Classification: BIOLOGICAL SCIENCES – EVOLUTION

Title: Immunogenetic novelty confers a selective advantage in host-pathogen coevolution

Short title: Selective advantage of immunogenetic novelty

Authors: Karl P. Phillips[a], Joanne Cable[b], Ryan S. Mohammed[c], Magdalena Herdegen-Radwan[d], Jarosław Raubic[a], Karolina J. Przesmycka[a], Cock van Oosterhout[e], Jacek Radwan[a,1]

Affiliations:

[a] Evolutionary Biology Group, Faculty of Biology, Adam Mickiewicz University, Poznań, Poland

[b] School of Biosciences, Cardiff University, Cardiff CF10 3AX, United Kingdom

[c] Department of Life Sciences, The University of the West Indies, St. Augustine, Trinidad and Tobago

[d] Department of Behavioural Ecology, Institute of Environmental Biology, Faculty of Biology, Adam Mickiewicz University, Umultowska 89, 61-614 Poznań, Poland

[e] School of Environmental Sciences, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, United Kingdom

[1] Correspondence to jradwan@amu.edu.pl

Keywords: host-pathogen coevolution; Red Queen coevolution; major histocompatibility complex; *Poecilia reticulata*; *Gyrodactylus*; frequency-dependent selection

Abstract

The major histocompatibility complex (MHC) is crucial to the adaptive immune response of vertebrates and is among the most polymorphic gene families known. Its high diversity is usually attributed to selection imposed by fast-evolving pathogens. Pathogens are thought to evolve to escape recognition by common immune alleles, and, hence, novel MHC alleles, introduced through mutation, recombination or gene flow, are predicted to give hosts superior resistance. Although this theoretical prediction underpins host-pathogen 'Red Queen' coevolution, it has not been demonstrated in the context of natural MHC diversity. Here, we experimentally tested whether novel MHC variants (both alleles and functional 'supertypes') increased resistance of guppies (*Poecilia reticulata*) to a common ectoparasite (*Gyrodactylus turnbulli*). We used exposure-controlled infection trials with wild-sourced parasites, and *Gyrodactylus*-naïve host fish that were F2 descendants of crossed wild populations. Hosts carrying MHC variants (alleles or supertypes) that were new to a given parasite population experienced a 35-37% reduction in infection intensity, but the number of MHC variants carried by an individual, analogous to heterozygosity in single-locus systems, was not a significant predictor. Our results provide direct evidence of novel MHC advantage, confirming a fundamental mechanism underpinning the exceptional polymorphism of this gene family, and highlighting the role of immunogenetic novelty in host-pathogen coevolution.

Significance

The major histocompatibility complex (MHC) is one of the most polymorphic gene families in the vertebrate genome, with natural selection actively promoting and maintaining variability. The exact mechanism/mechanisms responsible for these characteristics remain unclear, but identifying them is fundamental to our understanding of host-pathogen dynamics. Using targeted crosses of the model Trinidadian guppy, a tractable parasite, and exposure-controlled infection trials, we show that novel MHC variants are associated with less severe infections. Uniquely, our experimental design separates novel variant advantage from other modes of selection and confounding variables, such as individual MHC variability and genomic background. We thus demonstrate a fundamental process

driving evolution of the vertebrate immune system, which helps explain the unique features of MHC genes.

\body

Introduction

Host-pathogen coevolution is thought to drive the maintenance of genetic variation in immune genes, with consequences for important evolutionary processes including the evolution of virulence, the maintenance of sex, and sexual selection (1-4). One of the most striking examples of genetic polymorphism thought to be maintained by such processes is the vertebrate major histocompatibility complex (MHC), where dozens to hundreds of alleles may segregate in natural populations (5-7). The high polymorphism of this important immune gene family, which codes for proteins that present pathogen-derived antigens to T-cell receptors, has been a subject of research for decades (8), and understanding the processes that maintain this diversity has implications for areas outside evolutionary biology, from human health (9) to conservation biology (10, 11).

Despite almost fifty years of investigation, the processes driving evolution at the MHC are not fully understood (12). At the molecular level, the exceptionally high ratio of non-synonymous (protein-altering) to synonymous nucleotide substitutions in MHC genes suggests that selection is not only maintaining polymorphism ('balancing selection'), but is also actively promoting new polymorphism ['positive selection' (13, 14)]. Several mutually non-exclusive mechanisms may contribute to these selective pressures: heterozygote advantage (recognizing a wider spectrum of antigens); frequency-dependent selection from fast-evolving pathogens that favors rare or novel MHC variants; and variable selection in space and time (12). Recent theoretical work has suggested that frequency-dependent selection resulting from Red Queen dynamics may be the more important process, with the advantage conferred by novel alleles being particularly important in generating the patterns of allelic diversity observed at the MHC (15, 16). Novel allele advantage is an old hypothesis in MHC research, dating to the earliest days of observing the MHC's extreme polymorphism (17). The mechanistic potential for novel MHC variants to confer adaptive advantage against pathogens

has been demonstrated experimentally only relatively recently, using congenic mice and artificially selected virus lineages (18, 19). However, the number of MHC alleles segregating in wild populations can be upwards of two orders of magnitudes higher than that of the mouse-virus system. It may be more difficult for pathogens to adapt to specific local variants, and novel variants will be competing in a much larger pool of alleles with potentially a wide range of antigen-binding properties. Experimentally testing novel variant advantage in more natural, ecological contexts is much harder, as the potential selective pressures acting on the MHC are notoriously hard to disentangle (12).

Here, we used direct experimentation to investigate how novel MHC class II alleles (which recognize extracellular pathogens) in tropical freshwater guppies (*Poecilia reticulata*) affected the infection trajectory of their monogenean parasite *Gyrodactylus turnbulli*. These ectoparasites are widespread across guppy populations, and exert significant selective pressure (20, 21). Heavy infections can kill hosts (22, 23), and some MHC class II genotypes have been linked to gyrodactylid infection in the wild (21; see also 24). This host-parasite system is highly tractable: host exposure is easy to control, and infections can be monitored through time without killing host or parasite (22, 23). Hosts in our experiments were F2 descendants of crosses between guppy populations that shared no MHC alleles (see methods), and gyrodactylid worms for each replicate cross were wild-caught and came from one of the populations used to found the respective cross. In addition to MHC novelty defined by amino acid sequences, we also considered novelty based on MHC 'supertypes', where MHC alleles are grouped into clusters with similar physicochemical properties (25-27). We predicted that hosts carrying MHC variants that were novel with respect to parasite origin would perform better in controlled gyrodactylid infection trials than hosts carrying 'local' MHC variants.

Results

Amongst guppies that survived to the end of the experiment (n = 209), fish carrying only novel alleles or supertypes (designated as N/N genotypes and N/N supergenotypes, respectively; see 'Materials and methods') experienced *G. turnbulli* infections that were significantly less severe than fish

carrying only 'local' alleles or supertypes (L/L). Infection severity was measure in 'worm days' (the area under a graph of number of worms against time), and analyses used AIC_C-based multi-model inference (see methods). The N/N genotypes (n = 44) and N/N supergenotypes (n =14) respectively experienced 35% and 37% fewer 'worm days' compared to the L/L genotype and supergenotype fish (n = 59 and 88, respectively; *P* = 0.003 and 0.012; Tables S2.1a,b, S2.2a,b). L/N genotypes (n = 106; i.e. fish carrying both novel and local alleles) experienced infections of comparable intensity to L/L genotypes (*P* = 0.65; Table S2.1b), whereas L/N supergenotypes (n = 107) experienced intermediate infection intensities that were only marginally non-significant relative to L/L (*P* = 0.055; Table S2.2b). Direct comparison of the best allele-based and supertype-based models of worm days, using a constant set of covariates, indicated that allele-based groupings produced the better fit (ΔAIC_C = 4.22; Tables S2.1a, S2.2a); however, fish carrying at least one novel supertype (n = 121) experienced 27.1% fewer worm days than fish with no novel supertypes but at least one novel allele (n = 29; top-ranked model; *P* = 0.02; Tables S2.3a,b; Fig. S2.1). We did not detect a significant interaction between replicate population and MHC genotype/supergenotype class (ΔAIC_C = +7.37/+5.71, *P* ≥ 0.13/0.24). Neither the number of alleles nor the number of supertypes carried by a host – measures used as analogues of heterozygosity – were significantly associated with the number of worm days experienced by hosts (ΔAIC_C = +1.82/+1.28, *P* = 0.52/0.32; Tables S2.1a,b, S2.2a,b). No genetic variables were significant predictors of host mortality (Tables S2.5a-c, S2.6a-c), despite worm load itself being a significant predictor of mortality from infection day 3 onwards (Table S2.4).

Discussion

Our results support the novel MHC variant hypothesis: N/N hosts experienced parasite infections that were significantly less severe than those of L/L hosts. This was the case whether novelty was defined by amino acid sequences (alleles) or by physicochemical functional groups (supertypes). Differences in parasite burden between the genotype classes did not translate into a detectable effect on host survival. However, this may reflect the relatively benign and stable conditions of the

experiment: in the wild, fish weakened by infection may be more susceptible to predation (28) and secondary infections (29), and to environmental stressors such as river spates (30) – even one additional worm can reduce a wild guppy's survival probability (30). Furthermore, besides reducing survival, parasites may reduce host fitness by affecting reproductive potential, as has previously been demonstrated in guppies (31).

The novel variant advantage that we observed could, in theory, result in either balanced polymorphism or fixation of the novel variant – i.e. it is consistent with both balancing and positive selection. When a novel variant is introduced into a natural population (by point mutation, microrecombination or introgression), both processes are co-occurring and indistinguishable, likely resulting in an increase of the novel allele's frequency. We explored this potential using computer simulations parameterized from our current data on the effects of novel alleles/supertypes on gyrodactylid load, and from the effect of gyrodactylid load on the survival of guppies in the field from a mark-recapture study (30). These simulations show that upwards of 11% of novel variants should successfully establish in a population, and upwards of 54% if the variant is a novel supertype (compared to <0.1% in neutral simulations; see Appendix S10).

In the long term, the same process that leads to novel variant advantage – adaptation of parasites to local MHC genotypes – should diminish the advantage of the variant, leading to balanced polymorphism via negative frequency-dependent selection (12, 16, 17, 32). Such dynamics, whereby a novel (or rare) allele increases in frequency, loses its advantage, and decreases in frequency again have yet to be demonstrated. Alternatively, in the absence of balancing selection, novel variants could spread to fixation in a population, but such a scenario is inconsistent with the high MHC polymorphism observed in most study systems being coupled with strong signatures of positive selection.

Novel variant advantage may also explain the striking trans-species polymorphism observed at the MHC (33): if such polymorphisms are derived from hybridization [as opposed to being ancestral; (34)], novel variant advantage may accelerate introgression and promote interspecies

148 sharing of polymorphism. Furthermore, novel allele advantage might also affect the evolution of
149 MHC-based mating preferences, an important factor in shaping MHC diversity (35-37). Our results
150 suggest that preferences for partners with MHC alleles that are novel in a population, rather than
151 those that just differ from self MHCs, should be strongly favored.

152 Negative frequency-dependent dynamics (favoring both novel and [sufficiently] rare alleles)
153 and heterozygote advantage are two important types of balancing selection maintaining MHC
154 diversity (12). Although frequency dependence and heterozygote advantage are not mutually
155 exclusive, their relative influences are notoriously hard to test independently (12) and may be
156 impossible to separate by observational studies alone (38). Other processes that can contribute to
157 MHC diversity further complicate the separation (2, 12, 39). However, the crosses in our experiment
158 produced genotypes/supergenotypes that would not normally be found in natural populations at the
159 time at which novelty enters/arises. In particular, we generated hosts that were 'homozygous' with
160 respect to MHC novelty (N/N) while controlling for genetic background. Hence, we were able to
161 separate the effects of novelty from the effects of simply carrying more MHC variants, and this
162 showed that the number of alleles or supertypes carried by an individual was not significantly
163 associated with infection intensity (Tables S2.1a,b, S2.2a,b). Furthermore, we did not detect an
164 overdominance-type advantage for L/N fish (Tables S2.1b, S2.2b). The simplicity of the guppy-
165 *Gyrodactylus* system may explain the lack of heterozygote advantage, which previous work has
166 shown to be particularly important in multi-pathogen systems (40). Importantly, however, our
167 results show that, against a natural host-pathogen genetic background, novel MHC variants can be
168 selectively advantageous by virtue of properties arising from their novelty/extreme rarity (17, 32),
169 rather than by simply being present in heterozygotes.

170 Direct comparison of the best allele-based and supertype-based models of infection
171 intensity indicated that allele-based groupings produced the better fit (Tables S2.1a, S2.2a). On the
172 other hand, fish with at least one novel supertype experienced infections that were significantly less
173 severe than fish with no novel supertypes but at least one novel allele (i.e. a novel amino acid

sequence variant within a shared supertype). This suggests that functional novelty may be more important than simple allelic novelty, whereby the unique binding properties of the novel supertypes are more likely to fill an immune response void (41, 42). The disparity between the statistical model and empirical observations on parasite loads of guppies with and without novel supertypes highlights that the relative fitness contributions of novel supertypes and novel alleles within supertypes remain to be determined. Despite this uncertainly, our experiment demonstrates a general advantage of novel MHC variants.

Whilst the breeding design of our study controls for population-level linkage between MHC class II genes and other genes that may affect immune responses (28 chromosome pairs (43), plus recombination when F1s reproduce), we cannot exclude possible effects from genes that may be in close physical linkage with the MHC without knockdown experiments or isogenic guppy lineages. Unlike for tetrapods, linkage with MHC class I can be ruled out for teleost fish (44) – a pertinent point because, although MHC class I usually targets intracellular pathogens, these genes have been co-opted into roles more typical of class II in some teleosts (45). Concerning the exact mechanism by which the MHC may influence responses to skin ectoparasites, one possibility is through antigen-presenting skin cells (dendritic cells) mediating production of pro-inflammatory cytokines. This has been demonstrated *in vitro* with zebrafish skin tissue (46), and implied by gene expression studies on salmonids infected with sea lice (47-49). That supertypes are associated with lower infection intensities also suggests a functional rather than linkage-based influence of MHC.

Our results suggest that novel variants rather than locally adapted variants are associated with lower levels of parasite infection, and the lack of a significant interaction with replicate population suggests this finding may be generalisable. This contrasts with several studies on stickleback (*Gasterosteus aculeatus*) MHC-parasite interactions. When comparing or crossing lake and river stickleback populations, these studies have variously shown local adaptation (24), immigrant advantage (50), and a mixture of signals (51). This variation likely reflects the highly divergent ecologies and parasite faunas of lake and river sticklebacks, which may exert a complex

suite of selection pressures (24, 50-54). Our study builds on the important insights from these studies by overcoming some of their limitations, such as the use of only a single population pair (24, 50); no experimental control of population-level linkage [(50); although this study's use of statistical control means it is uniquely able to demonstrate effects of such linkage]; and the stock caveats of snapshot observational studies [(51); e.g., uncertainty regarding where an allele may be in a frequency-dependent dynamic, and the difficulty of separating effects of rare alleles from heterozygosity (12)]. Also, these previous studies did not test for effects of supertypes. The most pertinent stickleback experiment to our result presents the simplest finding: MHC variants that confer resistance, irrespective of being rare or common, tend to increase in frequency (55). Many traits and circumstances may make a variant resistant; our study shows that novelty is likely to be one of them.

Based on our results, we predict that novel variants, including those coming from immigrants, have a reasonable probability of establishing and spreading. However, inferring the consequences of novel variant advantage from MHC-based population genetic structure is more challenging, especially in populations connected by gene flow (42). Furthermore, genuine novelty of a variant can only be ascertained by exhaustively sampling a population's MHC diversity over an extended period of time. Introgression of novel alleles may be easier to observe if it occurs among well-diversified populations coming into secondary contact, including between species (e.g. 56) or between allopatric populations (such as those we investigated). We suspect that the small number of alleles shared between Trinidad and Tobago (Appendix S5) might come from fish introduced to Tobago by humans, as most of these alleles were found in a population close to human settlements and communication hubs. If so, our simulations suggest that these alleles are likely to spread across Tobago in the nearby future, subject to migration rate to other Tobagonian populations.

Our data constitute empirical demonstration of a long-posited, fundamental model of the evolution of MHC variability: novel immune variants confer a selective advantage, consistent with Red Queen scenarios in which parasites adapt to local host immune genotypes (8, 15, 16). We show

this effect using wild-sourced host and pathogen genetic variation, indicating that demonstrations of novel MHC advantage in congenic laboratory systems (18, 19) may be applicable in the context of natural MHC diversity. Furthermore, because we show this with replicate population crosses while controlling for population genetic background, our finding should be pertinent regardless of the source of novelty (e.g. mutation, recombination, introgression) or the rarity with which novelty enters a system. Looking beyond the MHC, although Red Queen dynamics have been shown in several non-vertebrate host-pathogen systems (57-59), examples of experimentally tested molecular mechanisms are rare, even for simple systems such as bacteria-phage interactions (60). In contrast, we explicitly link phenotype (infection intensity) to genotype (novel/local MHC) with an *a priori* hypothesis, using a gene family with a well-characterized immunological function. Overall, our results show that the advantage of immunogenetic novelty is a key selective agent underpinning Red Queen coevolution, and that, despite skepticism (1, 61), Red Queen coevolution may be an important force in shaping the immune genes of complex organisms.

Materials and methods

All methods are described in more detail in Appendix S1. At our field laboratory in Tobago, we conducted controlled gyrodactylid infection experiments on five replicate guppy populations. Each population was the F2 descendant of a cross between a wild Trinidad guppy population and a wild Tobago guppy population. Sampling locations and numbers of founding pairs per replicate are given in Appendix S3. We reared the founding females for each replicate from wild-caught juveniles to ensure they were A) virgins and B) free from gyrodactylids before being used for breeding (husbandry/screening details in Appendix S1.1). The males were freshly-caught wild adults. To ensure a 1:1 sex ratio at foundation, and to minimize the risk of parasite transfer from males to females, we used artificial insemination (62) to make the crosses. Between-island crosses allow a more powerful proof-of-principle test of the novel variant hypothesis than within-island crosses because they minimize the risk of crossed guppy populations having shared/exchanged MHC alleles

252 or parasites in the recent past (between-island population structure is stronger than within-island
253 structure; (42, 63); Appendix S5). Between-island crosses also removed a hierarchical factor ('island')
254 that we would have struggled to achieve adequate replication to control. All males and females were
255 fin-clipped for DNA (caudal fin, 2-4 mm^2; preserved in 0.3 ml 97% ethanol).
256 After insemination, we released the females to 800 L mesocosms (one per crossed
257 population) and provided supplementary feeding (Appendix S1.1). One month after observing the
258 first F1s in each mesocosm, we removed all surviving founding females and allowed the F1s to
259 mature and mate among themselves. One month after observing the first F2s in each mesocosm, we
260 removed and fin-clipped all F1s. We did not attempt to control the mating of F1s because of
261 logistical constraints, and because population-level replication coupled with the testing of a very
262 general allelic property (novelty) should minimize potential family effects. However, as a precaution,
263 we tested whether F_{IS} among F2s differed significantly from zero (see below). At both removal
264 stages, we verified that the fish were free from ectoparasites (64). We found no endoparasites in a
265 subsample of 4-8 fish per mesocosm that were dissected fresh after the experiment.
266 Three months after observing the first F2s, we began controlled gyrodactylid ('gyros')
267 infection experiments on these parasite-naïve fish. For each experimental population, we caught
268 fresh wild guppies from one founder population (Table S3.2), screened these for gyros, and used fish
269 with suitably large infections as parasite donors. Heavier donor infections are both easier to work
270 with (gyros jump to new hosts more readily when the donor has a higher gyro load) and likely to be
271 important sources of new infections in the wild. To minimize the handling of infected fish and the
272 time for which gyros were held in captivity prior to making experimental infections, we did not
273 attempt to identify gyros to species level before the experiments, despite the likelihood of multiple
274 species being present (see molecular identification below). We did not captive-breed individual
275 lineages because we wanted wild parasite genetic variation, free from artificial selection (65). Under
276 weak anesthesia (MS-222), each recipient F2 fish was infected with two gyros. Full details of the
277 infection protocol are given in Appendix S1.2. After being revived, each recipient fish was transferred

278 to a separate 400 ml isolation container. After infection on day 0, we briefly anaesthetized all
279 infected fish and counted the number of gyros on day 1, and then every other day thereafter for 17
280 days (22, 23). Isolation containers were kept at ambient outside temperature in shade (husbandry
281 details in Appendix S1.2). Any fish found dead (checked every 3-12 h, depending on infection
282 intensity) was promptly preserved whole in 1 ml 97% ethanol, with the ethanol replaced after 6 h
283 and again after 24 h. Fish that survived until day 17 were fin-clipped.

284 The work conducted in these experiments was approved by Cardiff University's animal ethics
285 committee and covered by UK Home Office Licence PPL 302876. All Tobago-sourced wild fish were
286 collected with permission from the Tobago House of Assembly (Permit #004/2014). No specific
287 permits are required on Trinidad, but we collected only from areas where guppies were reasonably
288 abundant.

289 From genomic DNA extracted from fin clips, we genotyped all P-generation and F2 fish at a
290 217 bp fragment of the MHC class II that codes for the highly polymorphic β-chain of the MHC
291 molecule's antigen binding groove (66). All genotypes used in downstream analyses had ≥ 300 allelic
292 reads (median = 1042 reads). Full details of primers, PCR conditions, and genotyping bioinformatics
293 (67, 68) parameters are given in Appendix S1.3.

294 We used custom Python scripts to assign MHC alleles in each replicate as coming from the
295 maternal or paternal founding populations. Assignment was unambiguous – there was only one
296 linkage block (1-2 allelic variants per haplotype), all variants within population crosses differed by at
297 least one non-synonymous mutation, and we observed no allele sharing between any crossed
298 population pair (Appendix S5). We then designated alleles as either 'L' (local) when they belonged to
299 the same host population as the worms used for that replicate, or 'N' (novel) when detected only in
300 the 'other' founding population, and allocated F2s to three genotype groups based on these
301 designations: N/N (two novel haplotypes); L/L (two local haplotypes); and L/N
302 (mixed/'heterozygous'). Although a nominally novel allele could be present in the 'local' population
303 at a frequency too low for our sample to detect, population genetic analyses on a larger dataset

304 suggested that this is likely to be extremely rare (e.g. only 8/214 alleles in our dataset were present
305 on both islands; Appendix S5). Because of the breeding design, all three genotype groups should
306 have the same average genetic background with respect to population hybridization and
307 heterozygosity (23 chromosome pairs (43), plus recombination when F1s reproduce).

308 Amino acid (AA) substitutions vary in their functional consequences for an MHC molecule's
309 antigen binding profile, such that alleles with different AA sequences may be functionally similar.
310 These MHC 'supertypes' are predicted to bind similar antigenic 'supermotifs', and may better
311 characterize the breadth of host defense than alleles (25-27). Using 15 guppy MHC codons
312 previously identified as being under positive selection (69), five physicochemical descriptors of each
313 AA (70), and discriminant analysis of principal components (71,72), we reduced the list of allele
314 sequences to 14 supertype clusters (full description in Appendix S1.3). Supertype designations were
315 used to assign fish into L/L, N/N and L/N 'supergenotypes' analogous to allele-based groupings,
316 except that supertypes shared between a pair of crossed populations (all pairs shared at least one
317 supertype; Appendix S5) were treated as 'local'. 'Supergenotypes' thus do not describe the
318 haplotype makeup of individuals – multi-allelic 'novel' haplotypes may contain one or more shared
319 supertypes. Only nine individuals, over three replicate populations, had a number of supertypes
320 lower than their number of alleles (i.e. they carried 2+ alleles of the same supertype). Of these, only
321 one changed novel/local categorization between allelic and supertype-based analyses (from L/N to
322 L/L). Novel and local alleles did not differ systematically in the supertype physicochemical parameter
323 space, and we thus concluded that the only 'special' property of novel variants was their novelty
324 (details in Appendix S6).

325 We used mitochondrial barcoding to identify 2-4 gyros/replicate to species level (full details
326 in Appendix S1.3). All sequences showed their strongest matches (98-100% identity) against
327 published *G. turnbulli* (*Gt*) sequences, except for those representing 35 fish from the AV/SS
328 population, all infected on the same day, which matched *G. bullatarudis* (*Gb*; 98-100%). These fish
329 also showed markedly different pathology, consistent with *Gb* (Appendix S1.3). Given the

taxonomical and pathological differences between *Gt* and *Gb* and the absence of indications of *Gb* elsewhere in the experiment, we excluded these fish from our analyses.

The analyses described in the next two paragraphs were performed separately for alleles and supertypes. We first tested if MHC group (L/L, L/N, N/N) predicted whether or not individual fish survived the experiment. We used corrected Akaike information criterion (AIC_C) ranking (73) of logistic regressions to explore all combinations of the following main effects: MHC group; number of MHC variants (alleles/supertypes; continuous, 1-4); standard length (continuous, z-transformed); age/sex (one variable: 'male' if in possession of a fully-shaped gonopodium, i.e. 'hook and hood' visible; 'female' if length > 13.0 mm and no gonopodium evident; 'juvenile' for all others); temperature (mean daily maximum over monitored period, z-transformed); and experimental population. We also included interactions between experimental population and both MHC group and number of MHC variants. We then examined models comprising the top two units of ranked AIC_C. For clarity of presentation in the main text, we only report the ΔAIC_C of the highest ranked model to include MHC group and the *P*-values (two-tailed) for certain contrasts, but in Appendix S2 we give a fully nuanced account of the analysis.

We used a similar process to test whether MHC group predicted infection burden among fish that survived the experiment. We used 'worm days' as the response variable, calculated as the total area under a fish's 17-day infection trajectory line. Worm days are tractable to analyze (no zero-inflation or random effects) and provide an ecologically relevant summary metric of an infection trajectory: fish that experience more worm days can reasonably be considered to have endured a greater parasite burden, and, in the wild, be more vulnerable to decreased condition and associated consequences (28-31). For this analysis, we used generalized linear models with negative binomial errors (log link function), and the same AIC_C approach and set of predictors used for analyzing probability of death. As a follow-up test of the relative importance of AA sequence novelty versus putative functional novelty, we used model ranking to test whether worm days differed significantly between fish carrying at least one novel supertype and fish carrying no novel supertypes

but at least one novel AA variant (the latter are novel alleles that belong to shared supertypes; Tables S2.3a-b; Fig. S2.1).

Twenty-eight fish from replicate population DR/SC were missing gyro counts for either day 7 or day 9 because of a 36 h power cut that rendered microscope work impossible. We dealt with this in the main analysis by dropping these same days from the worm-day calculation for all fish, rather than excluding the affected fish. However, N/N fish still experienced significantly fewer worm days than L/L fish when we used the full area data restricted to only complete cases (Appendix S7).

F_{IS} was significantly different from zero (0.216, SE = 0.060, P = 0.001) in HC/Guan, suggesting family effects may be distorting haplotype frequencies in this replicate population. Repeating the main analyses without this replicate did not result in a different interpretation (Appendix S8).

To visualize average infection trajectories for each MHC genotype/supergenotype, we analyzed the number of gyros per fish for each infection day separately, using AIC_C to find the highest ranked model to include MHC group as a predictor. From these models, we generated predicted daily worm loads for each MHC group that we then used A) to plot Fig. 2; B) to estimate the percentage reduction in number of worm days of N/N fish relative to L/L fish that was minimally affected by the incomplete cases; and C) as a *post-hoc* exploration of which MHC groups had the lowest/highest loads on which days, and which fish were the most likely to clear their infections (full details in Appendix S9).

Acknowledgements

We thank staff of the Environmental Research Institute Charlotteville ('ERIC'), Tobago, for support in the field, P Turpin for renting us the field station, A Pilastro for training in guppy artificial insemination, members of the Cardiff University parasite research group who helped with parasite work, J Lighten for sharing an as-then unpublished manuscript, and M Migalska for Fig. 1. We thank M Konczal, M Migalska, DS Richardson, J Kaufmann, LG Spurgin, W Babik, CA Morrison, J Lighten, W Smallbone, AR Ellison, M Hammers, MJG Gage, and three anonymous reviewers for comments on

earlier manuscript versions. This project was funded by a Polish National Science Centre Harmony grant to JRdwn, CvO and JC (grant number UMO-2013/10/M/NZ8/00253). JRdwn, CvO and JC conceived the study, with input from KPP, MHR and RSM on experimental design. RSM and KPP collected and reared the fish; MHR and JRdwn performed artificial inseminations; KPP performed infection trials, with assistance from JC, KJP and JRbc; JRbc performed molecular work, with assistance from MHR and KJP; KPP analyzed the data and performed computer simulations; JRdwn and KPP wrote the manuscript, with input from all co-authors. The authors declare no competing interests.

Data accessibility

All data and scripts will be made available in accordance with the publisher's requirements.

References

1. Woolhouse MEJ, Webster JP, Domingo E, Charlesworth B, Levin BR (2002) Biological and biomedical implications of the co-evolution of pathogens and their hosts. *Nature Genetics* 32:569-577.

2. Milinski M (2006) The major histocompatibility complex, sexual selection, and mate choice. *Annu Rev Ecol Evol Syst* 37:159-186.

3. Brockhurst MA, et al. (2014) Running with the Red Queen: the role of biotic conflicts in evolution. *Proc R Soc B* 281:20141382.

4. Ellison A, et al. (2012) Maintaining functional major histocompatibility complex diversity under inbreeding: the case of a selfing vertebrate. *Proc R Soc B* 279:5004-5013.

5. de Bellocq JG, Charbonnel N, Morand S (2008) Coevolutionary relationship between helminth diversity and MHC class II polymorphism in rodents. *J Evol Biol* 21:1144-1150.

6. Cano P, et al. (2007) Common and well-documented HLA alleles: report of the ad-hoc committee of the American Society for Histocompatiblity and Immunogenetics. *Hum Immunol* 68:392-417.

7. Apanius V, Penn D, Slev PR, Ruff LR, Potts WK (1997) The nature of selection on the major histocompatibility complex. *Crit Rev Immunol* 17:179-224.

8. Snell GD (1968) The H-2 locus of the mouse: observations and speculations concerning its comparative genetics and its polymorphism. *Folia Biol* 14:335-358.

9. Trowsdale J, Knight JC (2013) Major Histocompatibility Complex genomics and human disease. *Annual Review of Genomics and Human Genetics, Vol 14*, eds Chakravarti A & Green E, pp. 301-323.

10. Radwan J, Biedrzycka A, Babik W (2010) Does reduced MHC diversity decrease viability of vertebrate populations? *Biol Conserv* 143:537-544.

11. Sommer S (2005) The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Front Zool* 2:16.

12. Spurgin LG, Richardson DS (2010) How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proc R Soc B* 277:979-988.

13. Hughes AL, Nei M (1989) Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *PNAS* 86:958-962.

14. Garrigan D, Hedrick PW (2003) Perspective: detecting adaptive molecular polymorphism, lessons from the MHC. *Evolution* 57:1707-1722.

15. Borghans JAM, Beltman JB, De Boer RJ (2004) MHC polymorphism under host-pathogen coevolution. *Immunogenetics* 55:732-739.

16. Ejsmond MJ, Radwan J (2015) Red Queen processes drive positive selection on Major Histocompatibility Complex (MHC) genes. *PLoS Comput Biol* 11:e1004627.

17. Bodmer W (1972) Evolutionary significance of the HL-A system. *Nature* 237:139-145.

18. Kubinak JL, et al. (2013) Experimental viral evolution reveals major histocompatibility complex polymorphisms as the primary host factors controlling pathogen adaptation and virulence. *Genes Immun* 14:365-372.

19. Kubinak JL, Ruff JS, Hyzer CW, Slev PR, Potts WK (2012) Experimental viral evolution to specific host MHC genotypes reveals fitness and virulence trade-offs in alternative MHC types. *PNAS* 109:3422-3427.

20. Bakke TA, Cable J, Harris PD (2007) The biology of gyrodactylid monogeneans: the "Russian-doll killers". *Adv Parasitol* 64:161-376.

21. Fraser A, Neff B (2010) Parasite mediated homogenizing selectionat the MHC in guppies *Genetica* 138:273-278.

22. Cable J, van Oosterhout C (2007) The impact of parasites on the life history evolution of guppies (*Poecilia reticulata*): The effects of host size on parasite virulence. *Int J Parasitol* 37:1449-1458.

23. Cable J, Van Oosterhout C (2007) The role of innate and acquired resistance in two natural populations of guppies (*Poecilia reticulata*) infected with the ectoparasite *Gyrodactylus turnbulli*. *Biol J Linn Soc* 90:647-655.

24. Eizaguirre C, Lenz TL, Kalbe M, Milinski M (2012) Divergent selection on locally adapted major histocompatibility complex immune genes experimentally proven in the field. *Ecol Lett* 15:723-731.

25. Sidney J, Grey HM, Kubo RT, Sette A (1996) Practical, biochemical and evolutionary implications of the discovery of HLA class I supermotifs. *Immunol Today* 17:261-266.

26. Doytchinova IA, Flower DR (2005) *In silico* identification of supertypes for class II MHCs. *J Immunol* 174:7085-7095.

27. Schwensow N, Fietz J, Dausmann K, Sommer S (2007) Neutral versus adaptive genetic variation in parasite resistance: importance of major histocompatibility complex supertypes in a free-ranging primate. *Heredity* 99:265-277.

28. Hatcher MJ, Dick JTA, Dunn AM (2006) How parasites affect interactions between competitors and predators. *Ecol Lett* 9:1253-1271.

29. Kanno T, Nakai T, Muroga K (1989) Mode of transmission of vibriosis among ayu *Plecoglossus altivelis*. *J Aquat Anim Health* 1:2-6.

30. Van Oosterhout C, et al. (2007) Selection by parasites in spate conditions in wild Trinidadian guppies (*Poecilia reticulata*). *Int J Parasitol* 37:805-812.

31. Kennedy CEJ, Endler JA, Poynton SL, McMinn H (1987) Parasite load predicts mate choice in guppies. *Behav Ecol Sociobiol* 21:291-295.

32. Potts WK, Slev PR (1995) Pathogen-based models favoring MHC genetic diversity. *Immunol Rev* 143:181-197.

33. Figueroa F, Gunther E, Klein J (1988) MHC polymorphism pre-dating speciation. *Nature* 335:265-267.

34. Wegner KM, Eizaguirre C (2012) New(t)s and views from hybridizing MHC genes: introgression rather than trans-species polymorphism may shape allelic repertoires. *Mol Ecol* 21:779-781.

35. Penn DJ, Potts WK (1999) The evolution of mating preferences and major hstocompatibility complex genes. *Am Nat* 153: 145-164.

36. Winternitz JC, et al. (2013) Sexual selection explains more functional variation in the mammalian major histocompatibility complex than parasitism. *Proc R Soc B* 280: 20131605.

37. EjsmondMJ, Radwan J, Wilson AB (2014) Sexual selection and the evolutionary dynamics of the major histocompatibility complex. *Proc R Soc* B 281: 20141662

38. Takahata N, Nei M (1990) Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of Major Histocompatibility Complex loci. *Genetics* 124:967-978.

39. van Oosterhout C (2009) A new theory of MHC evolution: beyond selection on the immune genes. *Proc R Soc B* 276:657-665.

40. Penn DJ, Damjanovich K, Potts WK (2002) MHC heterozygosity confers a selective advantage against multiple-strain infections. *PNAS* 99:11260-11264.

41. Wakeland EK, et al. (1990) Ancestral polymorphisms of MHC class II genes: Divergent allele advantage. *Immunol Res* 9:115-122.

42. Lighten J, et al. (2017) Evolutionary genetics of immunological supertypes reveals two faces of the Red Queen. *Nat Commun* 8:1294

43. Lodi E (1978) Chromosome complement of the guppy, *Poecilia reticulata* Peters (Pisces, Osteichtyhyes). *Caryologia* 31:475-477.

44. Sato A, et al. (2000) Nonlinkage of major histocompatibility complex class I and class II loci in bony fishes. *Immunogenetics* 51:108-116.

45. Star B, et al. (2011) The genome sequence of Atlantic cod reveals a unique immune system. *Nature* 477:207-210.

46. Lugo-Villarino G, et al. (2010) Identification of dendritic antigen-presenting cells in the zebrafish. *PNAS* 107:15850-15855.

47. Braden LM, Barker DE, Koop BF, Jones SRM (2012) Comparative defense-associated responses in salmon skin elicited by the ectoparasite *Lepeophtheirus salmonis*. *Comp Biochem Physiol Part D: Genomics and Proteomics* 7:100-109.

48. Braden LM, Barker DE, Koop BF, Jones SRM (2015) Differential modulation of resistance biomarkers in skin of juvenile and mature pink salmon, *Oncorhynchus gorbuscha* by the salmon louse, *Lepeophtheirus salmonis*. *Fish Shellfish Immunol* 47:7-14.

49. Braden LM, Koop BF, Jones SRM (2015) Signatures of resistance to *Lepeophtheirus salmonis* include a TH2-type response at the louse-salmon interface. *Dev Comp Immunol* 48:178-191.

50. Bolnick DI, Stutz WE (2017) Frequency dependence limits divergent evolution by favouring rare immigrants over residents. *Nature* 546:285-288.

51. Stutz WE, Bolnick DI (2017) Natural selection on MHC IIβ in parapatric lake and stream stickleback: Balancing, divergent, both or neither? *Mol Ecol* :10.1111/mec.14158.

52. Eizaguirre C, Yeates SE, Lenz TL, Kalbe M, Milinski M (2009) MHC-based mate choice combines good genes and maintenance of MHC polymorphism. *Mol Ecol* 18:3316-3329.

53. Eizaguirre C, et al. (2011) Parasite diversity, patterns of MHC II variation and olfactory based mate choice in diverging three-spined stickleback ecotypes. *Evol Ecol* 25:605-622.

54. Kaufmann J, Lenz TL, Kalbe M, Milinski M, Eizaguirre C (2017) A field reciprocal transplant experiment reveals asymmetric costs of migration between lake and river ecotypes of three-spined sticklebacks (*Gasterosteus aculeatus*). *J Evol Biol* 30:938-950

55. Eizaguirre C, Lenz TL, Kalbe M, Milinski M (2012) Rapid and adaptive evolution of MHC genes under parasite selection in experimental vertebrate populations. *Nat Commun* 3:621.

56. Nadachowska-Brzyska K, Zieliński P, Radwan J, Babik W (2012). Interspecific hybridization increases MHC class II diversity in two sister species of newts. *Mol Ecol* 21: 887-906.

57. Decaestecker E, et al. (2007) Host-parasite 'Red Queen' dynamics archived in pond sediment. *Nature* 450:870-873.

58. Koskella B, Lively CM (2009) Evidence for negative frequency-dependent selection during experimental coevolution of a freshwater snail and a sterilizing trematode. *Evolution* 63:2213-2221.

59. Morran LT, Schmidt OG, Gelarden IA, Parrish RC, Lively CM (2011) Running with the Red Queen: Host-Parasite Coevolution Selects for Biparental Sex. *Science* 333:216-218.

60. Paterson S, et al. (2010) Antagonistic coevolution accelerates molecular evolution. *Nature* 464:275-278.

61. Little TJ (2002) The evolutionary significance of parasitism: do parasite-driven genetic dynamics occur *ex silico*? *J Evol Biol* 15:1-9.

62. Evans JP, Zane L, Francescato S, Pilastro A (2003) Directional postcopulatory sexual selection revealed by artificial insemination. *Nature* 421:360-363.

63. Barson NJ, Cable J, van Oosterhout C (2009) Population genetic analysis of microsatellite variation of guppies (*Poecilia reticulata*) in Trinidad and Tobago: evidence for a dynamic source-sink metapopulation structure, founder events and population bottlenecks. *J Evol Biol* 22:485-497.

64. Schelkle B, Shinn A, Peeler EJ, Cable J (2009) Treatment of gyrodactylid infections in fish. *Dis Aquat Organ* 86:65-75.

65. Stewart A, et al. (2017) Hook, line and infection: a guide to culturing parasites, establishing nfections and assessing immune responses in the three-spined stickleback. *Adv Parasitol*: DOI: 10.1016/bs.apar.2017.1007.1001.

66. Janeway CA, Travers P, Walport M, Shlomchik MJ (2005) *Immunobiology: The Immune System in Health and Disease* (Garland Science, New York) 6th Ed.

67. Biedrzycka A, Sebastian A, Migalska M, Westerdahl H, Radwan J (2017) Testing genotyping strategies for ultra-deep sequencing of a co-amplifying gene family: MHC class I in a passerine bird. *Mol Ecol Res* 17:642-655.

68. Sebastian A, Herdegen M, Migalska M, Radwan J (2016) amplisas: a web server for multilocus genotyping using next-generation amplicon sequencing data. *Mol Ecol Res* 16:498-510.

69. Lighten J, Van Oosterhout C, Paterson IG, McMullan M, Bentzen P (2014) Ultra-deep Illumina sequencing accurately identifies MHC class IIb alleles and provides evidence for copy number variation in the guppy (*Poecilia reticulata*). *Mol Ecol Res* 14:753-767.

70. Sandberg M, Eriksson L, Jonsson J, Sjöström M, Wold S (1998) New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids. *J Med Chem* 41:2481-2491.

71. Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24:1403-1405.

72. Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 27:3070-3071.

73. Bartoń K (2016) MuMIn: multi-model inference. R package v. 1.15.6. https://CRAN.R-project.org/package=MuMIn).

Figure Legends

Fig. 1. Schematic of the experiment. (A) Breeding design. Wild fish from two populations (P-generation) were crossed to produce F1s that were heterozygous across the genome with respect to population of origin. These were allowed to mate at random to produce F2s that segregated into heterozygotes and two types of homozygotes at the focal MHC class II locus, while having, on average, 'mixed' genetic backgrounds (23 chromosome pairs (40), plus crossing-over when F1s reproduce). (B) Controlled experimental infections. Two gyrodactylid worms from one of the P-generation source streams were inoculated on to the caudal fin of each F2 fish. Each infected fish was then kept in isolation and its infection monitored every other day for 17 days (22)

Fig. 2. Summary of Gyrodactylus turnbulli trajectory data for different MHC genotypes (i.e. allele-based; A) and supergenotypes (B). Points and lines apply only to fish that survived the experiment, and represent the mean infection intensity for each MHC genotype/supergenotype group on the focal day, controlling for other factors affecting infection intensity. Local/local = all MHC variants from the same stream as the worms; novel/novel = all variants from the 'other' stream; local/novel = mixed. Inset boxplots summarize the worm-day data as analyzed in Tables S2.1a,b (genotypes) and

S2.2a,b (supergenotypes). Pie charts show the proportion of fish of each MHC group that are dead (black), still infected (colored), or have cleared their infection (grey) at each infection day. In both graphs, the area under the novel/novel curve is significantly smaller than that for local/local, but death rate (black area of pies) does not differ significantly by genotype/supergenotype class. Error bars are 95% CIs of infection intensity on the focal day.