*Homo Sapien*	Primates
ElelDBB	XM_026995774	Electric eel	*Electrophorus electricus*	Gymnotiformes
DareDBB	XM_005159194	Zebrafish	*Danio rerio*	Cypriniformes
ChhaDBB	KX398847	Antarctic icefish	*Chionodraco hamatus*	Perciformes
LacaDBB	XP_018541680.1	Asian Seabass	*Lates calcarifer*	Perciformes
OrniDBB	XM_019365082	Nile Tilapia	*Oreochromis niloticus*	Perciformes
CtidDAA	EU186148	Grass Carp	*Ctenopharyngodon idella*	Cypriniformes
OntsDAA	XM_024408463	Chinook salmon	*Oncorhynchus tshawytscha*	Salmoniformes
OnneDAA	XM_029652725	Sockeye Salmon	*Oncorhynchus nerka*	Salmoniformes
PaolDAA	EU514684	Japanese flounder	*Paralichthys olivaceus*	Pleuronectiformes
CyseDAA	KR024711	Tongue sole	*Cynoglossus semilaevis*	Pleuronectiformes
SpauDAA	DQ019401	Gilthead seabream	*Sparus aurata*	Perciformes
PamaDAA	AY698064	Red seabream	*Pagrus major*	Perciformes
MimiDAA	GU936787	Miiuy Croaker	*Miichthys miiuy*	Perciformes
DilaDAA	DQ821109	European seabass	*Dicentrarchus labrax*	Perciformes
PlalDAA	KX118043	Sweetfish	*Plecoglossus altivelis*	Osmeriformes
OsmoDAA	BT075291	Rainbow smelt	*Osmerus mordax*	Osmeriformes
MumuDAA	K01923	House mouse	*Mus musculus*	Rodentia
PaolDAB	HQ634973	Japanese flounder	*Paralichthys olivaceus*	Pleuronectiformes
CyseDAB	NM_131476	Tongue sole	*Cynoglossus semilaevis*	Pleuronectiformes
DilaDAB	DQ821110	European seabass	*Dicentrarchus labrax*	Perciformes
PamaDAB	AY848956	Red seabream	*Pagrus major*	Perciformes
MimiDAB	HM236158	Miiuy Croaker	*Miichthys miiuy*	Perciformes
MeamDAB	KF193866	Wuchang Bream	*Megalobrama amblycephala*	Cypriniformes
LeloDAB	FJ797955	Longsnout catfish	*Leiocassis longirostris*	Silruriformes
MumuDAB	AAF21606	House mouse	*Mus musculus*	Rodentia

Table 1.S1. Sequences used in phylogenetic analysis and alignments of the MHCII DAA, DAB and DBB genes. Information includes: GenBank Accession Numbers, scientific and common species names and classification to the order level.

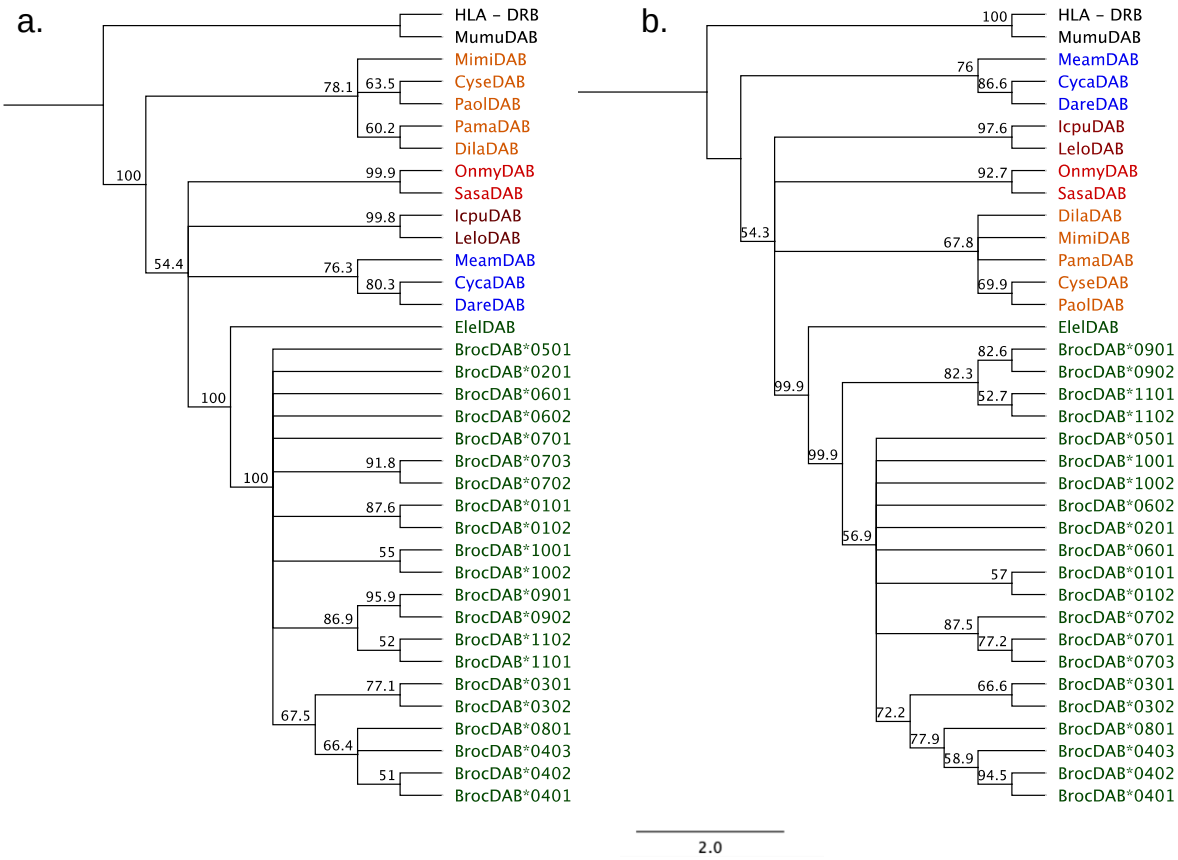

Fig 1. S1. Phylogenetic analysis of MHCII DAB sequences in *B. occidentalis* and other vertebrate groups. Neighbor-Joining trees were constructed using either the (a) amino acid sequences of the β1 domain, or (b) complete amino acid sequences. Numbers by each branch represent the percentage nodal support based on 10,000 bootstrap replicates. Accession numbers in Table A1.

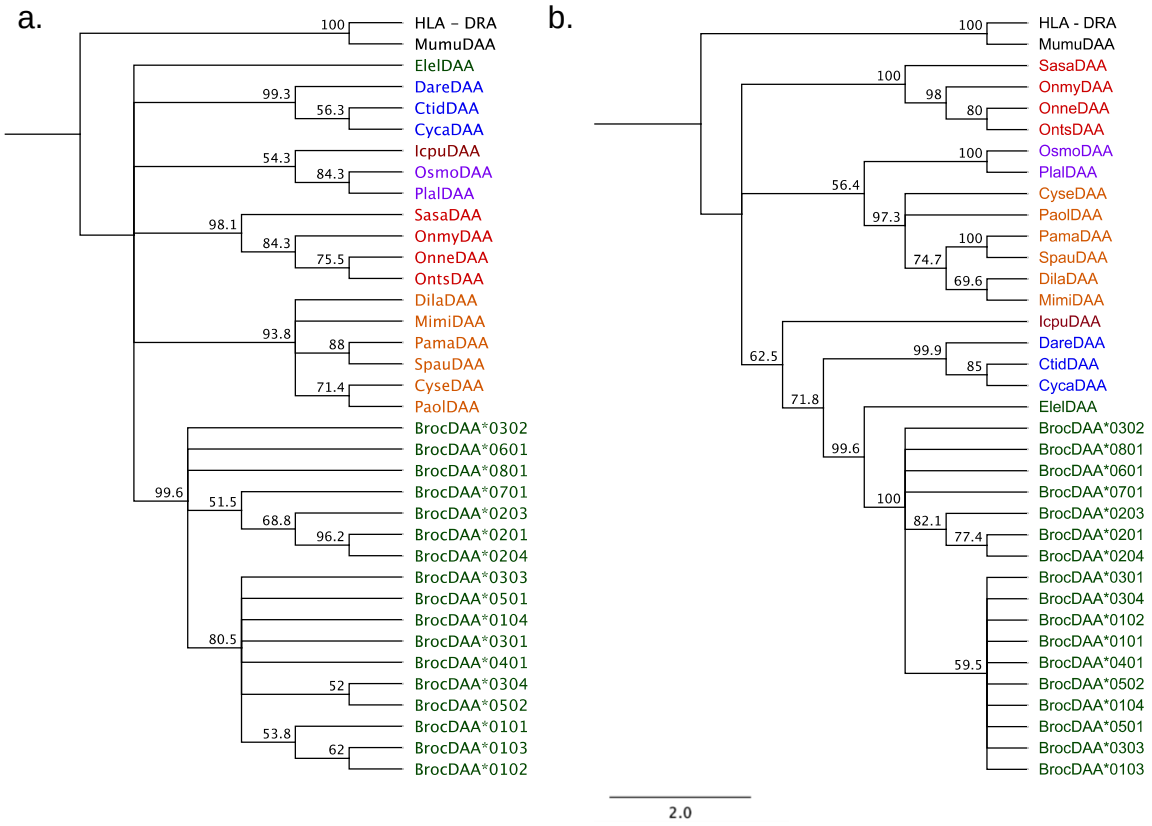

Fig 1.S2. Phylogenetic analysis of MHCII DAA sequences in *B. occidentalis* and other vertebrate groups. Neighbor-Joining trees were constructed using either the (a) amino acid sequences of the α1 domain, or (b) complete amino acid sequences. Numbers by each branch represent the percentage nodal support based on 10,000 bootstrap replicates. Accession numbers in Table A1.

CHAPTER 2: EXPLORING DRIVERS OF MHC DIVERSITY IN A WEAKLY ELECTRIC FISH, *BRACHYHYPOPOMUS OCCIDENTALIS*

Introduction

A central goal of evolutionary studies is to understand how selection pressures drive genetic variation in natural populations. Accordingly, the major histocompatibility complex (MHC) has been extensively studied for its extraordinary genetic diversity within and across lineages (Malmstrøm et al. 2016; Hedrick 2002; Miller, Allendorf, and Daugherty 2010; Christophe Eizaguirre et al. 2009). Found in all vertebrates, this set of genes is responsible for pathogen recognition and plays a key role in the adaptive immunity of an individual. MHC genes encode for cell surface proteins that bind to foreign peptides and present them to nearby T-cells, thus inducing an appropriate response (Altmann and Trowsdale 1989; Doherty and Zinkernagel 1975). In addition, this complex can be divided into two classes; Class I molecules are found on all nucleated cells and bind to viral peptide chains (Bjorkman & Parham, 1990), while Class II are localized solely on specialized antigen presenting cells (macrophages, dendric and B-cells) and work with extracellular antigens (Alfonso & Karlsson, 2000; Hughes & Nei, 1989).

High levels of polymorphism at these loci are thought to be maintained by several mutually non-exclusive processes. Pathogen driven selection is frequently cited as the main source influencing MHC diversity, specifically through rare allele advantage (Takahata and Nei 1990), heterozygote advantage (Doherty and Zinkernagel 1975; J. Klein and

O'Huigin 1994), and fluctuating selection (Gandon & Nuismer, 2009; Summers et al., 2003; Wegner et al., 2003). In the frequency dependent hypothesis, pathogens and hosts engage in an arms race of adaptations to outcompete each other. Rare alleles are advantageous and become less rare, while pathogens eventually adapt, and a new allele is favored (Christophe Eizaguirre et al. 2012; Takahata and Nei 1990; Clarke and Kirby 1966). Genetic diversity can also be promoted through heterozygote advantage, where diverse heterozygotes are better able to detect a wider variety of pathogens than homozygotes (Clarke & Kirby, 1966; Wegner, 2004). Lastly, pathogen communities may differ across space and time and populations may be under differing selective pressures (Summers et al. 2003; K Mathias Wegner, Reusch, and Kalbe 2003; Gandon and Nuismer 2009). With little to no gene flow, fluctuating selection could thus drive diversification within a lineage.

Experimental studies have strongly supported these theories; for example, Eizaguirre et al., (2012) infected several generations of three-spined stickleback with parasites, which resulted in specific pathogen resistant MHC alleles increasing in frequency over a relatively short amount of time. Testing this in natural populations, though, has proven challenging. Peng et al., (2021) found that while certain MHC alleles correlated with parasite resistance, neutral processes best explained genetic differences among wild populations of three-spined stickleback. While balancing selection has dominated the ever-on-going debate of what influences and maintains MHC diversity, genetic drift has been found to frequently outweigh selection in driving variation in natural populations (Miller, Allendorf, and Daugherty 2010). Inconsistencies in empirical studies are most likely due to the many intertwined variables that shape genetic diversity in natural

populations. Variables such as the evolutionary history and size of a population can significantly impact the ability of balancing selection to overcome drift in promoting MHC diversification (Ardern et al. 1997; Ejsmond and Radwan 2011).

Here, we use a weakly electric neotropical fish, *Brachyhypopomus occidentalis*, to explore potential factors driving MHC diversity. Electric fishes are known for their ability to produce species specific electric organ discharges (EODs) primarily used for navigation and communication (Albert 2001; Hagedorn 2008; Fugère, Ortega, and Krahe 2011). They are widespread throughout South America (W. G. Crampton 1996; Albert 2001) and a handful of species are found in lower Central America (Albert and Crampton 2006; Alda et al. 2013). It is thought that *B. occidentalis* colonized Panama in several colonization events (Picq et al. 2014), leading to isolated populations and increasing genetic divergence (Aguilar et al. 2018; Picq et al. 2014). Previous research on *B. occidentalis* suggests that interpopulation EOD differences are primarily due to genetic drift and isolation by distance (Picq et al. 2016); however, there is some evidence that the MHC loci could be under positive selection (Silva et al. to be submitted). Thus, *B. occidentalis* makes a particularly good system to investigate the variables associated with MHC diversity in wild populations. In this study, we use available genetic (MHC and COI) data, as well as information on a subset of parasite communities sampled, to explore **(1) the variation of the classical MHCIIβ loci in populations of *B. occidentalis* throughout the isthmus of Panama; (2) if natural selection, more so than drift, is primarily responsible for genetic variation across populations**, and **(3) potential associations between MHCIIβ variation and parasite communities.**

Materials and Methods

Data Collection

Fish were located and caught as described in Chapter 1. All samples were frozen in -20°C until dissections or genetic sequencing could take place. Sampling took place throughout the dry season of 2018 and 2019 (Jan – May) in a total of 19 rivers across 10 drainages of Panama. (Fig 1). We used available genetic data from 15 rivers however, only three sites overlapped with sites sampled for parasites. An additional three sites were sampled solely for parasite dissections, although rivers were not identical, the drainages they fed into overlapped with available genetic data. When possible, associations between populations at the river level was used (genetic analysis); broader scale between drainage comparisons were used whenever river sites did not match (parasite analysis). Summary of data used can be found on Table 2.

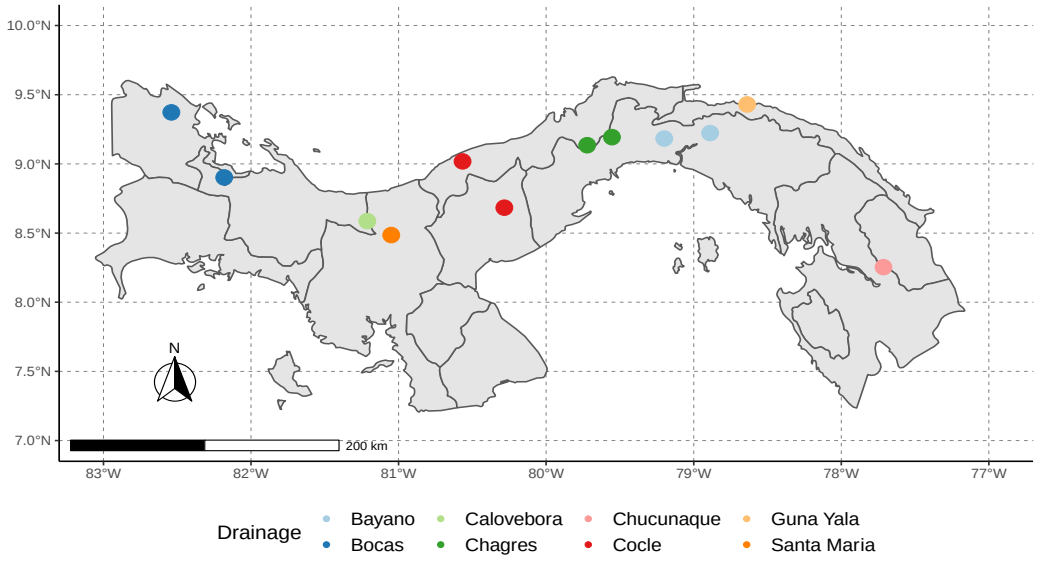

Fig 2.1. Map of drainages sampled within Panama

Parasite identification and diversity

Dissections for parasites took place after thawing frozen fish (approx. 15mins) and based on Kalbe et al. (2002). In summary, all ectoparasitic ciliates on the external right side of the fish were tallied under a dissecting microscope. Internal organs such as stomach, intestines and gallbladder were removed, and any nematodes were carefully spread and measured. The three right gill arches were also extracted, and parasitic monogeneans or protozoans were counted under a compound microscope (40X – 400X). Identification of parasites was made to the lowest taxonomic level possible by comparing key features seen in photographs with commonly known parasite families (Timofeeva, Gerasev, and Gibson 1997; P. Reed et al. 2009; Scholz, Aguirre-Macedo, and Salgado-Maldonado 2001; Yanong 2002; Kearn 1986). We calculated several metrics to estimate parasite diversity based on commonly used methods (Rózsa et al., 2000; Shaw et al., 2018). Namely, we used number of parasite families per fish (diversity), percentage of fish parasitized out of all fish dissected (prevalence), and number of parasites per infected host (intensity). We tested for fluctuating selection by first establishing that there was a significant difference in the parasitic communities within and between drainages. We repeated these calculations at the river and drainage level and tested for significance within both using a Wilcox t-test and the Holm-Bonferroni method (Holm 1979) to adjust p-values for multiple comparisons.

Estimating MHC and COI Alleles

The entire sequence of the classical MHCIIB and the mitochondrial cytochrome c oxidase I (COI) genes were obtained from the annotated genome of *B. occidentalis*. The reference genome was used to extract the MHCIIB and COI region from the NGS data of

the 62 fish sampled. Multiple sequence alignments were carried out through Geneious (v1.2) and only the coding regions of both genes were kept. We further confirmed the identity of this region by utilizing the BLAST function in the NCBI database. Alleles were assigned based on clusters identified in alignments; to minimize the influence of rare MHCII alleles, individuals were considered to have the same multilocus genotype (MLG) if at least 99% of sequence overlapped. Visual representation of allelic clustering was done through neighbor-joining trees (Saitou and Nei, 1987) and constructed in Geneious (v1.2) using 10,000 bootstrap replicates.

MHC Diversity

Patterns of MHCIIB variation within and among populations were estimated using R version 4.02 (R Development Core Team 2022) packages *poppr* (Kamvar, Tabima, and Grünwald 2014) and *adegenet* (Jombart 2015). Overall allelic diversity was calculated using the Simpson Index (λ), Shannon-Weiner Diversity Index (H), and standardized index of association (IA) with 1,000 permutations. Genetic diversity is influenced by both, the number of variants per locus (genotypic richness) and how well genotypes are distributed within a sample (genotypic evenness). Due to varying sample sizes, we estimated the expected number of multilocus genotypes (eMLG) as a measure of genetic richness per site. Distribution of alleles were calculated using an evenness index ($E_{.5}$); where 0 represents a fixed allele in a population, and 1 indicates equally abundant genotypes (Ludwig and Reynolds 1988; Grünwald et al. 2003). As opposed to other metrices of

evenness, $E_{.5}$ is the ratio of abundant to rare genotypes in a sample, and it is therefore less easily influenced by sample size (Grünwald et al., 2003).

Population structure and geographic isolation

To assess population structure based on both MHC and COI sequence data, we calculated genetic variation within and between rivers and drainages using an analysis of molecular variance (AMOVA). Specifically, we tested if genetic diversity was highest within sites, between sites of the same drainage, or between drainages. Significance of the AMOVAs was established by performing randomization tests as described by Excoffier et al., (1992). Random data was produced using a 10,000 iteration Monte Carlo simulation based on farthest neighbor distances. These analyses were performed to evaluate if hierarchal structures differed amongst the two genes, and whether different factors (e.g., selection or drift) may be influencing genetic variation at the MHC and COI loci. To evaluate the effect of geographic isolation or drift on genetic diversity, a pairwise genetic distance matrix was constructed using Nei's Gst for both the MHCII and COI genes. High Gst values indicate low heterozygosity and fixed alleles within a population, while a low Gst score (~ 0.15 to 0.3) generally suggests high amounts of genetic variation. Euclidean distances between sites were calculated using the *stats* package. We tested whether genetic variation was associated with physical distances between rivers by comparing matrixes using a Mantel test within the *vegan* package (Oksanen, Simpson, and Blanchet 2022).

Selection Tests

We calculated expected (Hexp) and observed (Hobs) heterozygosity for the MHCIIβ, and tested if allele frequencies were within Hardy-Weinberg (HW) expectations using the *pegas* package (Paradis and Barrett 2010). Further tests for selection were conducted using SLAC (Single Likelihood Ancestor Counting), MEME (Mixed Effects Model of Evolution) and GERD (Genetic Algorithm for Recombination Detection) methods in the HyPhy software (Kosakovsky Pond et al., 2005a) through the Datamonkey webserver (Kosakovsky Pond and Frost 2005b). MEME calculates two mutation rates (ω) per site, and the probability that each site evolved under diversifying ($\omega > 1$) or purifying ($\omega < 1$) selection. Only sites with a probability greater than 95% chance ($p < 0.05$) of occurring were considered to be under selection. Since recombination events can greatly influence selection results, we first used GERD (Genetic Algorithm for Recombination Detection) analysis to test for recombination (Pond et al. 2006). To test whether selection or drift is influencing genetic variation in *B. occidentalis*, we calculated synonymous (dS) and non-synonymous (dN) substitution rates for each site within the MHCIIβ and COI genes. The dN/dS rate of the MHCIIβ was compared to that of the neutral COI gene using a two tailed t-test.

Associations between parasites and MHC diversity

To evaluate any associations between parasitic pressures and genetic variation, we compared parasite diversity metrics (prevalence and intensity) and MHC variables (dN/dS and eMLG). We implemented a spearman's rank correlation between each potential

combination of parasite and MHC diversity metrics. In this analysis, we examined the parasitic communities of two rivers within the Bayano, Chagres and Santa Maria drainages respectively. Within each drainage, one river had genetic data available, but the other did not. Since parasite and genetic data were not all sampled from identical rivers, we calculated a mean genetic diversity variable from within each drainage to represent the missing sites. A Shapiro-Wilk test indicated that data was not normally distributed, therefore we used non-parametric tests in this analysis.

Results

Parasite Diversity

Of the 101 fish that were dissected, 78 were found to be infected with at least one type of parasite family. The parasites identified included: ectoparasitic protozoans (Ichthyophthiriidae; Fig 2a), round worms (Ascarididae; Fig 2c), monogenean flukes (Dactylogyrus Diesing 1850; Figs 2b,d,f), and internal ciliated protozoans (Trichodinidae, Fig 2e). The latter two appeared exclusively in the gills, and encompassed the majority of parasites found (86%). Monogenean adult flukes and eggs were especially abundant and found within all rivers sampled. On the other hand, round worms were only detected in very few (< 5%) fish within the Tapagrilla river, and therefore excluded from prevalence or intensity comparisons. Trichodinidae were also excluded from all but diversity analysis, given that it was only found in one fish from the Rio Toare Impressively, this fish suffered from an intense infection, with 102 Trichodinidae counted within its gills.

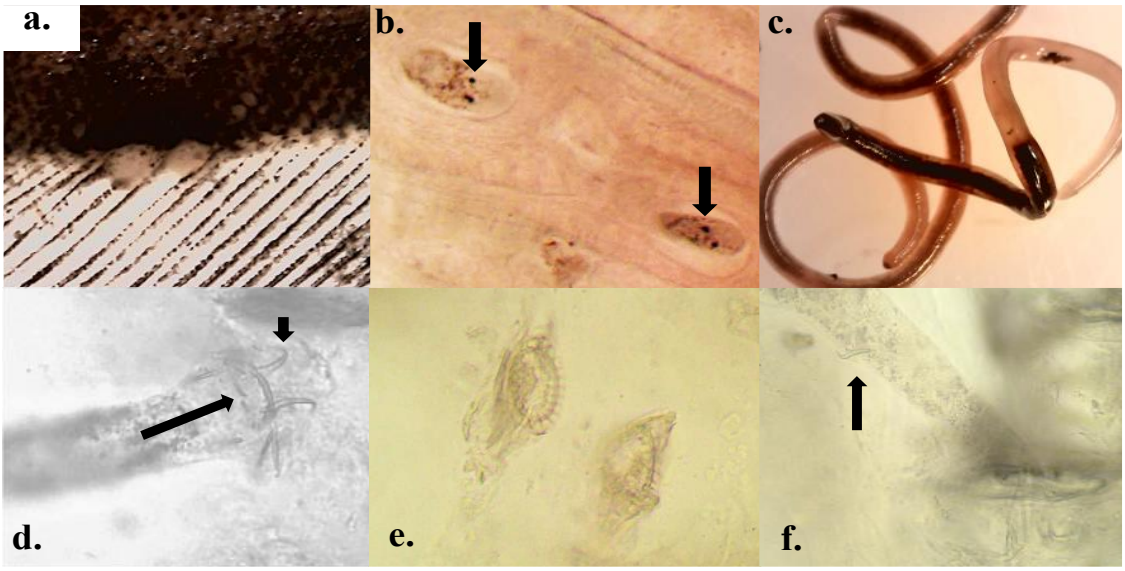

Fig 2.2 Common parasites found on and in *B. occidentalis.* arrows point toward key features of the Dactylogyrus taxon. (a) 10X Magnification: two *Ichthyophthirius multifiliis* on ventral fin (b) 400X: Dactylogyrus eggs between gill filaments, arrows pointing toward characteristic eye spots forming (c) 10X: uncoiled Ascarididae (d) 400X: Mature Dactylogyrus with opisthaptor organ attached to gill, anchor hooks (short arrow) and transverse bar (long arrow) are shown (e) 400X: two Trichodinidae seen in gills (f) 400X: Mature Dactylogyrus with accessory piece to copulatory organ shown

Prevalence of parasites within the six rivers sampled (Fig 3) varied from the Quebrada Honda (57.2% infected) to Flor de Laguna (100% parasitized). Flor de Laguna was unique because all fish were infected with at least one parasite and the intensity of the infection was remarkably high (mean 26.5 parasites per infected fish), but overall diversity was low (2 parasite families represented). In contrast, the Tapagrilla is within the same Bayano drainage as Flor de Laguna but showed opposite results; diversity was high (4 parasitic families), but intensity was low (mean of 7.5 parasites per infected fish). Significant intra-drainage variation in total parasites was only found within the Bayano (p

= 0.009), while inter-drainage differences were detected between the Santa Maria and Bayano (p = 0.039).

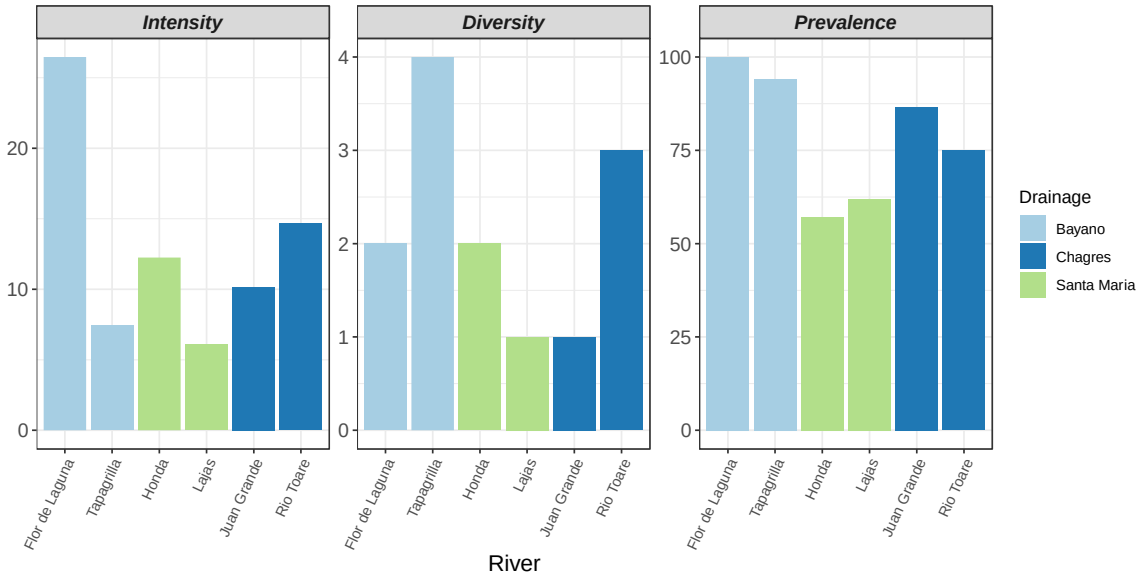

Fig 2.3. Parasitic pressures across rivers and drainages of Panama. Intensity (parasites per fish infected), Diversity (total parasite families) and Prevalence (percentage of fish parasitized).

MHC Diversity

The 723bp coding sequence of the MHCIIβ gene translates to a putative peptide chain of 241 amino acids (aa). Through BLAST searches, these sequences were identified as the β gene of the classical MHCII DA- lineage, and thus alleles were named *Broc*DAB*. Aligning all sequences to the annotated reference allowed us to identify key features of this gene, such as: exon/intron locations, peptide binding regions (PBR), conserved cysteines sites, and predicted N-glycosylation sites. As expected, this region was extremely diverse and highly variable across populations. A total of 91 unique *Broc*DAB sequences were

obtained from 62 diploid fish (length = 11.52cm ± 4.69cm; weight = 5.98g ± 4.34g), with only 9 genotypes present across more than one river (Table 1). Interestingly, no genotypes were shared amongst more than two rivers, and there were more similarities across drainages than within. The Bayano drainage (especially the Tugamantí river) was the most well represented, sharing alleles with the Chagres, Cocle, and Guna Yala drainages. The Bayano and Chagres drainages also had common alleles shared between rivers.

MLG	n	Rivers	
35	3	Tugamantí	Azucar
39	2	Tugamantí	Tambo
54	3	Juan Grande	La Puente
68	2	Tugamantí	Tapagrilla
74	3	Tapagrilla	La Puente
84	2	Tugamantí	Tapagrilla
109	6	La Puente	Pena Bijagual
111	2	La Puente	Pena Bijagual
112	3	Tugamantí	Juan Grande

Table 2.1. MHCII alleles shared between rivers. The id number of the shared genotype (MLG), total number of individuals with this genotype (n), and rivers where it was detected, is shown.

Genetic variation of the MHCIIβ was calculated by river populations using several variables (Table 2). Diversity indices such as the Shannon-Weiner Index (H) ranged from 0.693 to 2.398 (mean H = 1.68, sd = ± .61). The evenness index was high across all populations (mean $E_{.5}$ = .91, sd = ± .062), indicating that genotypes were similarly distributed throughout rivers. Since differing sample sizes could be skewing the number

of alleles observed, we calculated expected number of genotypes (eMLG) per population (mean eMLG = 6.39, sd = ± 3.10). Diversity was highest in the Bocas drainage (eMLG = 10, $E_{.5}$ = 1, H = 2.3), with all unique sequences being equally represented within the two rivers sampled. The lowest genetic variation was found in the Chagres, Cocle and Atrato drainages (Fig 4).

Drainage	River	Data	n	eMLG	E.5	H	Prevalence	Intensity	Diversity
Santa Maria	Honda	Parasites	14				85.714	10.5	2
	Rio Lajas	Parasites	21				61.905	6.076	1
		Genetic	5	7	0.87	1.834			
Bocas	Camaron	Genetic	5	10	1.00	2.302			
	Segla	Genetic	5	10	1.00	2.302			
Bayano	Flor de Laguna	Parasites	14				100.000	28.357	2
	Tapagrilla	Parasites	17				94.118	7.438	4
		Genetic	5	7	0.93	1.887			
	Tugamantí	Genetic	5	9	0.91	2.398			
Atrato	Leon	Genetic	2	3	0.92	1.039			
	Atrato	Genetic	2	2	1.00	0.693			
Chagres	Juan Grande	Parasites	15				86.667	10.153	1
		Genetic	4	5	0.95	1.560			
	Rio Toare	Parasites	12				75.000	14.667	3
	La Puente	Genetic	5	7	0.87	1.834			
Cocle	Cocle del Norte	Genetic	3	2	0.89	0.637			
	Tambo	Genetic	2	2	1.00	0.693			
Calovebora	Calovebora	Genetic	5	6	0.78	1.748			
Chucunaque	Peña Bijagual	Genetic	5	7	0.84	1.973			
Guna Yala	Azucar	Genetic	5	8	0.95	2.164			
San Juan	San Juan	Genetic	4	6	0.88	2.079			

Table 2.2. Information on data used and diversity calculated per river. Including total fish (n), expected multilocus genotypes (eMLG), evenness index (*E*.5), Shannon – Weiner diversity index (H) based on genetic information available. Parasite diversity variables (prevalence, intensity, and diversity) also included for a subset of rivers sampled.

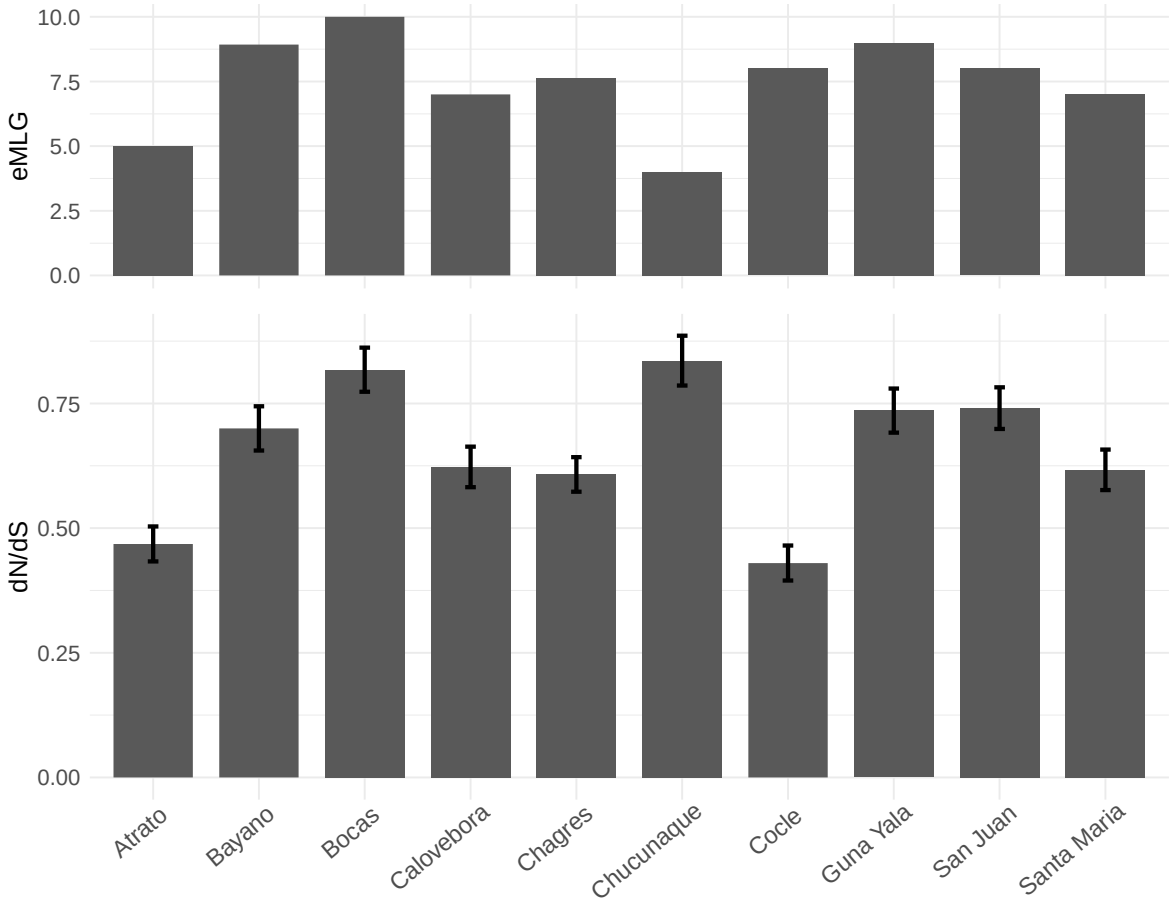

Fig 2.4. Diversity of the MHCIIβ in *B. occidentalis* among drainages. Included are the estimated number of multilocus genotypes (eMLG), and average non-synonymous to synonymous mutation rate (dN/dS) with standard error bars shown.

Selection or Drift

All *Broc*DAB alleles failed to meet the expectations of the HW equilibrium ($p < .001$), indicating that drift or selection may be influencing genetic variation. We then compared the dN/dS rate of the *Broc*DAB to that of the neutral COI gene (Table 3), with the assumption that rates would be equal under neutral processes. The average dN of all sites within the DAB gene was significantly higher than within the COI ($p = 0.0001$), while the dS showed no significant difference between the two ($p = 0.31$). Although non-

synonymous mutations occurred at significantly higher rates within the MHCIIβ than within the COI ($p = 0.0001$), both loci had sites that were predicted to be under selection. Based on the MEME analysis, a total of 31 sites within the DAB are likely to be under positive, or diversifying ($\omega > 1$) selection, while 28 sites within the COI are under negative, or purifying ($\omega < 1$) selection. Remarkably, 18 of the 31 MHCIIβ sites predicted to be under selection were detected within the PBR of exon 2.

Gene	N	dN	dS	dN/dS	ω > 1	ω < 1
DAB	241	2.65 ± .414	1.04 ± .2156	2.54	31	0
COI	519	0.0366 ± .0095	0.746 ± .0448	0.0491	0	28
		***	ns	***		

Table 2.3. Comparison of mutation rates between DAB and COI genes of *B. occidentalis*. Number of sites per gene (N), as well as the mean non-synonymous (dN) and synonymous (dS) substitution rates (± Standard Error) for all sites within the DAB and COI loci of *B. occidentalis*. Number of sites predicted to be under diversifying (ω > 1) or purifying (ω < 1) selection. Significant differences between selection rates were calculated using t-tests (* = p < .0005)**

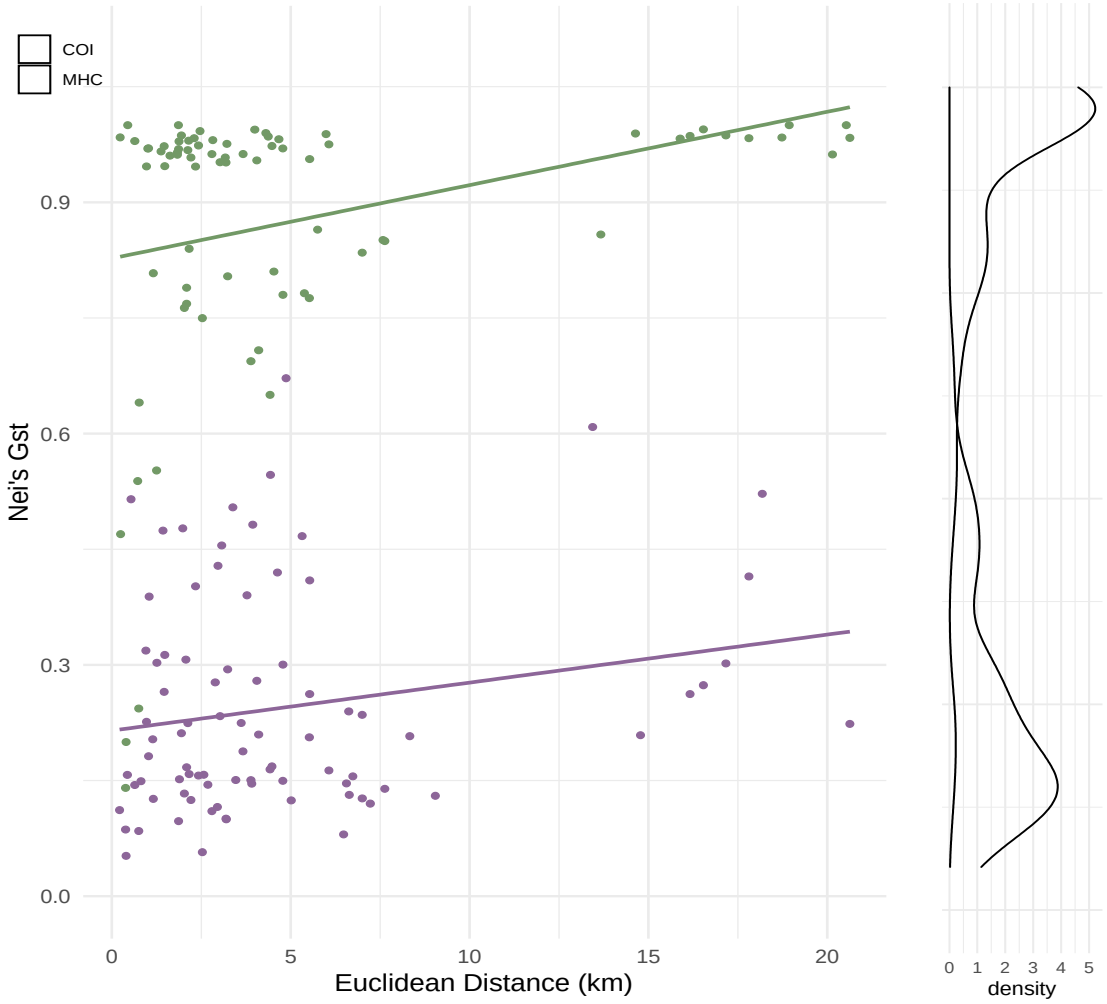

Fig 2.5. Pairwise genetic (Nei's Gst) and physical distance correlations. Pairwise distances of MHCIIβ (purple) and COI (green) of *B. occidentalis* are shown. A general linear model with 95% confidence interval is shaded in. A density graph shows the distribution of Gst values.

Overall, the analyses of isolation by distance showed some positive associations between pair-wise genetic distances (Gst) and geographic distance for both MHCIIβ and COI genes. However, the Mantel test showed that this association was only statistically significant in the COI gene ($r = 0.312$, $p = 0.019$), and not for the MHCIIβ gene ($r = 0.198$, $p = 0.16$). Though Gst values can never reach 1 in a multi allelic locus, values calculated based on the COI sequences came close (mean Gst = 0.88 ± 0.18). However, the overall range of Gst values were low for *Broc*DAB (mean Gst = 0.24 ± 0.14) and implies there is little population structure. The AMOVA showed that variation was highest within populations ($\sigma = 20.63$, % = 71) for the *Broc*DAB, further corroborating a lack of hierarchal structure. Opposite results were seen in the COI genes, with most of the variation being detected across drainages ($\sigma = 73.27$, % = 93).

Parasite and MHC associations

After establishing that selection, not drift, is driving genetic variation at the MHCII loci of *B. occidentalis* we tested for associations between parasite community metrics and genetic diversity (Fig 5). Intensity of parasite infection was not strongly associated with either genetic variable, and estimated number of MHC alleles also showed little correlation with parasite metrics. However, the dN/dS rate was significantly associated with prevalence of parasites ($rho = 0.956$, $p = 0.003$). That is, drainages where parasites were more prevalent also showed higher rates of non-synonymous to synonymous mutations within the MHCIIβ.

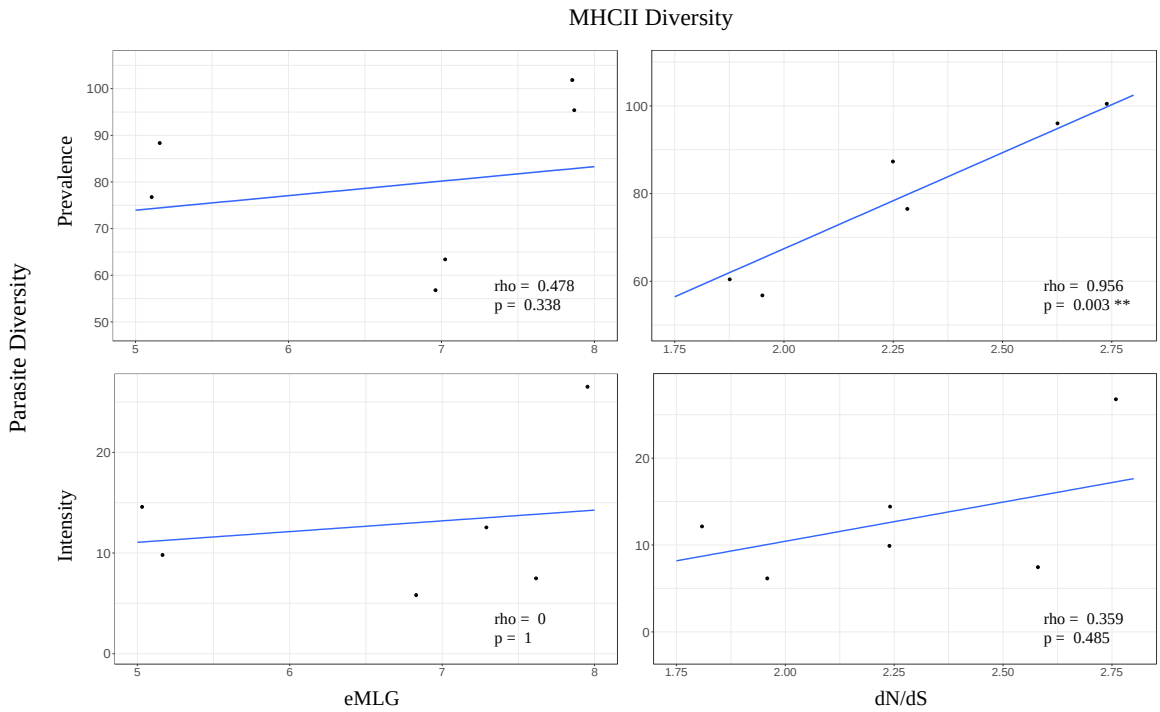

Fig 2.6. Associations between parasite and genetic diversity. Spearman's Rank Correlations (rho) between Genetic and Parasitic Diversity. Intensity and Prevalence of parasitic infections was compared to number of alleles (eMLG) and mutation rate (dN/dS) of the *Broc*DAB. Spearman's Rho was calculated for all associations, with a general linear model and 95% confidence interval shaded.

Discussion

Parasite Diversity

Monogenean flukes are one of the most frequently found parasites in freshwater fish (Reed et al. 2009; Timofeeva, Gerasev, and Gibson 1997) and as expected, were the most abundant type detected in *B. occidentalis*. Based on marginal hooks, presence of eye spots, and visible internal eggs (Figs 2 & 7), we determined the gill trematodes to belong to the Dactylogyrus family (Reed et al. 2009). Overall physical similarities were noted, but it is nearly impossible to visually confirm all features in all flukes counted, thus it is entirely

possible that some parasites were misidentified. Differences between Dactylogyrus and other monogenean families are marginal; all have similar body morphs and affect the host in parallel ways, therefore underestimating fluke diversity should not change our results.

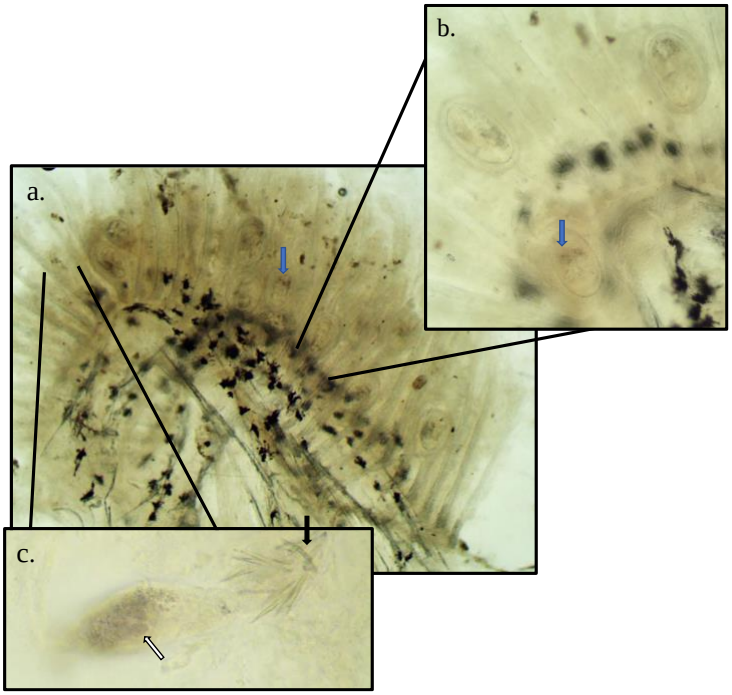

Fig 2.7. Monogenean eggs and adults parasitizing gill of *B. occidentalis.* Arrows point toward developing eye spots (blue arrow), transverse bar (black arrow), and immature eggs (white arrow). (a) 40X Magnification: Heavily parasitized gill arch; (b) 400X: three mature eggs; (c) 400X: adult fluke with developing eggs

Unlike some monogeneans, Dactylogyrus are oviparous and adults will lay 4-10 eggs every few days (Reed et al. 2009). Of the fish we sampled, 62 had monogenean infections, and 41 of those had eggs as well as adult flukes present within the gill filaments. This was interesting to note since Dactylogyrus eggs typically make their way out of the original host and settle on the bottom of the water column; however, immature eggs can occasionally be seen during microscopic analysis (Reed et al. 2009). While most rivers had

an expected ratio of adult flukes to underdeveloped eggs, Flor de Laguna had notably heavy infestations of late-stage monogenean eggs (mean = 25.8 eggs per adult, sd = 18.2). The overabundance of mature eggs within the gills was unique and not described in literature. Dactylogyrus is one of the most diverse families of parasites; more than 900 species have been described (Rehulková, Benovics, and Šimková 2020), and diversity is likely greatly underestimated (Timofeeva, Gerasev, and Gibson 1997). It is therefore feasible that the flukes seen in Flor de Laguna are unique in laying eggs within the gill filaments. Another possibility exists that the eggs seen do not belong to the same adult monogenean flukes also detected. The metacercaria of digeneans can resemble eggs of monogeneans; however, the former have complicated life cycles and often attach to fish as a final host in their adult form (Reed et al. 2009; Scholz, Aguirre-Macedo, and Salgado-Maldonado 2001).

MHC Diversity

Overall, there is strong support for balancing selection acting on the classical MHCIIβ genes of *B. occidentalis.* Variation was highest within rivers than between drainages and populations lacked structure, suggesting that internal selective pressures are driving genetic differences. All sites had higher than expected heterozygosity under the HW equilibrium, even when accounting for unequal sample sizes. Though calculating estimated alleles (eMLG) certainly helped mitigate the challenges of uneven data distribution, the differing sample sizes most likely still influenced this analysis.

Selection and Drift

In contrast, the COI was less polymorphic, and variation was heavily associated with geographic distance. There was clear population structure at the COI loci, which could be seen in the phylogenetic clustering of alleles and suggests isolation by distance is mostly driving diversity. Interestingly, selection was detected in 28 of the 519 sites of the COI gene, but all were predicted to be under purifying selection (Table 3). Negative, or purifying selection, decreases genetic diversity by selecting against adverse alleles and promoting fixation of beneficial alleles (Kosakovsky Pond and Frost 2005a; Decker, Stewart, and Lehman3 2002). Within the MHCIIβ, 31 of the 241 total sites were predicted to be under diversifying selection. The PBR only makes up about 9% of the MHC, but 58% of sites under selection were found in this region. Our results are consistent with other studies that often find higher levels of diversity within the PBR than the rest of the MHC (Miller, Allendorf, and Daugherty 2010). Selection may be stronger within the PBR due to the direct binding interactions between these sites and foreign peptides derived from bacterial, fungal, or parasitic infections (Hughes and Nei 1989; Alfonso and Karlsson 2000).

Parasites and MHC Variation

Parasite mediated selection is hypothesized to be a major driving force in influencing MHC diversity (J. Klein and O'Huigin 1994; K Mathias Wegner 2004; Nevo and Beiles 1992; C. Eizaguirre and Lenz 2010; Christophe Eizaguirre et al. 2012; K Mathias Wegner, Reusch, and Kalbe 2003). To test for this, we calculated diversity metrics (intensity and prevalence) based on tallied parasites found on the skin, gills, and internal

cavity of *B. occidentalis* and tested for any associations with metrics of genetic diversity (eMLG and dN/dS). We found that signals of parasite mediated selection on the MHC loci of *B. occidentalis* were weaker than expected, but there was a strong association between mutation rate and prevalence of parasites within rivers (Fig 6). Although more research is certainly needed, this pattern suggests that parasites could, at least partially, be affecting genetic variation in the immune genes of *B.* occidentalis. Interestingly, the intensity of parasite infection showed little to no correlation with number of MHCIIβ alleles or mutation rate. Since data was compared at a population level, it follows that overall percentage of fish parasitized (prevalence) was a more accurate representation of parasite diversity than intensity. Moreover, intensity of infections differ greatly between individuals and through time (Reed et al. 2009; Zargar et al. 2012). Thus, to get a better understanding of the contribution of parasite intensity on MHC diversity, the same individuals and a larger sampling size need to be analyzed. For instance, while not significant, there was a slight correlation between number of MHCIIβ alleles and prevalence (rho = 0.478, p = 0.338). Using estimated or observed alleles did little to change this result, but our small sample size most likely underestimated allelic diversity within populations.

Conclusion

It is important to note, that we also investigated the diversity of a non-classical MHCIIβ gene in *B.occidentalis.* Remarkably, all 124 sequences were identical and lacked variation across individuals, suggesting a lack of variation in this gene. and therefore, the analysis could not continue further than the sequence alignment step. While non-classical

MHC molecules can influence the peptide-binding process, they often lack direct binding capabilities to antigens (Alfonso and Karlsson 2000). It is therefore expected that the non-classical genes are not as diverse at the classical MHC, but the extreme homozygosity seen indicates that the *Broc*DBB allele was fixed within this species prior to the colonization of Panama.

Overall, this study provides novel information on the MHCIIβ diversity of several natural populations of *B.occidentalis*. With very little parasitic studies focused in the neotropics, this research also provides exploratory data on parasite communities within a freshwater fish of Panama. We show that selection, over genetic drift, is clearly driving variation at the MHC loci. Although parasitism may be influencing some of the diversity seen, further research is needed to investigate other variables. Studies have recently started to focus on how multiple direct and indirect interactions influence MHC variation. Indirectly, microbial communities could play a role in selecting against MHC diversity (Bonder et al. 2016; Bolnick et al. 2014). The theory being that individuals with high MHC heterozygosity would select against some microbial communities. While further research is needed on this topic, our study provides exploratory data that expands our understanding of what is driving diversification within the neotropical electric fishes.

REFERENCES

Aguilar, Celestino, Matthew J Miller, Jose R Loaiza, Rüdiger Krahe, and Luis F de 4 León. 2018. "Mitochondrial Genomes and Phylogenetic Analysis of Central American Weakly-Electric." https://doi.org/10.1101/353565.

Aguilar, Celestino, Matthew J. Miller, Jose R. Loaiza, Rüdiger Krahe, and Luis F. de León. 2019. "Mitogenomics of Central American Weakly-Electric Fishes." *Gene* 686: 164–70. https://doi.org/10.1016/J.GENE.2018.11.045.

Albert, James S. 2001. "Species Diversity and Phylogenetic Systematics of American Knifefishes (Gymnotiformes, Teleostei)." *Miscellaneous Publications*, no. 190.

Albert, James S., and William G. R. Crampton. 2006. *Diversity and Phylogeny of Neotropical Electric Fishes (Gymnotiformes)*. *Electroreception*. https://doi.org/10.1007/0-387-28275-0_13.

Alda, Fernando, Sophie Picq, Luis Fernando de León, Rigoberto González, Henriette Walz, Eldredge Bermingham, and Rüdiger Krahe. 2013. "First Record of Gymnotus Henni (Albert, Crampton and Maldonado, 2003) in Panama: Phylogenetic Position and Electric Signal Characterization." *Check List* 9 (3): 655–59. https://doi.org/10.15560/9.3.655.

Alfonso, Christopher, and Lars Karlsson. 2000. "NonClassical MHC Class II Molecules." *Annu. Rev. Immunol* 18: 113–42. www.annualreviews.org.

Altmann, D. M., and J. Trowsdale. 1989. "Major Histocompatibility Complex Structure and Function." *Current Opinion in Immunology* 2 (1): 93–98. https://doi.org/10.1016/0952-7915(89)90103-9.

Alves de Sá, André Luiz, Breanna Breaux, Tibério Cesar Tortola Burlamaqui, Thaddeus Charles Deiss, Leonardo Sena, Michael Frederick Criscitiello, and Maria Paula Cruz Schneider. 2019. "The Marine Mammal Class II Major Histocompatibility Complex Organization." *Frontiers in Immunology* 10: 1–14. https://doi.org/10.3389/fimmu.2019.00696.

Amadou, Claire, Attila Kumanovics, Elsy P Jones, Doris Lambracht-Washington, Masayasu Yoshino, and Kirsten Fischer Linduhl. 1999. "The Mouse Major Histocompatibility Complex: Some Assembly Required." *Immunologicat Reviews* 167: 211–21. https://doi.org/10.1155/2014/159247.

Ardern, S L, D M Lambert, A G Rodrigo, and I G Mclean. 1997. “The Effects of Population Bottlenecks on Multilocus DNA Variation in Robins The Effects of Bottlenecks on Genetic Variation in Two Wild Populations of New Zea-Land Robins (Petrolca Auatralia Auatralls) Were Examined Using Multilocus Minisat-Eiiite DNA Probes.” *Journal of Heredity*. https://academic.oup.com/jhered/article/88/3/179/2186556.

Arias, Carlos F, Rebecca B Dikow, W Owen Mcmillan, and Luis F de Le On. 2021. “De Novo Genome Assembly of the Electric Fish Brachyhypopomus Occidentalis (Hypopomidae, Gymnotiformes) GBE.” *Genome Biol. Evol* 13 (10). https://doi.org/10.1093/gbe/evab223.

Beck, S., D. Geraghty, H. Inoko, L. Rowen, B. Aguado, S. Bahram, R. D. Campbell, et al. 1999. “Complete Sequence and Gene Map of a Human Major Histocompatibility Complex.” *Nature* 401 (6756): 921–23. https://doi.org/10.1038/44853.

Bjorkman, P. J., and P. Parham. 1990. “Structure, Function, and Diversity of Class I Major Histocompatibility Complex Molecules.” *Annual Review of Biochemistry* 59: 253–88. https://doi.org/10.1146/ANNUREV.BI.59.070190.001345.

Bolnick, Daniel I., Lisa K. Snowberg, J. Gregory Caporaso, Chris Lauber, Rob Knight, and William E. Stutz. 2014. “Major Histocompatibility Complex Class IIb Polymorphism Influences Gut Microbiota Composition and Diversity.” *Molecular Ecology* 23 (19): 4831–45. https://doi.org/10.1111/mec.12846.

Bonder, Marc Jan, Alexander Kurilshikov, Ettje F Tigchelaar, Zlatan Mujagic, Floris Imhann, Arnau Vich Vila, Patrick Deelen, et al. 2016. “The Effect of Host Genetics on the Gut Microbiome.” *Nature GeNetics VOLUME* 48. https://doi.org/10.1038/ng.3663.

Carroll, Lara S., Dustin J. Penn, and Wayne K. Potts. 2002. “Discrimination of MHC-Derived Odors by Untrained Mice Is Consistent with Divergence in Peptide-Binding Region Residues.” *Proceedings of the National Academy of Sciences of the United States of America* 99 (4): 2187–92. https://doi.org/10.1073/PNAS.042244899/ASSET/2809040F-5485-49D8-B4C6-7DBA38EE52DE/ASSETS/GRAPHIC/PQ0422448002.JPEG.

Clarke, Bryan, and D.R.S Kirby. 1966. “Maintenance of Histocompatibility Polymorphisms.” *Nature* 211 (185): 999–1000. https://doi.org/10.1038/hdy.1968.17.

Cosson, Pierre, and Juan S Bonifacino. 1992. “Role of Transmembrane Domain Interactions in the Assembly of Class II MHC Molecules.” *Cell Biology and Metabolism Branch*. Vol. 150. https://doi.org/10.1126/science.1329208.

Crampton, W. G. 1996. “Gymnotiform Fish: An Important Component of Amazonian Floodplain Fish Communities.” *Journal of Fish Biology* 48 (2): 298–301. https://doi.org/10.1006/jfbi.1996.0029.

Crampton, William G.R. 2019. “Electroreception, Electrogenesis and Electric Signal Evolution.” *Journal of Fish Biology* 95 (1): 92–134. https://doi.org/10.1111/JFB.13922.

Cuesta, Alberto, Maria Ángeles Esteban, and José Meseguer. 2006. “Cloning, Distribution and up-Regulation of the Teleost Fish MHC Class II Alpha Suggests a Role for Granulocytes as Antigen-Presenting Cells.” *Molecular Immunology* 43 (8): 1275–85. https://doi.org/10.1016/j.molimm.2005.07.004.

Davis, Daniel J., Jenna Klug, Miriam Hankins, Holly M. Doerr, Stephanie R. Monticelli, Ava Song, Catherine H. Gillespie, and Elizabeth C. Bryda. 2015. “Effects of Clove Oil as a Euthanasia Agent on Blood Collection Efficiency and Serum Cortisol Levels in Danio Rerio.” *Journal of the American Association for Laboratory Animal Science : JAALAS* 54 (5): 564. /pmc/articles/PMC4587626/.

Decker, D.J., B.S. Stewart, and N. Lehman3. 2002. “Major Histocompatibility Complex Class II DOA Sequences from Three Antarctic Seal Species Verify Stabilizing Selection on the DO Locus.” *Tissue Antigens* 60 (6): 534–38. https://doi.org/10.1034/J.1399-0039.2002.600610.X.

Dick, Tobias P, Naveen Bangia, David R Peaper, and Peter Cresswell. 2002. “Disulfide Bond Isomerization and the Assembly of MHC Class I-Peptide Complexes Be Subject to Redox Catalysis. One Pair Forms a Disulfide Bond in the Membrane-Proximal 3 Domain; the Other Pair Is Part of the Peptide Binding Groove, Forming a.” *Immunity* 16: 87–98.

Dobson, Andy, Kevin D. Lafferty, Armand M. Kuris, Ryan F. Hechinger, and Walter Jetz. 2008. “Homage to Linnaeus: How Many Parasites? How Many Hosts?” *Proceedings of the National Academy of Sciences of the United States of America* 105 (SUPPL. 1): 11482–89. https://doi.org/10.1073/pnas.0803232105.

Doherty, Peter C., and Rolf M. Zinkernagel. 1975. “Enhanced Immunological Surveillance in Mice Heterozygous at the H-2 Gene Complex.” *Nature* 256: 50–52. https://doi.org/https://doi.org/10.1038/256050a0.

Eizaguirre, C., and T. L. Lenz. 2010. “Major Histocompatability Complex Polymorphism: Dynamics and Consequences of Parasite-Mediated Local Adaptation in Fishes.” *Journal of Fish Biology* 77 (9): 2023–47. https://doi.org/10.1111/j.1095-8649.2010.02819.x.

Eizaguirre, Christophe, Tobias L. Lenz, Martin Kalbe, and Manfred Milinski. 2012. "Rapid and Adaptive Evolution of MHC Genes under Parasite Selection in Experimental Vertebrate Populations." *Nature Communications* 3. https://doi.org/10.1038/ncomms1632.

Eizaguirre, Christophe, Tobias L. Lenz, Ralf D. Sommerfeld, Chris Harrod, Martin Kalbe, and Manfred Milinski. 2011. "Parasite Diversity, Patterns of MHC II Variation and Olfactory Based Mate Choice in Diverging Three-Spined Stickleback Ecotypes." *Evolutionary Ecology* 25 (3): 605–22. https://doi.org/10.1007/s10682-010-9424-z.

Eizaguirre, Christophe, Tobias L Lenz, Arne Traulsen, and Manfred Milinski. 2009. "Speciation Accelerated and Stabilized by Pleiotropic Major Histocompatibility Complex Immunogenes." *Ecology Letters* 12 (1): 5–12. https://doi.org/10.1111/j.1461-0248.2008.01247.x.

Ejsmond, Maciej Jan, and Jacek Radwan. 2011. "MHC Diversity in Bottlenecked Populations: A Simulation Model." *Conservation Genetics* 12 (1): 129–37. https://doi.org/10.1007/s10592-009-9998-6.

Engelhard, Victor H. 1994. "Structure of Peptides Associated with Class I and Class II MHC Molecules." *Annual Review of Immunology* 12: 181–207. https://doi.org/10.1146/annurev.iy.12.040194.001145.

Excoffier, Laurent, Peter E Smouse, and Joseph M Quattro. 1992. "Analysis of Molecular Variance Inferred From Metric Distances Among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data." https://academic.oup.com/genetics/article/131/2/479/6007328.

Figueroa, Felipe, Eberhardt Gúnther, and Jan Klein. 1988. "MHC Polymorphism Pre-Dating Speciation." *Nature 1988 335:6187* 335 (6187): 265–67. https://doi.org/10.1038/335265a0.

Figueroa, Felipe, Akie Sato, William S Talbot, Stephen L Johnson, Michael Gates, John H Postlethwait, and Jan Klein. 1997. "Mapping of Mhc Class I and Class II Regions to Different Linkage Groups in the Zebrafish, Danio Rerio."

Fugère, Vincent, Hernán Ortega, and Rü Diger Krahe. 2011. "Electrical Signalling of Dominance in a Wild Population of Electric Fish." *Biology Letters* 7 (2): 197–200. https://doi.org/10.1098/RSBL.2010.0804.

Gallant, Jason R., Lindsay L. Traeger, Jeremy D. Volkening, Howell Moffett, Po Hao Chen, Carl D. Novina, George N. Phillips, et al. 2014. "Genomic Basis for the Convergent Evolution of Electric Organs." *Science* 344 (6191): 1522–25. https://doi.org/10.1126/science.1254432.

Gandon, Sylvain, and Scott L Nuismer. 2009. “Interactions between Genetic Drift, Gene Flow, and Selection Mosaics Drive Parasite Local Adaptation” 173 (2). https://doi.org/10.1086/593706.

Gerdol, Marco, Daniela Lucente, Francesco Buonocore, Elia Poerio, Giuseppe Scapigliati, Simonetta Mattiucci, Alberto Pallavicini, and Roberta Cimmaruta. 2019. “Molecular and Structural Characterization of MHC Class II β Genes Reveals High Diversity in the Cold-Adapted Icefish Chionodraco Hamatus.” *Scientific Reports* 9 (1): 1–14. https://doi.org/10.1038/s41598-019-42003-5.

Glaberman, Scott, Maria A Moreno, and Adalgisa Caccone. 2009. “Characterization and Evolution of MHC Class II B Genes in Galápagos Marine Iguanas (Amblyrhynchus Cristatus).” *Developmental & Comparative Immunology* 33 (8): 939–47. https://doi.org/10.1016/j.dci.2009.03.003.

Grimholt, Unni, Andrew Getahun, Trudi Hermsen, and René J.M. Stet. 2000. “The Major Histocompatibility Class II Alpha Chain in Salmonid Fishes.” *Developmental and Comparative Immunology* 24 (8): 751–63. https://doi.org/10.1016/S0145-305X(00)00034-3.

Grünwald, Niklaus J, Stephen B Goodwin, Michael G Milgroom, and William E Fry. 2003. “Analysis of Genotypi Diversity Data for Populations of Microorganisms.” *Phytopathology*. Vol. 93.

Hagedorn, Mary. 2008. “Ecology and Behavior of a Pulse-Type Electric Fish , Hypopomus Occidentalis (Gymnotiformes , Hypopomidae), in a Fresh-Water Stream in Panama” 1988 (2): 324–35.

Hashimoto, K., T. Nakanishi, and Y. Kurosawa. 1990. “Isolation of Carp Genes Encoding Major Histocompatibility Complex Antigens.” *Proceedings of the National Academy of Sciences of the United States of America* 87 (17): 6863–67. https://doi.org/10.1073/PNAS.87.17.6863.

Hedrick, Philip W. 2002. “Pathogen Resistance and Genetic Variation at MHC Loci.” *Evolution* 56 (10): 1902–8. https://doi.org/10.1111/j.0014-3820.2002.tb00116.x.

Holm, Sture. 1979. “A Simple Sequentially Rejective Multiple Test Procedure.” *Scandinavian Journal of Statistics* 6 (2): 65–70.

Horton, Roger, Laurens Wilming, Vikki Rand, Ruth C. Lovering, Elspeth A. Bruford, Varsha K. Khodiyar, Michael J. Lush, et al. 2004. “Gene Map of the Extended Human MHC.” *Nature Reviews Genetics 2004 5:12* 5 (12): 889–99. https://doi.org/10.1038/nrg1489.

Hughes, Austin L, and Masatoshi Nei. 1989. "Nucleotide Substitution at Major Histocompatibility Complex Class II Loci: Evidence for Overdominant Selection (DNA Evolution/Polymorphism/Nonsynonymous Substitution/Gene Conversion)." *Proc. Nati. Acad. Sci. USA* 86: 958–62.

Jeon, Hyung Bae, Hari Won, and Ho Young Suk. 2019. "Polymorphism of MHC Class IIB in an Acheilognathid Species, Rhodeus Sinensis Shaped by Historical Selection and Recombination." *BMC Genetics* 20 (1): 74. https://doi.org/10.1186/S12863-019-0775-3.

Jombart, Thibaut. 2015. "An Introduction to Adegenet 2.0.0." *R Package*. http://adegenet.r-forge.r-project.org/files/tutorial-basics.pdf.

Kalbe, M., K. M. Wegner, and T. B.H. Reusch. 2002. "Dispersion Patterns of Parasites in 0+ Year Three-Spined Sticklebacks: A Cross Population Comparison." *Journal of Fish Biology* 60 (6): 1529–42. https://doi.org/10.1006/jfbi.2002.2013.

Kamvar, Zhian N., Javier F. Tabima, and Niklaus J. Grünwald. 2014. "Poppr: An R Package for Genetic Analysis of Populations with Clonal, Partially Clonal, and/or Sexual Reproduction." *PeerJ* 2014 (1): 1–14. https://doi.org/10.7717/PEERJ.281/TABLE-6.

Kearn, Graham C. 1986. "The Eggs of Monogeneans." *Advances in Parasitology* 25 (C): 175–273. https://doi.org/10.1016/S0065-308X(08)60344-9.

Klein, J., and C. O'Huigin. 1994. "MHC Polymorphism and Parasites." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 346 (1317). https://doi.org/10.1007/978-94-009-0077-6_9.

Klein, Jan, Akie Sato, Sandra Nagl, and Colm O'huigín. 1998. "Molecular Trans-Species Polymorphism." *Annu. Rev. Ecol. Syst* 29: 1–21. www.annualreviews.org.

Klein, Jan, Akie Sato, and Nikolas Nikolaidis. 2007. "MHC, TSP, and the Origin of Species: From Immunogenetics to Evolutionary Genetics." *Annual Review of Genetics*. https://doi.org/10.1146/annurev.genet.41.110306.130137.

Kosakovsky Pond, Sergei L, Simon D W Frost, and Spencer v Muse. 2005. "HyPhy: Hypothesis Testing Using Phylogenies." *BIOINFORMATICS APPLICATIONS NOTE* 21 (5): 676–79. https://doi.org/10.1093/bioinformatics/bti079.

Kosakovsky Pond, Sergei L., and Simon D.W. Frost. 2005a. "Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection." *Molecular Biology and Evolution* 22 (5): 1208–22. https://doi.org/10.1093/MOLBEV/MSI105.

———. 2005b. “Datamonkey: Rapid Detection of Selective Pressure on Individual Sites of Codon Alignments.” *Bioinformatics* 21 (10): 2531–33. https://doi.org/10.1093/BIOINFORMATICS/BTI320.

Lekstrom-Himes, Julie A., Patricia Hohman, Terri Warren, Anna Wald, Jun Mo Nam, Toni Simonis, Lawrence Corey, and Stephen E. Straus. 1999. “Association of Major Histocompatibility Complex Determinants with the Development of Symptomatic and Asymptomatic Genital Herpes Simplex Virus Type 2 Infections.” *The Journal of Infectious Diseases* 179 (5): 1077–85. https://doi.org/10.1086/314729.

Li, Hongjun, Lianxin Jiang, Jiabo Han, Hao Su, Qing Yang, and Chongbo He. 2011. “Major Histocompatibility Complex Class IIA and IIB Genes of the Spotted Halibut Verasper Variegatus: Genomic Structure, Molecular Polymorphism, and Expression Analysis.” *Fish Physiology and Biochemistry* 37 (4): 767–80. https://doi.org/10.1007/S10695-011-9476-1.

Lin, Tian Long, Theodore G. Clark, and Harry Dickerson. 1996. “Passive Immunization of Channel Catfish (Ictalurus Punctatus) against the Ciliated Protozoan Parasite Ichthyophthirius Multifiliis by Use of Murine Monoclonal Antibodies.” *Infection and Immunity* 64 (10): 4085–90. https://doi.org/10.1128/IAI.64.10.4085-4090.1996.

Lohm, Jakob, Mats Grahn, Åsa Langefors, Øivind Andersen, Arne Storset, and Torbjörn von Schantz. 2002. “Experimental Evidence for Major Histocompatibility Complexallelespecific Resistance to a Bacterial Infection.” *Proceedings of the Royal Society of London. Series B: Biological Sciences* 49 (191): 2028–33. https://doi.org/10.1098/RSPB.2002.2114.

Ludwig, John A., and James F. Reynolds. 1988. “Statistical Ecology : A Primer on Methods and Computing,” 337.

Málaga-Trillo, Edward, Zofia Zaleska-Rutczynska, Brendan McAndrew, Vladimir Vincek, Felipe Figueroa, Holger Sültmann, and Jan Klein. 1998. “Linkage Relationships and Haplotype Polymorphism Among Cichlid Mhc Class II B Loci.” *Genetics* 149 (3): 1527–37. https://doi.org/10.1093/GENETICS/149.3.1527.

Malmstrøm, Martin, Michael Matschiner, Ole K. Tørresen, Bastiaan Star, Lars G. Snipen, Thomas F. Hansen, Helle T. Baalsrud, et al. 2016. “Evolution of the Immune System Influences Speciation Rates in Teleost Fishes.” *Nature Genetics* 48 (10): 1204–10. https://doi.org/10.1038/ng.3645.

Miller, Hilary C., Fred Allendorf, and Charles H. Daugherty. 2010. “Genetic Diversity and Differentiation at MHC Genes in Island Populations of Tuatara (Sphenodon Spp.).” *Molecular Ecology* 19 (18): 3894–3908. https://doi.org/10.1111/j.1365-294X.2010.04771.x.

Milner, Caroline M, and R Duncan Campbell. 2001. "GENETIC ORGANIZATION OF THE HUMAN MHC CLASS III REGION." *Frontiers in Bioscience* 6: 914–26. http://www.sanger.ac.uk/HGP/.

Murray, Brent W, Seikou Shintani, Holger Sültmann, Jan Klein, B W Murray, S Shintani, H Sültmann, and J Klein. 2000. "Major Histocompatibility Complex Class II A Genes in Cichlid Fishes: Identification, Expression, Linkage Relationships, and Haplotype Variation." *Immunogenetics 2000 51:7* 51 (7): 576–86. https://doi.org/10.1007/S002510000177.

Murrell, Ben, Joel O. Wertheim, Sasha Moola, Thomas Weighill, Konrad Scheffler, and Sergei L. Kosakovsky Pond. 2012. "Detecting Individual Sites Subject to Episodic Diversifying Selection." *PLOS Genetics* 8 (7): e1002764. https://doi.org/10.1371/JOURNAL.PGEN.1002764.

Nevo, E., and A. Beiles. 1992. "Selection for Class IIMhc Heterozygosity by Parasites in Subterranean Mole Rats." *Experientia 1992 48:5* 48 (5): 512–15. https://doi.org/10.1007/BF01928177.

Oksanen, J., Gavin L Simpson, and F Guillaume Blanchet. 2022. "Vegan: Community Ecology Package. R Package Version 2.6-2."

Onott, Hideki, Dagmar Klein, Vladimir Vincek, Felipe Figueroat, Colm O'huigint, Herbert Tichyt, and Jan Kleint. 1992. "Major Histocompatibility Complex Class II Genes of Zebrafish." *Proc. Nail. Acad. Sci. USA*. Vol. 89.

Ottová, Eva, Andrea Šimková, Jean François Martin, Joëlle Goüy De Bellocq, Milan Gelnar, Jean François Allienne, and Serge Morand. 2005. "Evolution and Trans-Species Polymorphism of MHC Class IIβ Genes in Cyprinid Fish." *Fish and Shellfish Immunology* 18 (3): 199–222. https://doi.org/10.1016/j.fsi.2004.07.004.

Paradis, Emmanuel, and Jeffrey Barrett. 2010. "Pegas: An R Package for Population Genetics with an Integrated-Modular Approach." *BIOINFORMATICS APPLICATIONS NOTE* 26 (3): 419–20. https://doi.org/10.1093/bioinformatics/btp696.

Persson, Anna-Carin, René J M Stet, Lars Pilström, A.-C Persson, L Pilström, and R J M Stet. 1999. "Characterization of MHC Class I and B2-Microglobulin Sequences in Atlantic Cod Reveals an Unusually High Number of Expressed Class I Genes." *Immunogenetics 1999 50:1* 50 (1): 49–59. https://doi.org/10.1007/S002510050685.

Picq, Sophie, Fernando Alda, Eldredge Bermingham, and Rüdiger Krahe. 2016. "Drift-Driven Evolution of Electric Signals in a Neotropical Knifefish." *Evolution; International Journal of Organic Evolution* 70 (9): 2134–44. https://doi.org/10.1111/evo.13010.

Picq, Sophie, Fernando Alda, Rüdiger Krahe, and Eldredge Bermingham. 2014. "Miocene and Pliocene Colonization of the Central American Isthmus by the Weakly Electric Fish Brachyhypopomus Occidentalis (Hypopomidae, Gymnotiformes)." *Journal of Biogeography* 41 (8): 1520–32. https://doi.org/10.1111/jbi.12309.

Pink, J. R.L., W. Droege, K. Hála, V. C. Miggiano, and A. Ziegler. 1977. "A Three-Locus Model for the Chicken Major Histocompatibility Complex." *Immunogenetics 1977 5:1* 5 (1): 203–16. https://doi.org/10.1007/BF01570477.

Pinto, Hudson Alves, and Alan Lane de Melo. 2012. "Metacercariae of Centrocestus Formosanus (Trematoda: Heterophyidae) in Australoherosfacetus (Pisces: Cichlidae) in Brazil." *Revista Brasileira de Parasitologia Veterinaria* 21 (3): 334–37. https://doi.org/10.1590/s1984-29612012000300032.

Podivinský, F. 1966. "Evolution and Ecology of MHC Molecules: From Genomics to Sexual Selection." *Stereotactic and Functional Neurosurgery* 28 (6): 443–53. https://doi.org/10.1159/000103795.

Pond, Sergei L.Kosakovsky, David Posada, Michael B. Gravenor, Christopher H. Woelk, and Simon D.W. Frost. 2006. "Automated Phylogenetic Detection of Recombination Using a Genetic Algorithm." *Molecular Biology and Evolution* 23 (10): 1891–1901. https://doi.org/10.1093/MOLBEV/MSL051.

Reed, Kent M., and Robert E. Settlage. 2021. "Major Histocompatibility Complex Genes and Locus Organization in the Komodo Dragon (Varanus Komodoensis)." *Immunogenetics 2021* 1: 1–13. https://doi.org/10.1007/S00251-021-01217-6.

Reed, Peggy, Ruth Francis-Floyd, Ruth Ellen Klinger, and Denise Petty. 2009. "Monogenean Parasites of Fish." *University of Florida IFAS Extension*, 1–5. http://edis.ifas.ufl.edu.

Rehulková, Eva, Michal Benovics, and Andrea Šimková. 2020. "Uncovering the Diversity of Monogeneans (Platyhelminthes) on Endemic Cypriniform Fishes of the Balkan Peninsula: New Species of Dactylogyrus and Comments on Their Phylogeny and Host-Parasite Associations in a Biogeographic Context." *Parasite* 27: 2020. https://doi.org/10.1051/parasite/2020059.

Rice, Peter, Lan Longden, and Alan Bleasby. 2000. "EMBOSS: The European Molecular Biology Open Software Suite." *Trends in Genetics* 16 (6): 276–77. https://doi.org/10.1016/S0168-9525(00)02024-2.

Rózsa, Lajos, Jenö Reiczigel, and Gábor Majoros. 2000. "Quantifying Parasites in Samples of Hosts." *Journal of Parasitology* 86 (2): 228–32. https://doi.org/10.1645/0022-3395(2000)086[0228:QPISOH]2.0.CO;2.

Sambrook, Jennifer G, Felipe Figueroa, and Stephan Beck. 2005. “A Genome-Wide Survey of Major Histocompatibility Complex (MHC) Genes and Their Paralogues in Zebrafish.” *BMC Genomics 2005 6:1* 6 (1): 1–10. https://doi.org/10.1186/1471-2164-6-152.

Sato, Akie, Felipe Figueroa, Brent W Murray, Edward Málaga-Trillo, Zofia Zaleska-Rutczynska, Holger Sültmann, Satoru Toyosawa, et al. 2000. “Nonlinkage of Major Histocompatibility Complex Class I and Class II Loci in Bony Fish.” *Immunogenetics*. Vol. 51. Springer-Verlag.

Savage, Anna E., and Kelly R. Zamudio. 2011. “MHC Genotypes Associate with Resistance to a Frog-Killing Fungus.” *Proceedings of the National Academy of Sciences of the United States of America* 108 (40): 16705–10. https://doi.org/10.1073/pnas.1106893108.

Scholz, T., M. L. Aguirre-Macedo, and G. Salgado-Maldonado. 2001. “Trematodes of the Family Heterophyidae (Digenea) in Mexico: A Review of Species and New Host and Geographical Records.” *Journal of Natural History* 35 (12): 1733–72. https://doi.org/10.1080/00222930152667087.

Shaw, Allison K., Julie Sherman, F. Keith Barker, and Marlene Zuk. 2018. “Metrics Matter: The Effect of Parasite Richness, Intensity and Prevalence on the Evolution of Host Migration.” *Proceedings of the Royal Society B: Biological Sciences* 285 (1891). https://doi.org/10.1098/rspb.2018.2147.

Start, Denon, and Benjamin Gilbert. 2016. “Host-Parasitoid Evolution in a Metacommunity.” *Proceedings of the Royal Society B: Biological Sciences* 283 (1831). https://doi.org/10.1098/rspb.2016.0477.

Stet, Rene J. M, Christopher Secombes, Peter Parham, Ronald P Hedrick, R. Barry Nehring, Benny P Shum, Lisbeth Guethlein, and Laura R Flodin. 2001. “Evolution Differ from the Primate Paradigm Modes of Salmonid MHC Class I and II.” *The Journal of Immunology* 166: 3297–3308. https://doi.org/10.4049/jimmunol.166.5.3297.

Stet, René J., Beja de Vries, Karin Mudde, Trudi Hermsen, Joost van Heerwaarden, Benny P. Shum, and Unni Grimholt. 2002. “Unique Haplotypes of Co-Segregating Major Histocompatibility Class II A and Class II B Alleles in Atlantic Salmon (Salmo Salar) Give Rise to Diverse Class II Genotypes.” *Immunogenetics* 54 (5): 320–31. https://doi.org/10.1007/s00251-002-0477-1.

Stoddard, Philip K. 1999. “Predation Enhances Complexity in the Evolution of Electric Fish Signals.” *Nature* 400 (6741): 254–56. https://doi.org/10.1038/22301.

Summers, Kyle, Sea McKeon, Jon Sellars, Mark Keusenkothen, James Morris, David Gloeckner, Corey Pressley, Blake Price, and Holly Snow. 2003. “Parasitic Exploitation as an Engine of Diversity.” *Biological Reviews of the Cambridge Philosophical Society*. https://doi.org/10.1017/S146479310300616X.

Takahata, Naoyuki, and Masatoshi Nei. 1990. “Allelic Genealogy Under Overdominant and Frequency-Dependent Selection and Polymorphism of Major Histocompatibility Complex Loci.” https://academic.oup.com/genetics/article/124/4/967/5999972.

Thursz, Mark R, Dominic Kwiatkowski, Catherine E.M. Allsopp, Brian M. Greenwood, Howard C. Thomas, and Adrian V.S. Hill. 1995. “Association between an MHC Class II Allele and Clearane of Hepatitis B Virus in the Gambia.” Vol. 332.

Timofeeva, Tatjana A, Pavel I Gerasev, and David I Gibson. 1997. “A Catalogue of the Nominal Species of the Monogenean Family Dactylogyridae Bychowsky, 1933 (Excluding Dactylogyrus Diesing, 1850).” *Systematic Parasitology* 38 (2): 153–58. https://doi.org/10.1023/A:1005889911627.

Traeger, Lindsay L, Jeremy D Volkening, Howell Moffett, Jason R Gallant, Po Hao Chen, Carl D Novina, George N Phillips, et al. 2015. “Unique Patterns of Transcript and MiRNA Expression in the South American Strong Voltage Electric Eel (Electrophorus Electricus).” *BMC Genomics* 16 (1): 243. https://doi.org/10.1186/s12864-015-1288-8.

Wang, Biao, Robert Ekblom, Tanja M. Strand, Silvia Portela-Bens, and Jacob Höglund. 2012. “Sequencing of the Core MHC Region of Black Grouse (Tetrao Tetrix) and Comparative Genomics of the Galliform MHC.” *BMC Genomics* 13 (1). https://doi.org/10.1186/1471-2164-13-553.

Wegner, K. M., M. Kalbe, G. Rauch, J. Kurtz, H. Schaschl, and T. B.H. Reusch. 2006. “Genetic Variation in MHC Class II Expression and Interactions with MHC Sequence Polymorphism in Three-Spined Sticklebacks.” *Molecular Ecology* 15 (4): 1153–64. https://doi.org/10.1111/j.1365-294X.2006.02855.x.

Wegner, K Mathias. 2004. “Major Histocompatibility Genes , Polymorphism and Balancing Selection - the Case of Parasites and Sticklebacks.” *Christian-Albrechts-Univ. Kiel*, 118. http://en.scientificcommons.org/20560700.

Wegner, K Mathias, T. B.H. Reusch, and M Kalbe. 2003. “Multiple Parasites Are Driving Major Histocompatibility Complex Polymorphism in the Wild.” *Journal of Evolutionary Biology* 16 (2): 224–32. https://doi.org/10.1046/j.1420-9101.2003.00519.x.

Yanong, Roy P E. 2002. “Nematode (Roundworm) Infections in Fish 1.” *University of Florida*: 1–9.

Youngerman, Eric D., Brooke E. Flammang, and George v. Lauder. 2014. “Locomotion of Free-Swimming Ghost Knifefish: Anal Fin Kinematics during Four Behaviors.” *Zoology* 117 (5): 337–48. https://doi.org/10.1016/j.zool.2014.04.004.

Yu, Suhong, Jingqun Ao, and Xinhua Chen. 2010a. “Molecular Characterization and Expression Analysis of MHC Class II α and β Genes in Large Yellow Croaker (Pseudosciaena Crocea).” *Molecular Biology Reports* 37 (3): 1295–1307. https://doi.org/10.1007/s11033-009-9504-8.

Yu, Suhong, Jingqun Ao, and Xinhua Chen. 2010b. “Molecular Characterization and Expression Analysis of MHC Class II α and β Genes in Large Yellow Croaker (Pseudosciaena Crocea).” *Molecular Biology Reports* 37 (3): 1295–1307. https://doi.org/10.1007/s11033-009-9504-8.

Zargar, U. R., M. Z. Chishti, A. R. Yousuf, and A. Fayaz. 2012. “Infection Level of Monogenean Gill Parasite, Diplozoon Kashmirensis (Monogenea, Polyopisthocotylea) in the Crucian Carp, Carassius Carassius from Lake Ecosystems of an Altered Water Quality: What Factors Do Have an Impact on the Diplozoon Infection?” *Veterinary Parasitology* 189 (2): 218–26. https://doi.org/10.1016/j.vetpar.2012.04.029.

Zhang, Lin, Dongmei Lin, Sen Yu, Junping Bai, Wanchun Jiang, Wenzheng Su, Yanyan Huang, Shaohua Yang, and Jiaqiang Wu. 2019. “Polymorphism of Duck MHC Class Molecules.” *Immunogenetics* 71 (1): 49–59. https://doi.org/10.1007/S00251-018-1076-0/TABLES/3.