

What evolutionary processes maintain MHC II β diversity within and among populations of stickleback?

Foen Peng¹  | Kimberly M. Ballare²  | S. Hollis Woodard³  | Stijn den Haan⁴ | Daniel I. Bolnick¹

¹Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, USA

²Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, CA, USA

³Department of Entomology, University of California, Riverside, CA, USA

⁴Central European University, Budapest, Hungary

Correspondence

Daniel I. Bolnick, Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, USA.
Email: daniel.bolnick@uconn.edu

Funding information

Howard Hughes Medical Institute; National Institute of Allergy and Infectious Diseases, Grant/Award Number: 1R01AI123659-01A1; David and Lucille Packard Fellowship for Science and Engineering

Abstract

Major Histocompatibility Complex (MHC) genes code for proteins that recognize foreign protein antigens to initiate T-cell-mediated adaptive immune responses. They are often the most polymorphic genes in vertebrate genomes. How evolution maintains this diversity remains of debate. Three main hypotheses seek to explain the maintenance of MHC diversity by invoking pathogen-mediated selection: heterozygote advantage, frequency-dependent selection, and fluctuating selection across landscapes or through time. Here, we use a large-scale field parasite survey in a stickleback metapopulation to test predictions derived from each of these hypotheses. We identify over 1000 MHC II β variants (alleles spanning paralogous genes) and find that many of them covary positively or negatively with parasite load, suggesting that these genes contribute to resistance or susceptibility. However, despite our large sample-size, we find no evidence for the widely cited stabilizing selection on MHC heterozygosity, in which individuals with an intermediate number of MHC variants have the lowest parasite burden. Nor do we observe a rare-variant advantage, or widespread fluctuating selection across populations. In contrast, we find that MHC diversity is best predicted by neutral genome-wide heterozygosity and between-population genomic divergence, suggesting neutral processes are important in shaping the pattern of metapopulation MHC diversity. Thus, although MHC II β is highly diverse and relevant to the type and intensity of macroparasite infection in these populations of stickleback, the main models of MHC evolution still provide little explanatory power in this system.

KEYWORDS

evolution, MHC, parasite, stickleback

1 | INTRODUCTION

Genetic variation is the raw material for natural selection to act upon. Hence, there is a long history of evolutionary studies on how genetic variation is maintained in natural populations (Dobzhansky, 1982), either invoking selection or neutral processes. One of the most dramatic cases of genetic variation concerns polymorphism

at Major Histocompatibility Complex (MHC) genes, which are commonly the most variable genes in vertebrate genomes (Sommer, 2005). Surprisingly, despite decades of research on MHC evolution, the evolutionary processes sustaining MHC polymorphism remain unclear, with inconsistent and often ambiguous support for competing hypotheses (Radwan et al., 2020). Here, we evaluate predictions of multiple competing MHC evolution models (adaptive and neutral),

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 The Authors. Molecular Ecology published by John Wiley & Sons Ltd.

using an exceptionally large data set of genetic variation and macroparasite infection in a metapopulation of threespine stickleback (*Gasterosteus aculeatus*) inhabiting lakes on Vancouver Island.

MHC genes play a key role in the adaptive immune system of vertebrates. The main function of these genes is to code for cell surface proteins, which are used to detect foreign molecules. These molecules are then presented to T cells to initiate appropriate immune responses. The MHC gene family consists of two classes of genes, MHC I and MHC II, each of which can be represented by multiple paralogue copies. MHC class I genes are expressed in all nucleated cells and typically bind to peptides derived from intracellular molecules, such as viral protein (Jensen, 2007). MHC class II genes (our focus here) are expressed only in antigen-presenting cells, such as macrophages, dendritic cells and B cells. These cells phagocytize extracellular material, digest the proteins, and if the MHC II proteins bind to the resulting fragments, these are presented on the cell surface to T cells, possibly to initiate an adaptive immune response (Jensen, 2007). The second exon of the MHC II gene is often used to characterize the variability of MHC II genes, because it contains the antigen-binding domain (Sommer, 2005).

The maintenance of MHC diversity is often attributed to pathogen-mediated selection, although intraspecific processes, such as mate choice (Penn & Potts, 1999), are also potential sources of selection. As reviewed in Spurgin and Richardson (2010), there are three nonmutually exclusive pathogen-mediated selection mechanisms to explain MHC diversity. (i) *Heterozygote advantage*. Vertebrate genomes contain multiple MHC loci. Individuals being heterozygous in more loci have higher overall diversity of MHC alleles. These individuals might be able to recognize a more diverse set of parasite species, and thus have fewer parasites (Doherty & Zinkernagel, 1975; Hughes & Nei, 1988). This advantage may lead to directional selection for ever-increasing diversity, but some widely cited studies have presented evidence for intermediate optima (Wegner, Kalbe, et al., 2003). (ii) *Negative frequency-dependent selection*. In a given population, common MHC alleles that protect against parasites impose selection on those parasites to change their antigens and evade recognition. Once parasites evolve strategies to evade these common alleles, individuals carrying the common alleles may be at a selective disadvantage. By contrast, a rare MHC allele imposes little selection on the parasites it recognizes (because it is rare), and so may tend to be more protective (Slade & McCallum, 1992; Takahata & Nei, 1990). For this negative frequency-dependence to work in the long run, it must on average hold that rare alleles are more protective (and hence more fit) than common alleles. (iii) *Fluctuating selection*. Parasite community and parasite genotypes may vary in space and time. Given this, different host MHC alleles may be selectively favoured in different locations and at different times. Gene flow between locations subject to divergent selection can sustain MHC diversity, as can temporally fluctuating selection (Hedrick, 2002; Hill, 1991).

Although many studies have examined the roles of the above-mentioned mechanisms in maintaining MHC diversity, the results are often mixed or even contradictory. For example, heterozygote

advantage was supported in one experimental infection study with inbred mice (Penn et al., 2002), but the effect was not found in another study with outbred mice (Ilmonen et al., 2007). Theoretically, negative frequency-dependent selection is more likely to explain the extraordinary diversity in MHC genes than heterozygote advantage (Borghans et al., 2004). However, empirically demonstrating negative frequency-dependent selection is more challenging, because the effect depends on many variables, such as temporal variation in the strength of selection, mutation rate, the generation time of host and parasite, etc. Also, the patterns arising from negative frequency-dependent selection are often consistent with other mechanisms. For example, in house sparrows, population-specific MHC alleles were linked to host resistance or susceptibility to malaria (Bonneaud et al., 2006), but this pattern could be explained either by negative frequency-dependent selection or spatially varying selection. A previous study of stickleback from three lake-stream population pairs (Stutz & Bolnick, 2017) tried to distinguish between negative frequency-dependent selection or spatially varying selection. The study took advantage of migration between parapatric populations with different parasite communities, which in principle would allow them to partition benefits of rarity within populations, from costs of being a (rare) immigrant to a foreign habitat. Yet, no overall trend was found to support either mechanism. Such inconsistent or ambiguous results are typical in the MHC evolution literature. Therefore, a consensus regarding the relative importance of these mechanisms in maintaining MHC diversity has not been reached (Radwan et al., 2020).

In contrast to these adaptive hypotheses, whether and how MHC diversity is shaped by neutral processes has been less well studied. In small and/or isolated populations, MHC diversity is often heavily influenced by genetic drift, rather than balancing selection (Hedrick et al., 2001; Miller & Lambert, 2004; Seddon & Ellegren, 2004). In large metapopulations, it is unclear what role neutral processes play in shaping MHC diversity.

MHC studies in threespine stickleback have generated some unique insights (Eizaguirre et al., 2012; Matthews et al., 2010; Mccairns et al., 2011; Milinski et al., 2005; Wegner, Kalbe, et al., 2003). For example, Wegner, Kalbe, et al. (2003) conducted a study on eight stickleback populations to examine the association between MHC diversity and 15 parasite species. They found populations exposed to a wider range of parasites had higher MHC allelic diversity, though MHC diversity also positively, albeit weakly, correlated with neutral diversity. Moreover, at the individual level, parasite burden was minimized in fish with an intermediate, rather than maximal, number of MHC II β alleles. This result supports the theoretical model of stabilizing selection on MHC heterogeneity (Nowak et al., 1992). The logic is that low-MHC-diversity individuals do not have the capacity to recognize diverse parasites, but high-MHC-diversity individuals deplete their T cells via negative selection on self-reactive T cells. The resulting paucity of T cell receptor diversity also limits their ability to recognize diverse parasites. However, while this intermediate-advantage hypothesis was widely cited and has gained some support in recent years (Ishigaki et al., 2020; Migalska

et al., 2019), it has not been replicated in many systems and remains controversial (Borghans et al., 2003). Another example of interesting insights from stickleback is the involvement of MHC alleles in mate recognition. It was found that female stickleback could choose their mate based on odour to optimize the number of MHC alleles in their offspring, though the molecular mechanism is unknown (Milinski et al., 2005).

In this study, we set out to test predictions associated with three main models of MHC polymorphism invoking parasite-mediated selection, using an extensive field survey of parasites in a meta-population of threespine stickleback. We did not find evidence for stabilizing or positive selection for MHC heterozygosity, contrary to Wegner et al. (2003). Nor did we observe fluctuating selection across populations. Although we found significant associations between specific MHC alleles and parasite species both within and across populations, neutral processes best explained within- and between-population MHC diversity in our data set. Note that our genotyping method (like many for MHC) did not allow us to distinguish whether distinct MHC sequences are from the same locus or from different loci. So, for the sake of clarity, from here on we call each distinct MHC sequence as an MHC “variant,” instead of a MHC “allele.”

2 | MATERIALS AND METHODS

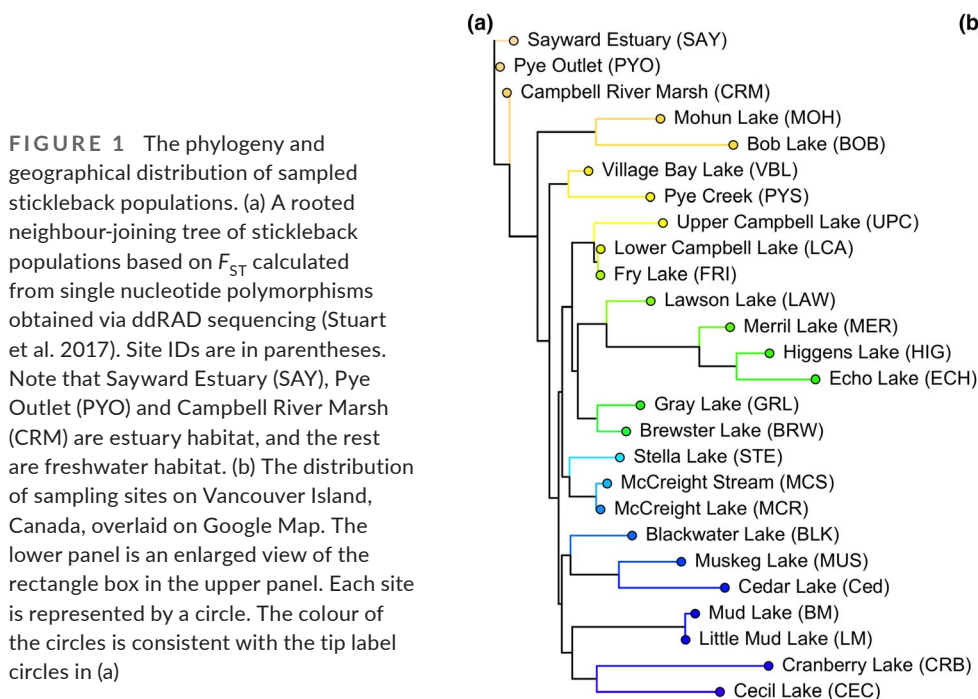
2.1 | Fish sampling and parasite identification

We used unbaited minnow traps to collect adult threespine stickleback from lake, river and estuary sites on Vancouver Island in late May and early June 2009. Collections were approved by

the University of Texas IACUC (07-032201) and a Scientific Fish Collection Permit from the Ministry of the Environment of British Columbia (NA07-32612). Fish were euthanized in MS-222 and then preserved in formalin, after the caudal fin of each individual fish was removed and preserved in pure ethanol for genotyping. Details are provided in Bolnick and Ballare (2020). These samples were used to examine within- and between-population variation in diet (Bolnick & Ballare, 2020), parasite community composition (Bolnick et al., 2020b), and parasite species richness (Bolnick et al., 2020a). The parasite infection data for this study are archived at the Dryad Digital Repository (Bolnick & Ballare, 2020). Here, we focus on a random subset of 26 sampling sites ($N = 1437$ stickleback, Figure 1) for which we also genotyped individuals for MHC II β . For each stickleback individual, we recorded sex and body length and then dissected each fish to count and identify macroparasites as described by Stutz and Bolnick (2017).

2.2 | Genotyping

The methods to sequence and genotype MHC II β variants were identical to Stutz and Bolnick (2017). Briefly, genomic DNA was extracted from fin clips using a Promega Wizard 96-well extraction kit. We used PCR (polymerase chain reaction) to amplify the second exon of MHC II β genes in each fish, with primers and PCR cycles as described in Stutz and Bolnick (2017). This exon contains the hyper-variable peptide-binding region that binds to possible parasite antigens (Sommer, 2005). Each specimen was barcoded with a unique combination of forward and reverse primer tags for multiplexing. We used Quant-iT PicoGreen kits (Invitrogen P11496) to quantify DNA concentrations of magnetic bead-purified (Agencourt AMPure XP



beads) PCR products, then pooled up to 400 samples in equimolar amounts to construct a library. We used Illumina Mi-Seq to sequence these multiplexed amplicon libraries. Then, we used a Stepwise Threshold Clustering (STC) program (Stutz & Bolnick, 2014), implemented in the AMPLISAS web software (<http://evobiolab.biol.amu.edu.pl/amplisat/index.php?amplisat>; Sebastian et al., 2016) to distinguish real sequence variants from sequencing error or PCR chimeras. The algorithm was originally validated by sequencing cloned amplicon products (Stutz & Bolnick, 2017). The software outputs a table of individual fish (rows) and unique MHC sequences (columns) with read depths.

To efficiently process data in AMPLISAS, we set the upper limit of read depth for each individual as 5000, which was sufficient to retrieve all the possible MHC variants (Stutz & Bolnick, 2014). Because low sequencing coverage could bias the number of MHC variants to be identified, individual fishes with coverage lower than 450 were excluded from this study (Figure S1). After excluding the 160 individuals with low coverage (leaving $N = 1277$ individuals in the subsequent analyses), there was no longer a significant linear relationship between sequencing coverage and the number of MHC variants ($t = 1.46$, $p = .14$). The number of unique MHC variants was inferred based on the translated protein sequences, thus merging distinct exonic sequences that produce identical amino acid sequences. Previous work (Stutz & Bolnick, 2017) showed that few MHC variants had statistically significant linkage, so we kept all the MHC variants for data analysis.

2.3 | Data analysis

2.3.1 | Heterozygote advantage and the selection for variant number

We tested whether MHC diversity is associated with parasite burden at both the individual and the population level. At the individual level, we used parasite richness (i.e., the number of parasite species) to measure parasite burden. Given that MHC genes locate in many genomic loci and often act in a dominant manner in parasite resistance, we used variant number (i.e., the total number of unique MHC variants found in an individual) in this study to evaluate the level of MHC heterozygosity. We used a mixed effect generalized linear model with Poisson distribution to evaluate the impact of MHC variants on parasite richness. The fixed effects were variant number and log body length, which is known to affect parasite richness in stickleback (Bolnick et al., 2020a) and in other fish species (Calhoun et al., 2018). We did not include sex as a main effect because we found no main effect of sex on parasite richness in a previous analysis of this data set (Bolnick et al., 2020a). Sample site was treated as a random effect. In addition, we also evaluated the optimizing hypothesis (i.e., MHC variant number is optimized at an intermediate level, driven by the negative selection pressure to decrease T cell depletion by self-recognition) by adding a quadratic MHC variant number term into the mixed

effect model. We considered the interaction between site and the linear and quadratic effects of MHC variant number, using a random slope effect across sites. Statistical support for the model terms was evaluated based on Akaike's information criterion (AIC).

At the population level, we calculated the mean value of parasite richness and the average number of MHC variants (per fish) in each site. We used a linear regression model to test if the mean value of parasite richness is associated with the average MHC variant number for each site.

2.3.2 | Frequency-dependent selection: Is there a rare variant advantage?

If MHC variants were under negative frequency-dependent selection, then on average rare variants must confer an advantage (in the form of lower parasite infection) compared to common variants. In contrast, MHC variants with high frequency would be less effective in parasite detection, because parasites would be evolving strategies to avoid detection by those variants. We therefore expect a positive relationship between the frequency of a variant and its overall effect on parasite infection, where a negative effect denotes protection and a positive effect suggests susceptibility or survivor bias. Because MHC II genes scatter throughout multiple loci in fish genomes (Kaufman, 2018), it is difficult to use allele frequency to estimate the abundance of MHC variants in the population. We used variant prevalence to describe the prevalence of a variant in a population, which was calculated as the percentage of individuals carrying a focal variant in a population. Note that variant prevalence had different properties from allele frequency; for example, variant prevalence values of all the variants in a population do not sum to 1. Similarly, we also calculated parasite prevalence in a population as the percentage of individuals infected by a focal parasite. Variants and parasites that are too rare (<0.05 prevalence) or too common (>0.95 prevalence) do not provide sufficient variance to estimate effects. For every moderately prevalent MHC-parasite combination (both prevalence variable ranging from 0.05 to 0.95) in every sampling site, we used a generalized linear model with negative binomial distribution to test if the presence/absence of the focal MHC variant influenced the infection intensity of the focal parasite in that stickleback population. We corrected p values for multiple-comparison with the BH method (Benjamini & Hochberg, 1995). The Z value of the regression models indicated the effect size of the particular MHC variant on the infection rate by a particular parasite, which was used as the dependent variable in the regression against variant prevalence. We proposed that variant prevalence predicted the focal variant's effect on a given parasite (an among-individual effect). The variant's effect on infection was measured by Z values, to determine (for each population) whether a given variant was protective or vulnerable (positive or negative effect size). This Z value approach allowed us to evaluate a variant's effect by comparing individuals with, vs. without, the variant within a given population. This

among-individual (within-lake) effect size element controlled for many of the other ecological factors that influenced both parasite prevalence and host immune state, including lake size, prey community, etc. We excluded the models disproportionately influenced by extreme values; that is, the absolute Z value of a model would change over 0.5 if excluding the largest data point from the model. After iterating this procedure for all qualifying MHC–parasite combinations, we used another mixed-effect linear regression model to examine if the estimated effect sizes (Z) of MHC variants were influenced by local variant prevalence. In this model, we treated both sampling site and focal parasite as random effects. Because many parasites and MHC variants were included in more than one model in the previous step, which inflated the number of data points entering into the regression against variant prevalence, we also ran the second-step regression by only including models with $p < .05$ (before multiple comparison correction).

2.3.3 | Frequency-dependent selection: Are variant effects inconsistent across lakes?

If stickleback and their parasites engage in a Red Queen race-style co-evolution, the efficacy of any single variant will shift through time. For hosts and parasites with limited dispersal, physically disconnected sites are unlikely to be in the same phase of the arms race. As a result, a given MHC variant may have different effects on a given parasite from one site to the next: effective against defence in some places/times, ineffective or susceptible at others. Alternatively, if the same variant has similar effects on the same parasite across different lakes (without gene flow), such fluctuating frequency-dependent selection is unlikely. To test these alternatives, we first identified the moderately prevalent MHC–parasite combinations that were present in more than one site. For each qualified combination, we used a generalized linear model with negative binomial distribution to examine if the infection intensity of the parasite is influenced by the MHC variant, sampling site, and the interaction between site and MHC variant (this differs from the GLMs described in 2.3.2, which were done separately for each sample site). We corrected p values for multiple-comparison with the BH method (Benjamini & Hochberg, 1995). We used the *anova* function in R to perform an analysis of deviance for each regression model. It reported the reductions in the residual deviance as each term of the formula was added in turn. We evaluated whether, across many MHC–parasite combinations, more variation was explained by the focal variant's main effect (implying consistent protection across populations), or variant \times population interactions (inconsistent protection). Parasites transmitted by birds could spread over a larger spatial scale, so they are less likely to be engaged in an evolutionary arms race. We used a Chi-squared test to test whether the parasite taxa, which have birds as final hosts, were more likely to be found in the models with significant main effect than in the models with a significant variant \times population interaction effect.

2.3.4 | Fluctuating selection

The parasite communities of stickleback differ significantly among the different sites on Vancouver Island (Bolnick et al., 2020a; Stutz & Bolnick, 2017; Stutz et al., 2015). This heterogeneity could select for different MHC genotypes at different sites. If it is true, we expect that the distributions of MHC genotypes and parasite species would be correlated across the stickleback metapopulation. The populations that have similar parasite communities should have similar MHC genotype compositions, and vice versa. To test this hypothesis, we first constructed a Euclidian distance matrix across all the sites for MHC variants and for parasite species, respectively. Then we used a Mantel test to examine if the two distance matrices were correlated. A previous analysis (Bolnick et al., 2020b) found no significant effect of as-the-crow-flies (i.e., Euclidean distance) or as-the-fish-swims (i.e., the shortest in-water path) geographical distance on parasite community structure, so we omit that as a covariate in this distance analysis.

2.3.5 | Neutral evolution

We previously generated reduced representation genomic sequence data for the populations in this study, using double digest restriction-site associated DNA sequencing (ddRADseq) following the protocol of Peterson et al. (2012) modified as described in Stuart et al. (2017). From these data we estimated genome-wide mean heterozygosity for each individual fish, and from this population-level mean heterozygosity. We also calculated pairwise Weir–Cockerham adjusted F_{ST} between each pair of populations.

If MHC polymorphism is strongly affected by bottlenecks or other neutral population genetic processes, then we may expect that MHC diversity would be positively correlated with genomic mean heterozygosity. We tested whether the average number of MHC variants (per fish) in each population varied as a function of mean heterozygosity. We might then also expect that between-population MHC divergence is primarily a reflection of shared ancestry due to colonization processes or ongoing gene flow. To test this, we used a Mantel test to evaluate correlations between among-population MHC distance matrix and the F_{ST} matrix.

3 | RESULTS

3.1 | Sequencing results and parasite information

We kept 1277 stickleback from 26 sites (Figure 1) for subsequent analyses, after excluding 160 individuals whose genotype calls may have under-represented their actual diversity due to low read depth (<450 reads each; Figure S1). These 26 sites were from seven different watersheds (sample site information in Table S1). The habitat types included 21 lakes, two rivers and three estuaries. Each sample site had an average sample size of about 50 fish,

with a range from 18 to 78 (mean = 49.11, SD = 14.7). We identified a total of 1115 unique MHC variants, 820 (73.54%) of which were private variants found in only one population. On average, each fish had seven distinct MHC variants, with a range from three to 15 (mean = 6.94, SD = 1.87). In total, 4.6% of the variation of MHC variant number could be explained by habitat type (ANOVA, habitat term, $p < 2e-16$), 1.7% could be explained by watershed (ANOVA, watershed term, $p = 4.38e-05$), and 23.0% was explained by sample site (ANOVA, watershed/site term, $p < 2e-16$). Fish from river (mean = 8.38, SD = 1.98) and estuary (mean = 7.1, SD = 1.81) habitat had higher MHC variant number than fish from lake habitat (mean = 6.77, SD = 1.82).

In total, we identified 33 parasite taxa (parasite information in Table S2). On average, each fish had 2.5 different parasite taxa, with a range from zero to 10 (mean = 2.53, SD = 1.74). Fish from river (mean = 1.94, SD = 1.6) and estuary (mean = 1.15, SD = 0.76) habitat had lower parasite burdens than fish from lake habitat (mean = 2.82, SD = 1.74). In total, 11.8% of the variation in parasite richness could be explained by habitat type, 4.9% could be explained by watershed (ANOVA, habitat term, $p < 2e-16$), while 28.3% could be explained by sample site (ANOVA, watershed/site term, $p < 2e-16$). Bolnick et al. (2020) provide further analysis of ecological factors structuring the parasite metacommunity (diversity, composition, co-occurrence) within and among lakes.

3.1.1 | Heterozygote advantage and the selection for variant number

We found no evidence for an association between stickleback MHC variant number and parasite diversity, at either the scale

of individual fish or among populations. At the individual level (Figure 2a), the best mixed effect generalized linear regression model was the base model (AIC: 4226.02) that only included the fixed effect of log fish body length (coefficient = 1.24, $z = 8.78$, $p < 2e-16$) and the random intercept of sampling site (SD: 0.47) conveying that populations differ in parasite diversity (as previously shown; Bolnick et al., 2020a). Adding the number of MHC variants (coefficient = 0.01, $z = 1.22$, $p = .22$) into the base model slightly decreased model fit (AIC: 4226.53). Adding the square of MHC variant number further decreased model fit (AIC: 4239.6). No additional explanatory power was given by site by MHC variant number interactions (linear or quadratic), so we have no support for the possibility that MHC variant number is under selection in some populations but not others (a summary of all the models described in this section is available in Table S3). This result suggested that MHC variant number did not have a significant impact on parasite richness.

In the model just described, the random effect of sampling site explained a large proportion of the total variance of parasite richness, as noted previously (Bolnick et al., 2020a). Populations ranged from as few as on average 0.5 macroparasite taxa per individual, to nearly five taxa per individual (Figure 2b). We also observed significant variation in MHC variant number between populations, ranging from as low as on average 5.31 variants per fish, to as high as 9.71 variants per fish (Figure 2b). Among populations, average MHC variant number was unrelated to average parasite richness (coefficient = 0.03, $F = 0.02$, $df = 24$, $p = .89$, Figure 2b). The result was similar when we used the total number of MHC variants in a population as the dependent variable and the total number of parasite species in a population as the independent variable (linear regression, $t = -0.19$, $p = .85$, Data S1).

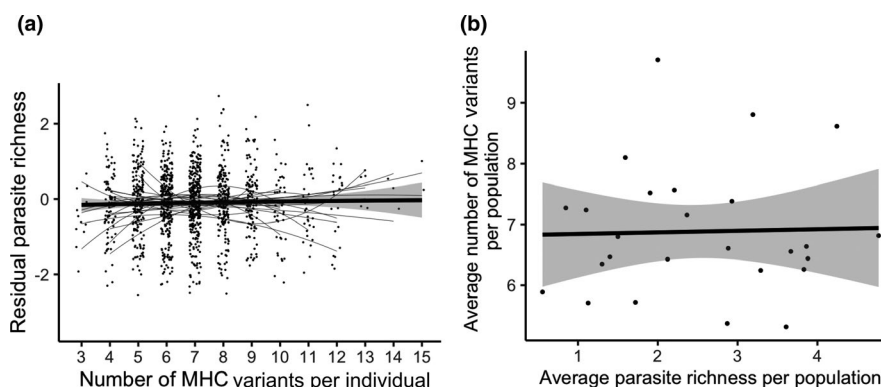


FIGURE 2 Testing predictions of heterozygote advantage. (a) The relationship between the number of MHC variants per individual and its residual parasite richness. For the purposes of plotting, the residual parasite richness for each fish was calculated by using a mixed-effect generalized linear model to control for the random effect of sampling site and the fixed effect of log body length (covariates in the statistical analysis are reported in the text). Each dot represents a fish. The quadratic fit using either the full data set or each population is represented by thick black line or thin grey lines, respectively. None of the quadratic fit using data from individual site was significant after correcting for multiple comparisons. (b) The relationship between the average number of MHC variants per fish in a given population and the average per-fish parasite richness in that population. Each dot represents a population. The grey area represents the 95% confidence interval. Note that x and y axes are reversed from panel (a) to (b). In (a) we are specifically testing whether individual genotype affects their parasite burden, as the reverse direction of causation is not plausible. In (b) we are considering the hypothesis that existing parasite diversity is the driver of evolution of MHC diversity

3.1.2 | Frequency-dependent selection: Is there a rare variant advantage

Although we found many MHC–parasite associations, we found no tendency for variants to be protective when rare, or vulnerable to infection when common. That is, the Z value of a given variant's effect on a given parasite in a given site was independent of that variant's prevalence in that focal site. Across all sites, there were a total of 5623 MHC–parasite combinations that were moderately frequent

and so included in the analysis. After excluding the models heavily influenced by outlier values with high leverage, we obtained 4130 linear regression models (see Figure 3a for an example of this result from one population). In total, 45 models had significant negative MHC–parasite associations and 45 models had significant positive MHC–parasite associations (see Figure 3b for an example). Some MHC variants, and parasites, are repeated across multiple regression models from different populations. None of the models were significant after we corrected for multiple-comparison with the BH

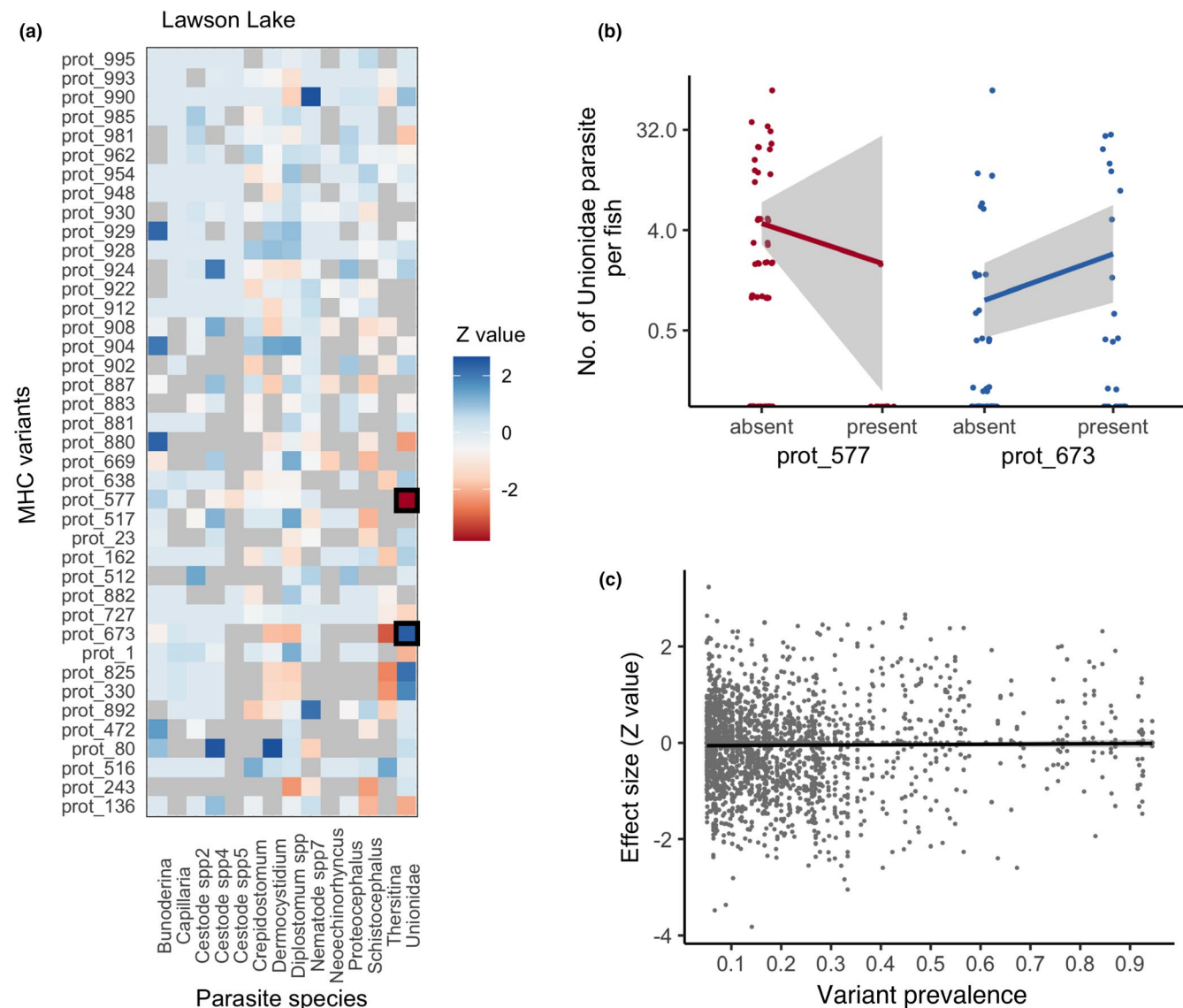


FIGURE 3 Testing predictions of frequency-dependent selection: rare variant advantage. (a) The Z values of all MHC–parasite regressions in one sample site (Lawson Lake), as an example. Each model used data from one population and regressed the infection intensity of a focal parasite against the presence/absence of a focal MHC variant. Each row in the grid represents an MHC variant, and each column represents a parasite taxon. The colour filling represents the effect size. The darker the blue colour indicates a stronger positive association (having the variant confers higher infection load). The darker the orange colour indicates a stronger negative association (e.g., resistance). The MHC–parasite regressions heavily biased by extreme values are filled as grey and excluded from further analysis. Two examples (highlighted in black boxes in (a)) are plotted in (b). (b) The infection intensity of parasite Unionidae is negatively and positively influenced by the presence/absence of MHC variant prot_577 and prot_673, respectively. Note the y axis of (b) is on a log scale. (c) The relationship between the variant prevalence of an MHC variant and its effect size (Z value) on the infection intensity of certain parasite. Each dot represents an MHC–parasite combination in a particular population. The solid line depicts the result of linear regression using all the data. The grey area surrounding the regression line indicates the 95% confidence interval

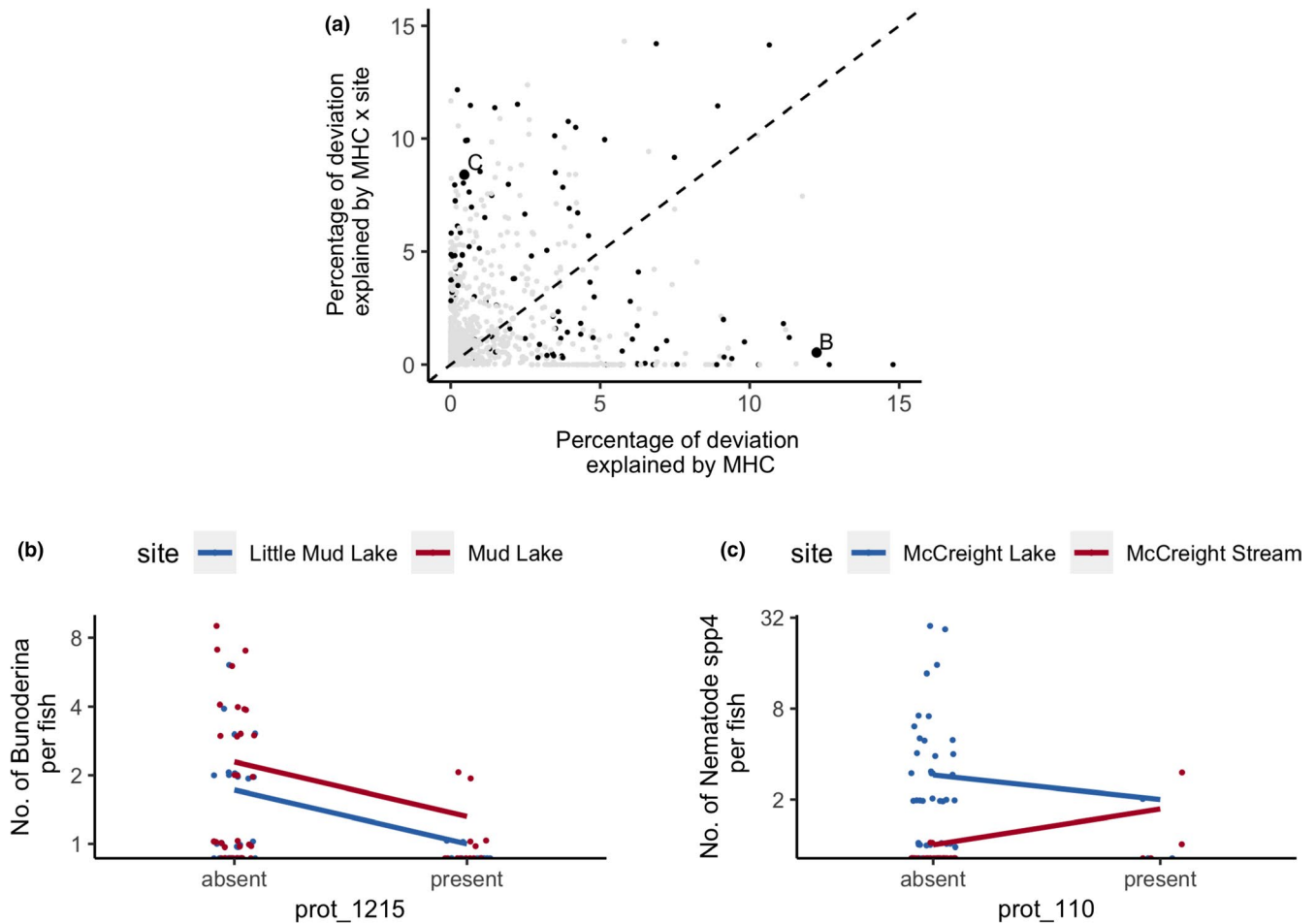


FIGURE 4 Testing predictions of frequency-dependent selection: inconsistent variant effects across sites. (a) The percentage of deviation residuals explained by the MHC term and the MHC \times sample site interaction term in the regression models, which regressed the infection load of a certain parasite against a particular MHC variant, sample site and the MHC \times sample site interaction. Each dot represents a MHC–parasite combination. The dotted line is a line of equal effect size (slope of 1, intercept of 0). The models with either term significant are plotted with black dots, and all the nonsignificant models are in grey. Two significant examples were highlighted in (a), and plotted in (b) and (c). (b) The effect of MHC variant prot_1215 on the infection load of parasite *Bunoderina* in Mud Lake and Little Mud Lake. (c) The effect of MHC variant prot_110 on the infection load of parasite *Nematode spp4* in McCreight Lake and McCreight Stream. Note that the y axes of (b) and (c) are on a log scale

method (Benjamini & Hochberg, 1995). A mixed effect linear model showed that the impact of variant prevalence on Z values was not significant (coefficient = 0.09, $t = 1.14$, $p = .27$; Figure 3c). The regression result was still not significant when we only included the Z values from models with $p < .05$ (before multiple comparison correction). Therefore, this result did not provide evidence for rare variant advantage of MHC variants on parasite infection intensities.

3.1.3 | Frequency-dependent selection: Are variant effects inconsistent across lakes?

We found some cases where the variant effects are inconsistent across lakes (significant MHC variant \times site interaction effects), suggesting there were potentially Red Queen style co-evolutionary arms races between stickleback populations and parasites. We obtained 788 generalized linear models for each moderately prevalent MHC–parasite

combination that was present in more than one site. In total, 53 models had a significant interaction term but an insignificant MHC term (11 of them involved parasites with birds as final hosts), suggesting inconsistent allelic effects across sites. For example, the variant prot_110 confers susceptibility (higher infection) in McCreight Stream, but has no effect in McCreight Lake (Figure 4c). Although none of the models were significant after correction for multiple-comparison with the BH method, the number of significant models was greater than a null expectation with a significance level of .05 ($792 \times .05 = 39.6$), so some of the model results were truly significant despite the issue of multiple comparison. In total, 66 models had a significant MHC term but an insignificant interaction term (13 of them involved parasites with birds as final hosts; Figure 4b), suggesting that those variants had consistent effects on the focal parasite across sites, and thus did not fit localized Red Queen race dynamics. Parasites with birds as final hosts were not more likely to be found in the models with a significant MHC term than in the models with a significant interaction term (Chi-squared test, $\chi^2 = 0.004$, $p = .95$).

3.1.4 | Fluctuating selection

Despite spatial heterogeneity in the distribution of parasite species (Bolnick et al., 2020b), we did not detect parasite-mediated spatially fluctuating selection. The Mantel test between the distance matrix of parasite community and the distance matrix of MHC variants across all populations suggested that the two matrices were not correlated ($r = .089$, $p = .22$, Figure 5). Thus, although both MHC variants and parasite communities differ strongly between lakes, there is no detectable relationship between them.

3.1.5 | Neutral evolution

We found that MHC diversity was significantly associated with genome-wide genetic diversity that should mostly reflect neutral processes. First, the average MHC variant number covaried positively with population mean heterozygosity (coefficient = 25.4, $F = 6.57$, $df = 24$, $p = .02$, Figure 6a). Second, there was a strong

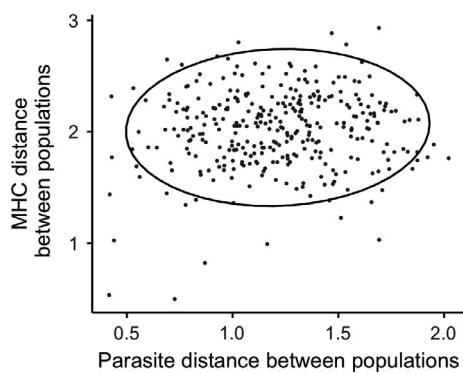


FIGURE 5 Testing predictions of spatially fluctuating selection. The parasite distance between populations is not detectably correlated with the MHC distance between populations. Each dot represents a pair of populations. The ellipse represents the 95% confidence interval prediction

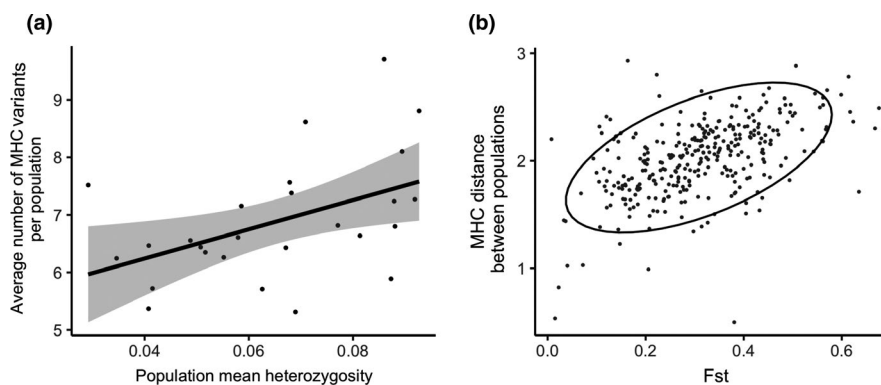


FIGURE 6 Testing predictions of neutral hypotheses. (a) The average number of MHC variants per population is positively associated with population mean heterozygosity. Each dot represents a population. The solid line represents the linear regression prediction. The grey area indicates the 95% confidence interval. (b) The relationship between MHC distance between populations and F_{ST} . Each dot represents a pair of populations. The ellipse indicates the 95% confidence interval prediction

positive correlation between MHC divergence between populations, and genome-wide divergence (Mantel test: $r = .51$, $p = .002$, Figure 6b). That is, more closely related populations tended to be more similar at their MHC loci as well.

4 | DISCUSSION

This study represents an effort to comprehensively evaluate multiple major hypotheses concerning the maintenance of MHC diversity, using observational data from 26 stickleback populations. Despite the large scale of sampling, we did not find strong support for any of the three popular parasite-mediated selection hypotheses, namely heterozygote advantage, frequency-dependent selection and fluctuating selection. In contrast, neutral processes seem to best explain MHC diversity in this system, in terms of both MHC variant number and between-population divergence. Spurgin and Richardson (2010) suggested that “MHC supertype” (MHC alleles clustered based on allelic divergence) better reflected functional differences among alleles than MHC alleles. We found that our conclusions were not changed even when we used MHC superotypes, instead of MHC alleles, to represent MHC variation (Data S1).

Consistent with previous studies in other vertebrates (Kaufman, 2018), MHC II β genes are highly polymorphic in the surveyed stickleback metapopulation. We found 1115 unique MHC II β variants from 1277 individuals, with a majority of variants only present in one population. However, although on average fish from lake habitats are infected with more diverse parasite species than those from river and estuary habitats (Bolnick et al., 2020a), lake stickleback have fewer MHC variants than others. The extremely high level of polymorphism in MHC genes is often attributed to parasite-mediated selection; however, the discrepancy between parasite richness and MHC variant number among habitat types suggests that parasite-mediated selection alone is unlikely to explain the high level of MHC polymorphism in the studied stickleback populations. This finding contradicts previous results from studies of stickleback inhabiting

different habitats in Germany (Wegner, Reusch, et al., 2003), where lake fish had higher parasite diversity (as in this study), and higher MHC diversity (unlike this study).

The above finding is further supported through testing specific predictions derived from the heterozygote advantage model. The heterozygote advantage hypothesis predicts that the number of MHC variants in the host genome would be *maximized* to recognize a wide range of parasites. For example, an experimental infection study in mice found that pathogen load was reduced by 41% in heterozygous mice with two different MHC resistant haplotypes, compared to MHC-congenic homozygous mouse strains (McClelland et al., 2003). In some lizard populations, the number of MHC alleles per individual was positively associated with tick numbers, which was a proxy for the pathogens the ticks transmit (Radwan et al., 2014). However, a meta-analysis found that whether the relationship between MHC diversity and parasite richness is positive or negative is taxon-dependent (Winternitz et al., 2013). Previous studies (Wegner, Kalbe, et al., 2003) in stickleback found support for the theoretical model that the number of MHC variants should be *optimized* at an intermediate level, rather than *maximized*. The argument is that too many MHC variants would reduce the T cell portfolio due to negative selection of self-recognition T cell receptors. In this study, we applied similar analytical approaches as Wegner et al. (2003) to a larger sample size and many more populations. Contrary to those previous results, we did not find support for either the *maximizing* or the *optimizing* hypothesis. This conclusion still holds when we used different diversity measures for MHC and parasite. (MHC variant richness in a population, controlling neutral genetic diversity, had little impact on average parasite load; Data S1.)

Negative frequency-dependent selection suggests that the frequency of MHC variants might experience a cyclical pattern through time, with different low-frequency variants playing the role of conferring resistance at different time points. Empirically demonstrating negative frequency-dependent selection is often challenging, but some studies support its role in MHC diversity maintenance. For example, in guppies, novel MHC variants were associated with less severe infection, which is consistent with negative frequency-dependent selection (Phillips et al., 2018). Another study found that retrovirus had higher fitness in mice with familiar MHC genotype (Kubinak et al., 2012). In this study, we found that MHC variant prevalence in a population is not associated with whether that variant is effective in parasite resistance. We further examined whether variants have inconsistent effects across sites, which would be predicted by negative frequency-dependent selection, because different populations are unlikely to be in the same phase of the host-parasite evolutionary arms race. We found inconsistent variant effects in some cases, but we also found consistent effects in a larger fraction of cases. So negative frequency-dependent selection might be able to explain a limited number of variant-parasite dynamics in our study. However, in most cases MHC variants confer equivalent protection in multiple isolated populations, inconsistent with cyclical changes due to coevolution. Note that only a small proportion of variants are shared by multiple sampling sites, which limits

our ability to test negative frequency-dependent selection broadly. Ideally, to demonstrate negative frequency-dependent selection, one would need to carry out a longitudinal study that samples both parasite genotypes and host genotypes through time. However, this type of study is very challenging to implement in a field setting, especially for annually reproducing animals.

Finally, fluctuating selection, either temporally or spatially, is frequently invoked to explain the maintenance of genetic diversity across a landscape (Osborne et al., 2017; Schemske & Bierzychudek, 2007). A long-term genetic survey in Soay sheep found that MHC diversity is probably maintained by the spatial and temporal fluctuation of its nematode parasite (Charbonnel & Pemberton, 2005). Our sampling sites on Vancouver Island differ in many physical parameters, such as water area, depth, temperature, flow speed, etc. It is also documented that parasite communities differ significantly among different sites (Bolnick et al., 2020a). This spatial heterogeneity in parasite communities is due to abiotic and biotic variables (e.g., lake size, fish diet; Bolnick et al. 2020), and could generate spatially fluctuating selection that maintains MHC diversity. However, we did not find significant correlation between the parasite distance matrix and the MHC distance matrix, suggesting that the spatial heterogeneity in parasite community could not explain the observed spatial variation in MHC diversity.

Interestingly, we found that neutral processes could contribute to MHC diversity in two different analyses: (i) populations with higher genomic heterozygosity also have higher average MHC variant number, and (ii) less divergent populations (low genomic F_{ST}) have similar MHC genotypes and frequencies. Similar to analysis (i), a previous study in stickleback (Wegner, Reusch, et al., 2003) also found a positive, albeit weak, association between neutral genetic diversity and population MHC allelic diversity. The role of neutral processes in the maintenance of MHC diversity has often been studied in the context of conservation, which is usually concerned with small and isolated populations (Miller & Lambert, 2004; Seddon & Ellegren, 2004). Our result suggested that at larger scales, neutral processes are also important factors in explaining MHC diversity within and between populations, even when there is no specific evidence for past bottlenecks.

Parasite-mediated selection is believed to have an important role in maintaining MHC polymorphism. Consistent with this, we found some associations between certain parasites and individual MHC variant. However, we did not find strong support for parasite-mediated selection hypotheses at the scale of metapopulations. There are several possible explanations. First, in this study, we examined the three hypotheses separately, but in reality, the mechanisms are not mutually exclusive. For example, low-frequency variants are more likely to be in a heterozygous state in a population. Therefore, if a variant were effective in resistance at low frequency, it would be consistent with both negative frequency-dependent selection and heterozygote advantage. Furthermore, if these mechanisms work in combination in different MHC-parasite pairs or in different populations, we may not detect a clear signal of each mechanism when combining data from multiple MHC-parasite pairs and many

populations. A simulation study showed that negative frequency-dependent selection could be weak and hard to detect, especially when it acts in combination with heterozygote advantage (Ejmond & Radwan, 2015). Second, parasites are not the only selective agents that act upon MHC loci. MHC genes are also involved in interactions with other species, such as symbionts comprising the gut microbiome (Bolnick et al., 2014), and within species interactions, such as mate choice (Milinski et al., 2005). It is unknown which MHC genes are relevant to which selective forces. If other selective forces are also at work, the influences of parasite-mediated selection on MHC diversity could be obscured. Third, we only measured parasite richness in this study to represent parasite diversity, so the details of parasite information, such as the differences in life history, phylogenetic relationship, etc., were not considered. In the future, would be of value to test the parasite-mediated selection hypotheses with diversity measurements that take into account the above nuances. Finally, our survey does not span multiple seasons, so we cannot evaluate how MHC variation is shaped across time by temporally fluctuating selection.

To summarize, we used a large-scale field survey to evaluate the three popular parasite-mediated selection hypotheses on MHC diversity, namely heterozygote advantage, negative frequency-dependent selection and fluctuating selection. We found that neutral processes best explain MHC diversity (allelic richness and population divergence), instead of those parasite-mediated selection mechanisms. Because MHC diversity can be influenced by many selective forces, and the outcome of this selection could be further shaped by spatial and temporal heterogeneity, our study suggests it may not be possible to parse out how each selective force influences MHC diversity at large scales directly from observational data. We propose that it is worthwhile to instead investigate how each selective force acts upon MHC diversity with experimental approaches (e.g., controlling other selective forces) at smaller scales. Combining insights from these small-scale controlled studies in a step-wise manner can be a fertile future direction to understand the complex process of MHC evolution and diversification.

ACKNOWLEDGMENTS

This work was supported by the David and Lucille Packard Fellowship for Science and Engineering (D.I.B.), the Howard Hughes Medical Institute (Early Career Scientist award to D.I.B.), and by NIH NIAID grant 1R01AI123659-01A1. Field sampling in 2009 was made possible by assistance from Todasporn Rodbumrung, Chris Harrison, Will Shim, Yuexin Jiang and Will Stutz. Will Shim provided the parasite information in Table S2.

AUTHOR CONTRIBUTIONS

D.I.B. conceived and designed this project with W.S. D.I.B. and W.S. performed the field survey. K.B. carried out the dissections. S.H.W. performed the M.H.C. sequencing, S.D.H. performed genotype scoring. F.P. performed data analysis and wrote the manuscript with input from D.I.B. All authors contributed in the revision of the manuscript.

DATA AVAILABILITY STATEMENT

MHC variant number and parasite species richness data, and population genomic heterozygosity data are deposited in Dryad: <https://doi.org/10.5061/dryad.1rn8pk0sk>. Sample site and parasite taxa information is included in Data S1. The code used to perform data analysis is available in the GitHub repository: https://github.com/foenpeng/MHC_analysis.

ORCID

Foen Peng  <https://orcid.org/0000-0001-5549-7753>

Kimberly M. Ballare  <https://orcid.org/0000-0002-1824-3364>

S. Hollis Woodard  <https://orcid.org/0000-0003-4948-1282>

REFERENCES

- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 289–300.
- Bolnick, D. I., & Ballare, K. M. (2020). Resource diversity promotes among-individual diet variation, but not genomic diversity, in lake stickleback. *Ecology Letters*, 23(3), 495–505. <https://doi.org/10.1111/ele.13448>
- Bolnick, D. I., Resetarits, E. J., Ballare, K., Stuart, Y. E., & Stutz, W. E. (2020a). Host patch traits have scale-dependent effects on diversity in a stickleback parasite metacommunity. *Ecography*, 43(7), 990–1002. <https://doi.org/10.1111/ecog.04994>
- Bolnick, D. I., Resetarits, E. J., Ballare, K., Stuart, Y. E., & Stutz, W. E. (2020b). Scale-dependent effects of host patch traits on species composition in a stickleback parasite metacommunity. *Ecology*, 101(12), e03181. <https://doi.org/10.1002/ecy.3181>
- Bolnick, D. I., Snowberg, L. K., Caporaso, J. G., Lauber, C., Knight, R., & Stutz, W. E. (2014). Major Histocompatibility Complex class IIb polymorphism influences gut microbiota composition and diversity. *Molecular Ecology*, 23(19), 4831–4845. <https://doi.org/10.1111/mec.12846>
- Bonneaud, C., Pérez-Tris, J., Federici, P., Chastel, O., & Sorci, G. (2006). Major histocompatibility alleles associated with local resistance to malaria in a passerine. *Evolution*, 60(2), 383–389. <https://doi.org/10.1111/j.0014-3820.2006.tb01114.x>
- Borghans, J. A. M., Beltman, J. B., & De Boer, R. J. (2004). MHC polymorphism under host-pathogen coevolution. *Immunogenetics*, 55(11), 732–739. <https://doi.org/10.1007/s00251-003-0630-5>
- Borghans, J. A. M., Noest, A. J., & De Boer, R. J. (2003). Thymic selection does not limit the individual MHC diversity. *European Journal of Immunology*, 33(12), 3353–3358. <https://doi.org/10.1002/eji.200324365>
- Calhoun, D. M., McDevitt-Galles, T., & Johnson, P. T. J. (2018). Parasites of invasive freshwater fishes and the factors affecting their richness. *Freshwater Science*, 37(1), 134–146. <https://doi.org/10.1086/696566>
- Charbonnel, N., & Pemberton, J. (2005). A long-term genetic survey of an ungulate population reveals balancing selection acting on MHC through spatial and temporal fluctuations in selection. *Heredity*, 95(5), 377–388. <https://doi.org/10.1038/sj.hdy.6800735>
- Dobzhansky, T. (1982). *Genetics and the origin of species* (Vol. 11). Columbia University Press.
- Doherty, P. C., & Zinkernagel, R. M. (1975). Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature*, 256(5512), 50–52. <https://doi.org/10.1038/256050a0>
- Eizaguirre, C., Lenz, T. L., Kalbe, M., & Milinski, M. (2012). Rapid and adaptive evolution of MHC genes under parasite selection in experimental vertebrate populations. *Nature Communications*, 3(1), 621. <https://doi.org/10.1038/ncomms1632>

- Ejsmond, M. J., & Radwan, J. (2015). Red queen processes drive positive selection on major histocompatibility complex (MHC) genes. *PLOS Computational Biology*, 11(11), e1004627. <https://doi.org/10.1371/journal.pcbi.1004627>
- Hedrick, P. W. (2002). Pathogen resistance and genetic variation at MHC loci. *Evolution; International Journal of Organic Evolution*, 56(10), 1902–1908. <https://doi.org/10.1111/j.0014-3820.2002.tb00116.x>
- Hedrick, P. W., Parker, K. M., & Lee, R. N. (2001). Using microsatellite and MHC variation to identify species, ESUs, and MUs in the endangered Sonoran topminnow. *Molecular Ecology*, 10(6), 1399–1412. <https://doi.org/10.1046/j.1365-294X.2001.01289.x>
- Hill, A. V. S. (1991). HLA associations with malaria in Africa: Some implications for MHC evolution. In J. Klein & D. Klein (Eds.), *Molecular evolution of the major histocompatibility complex* (pp. 403–420). Springer. https://doi.org/10.1007/978-3-642-84622-9_33
- Hughes, A. L., & Nei, M. (1988). Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature*, 335(6186), 167–170. <https://doi.org/10.1038/335167a0>
- Ilmonen, P., Penn, D. J., Damjanovich, K., Morrison, L., Ghotbi, L., & Potts, W. K. (2007). Major histocompatibility complex heterozygosity reduces fitness in experimentally infected mice. *Genetics*, 176(4), 2501–2508. <https://doi.org/10.1534/genetics.107.074815>
- Ishigaki, K., Lagattuta, K., Luo, Y., James, E., Buckner, J., & Raychaudhuri, S. (2020). HLA autoimmune risk alleles restrict the hypervariable region of T cell receptors. *MedRxiv*, 2020.11.08.20227983. doi: 10.1101/2020.11.08.20227983
- Jensen, P. E. (2007). Recent advances in antigen processing and presentation. *Nature Immunology*, 8(10), 1041–1048. <https://doi.org/10.1038/ni1516>
- Kaufman, J. (2018). Unfinished business: Evolution of the MHC and the adaptive immune system of jawed vertebrates. *Annual Review of Immunology*, 36, 383–409. <https://doi.org/10.1146/annurev-immunol-051116-052450>
- Kubinak, J. L., Ruff, J. S., Hyzer, C. W., Slev, P. R., & Potts, W. K. (2012). Experimental viral evolution to specific host MHC genotypes reveals fitness and virulence trade-offs in alternative MHC types. *Proceedings of the National Academy of Sciences of the United States of America*, 109(9), 3422–3427. <https://doi.org/10.1073/pnas.1112633109>
- Matthews, B., Harmon, L. J., M'Gonigle, L., Marchinko, K. B., & Schaschl, H. (2010). Sympatric and allopatric divergence of MHC genes in threespine stickleback. *PLoS One*, 5(6), e10948. <https://doi.org/10.1371/journal.pone.0010948>
- Mccairns, R. J. S., Bourget, S., & Bernatchez, L. (2011). Putative causes and consequences of MHC variation within and between locally adapted stickleback demes. *Molecular Ecology*, 20(3), 486–502. <https://doi.org/10.1111/j.1365-294X.2010.04950.x>
- McClelland, E. E., Penn, D. J., & Potts, W. K. (2003). Major histocompatibility complex heterozygote superiority during coinfection. *Infection and Immunity*, 71(4), 2079–2086. <https://doi.org/10.1128/IAI.71.4.2079-2086.2003>
- Migalska, M., Sebastian, A., & Radwan, J. (2019). Major histocompatibility complex class I diversity limits the repertoire of T cell receptors. *Proceedings of the National Academy of Sciences of the United States of America*, 116(11), 5021–5026. <https://doi.org/10.1073/pnas.1807864116>
- Milinski, M., Griffiths, S., Wegner, K. M., Reusch, T. B. H., Haas-Assenbaum, A., & Boehm, T. (2005). Mate choice decisions of stickleback females predictably modified by MHC peptide ligands. *Proceedings of the National Academy of Sciences of the United States of America*, 102(12), 4414–4418. <https://doi.org/10.1073/pnas.0408264102>
- Miller, H. C., & Lambert, D. M. (2004). Genetic drift outweighs balancing selection in shaping post-bottleneck major histocompatibility complex variation in New Zealand robins (Petroicidae). *Molecular Ecology*, 13(12), 3709–3721. <https://doi.org/10.1111/j.1365-294X.2004.02368.x>
- Nowak, M. A., Tarczy-Hornoch, K., & Austyn, J. M. (1992). The optimal number of major histocompatibility complex molecules in an individual. *Proceedings of the National Academy of Sciences of the United States of America*, 89(22), 10896–10899. <https://doi.org/10.1073/pnas.89.22.10896>
- Osborne, M. J., Pilger, T. J., Lusk, J. D., & Turner, T. F. (2017). Spatio-temporal variation in parasite communities maintains diversity at the major histocompatibility complex class IIβ in the endangered Rio Grande silvery minnow. *Molecular Ecology*, 26(2), 471–489. <https://doi.org/10.1111/mec.13936>
- Penn, D. J., Damjanovich, K., & Potts, W. K. (2002). MHC heterozygosity confers a selective advantage against multiple-strain infections. *Proceedings of the National Academy of Sciences of the United States of America*, 99(17), 11260–11264. <https://doi.org/10.1073/pnas.162006499>
- Penn, D. J., & Potts, W. K. (1999). The evolution of mating preferences and major histocompatibility complex genes. *The American Naturalist*, 153(2), 145–164. <https://doi.org/10.1086/303166>
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP Discovery and genotyping in model and non-model species. *PLoS One*, 7(5), e37135. <https://doi.org/10.1371/journal.pone.0037135>
- Phillips, K. P., Cable, J., Mohammed, R. S., Herdegen-Radwan, M., Raubic, J., Przesmycka, K. J., van Oosterhout, C., & Radwan, J. (2018). Immunogenetic novelty confers a selective advantage in host–pathogen coevolution. *Proceedings of the National Academy of Sciences of the United States of America*, 115(7), 1552–1557. <https://doi.org/10.1073/pnas.1708597115>
- Radwan, J., Babik, W., Kaufman, J., Lenz, T. L., & Winternitz, J. (2020). Advances in the evolutionary understanding of MHC polymorphism. *Trends in Genetics*, 36(4), 298–311. <https://doi.org/10.1016/j.tig.2020.01.008>
- Radwan, J., Kuduk, K., Levy, E., LeBas, N., & Babik, W. (2014). Parasite load and MHC diversity in undisturbed and agriculturally modified habitats of the ornate dragon lizard. *Molecular Ecology*, 23(24), 5966–5978. <https://doi.org/10.1111/mec.12984>
- Schemske, D. W., & Bierzychudek, P. (2007). Spatial differentiation for flower color in the desert annual *Linanthus parryae*: Was wright right? *Evolution*, 61(11), 2528–2543. <https://doi.org/10.1111/j.1558-5646.2007.00219.x>
- Sebastian, A., Herdegen, M., Migalska, M., & Radwan, J. (2016). amplis: A web server for multilocus genotyping using next-generation amplicon sequencing data. *Molecular Ecology Resources*, 16(2), 498–510. <https://doi.org/10.1111/1755-0998.12453>
- Seddon, J. M., & Ellegren, H. (2004). A temporal analysis shows major histocompatibility complex loci in the Scandinavian wolf population are consistent with neutral evolution. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 271(1554), 2283–2291. <https://doi.org/10.1098/rspb.2004.2869>
- Slade, R. W., & McCallum, H. I. (1992). Overdominant vs. frequency-dependent selection at MHC loci. *Genetics*, 132(3), 861–864.
- Sommer, S. (2005). The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Frontiers in Zoology*, 2, 16. <https://doi.org/10.1186/1742-9994-2-16>
- Spurgin, L. G., & Richardson, D. S. (2010). How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings of the Royal Society B: Biological Sciences*, 277(1684), 979–988. <https://doi.org/10.1098/rspb.2009.2084>
- Stuart, Y. E., Veen, T., Weber, J. N., Hanson, D., Ravinet, M., Lohman, B. K., Thompson, C. J., Tasneem, T., Doggett, A., Izen, R., Ahmed, N., Barrett, R. D. H., Hendry, A. P., Peichel, C. L., & Bolnick, D. I. (2017). Contrasting effects of environment and genetics generate a continuum of parallel evolution. *Nature Ecology & Evolution*, 1(6), 1–7. <https://doi.org/10.1038/s41559-017-0158>

- Stutz, W. E., & Bolnick, D. I. (2014). Stepwise threshold clustering: A new method for genotyping MHC loci using next-generation sequencing technology. *PLoS One*, 9(7), e100587. <https://doi.org/10.1371/journal.pone.0100587>
- Stutz, W. E., & Bolnick, D. I. (2017). Natural selection on MHC II β in parapatric lake and stream stickleback: Balancing, divergent, both or neither? *Molecular Ecology*, 26(18), 4772–4786. <https://doi.org/10.1111/mec.14158>
- Stutz, W. E., Schmerer, M., Coates, J. L., & Bolnick, D. I. (2015). Among-lake reciprocal transplants induce convergent expression of immune genes in threespine stickleback. *Molecular Ecology*, 24(18), 4629–4646. <https://doi.org/10.1111/mec.13295>
- Takahata, N., & Nei, M. (1990). Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics*, 124(4), 967–978.
- Wegner, K. M., Kalbe, M., Kurtz, J., Reusch, T. B. H., & Milinski, M. (2003). Parasite selection for immunogenetic optimality. *Science*, 301(5638), 1343. <https://doi.org/10.1126/science.1088293>
- Wegner, K. M., Reusch, T. B. H., & Kalbe, M. (2003). Multiple parasites are driving major histocompatibility complex polymorphism in the wild. *Journal of Evolutionary Biology*, 16(2), 224–232. <https://doi.org/10.1046/j.1420-9101.2003.00519.x>
- Winternitz, J. C., Minchey, S. G., Garamszegi, L. Z., Huang, S., Stephens, P. R., & Altizer, S. (2013). Sexual selection explains more functional variation in the mammalian major histocompatibility complex than parasitism. *Proceedings. Biological Sciences*, 280(1769), 20131605. <https://doi.org/10.1098/rspb.2013.1605>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Peng F, Ballare KM, Hollis Woodard S, Haan S, Bolnick DI. What evolutionary processes maintain MHC II β diversity within and among populations of stickleback?. *Mol Ecol*. 2021;30:1659–1671. <https://doi.org/10.1111/mec.15840>