npg

ORIGINAL ARTICLE

Selection at the MHC class IIB locus across guppy (*Poecilia reticulata*) populations

BA Fraser[1], IW Ramnarine[2] and BD Neff[1]

[1]*Department of Biology, University of Western Ontario, London, Ontario, Canada and* [2]*Department of Life Sciences, University of the West Indies, St Augustine, Trinidad, West Indies*

The highly diverse genes of the major histocompatibility complex (MHC) are important in the adaptive immune system and are expected to be under selection from pathogens. Thus, the MHC genes provide an exceptional opportunity to investigate patterns of selection within and across populations. In this study, we analyzed genetic variation at the MHC class IIB gene and six microsatellite loci across 10 populations of guppies (*Poecilia reticulata*) in the northern range of Trinidad. We found a high level of diversity at the MHC, with a total of 43 alleles in 142 individuals. At the population level, we found that neutral evolution could not fully account for the variability found at the MHC. Instead, we found that MHC F_{ST} statistics were lower than F_{ST} derived from the microsatellite loci; 33 of 45 population pairwise estimates for the MHC were significantly lower than those for the microsatellite loci, and MHC F_{ST} estimates were consistently lower than those predicted by a coalescent model of neutral evolution. These results suggest a similar selection acting across populations, and we discuss the potential roles of directional and balancing selection. At the sequence level, we found evidence for both positive and purifying selection. Furthermore, positive selection was detected within and adjacent to the putative peptide-binding region (PBR) of the MHC. Surprisingly, we also found a purifying selection at two sites within the putative PBR. Overall, our data provide evidence for selection for functional diversity at the MHC class IIB gene at both the population and nucleotide levels of guppy populations.
Heredity (2010) **104,** 155–167; doi:10.1038/hdy.2009.99; published online 29 July 2009

Introduction

Identifying the processes that maintain genetic diversity within and among populations is a central goal of modern evolutionary genetics (Richman, 2000). Although the neutral theory of molecular evolution predicts that most genetic variation is selectively neutral, this theory cannot explain the vast amount of diversity found in functionally important traits (Kimura, 1983). Advances in molecular biology now allow researchers to directly examine selection at genes that underlie functional traits. One important example comes from the genes of the major histocompatibility complex (MHC). These genes are found in all jawed vertebrates and play a critical role in an organism's immune response (Klein, 1990). The MHC genes are among the most polymorphic functional genes in vertebrate genomes and can therefore provide a strong test of theories of adaptive variation. Several non-neutral mechanisms are thought to maintain genetic diversity at the MHC (Bernatchez and Landry, 2003; Piertney and Oliver, 2006). These mechanisms include balancing selection involving either overdominance, in which case MHC heterozygous individuals have increased survivorship because they are able to

present a broader array of antigens and thereby resist a broader array of pathogens (Doherty and Zinkernagel, 1975; Hughes and Nei, 1988; for example, Oliver *et al.*, 2009), or negative frequency-dependent selection, in which case a coevolutionary arms race between specific MHC alleles and parasites can lead to the cycling in frequency of alleles (Clarke and Kirby, 1966; for example, Paterson *et al.*, 1998). In addition, across the landscape, pathogen communities can vary spatially or temporally, which can lead to local adaptations in a host–parasite arms race (for example, Hill, 1991; Westerdahl *et al.*, 2004). In addition, other non-adaptive mechanisms can maintain diversity at the MHC including selection against mutational load thought to accumulate near MHC loci (van Oosterhout, 2009).

The MHC proteins act as the interface between pathogens and the host's adaptive immune system by coding for proteins that bind pathogen-derived peptides and presenting them to T-cell receptors, which then initiate an immune response (Klein, 1990). The MHC class I proteins present peptides derived from the cell cytoplasm to the T cells, whereas the class II proteins present exogenously derived peptides to T cells (Klein, 1990). For MHC class II proteins, peptide recognition occurs in the peptide-binding groove, which is a structure consisting of two α-helices that border a β-sheet and is formed by a heterodimer consisting of an α and β chain (Brown *et al.*, 1993). Both the α and β chains have specific sites that recognize the foreign peptide known as the peptide-binding region (PBR;

Correspondence: *Dr BA Fraser, Department of Biology, University of Western Ontario, 1151 Richmond Street, London, Ontario, Canada N6A 5B7.*
E-mail: *fraser.bonnie8@gmail.com*

Klein, 1990). The PBR of the MHC genes are typically highly polymorphic in populations, which affects the peptide-binding specificity of the MHC protein, and therefore individuals can differ in their level of resistance to specific pathogens (Klein, 1990). Thus, genetic variation at the PBR should be under strong selection mediated by the pathogen community a population encounters.

Local selection at the MHC can be inferred from population differentiation statistics. Homogeneity in selection pressures can lead to low population differentiation, whereas heterogeneity in selection pressures can lead to high genetic differentiation across populations. Many studies have attempted to detect selection at the MHC by comparing population differentiation at the MHC with that observed at neutral loci. For example, Sommer (2003) found that population differentiation at the MHC was lower than that observed at neutral loci in two populations of Malagasy giant jumping rats (*Hypogeomys antimena*) and inferred that balancing selection was occurring at the MHC. Balancing selection is thought to make populations more similar by selecting for rare migrants, and therefore increasing the effective migration rate (Schierup *et al.*, 2000; Muirhead, 2001). Conversely, Landry and Bernatchez (2001) found populations of Atlantic salmon (*Salmo salar*) occurring in different types of habitats had higher than expected population divergence at the MHC when compared with microsatellite loci, and concluded that varying selection pressure from the different habitat types was occurring at the MHC. Yet other studies have found no difference in estimates of population structure between MHC and neutral loci, which suggests minimal selection at the MHC (see Muirhead, 2001). Comparing population differentiation estimates is therefore a useful approach to investigate selection at the MHC across populations.

In addition, selection can be examined at the sequence level. In this case, evidence for different types of selection is determined by comparing synonymous (silent substitutions, referred to as 'dS') with nonsynonymous (amino acid changing, referred to as 'dN') substitution rates in protein-coding gene sequences (Yang and Bielawski, 2000). A dN/dS ratio close to 0 indicates purifying selection (nonsynonymous substitutions occur at a lower rate than synonymous substitutions), a dN/dS close to 1 indicates neutral evolution (synonymous and nonsynonymous substitutions occur at equal rates) and a dN/dS ratio above 1 indicates positive selection (nonsynonymous substitutions occur at a higher rate than synonymous substitutions). Selection at the sequence level of the MHC is well documented. The genes of the MHC are well known to have dN/dS ratios greater than 1, providing evidence for positive selection (Bernatchez and Landry, 2003; for example, Landry and Bernatchez, 2001; Dionne *et al.*, 2007). In addition, the PBR of the MHC has a higher dN/dS ratio than non-PBR sites in most studies in which it has been examined (Bernatchez and Landry, 2003; for example, Hughes and Yeager, 1998). Less attention has been given to selection at individual sites within these domains. However, recent advances in theory now make it feasible to detect differing dN/dS ratios across individual codons (Yang and Bielawski, 2000). These applications allow a more refined analysis to detect differing types of selection

across the sequence itself (for example, Consuegra *et al.*, 2005; Schaschl *et al.*, 2005; Blais *et al.*, 2007).

In this study, we determine whether selection is acting at the MHC class IIB gene in 10 wild guppy (*Poecilia reticulata*) populations in northern Trinidad. The guppy is a tropical freshwater fish native to Trinidad and neighboring portions of north-eastern South America (Magurran, 2006). Guppies in northern Trinidad are used as a model system in ecology and evolution because their rivers are sufficiently isolated such that they can be regarded as independent replicates in a natural selection experiment (Magurran, 2006). High population differentiation has been supported by a variety of population genetic studies using microsatellites (Barson *et al.*, 2009; Suk and Neff, 2009), mitochondrial DNA (Fajen and Breden, 1992; Alexander *et al.*, 2006) and allozymes (Carvalho *et al.*, 1991). Within each river, upper and lower populations are separated by physical barriers, such as waterfalls that substantially reduce gene flow, and create distinct genetic clusters (Crispo *et al.*, 2006). The two largest drainages in northern Trinidad, the Caroni and Oropouche, likely diverged some 2.5 million years ago (Magurran, 2006). However, the drainages are in close proximity and are likely to have similar pathogen communities. Specifically, the ectoparasitic monogeneans, *Gyrodactylus turnbulli* and *G. bullatarudis*, are known to cause detrimental effects to guppies (Cable and van Oosterhout, 2007a, b) and are widespread across northern Trinidad (Lyles, 1990; Martin and Johnsen, 2007). In addition, high selection coefficients associated with gyrodactylus infection have been estimated in wild populations by van Oosterhout *et al.* (2007). Furthermore, Hedrick *et al.* (2001) have shown that MHC genotype may be associated with reduced gyrodactylus infections in the closely related Gila topminnow (*Poeciliopsis o. occidentalis*). Thus, gyrodactylus could be an important selective agent on the MHC in guppies, and the guppy populations of northern Trinidad should provide an ideal system to test pathogen-induced selection at the MHC.

In guppies, the MHC class II has undergone at least one duplication event (McConnell *et al.*, 1998; van Oosterhout *et al.*, 2006a, b), which increases the possibility of high diversity. Recently, van Oosterhout and colleagues (2006b) surveyed two populations of guppies from one river in northern Trinidad and found that the populations were more similar in their MHC class IIB alleles than predicted by neutrality. Using a computer simulation to detect the selection pressure needed to maintain the observed frequencies of MHC alleles, with demography parameters estimated by an isolation-with-migration model (Hey and Nielson, 2004) from microsatellite loci data, they showed that the observed allele frequencies at the MHC were most likely maintained by overdominant selection. van Oosterhout (2009) recently reanalyzed these data and argued that MHC variation may instead be maintained by selection against mutational load thought to surround MHC loci. In this study, we expand on this work by surveying guppies from across a wider distribution in the northern range of Trinidad. Specifically, we examine the genetic diversity of the MHC class IIB exon across 10 populations that encompass five rivers and three drainages. Thus, our study considers MHC diversity at a much broader landscape and temporal scale, including populations

that diverged some 2.5 million years ago. We compare population differentiation at the MHC to differentiation at six microsatellite loci. We also examine selection at the sequence level and use these data to infer the modes of selection both within and between populations. If similar selection across populations is indeed acting at the MHC across guppy populations (sensu van Oosterhout et al., 2006b), then we predict that populations across rivers and drainages should also be less differentiating at the MHC than neutral markers. Our study provides a particularly strong test for such selection because we compare populations across the highly diverged drainages of northern Trinidad.

Materials and methods

Population samples
During May 2006, 10 populations of guppies from across northern Trinidad were sampled (Figure 1 and Table 1). The populations comprised upstream and downstream locations in the Aripo and Guanapo rivers in the Caroni drainage, the Quare and Turure rivers in the Oropouche drainage and the Yarra river on the northern coast. Fish were collected from each population using seine and dip nets. Fish were transported back to the lab at the University of the West Indies, St Augustine, where body tissue samples were preserved in 95% ethanol for genetic analysis. Of note is that about 50 years ago, guppies from the lower Guanapo were transferred to the upper Turure population as part of an introduction experiment (Haskins and Haskins, 1954). These two populations remain genetically similar as measured by microsatellite loci (Suk and Neff, 2009).

PCR and cloning
A total of 11 to 17 adults from each population were sequenced at the MHC class IIB (exon 2). This exon makes up half the peptide-binding groove and encodes most of the polymorphic PBR (Brown et al., 1993). First, DNA was extracted using a Wizard Genomic DNA purification kit (Promega, Madison, WI, USA). The 230-bp exon was amplified with primers published in van Oosterhout et al. (2006b) using the following PCR reaction mixture: $1 \times$ PCR buffer (Invitrogen Life Technologies, San Diego, CA, USA) 2.5 mM $MgCl_2$, 0.25 mM each dNTPs, 10 μM of both the sense and antisense primers (Invitrogen Life Technologies) and 5 U of Taq DNA polymerase (Invitrogen Life Technologies) in a total volume of 50 μl. The PCR conditions were 94 °C for 1 min, followed by 35 cycles of 30 s at 92 °C, 30 s at 57 °C, 30 s at 72 °C and a final extension for 7 min at 72 °C.

Briefly, samples were first analyzed using single strand conformation polymorphism (Amersham Biosciences, Fairfield, CT, USA). At least one sample that corresponded to each unique banding pattern was re-amplified and the PCR product was inserted into a plasmid vector (Promega) following the manufacturer's instructions. The vector was then transformed into DH5α competent bacteria (Invitrogen Life Technologies) and grown on LB agar plates with ampicillin (Sigma-Aldrich, St Louis, MO, USA). Up to 10 clones were sequenced for each individual and a total of 691 clones were analyzed. The probability that an allele was missed in an individual based on 10 clones is only 5.6% when the

allele is present in a single copy, 0.1% when present in two copies and less than 0.0001% when present in three copies. Furthermore, the cloning effort was the same per population ($F_{9,142} = 1.37$, $P = 0.20$), and therefore any potentially missed alleles due to insufficient cloning should not bias the results from population comparisons. The corresponding chromatograms were read in BioEdit Sequence Alignment Editor (ver. 7.0.5.3, Carlsbad, CA, USA) and sequences were aligned using MEGA (ver. 4.0.1; Tamura et al., 2007). Following Lukas and Vigilant (2005), alleles were identified as those that occurred in at least two independent PCR reactions; this approach was used to avoid overestimation of diversity due to Taq polymerase errors, sequencing errors and heteroduplex mismatch. Finally, to test whether or not the primers we used preferentially amplified certain alleles, a second set of primers were designed that were located within the 230-bp exon. A subset of individuals ($n = 15$) was evaluated a second time with the internal primers, and genotypes were compared. In addition, 19–39 individuals from each population were genotyped at six microsatellite loci as part of a larger study of population differentiation in guppies (Suk and Neff, 2009); some of these individuals were not the same as those used for the MHC genotyping. We used three dinucleotide microsatellite loci (Pr39, Pr92 and Pr171; Becher et al., 2002) and three tetranucleotide microsatellite loci (Pre8, Pre9 and Pre15; Paterson et al., 2005). In the lower Guanapo and upper Yarra populations, the Pr171 locus could not be amplified.

Sequence analysis
The MHC class IIB reading frame was determined by aligning our sequences with the reading frame identified for the guppy in van Oosterhout et al. (2006b) and three species of African cichlids (Haplochromis lividus, Oreochromis alcalicus and Aulonocara hansbaenschi) (Ono et al., 1993; Figueroa et al., 2000). The amplified 230 bp were subsequently trimmed to 218 bp that constituted an open reading frame. MEGA was then used to calculate pairwise nucleotide distances between alleles (using a Jukes–Cantor correction) and to construct a neighbor-joining tree (Saitou and Nei, 1987). The correction accounts for multiple substitutions and is particularly important in the analyses of highly variable sequences (Jukes and Cantor, 1969). The putative PBR was then identified by aligning our observed amino acid sequences with those of the HLA-DR1 (human MHC class II) for which the PBR has been determined through X-ray crystallography (Brown et al., 1993).

To examine selection across codons, we used the random sites codon-model-based approach in PAML ver. 4 (Yang, 2007). In these models, codons can vary according to estimated rates of synonymous to non-synonymous substitutions (dN/dS ratios, ω). In this study, we evaluated four codon models and the likelihood of each model was determined using a nested log-likelihood ratio test. The first model (M1a in Yang, 2007) is a null model and is referred to as the 'nearly neutral' model, and it estimates a proportion of codons, p_0, that are undergoing purifying selection ($0 < \omega_0 < 1$) and a proportion of codons p_1 ($= 1-p_0$) that are undergoing neutral evolution ($\omega_1 = 1$). The second model
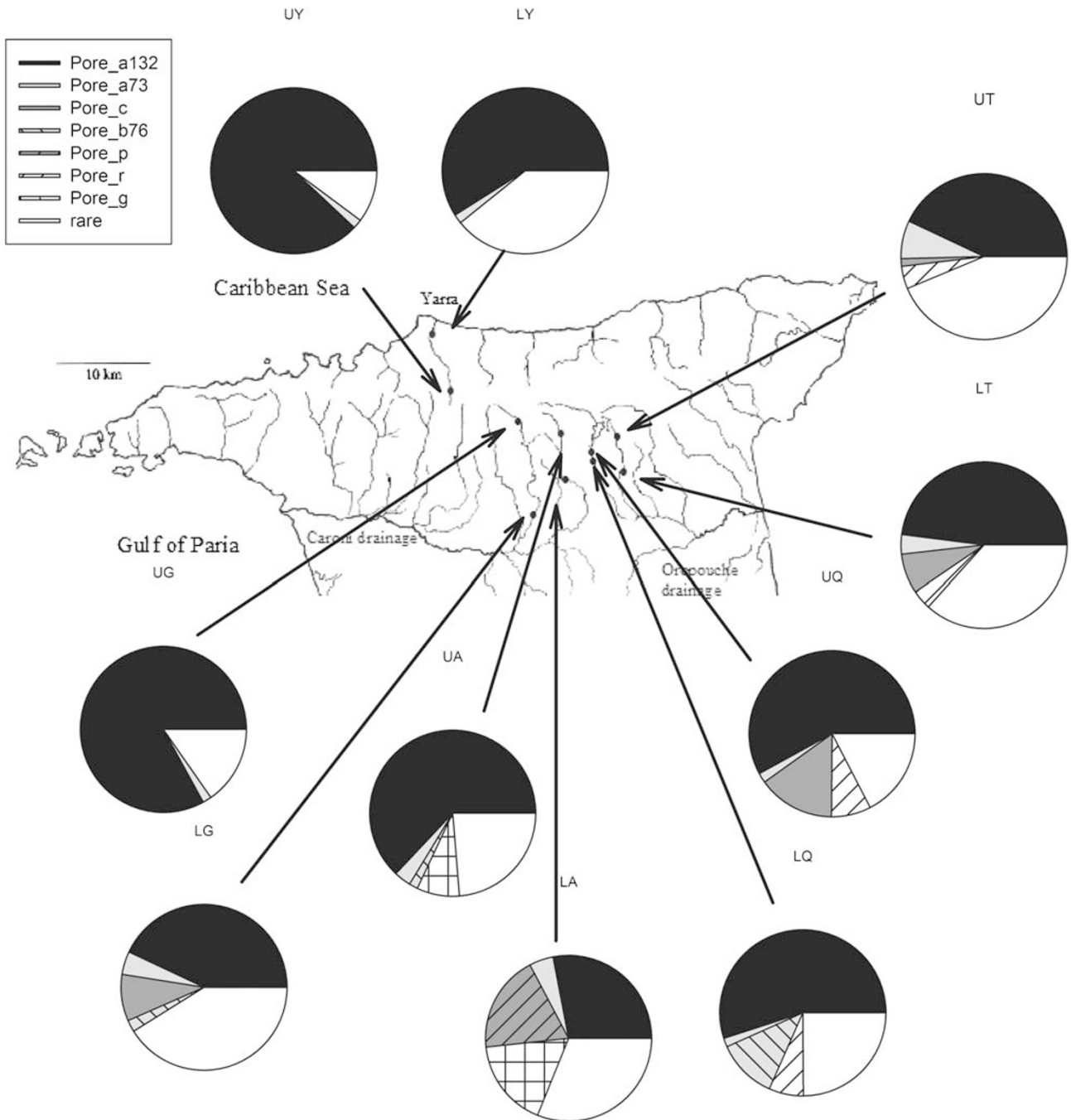
Figure 1 Location and MHC class IIB allele frequencies of 10 guppy (*Poecilia reticulata*) sampling sites in northern Trinidad. Populations comprise, from the north slope: Lower Yarra (LY) and Upper Yarra (UY); from the Oropouche drainage: Lower Quare (LQ), Upper Quare (UQ), Lower Turure (LT) and Upper Turure (UT); and from the Caroni drainage: Lower Guanapo (LG), Upper Guanapo (UG), Lower Aripo (LA) and Upper Aripo (UA). Frequencies are shown for the seven most common alleles, whereas the remaining alleles were collapsed to a rare category for each population.

(M2a) is referred to as the 'positive selection' model and it is equivalent to M1a but includes a third class of codons defined by the proportion p_2 $(=1-p_0-p_1)$ and $\omega_2 > 1$. Thus, this model allows for some codons to be undergoing positive selection. The final two models used a less restrictive definition for ω values between 0 and 1 by implementing a β distribution. The beta distribution is a flexible probability density function that is estimated directly from the data and it can capture more of the variation in ω across sites (for further discussion see Yang *et al.*, 2000). Our third model (M7) served as a null β model, with $0 < \omega_0 < 1$ modeled according to a β distribution defined by the shape parameters p and q. The fourth model (M8) is referred to the 'positive selection plus β' model and it is equivalent to M7 but includes a proportion of codons p_1, undergoing positive selection ($\omega_1 > 1$). Assigning codons to different selection classes was done using the Bayes Empirical Bayes approach and a 95% posterior probability cutoff (Yang *et al.*, 2005). For

Table 1 Summary of the MHC class IIB locus and six microsatellite loci sampled in 10 guppy (*Poecilia reticulata*) populations

	LA	UA	LG	UG	LQ	UQ	LT	UT	LY	UY
Location	PS 938 786	PS 931 817	PS 909 744	PS 893 848	PS 971 805	PS 970 806	QS 002 784	QS 003 809	PS 807 948	PS 833 876
MHC										
N	16	15	11	14	15	13	13	17	14	14
K	14	12	13	4	6	8	9	15	9	5
Alleles	1.8	1.6	1.8	1.2	1.4	1.5	1.4	1.8	1.5	1.2
H_E	0.85	0.59	0.78	0.29	0.66	0.64	0.74	0.80	0.63	0.23
Ho	0.56	0.53	0.64	0.14	0.33	0.54	0.38	0.47	0.43	0.21
Π_{JC}	0.14	0.082	0.12	0.002	0.069	0.12	0.12	0.13	0.07	0.05
Microsatellites loci										
Mean across all loci										
N	21	19	19	19	38	39	21	19	20	19
K	9.5	5.0	9.6	2.0	11.3	7.8	9.8	7.5	7.5	2.8
H_E	0.78	0.46	0.85	0.15	0.79	0.70	0.78	0.78	0.69	0.35
Pre9										
K	12	7	10	1	11	9	11	11	9	4
H_E	0.91	0.72	0.86	0	0.85	0.79	0.84	0.83	0.85	0.50
Pr39										
K	6	2	6	1	7	6	6	5	4	2
H_E	0.76	0.31	0.82	0	0.66	0.61	0.77	0.75	0.51	0.15
Pre8										
K	12	9	11	2	19	9	17	11	12	3
H_E	0.89	0.80	0.88	0.15	0.92	0.71	0.91	0.88	0.90	0.48
Pr92										
K	4	1	6	1	6	4	3	3	3	1
H_E	0.59	0	0.77	0	0.71	0.57	0.42	0.53	0.66	0
Pre15										
K	19	9	15	6	17	13	13	9	15	4
H_E	0.95	0.82	0.94	0.77	0.93	0.85	0.91	0.87	0.93	0.62
Pr171										
K	4	2	—	1	8	6	9	6	2	—
H_E	0.56	0.10	—	0	0.68	0.67	0.85	0.82	0.30	—

For each population, data include map coordinates, number of individuals (N), number of MHC class IIB or average number of microsatellite alleles (K), expected heterozygosity (H_E), and for the MHC only, average number of alleles per individual, observed heterozygosity (H_O), and nucleotide diversity following a Jukes–Cantor correction (π_{JC}). Number of alleles and expected heterozygosities are also presented for each microsatellite loci separately. The populations are lower Aripo (LA), upper Aripo (UA), lower Guanapo (LG), upper Guanapo (UG), lower Quare (LQ), upper Quare (UQ), lower Turure (LT), upper Turure (UT), lower Yarra (LY) and upper Yarra (UY).

example, sites that are undergoing positive selection were identified as those that had a posterior probability of greater than 95% belonging to the positive selection class.

Population analysis

From clone data, genotypes were determined probabilistically, choosing the most likely genotype using multinomial probabilities. Population diversity measures, comprising expected heterozygosity, H_E, nucleotide diversity using a Jukes–Cantor correction and number of alleles per population were calculated using SPAGeDi ver 1.2 (Hardy and Vekemans, 2002) or DnaSP ver 4.0 (Rozas *et al.*, 2003). The population differentiation measure F_{ST} was determined using a polysomic polyploids model (Ronfort *et al.*, 1998), which reflects a situation similar to unlinked duplicated genes as in the MHC (McConnell *et al.*, 1998), that is, when alleles cannot be assigned to loci and all possible pairwise combinations are assumed to be equally likely to be inherited (Hardy and Vekemans, 2002). Evidence for unlinked MHC class IIB genes in the guppy comes from aligning MHC sequences found in the guppy to *Xiphophorus* species with recombination information (McConnell *et al.*, 1998). There is yet no direct evidence for unlinked MHC class IIB alleles in the guppy, and thus the F_{ST}

measure should be interpreted with some caution. Consequently, we also compared our results with the more general population differentiation measure G_{ST} (Pons and Petit, 1995). The G_{ST} estimator does not make any assumption about linkage of the genes. A Mantel test revealed that these two measures were highly correlated ($r^2 = 0.98$, $n = 45$, $P = 0.005$) and we therefore focus most of our results on the more commonly presented F_{ST} estimator. F_{ST} measures for microsatellites were calculated using Weir and Cockerham (1984). Mantel correlations were conducted in FSTAT 2.9.3 (Goudet, 2001).

Isolation by distance was examined by correlating distance and genetic differentiation ($1/(1 - F_{ST})$) using SPAGeDi. Two distances were considered, the direct geographical distance and the geographical distance along the length of the stream (stream + coastal contour; see Suk and Neff, 2009) using Mantel tests. Statistical significance was determined by 20 000 permutations of the columns and rows in the distance matrix. An analysis of molecular variance was used to partition molecular variation to within populations, among populations within rivers, and among rivers or among drainages using Arlequin 2.0 (Schneider *et al.*, 2000). The analysis employed a nucleotide distance matrix using a Jukes–Cantor correction and significance was determined by 1000 permutations of haplotypes.

We compared F_{ST} of MHC class IIB to previous estimates of F_{ST} from six microsatellites loci (Suk and Neff, 2009) using both empirical and model-based approaches. The empirical approach generated 95% confidence intervals (CI) for MHC, and microsatellite loci F_{ST} estimates through resampling individuals in populations with replacement for a total of 1000 runs (BD Neff and BA Fraser, unpublished). Significance was determined by comparing the pairwise values for microsatellites and MHC in each randomization run, and determining the overall proportion in which the MHC estimate was lower than the microsatellite estimate.
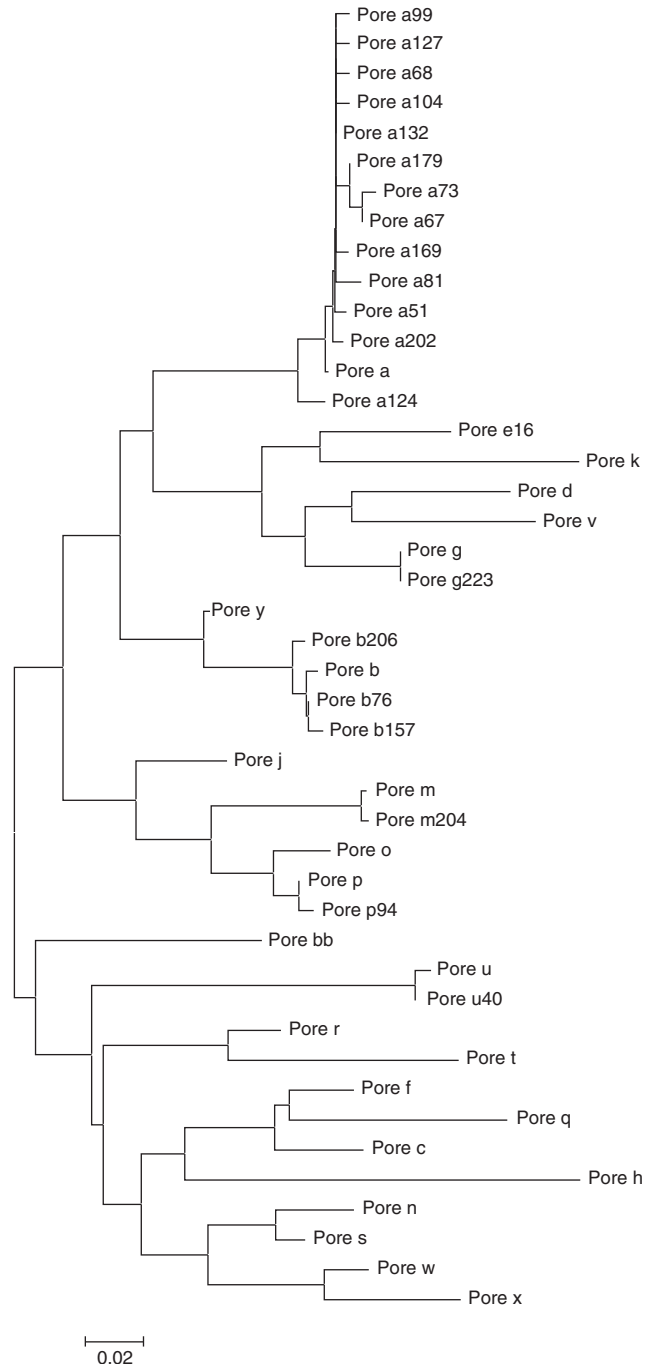
The model-based approach followed the methods outlined in Beaumont and Nichols (1996) as implemented in fdist2. This method builds an expected neutral distribution of F_{ST} versus heterozygosity based on coalescent simulations. Using this distribution, outliers are determined as the empirically estimated F_{ST} and H_E values that fall outside of the 95% CI. To simulate the F_{ST} and H_E distribution for our data, we ran 50 000 simulations with parameters set suggested by Beaumont and Nichols (1996). Following Beaumont and Nichols (1996) and Beaumont and Balding (2004), loci with a heterozygosity of <0.01 were omitted and the starting F_{ST} was then calculated as the average of all remaining F_{ST} values. We used the infinite alleles model for the mode of mutation because it is more conservative than the stepwise mutation model in this analysis. In the stepwise mutation model, F_{ST} decreases more rapidly with higher heterozygosity and therefore the model is more likely to erroneously identify outliers. Simulations were run for each drainage separately because we could not fit an expected distribution to the majority of alleles when all populations were analyzed together most likely due to the large range of F_{ST} values observed.

Results

Sequence analysis

We found 43 different MHC class IIB alleles in 142 individuals across the 10 populations (Figure 2). Only four of the alleles were found to be identical to other guppy MHC class IIB sequences published in Genbank (van Oosterhout *et al.*, 2006b; Supplementary Figure 1). No indels or stop codons were found in our sequences, which suggests that it is unlikely that any of them represent pseudogenes. Of the 218 coding sites examined, 117 (54%) were polymorphic and the overall mean pairwise nucleotide difference was 0.183. The 43 alleles gave rise to 39 different amino acid sequences. Of the 72 codons identified in the open reading frame, 50 (69%) were variable (Supplementary Figure A1). The 15 individuals that were analyzed with the internal primers showed the same genotypes as the original analysis with the published primers.

Examining selection on the MHC, we found that the positive selection model (M2a) was significantly more likely than the nearly neutral model (M1a) (LnLRT = 42.4, $P < 0.001$; Table 2), which indicates that

Figure 2 Neighbor-joining tree of MHC class IIB alleles found in the guppy (*Poecilia reticulata*). The scale refers to base pair differences among alleles and incorporates the Jukes–Cantor correction.

Table 2 Summary of models that were used to detect selection across codons of the MHC class IIB gene in the guppy (*Poecilia reticulata*)

Model	Parameters	LnL	Parameter estimates
M1a nearly neutral	90	−2220.15	$\omega_0 = 0.13$, $p_0 = 0.61$ $\omega_1 = 1$, $p_1 = 0.39$
M2a positive selection	92	−2198.93	$\omega_0 = 0.11$, $p_0 = 0.51$ $\omega_1 = 1$, $p_1 = 0.35$ $\omega_2 = 3.17$, $p_2 = 0.14$
M7 null+β	90	−2219.14	$p = 0.30$ $q = 0.39$
M8 positive selection+β	92	−2197.56	$p_0 = 0.83$, $p = 0.40$, $q = 0.68$ $\omega_1 = 2.62$, $p_1 = 0.17$

Data are from four models and comprise number of parameters estimated in the model, the logarithm of the likelihood (LnL) and parameter estimates.

positive selection has occurred at the MHC class IIB. On the basis of the p_2 parameter from the M2a model, 14% of codons carried the positive selection signature, whereas the remaining 86% of codons carried signatures of purifying selection or neutral evolution. Moreover, the estimated ω for the positive selection class was much higher than 1 at $\omega_2 = 3.17$ (Table 2). The positive selection plus β distribution model (M8) was also significantly more likely than the null plus β distribution model (M7) (LnLRT = 43.2, $P < 0.001$; Table 2). Similar to the previous model, 17% of the codons carried the positive selection signature and the ω associated with positive selection was much higher than 1 ($\omega_2 = 2.62$). For the remainder of

the codons, the M8 model generated p and q estimates that indicated a slight U-shaped probability distribution for ω between 0 and 1, which suggests that purifying or neutral selection is marginally more likely at these codons.

For both the M2a and M8 models, the Bayes Empirical Bayes approach indicated that five codons were under positive selection (numbers 15, 17, 24, 54 and 59, with posterior probabilities of 1.00, 1.00, 0.98, 1.00 and 0.98, respectively). Of these five sites, three are in the putative PBR and the other two are immediately adjacent to the PBR (Figure 3a). For the M8, the PBR had a high weighted mean ω of 1.32, whereas the non-PBR
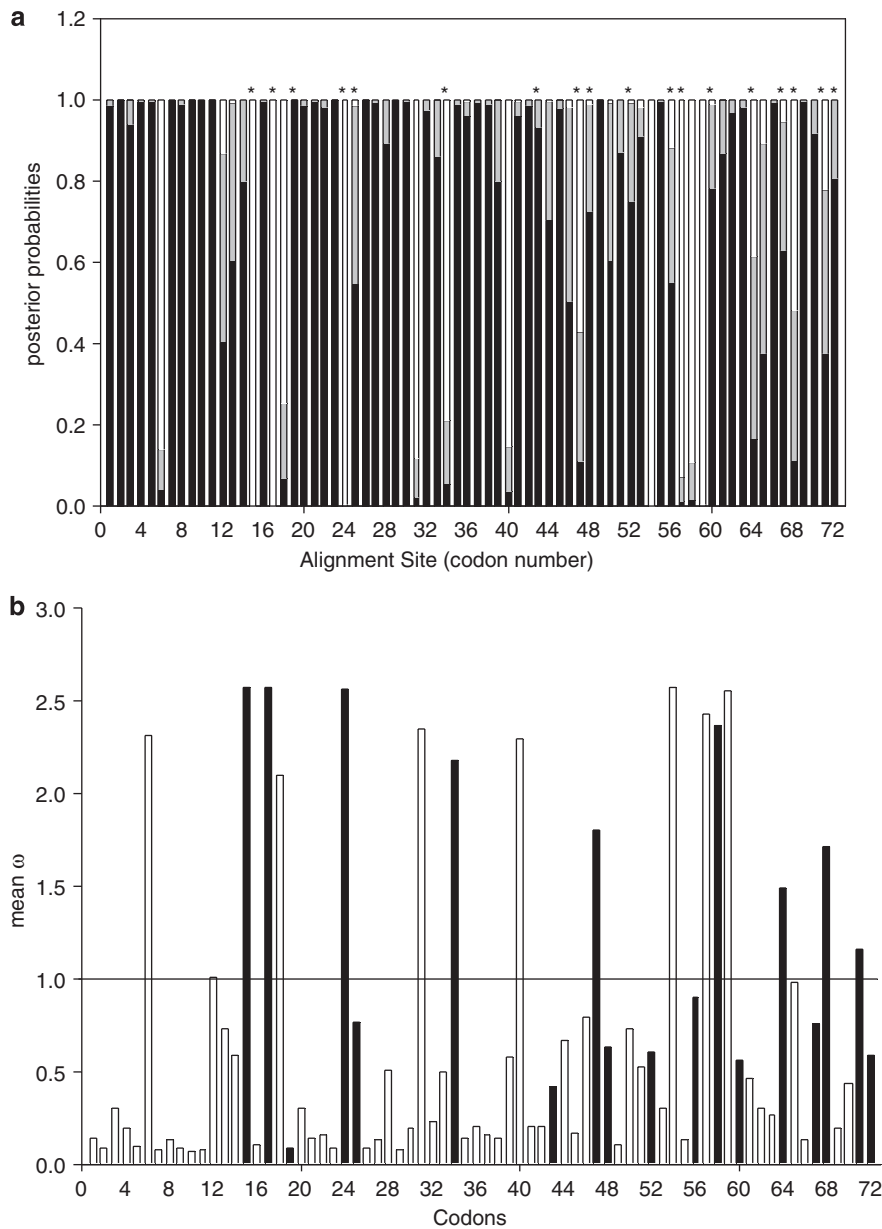

Figure 3 Selection occurring at the MHC class IIB at the sequence level in the guppy (*Poecilia reticulata*). (**a**) Codons of the MHC class IIB allele and their posterior probability of falling into three selection classes: black, purifying selection ($\omega = 0$–0.67); gray, neutral selection ($\omega = 0.83$–0.96) and white, positive selection ($\omega = 2.62$). Asterisks indicate putative peptide-binding regions. (**b**) Weighted mean ω for each codon across the MHC class IIB allele. Codons in the putative peptide-binding region are indicated in black, whereas those outside the peptide-binding region are in white.

had a lower weighted mean of 0.57 (Figure 3b). Using a randomization of weighted means and sites (non-PBR versus PBR), we found that the PBR sites had a significantly higher ω than the non-PBR sites ($P < 0.001$).

Population analysis

Populations differed in their MHC class IIB diversity (Table 1). Mean number of MHC class II alleles was 9.5 and ranged from 4 to 15, mean expected heterozygosity was 0.62 and ranged from 0.23 to 0.85, and mean nucleotide diversity was 0.09 and ranged from 0.002 to 0.14. The lower Aripo and upper Turure populations were the most diverse, were found to have 14 and 15 different MHC class IIB alleles, have nucleotide diversity indices of 0.14 and 0.13, respectively, and similar expected heterozygosity measures (0.85 and 0.80, respectively). The upper Guanapo population was the least diverse, although a similar number of alleles and expected heterozygosity was found in the upper Yarra population. The upper Guanapo population had much lower nucleotide differences ($\pi_{JC} = 0.002$) than the upper Yarra ($\pi_{JC} = 0.05$; Table 1), which indicates that the alleles found in the upper Guanapo were similar in sequence. The average observed heterozygosity was 0.42 and ranged from 0.64 to 0.14. These values were consistently lower than the expected heterozygosity in each population. The average number of MHC alleles found in an individual was 1.5 and ranged from 1.2 to 1.8. Expected heterozygosity was higher in MHC than the average expected heterozygosity of the microsatellites in all populations except for the upper Aripo and upper Quare (Table 1). The number of MHC alleles was also higher than the average number of microsatellite alleles in all populations except the lower and upper Quare, but the average expected heterozygosity measures were similar (Table 1). Expected heterozygosities of MHC and microsatellites were highly correlated ($r^2 = 0.91$, $n = 10$, $P = 0.0003$), but the number of alleles was not correlated ($r^2 = 0.17$, $n = 10$, $P = 0.24$).

Although we found that the MHC class IIB is diverse in guppies, a single allele (Pore_a132) dominated in most populations with a frequency between 28 and 83% (Figure 1, Supplementary Table A1). The next most common allele within any one population was 19% and only one other allele, Pore_a73, occurred in all populations. In total 19 alleles occurred in only one population, including the Pore_p allele that was found at a high frequency (19%) in the lower Aripo. There was no strong clustering of alleles within rivers or drainages; six alleles were unique to a river, including Pore_g that was found at relatively high frequencies (8%) in the Aripo river and five alleles were unique to a drainage, including Pore_r, which had a frequency of 8% in the Oropouche drainage (Figure 1, Supplementary Table A1).

The populations were significantly differentiated at the MHC class IIB (global $F_{ST} = 0.0489$, $P = 0.0004$; Table 3). There was no relationship between MHC F_{ST} and either the direct distance ($r^2 < 0.01$, $n = 45$, $P = 0.83$; Figure 4) or the stream-coast distance ($r^2 < 0.01$, $n = 45$, $P = 0.95$). Similarly, no relationship was found between G_{ST} and either direct or stream-coast distance ($r^2 < 0.01$, $n = 45$, $P = 0.89$; $r^2 < 0.01$, $n = 45$, $P = 0.98$, respectively). The analysis of molecular variance analysis indicated that

Table 3 Summary of genetic differentiation parameters for pairwise comparisons of 10 guppy (*Poecilia reticulata*) populations

| | Caroni | | | Oropouche | | Northern | | | |
	LA	UA	LG	UG	LQ	UQ	LT	UT	LY	UY
LA		**0.16**	**0.09**	**0.45**	**0.13**	**0.20**	**0.12**	**0.13**	**0.13**	**0.41**
UA	0.08			**0.27**	**0.68**	**0.26**	**0.35**	**0.34**	**0.37**	**0.34** **0.60**
LG	0.02	0.01		**0.42**	**0.12**	**0.18**	**0.06**	**0.06**	**0.15**	**0.41**
UG	**0.22**	0.03	**0.11**		**0.42**	**0.48**	**0.32**	**0.38**	**0.52**	**0.74**
LQ	**0.08**	0.01	−0.003	**0.086**		**0.08**	**0.12**	**0.13**	**0.18**	**0.38**
UQ	**0.08**	0.03	−0.011	0.071	0.006		**0.17**	**0.17**	**0.24**	**0.44**
LT	0.04	0.06	−0.026	**0.097**	0.006	−0.01		0.001	**0.20**	**0.42**
UT	0.03	**0.019**	−0.020	**0.12**	0.002	−0.004	−0.017		**0.19**	**0.43**
LY	**0.08**	−0.006	0.002	**0.053**	0.011	0.003	−0.002	0.01		**0.36**
UY	**0.25**	0.049	**0.154**	−0.022	**0.110**	0.095	0.12	**0.14**	0.07	

Estimates of F_{ST} for MHC class IIB appear below the diagonal and estimates of F_{ST} for the microsatellites appear above the diagonal. Populations are grouped by drainage. Significant values at $P < 0.05$ are presented in bold.
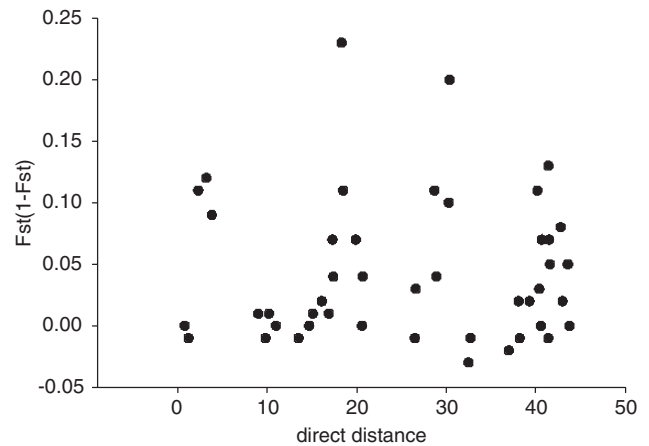

Figure 4 The relationship between the F_{ST} of MHC class IIB alleles (corrected to $1/(1-F_{ST})$) and direct geographical distance among 10 populations of the guppy (*Poecilia reticulata*).

most of the molecular variation was found within populations (89%, $F_{ST} = 0.132$, $d.f. = 531, 540$, $P < 0.001$), and to a lesser extent among populations within rivers (13%; $F_{ST} = 0.11$, $d.f. = 5, 540$, $P < 0.001$). The rivers did not differ significantly and no variation was found to partition among rivers (-2.4%; $F_{ST} = -0.024$, $d.f. = 4$, 540, $P = 0.75$). Similarly, no variation was found to partition among drainages, when populations were grouped into drainages instead of rivers (data not shown).

F_{ST} of MHC class IIB and F_{ST} of microsatellites were significantly positively correlated ($r^2 = 0.10$, $n = 45$, $P = 0.028$). Similarly, G_{ST} of the MHC was significantly related to G_{ST} of microsatellites ($r^2 = 0.13$, $n = 45$, $P = 0.019$). The majority of F_{ST} values estimated from MHC was less than those calculated from microsatellites (Figure 5). In total, 33 of the 45 (73%) pairwise MHC F_{ST} estimates were significantly lower than F_{ST} estimates for the microsatellite loci ($P < 0.05$). By comparing G_{ST} of MHC to G_{ST} of microsatellites we detected a similar pattern, with 32 (71%) pairwise estimates for the MHC being significantly lower than the estimates for the microsatellite loci.
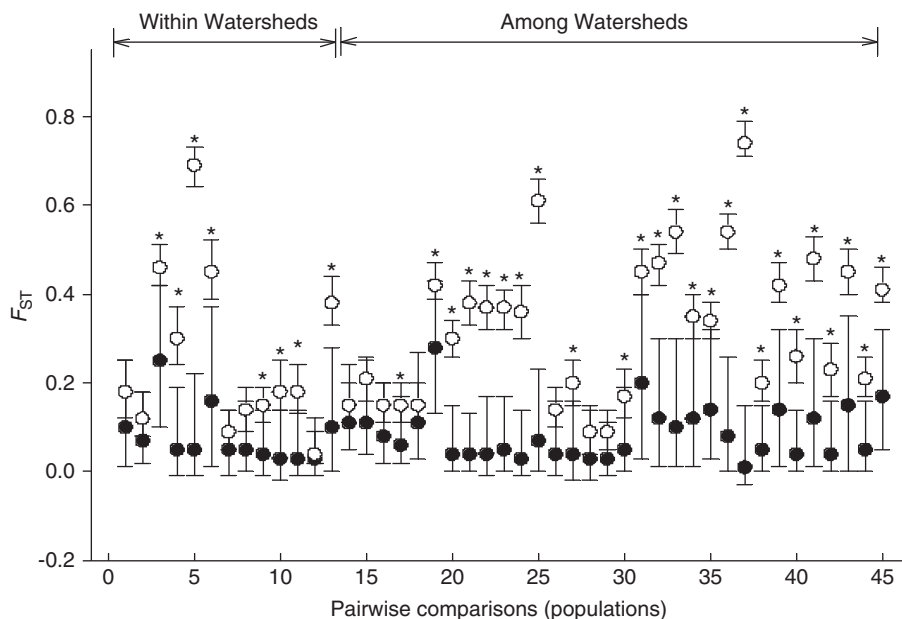
Figure 5 Comparison of F_{ST} values of neutral loci (six microsatellite loci) to MHC class IIB for each pairwise comparison of 10 populations of the guppy (*Poecilia reticulata*). Median F_{ST} estimates of the microsatellite loci are indicated with open circles and median F_{ST} estimates of the MHC class IIB are indicated with closed circles. Error bars indicate 95% confidence intervals and were estimated by resampling individuals within a population with replacement 1000 times ($P < 0.05$ indicated with an asterisk). Pairwise estimates are grouped by within drainage and among drainages comparisons.

The model-based approach outlined by Beaumont and Nichols (1996) identified the MHC class IIB F_{ST} as lower than the expected neutral F_{ST} in the Caroni drainage, ($H_E = 0.67$, $F_{ST} = 0.082$ test statistic = -2.57, $P = 0.006$; Figure 6a) and the Oropouche drainage ($H_E = 0.73$, $F_{ST} = 0.0001$, test statistic = -3.71, $P < 0.001$; Figure 6b). Similar results were found when using G_{ST} ($H_E = 0.67$, $G_{ST} = 0.11$, test statistic = -2.18, $P = 0.018$ and $H_E = 0.73$, $G_{ST} = 0.035$, test statistic = -1.78, $P = 0.045$ for Caroni and Oropouche, respectively). In the Yarra drainage, F_{ST} for MHC was not significantly lower than the expected F_{ST} ($H_E = 0.57$, $F_{ST} = 0.087$, test statistics = -0.10, $P = 0.18$; Figure 6c) or G_{ST} ($He = 0.57$, $G_{ST} = 0.13$, test statistic = -0.68, $P = 0.27$). F_{ST} at one of the microsatellite loci (Pre15) was also found to fall below the 0.005 quantile in both the Caroni and Oropouche drainages (Caroni: $H_E = 0.97$, $F_{ST} = 0.10$, test statistic = -5.00, $P < 0.001$; Oropouche: $H_E = 0.93$, $F_{ST} = 0.049$, test statistic = -1.78, $P = 0.045$; Figure 6). This microsatellite was removed along with Pr171 from the Yarra analysis because of its low heterozygosity (using the recommended cutoff of 0.01 from Beaumont and Nichols, 1996). The low heterozygosity relative to F_{ST} of Pre15 may indicate that Pre15 is linked to a region that is undergoing selection lowering the population differentiation at this locus. All other microsatellites were not significantly different from the expected F_{ST} ($P > 0.5$) and fell within the 95% quantiles. Similar results were found when the stepwise mutation model was used for the mode of mutation.

Discussion

Despite the known high differentiation between guppy populations in northern Trinidad, especially those occurring in different drainages, we found that popula-

tions were genetically similar at the MHC class IIB locus. The majority of population pairwise estimates for the MHC was significantly lower than those for the microsatellite loci, and MHC F_{ST} estimates were consistently lower than those predicted by a coalescent model of neutral evolution. This is a remarkable finding given that populations in different drainages are estimated to have diverged some 2.5 million years ago (Magurran, 2006). We also found that neutral evolution could not account for the low population differentiation. The lack of differentiation may instead be due to similar selection acting across the populations. It is unlikely that sampling error can account for our observations because we employed a novel bootstrapping approach that is ideal for incorporating small sample sizes. This approach involved resampling individuals within populations for both the MHC and microsatellite data, which reduce the sensitivity of the analysis to rare alleles. Thus, the change at the MHC cannot be accounted for by sampling artifact of rare alleles. Furthermore, our application of the model-based approach of Beaumont and Nichols (1996) showed a similar lack of differentiation at the MHC locus, and although this test loses some statistical power with small sample sizes, it also becomes more conservative with lower sample sizes.

Neutral evolution seems to have a minor role in MHC class IIB diversity in guppies. We found positive correlations between expected MHC heterozygosity and microsatellite heterozygosity, and between F_{ST} estimated for the MHC and microsatellites. We also found that approximately 14% of the molecular variation at the MHC was attributed to differences among populations within rivers. Indeed, 19 alleles (42% of all alleles) were unique to individual populations and contributed to the variation among populations within rivers. This large number of unique alleles may be due to genetic drift or
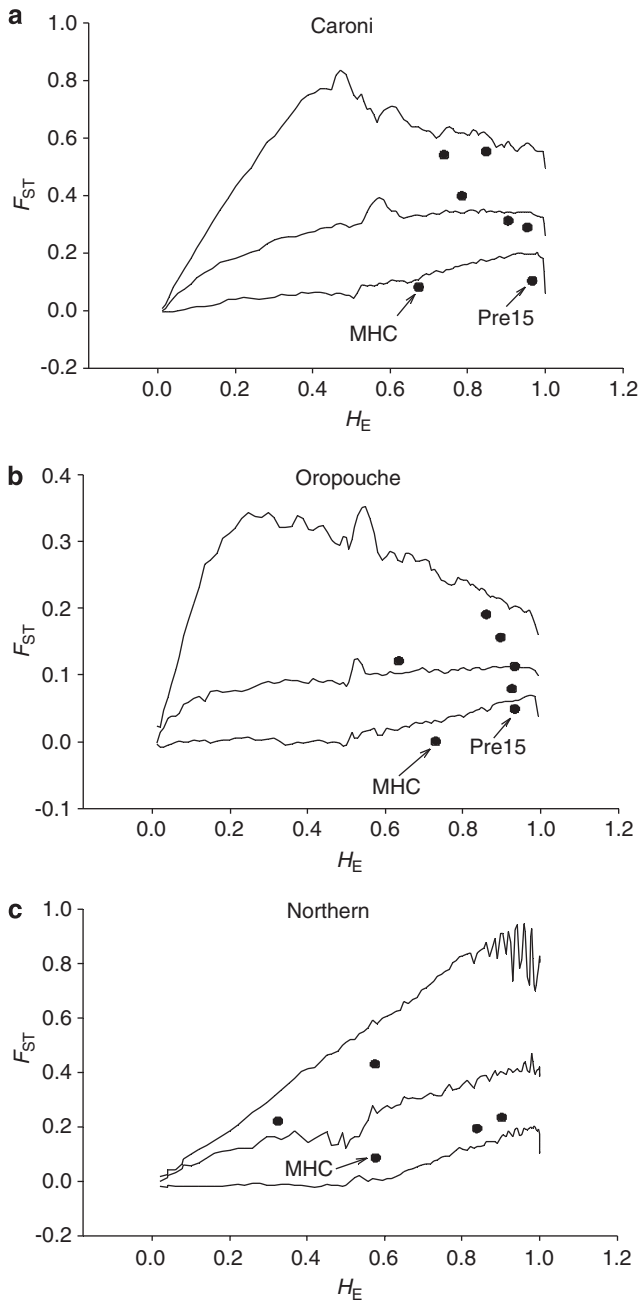
Figure 6 Simulated F_{ST} as a function of expected heterozygosity using the coalescent theory in the guppy (*Poecilia reticulata*). Results are shown for each drainage studied and comprise (**a**) Caroni, (**b**) Oropouche and (**c**) a northern drainage. Lines denote 0.995, 0.5 and 0.005 quantiles of the expected distribution of F_{ST} as a function of H_E. Black circles indicate the observed (empirical) F_{ST} and H_E values for each locus. MHC class IIB and microsatellite loci that fell outside the quantiles are labeled.

random effects associated with colonization. Given that a majority of the unique alleles were quite divergent from other alleles within the populations, it is unlikely that the unique alleles arose by mutation after population divergence. On the other hand, most populations were more similar at the MHC than expected based on microsatellite loci; F_{ST} based on the MHC was only significantly different from zero for 13 of the 45 (29%) pairwise comparisons as compared with that for 44 (98%)

of the pairwise comparisons based on the microsatellites. Furthermore, there was no relationship between the number of alleles at the MHC and average number of alleles at the microsatellite loci, no significant differentiation at the MHC among rivers or drainages, and no relationship between MHC differentiation and the geographical distance between populations. Thus, taken together these data suggest that neutral evolution cannot explain a majority of the variation at the MHC in the guppy.

Instead, we found that selection is likely to be the main force affecting diversity at the MHC class IIB locus in guppies. We found that the F_{ST} of the MHC was consistently lower than the F_{ST} of the microsatellites. Indeed, in 33 of the 45 (73%) population pairwise comparisons, the F_{ST} of MHC were significantly lower than the F_{ST} of the microsatellites. Similarly, 33 of the 45 (73%) G_{ST} estimates for the MHC were significantly lower than those for the microsatellite loci. Population differentiation was also found to be lower at the MHC in two of the three drainages when compared with a simulated distribution from neutral expectations. Analogous results were found for an analysis of just two populations by van Oosterhout *et al.* (2006b), who concluded that reduced population differentiation at the MHC was due to balancing selection at the MHC or selection on closely linked loci (also see van Oosterhout, 2009). Selection that results in reduced population differentiation at the MHC has been shown in other taxa (for example, Miller *et al.*, 2001; Sommer, 2003) and the authors of those studies argued for directional or balancing selection. Our results similarly suggest that directional or balancing selection is operating widely across guppy populations in northern Trinidad preventing them from differentiation through other stochastic forces.

It is possible that directional selection is reducing among population differences at the MHC class IIB in guppies. We found that one allele (Pore_a132) was present in all populations and the allele typically had a high frequency (range 28–83%). Conceivably, the Pore_a132 allele may be associated with resistance to a pervasive pathogen, which would lead to directional selection on the allele. Such a claim has been made in sockeye salmon in which Miller *et al.* (2001) hypothesized that MHC allele uniformity across some populations was due to directional selection from a pathogen. Direct associations between specific MHC alleles and resistance to pathogens have been uncovered in a variety of taxa. For example, Langefors *et al.* (2001) found that Atlantic salmon experimentally infected with a bacterium (*Aeromonas salmonicida*) had higher resistance if they carried specific MHC class II alleles, and Meyer-Lucht and Sommer (2005) found that both susceptibility and resistance against a common nematode was conferred by a single MHC class II allele in wild populations of yellow-necked mice (*Apodemus flavicollis*) (also see, Hill *et al.*, 1991; Lohm *et al.*, 2002; Schad *et al.*, 2005). Thus, directional selection from a prevalent pathogen could explain the high frequency of the Pore_a132 allele and the low population differentiation at the MHC across the guppy populations that we studied.

The lack of differentiation at the MHC class IIB across guppy populations may also be explained by balancing selection. Balancing selection maintains genetic

polymorphisms through frequency-dependent selection or overdominance. First, frequency-dependent selection favors rare alleles and can lead to a cycling pattern as different alleles become advantageous. Frequency-dependent selection is a key component of models of host–parasite coevolution (for example, Hamilton and Zuk, 1982) and has been implicated as a mechanism maintaining diversity at the MHC in several other species (for example, Hill *et al.*, 1991). Second, overdominance favors heterozygotes and can select for rare alleles in populations because rare alleles mostly appear in heterozygous genotypes (Takahata and Nei, 1990). Both processes lead to selection on rare alleles, and therefore can reduce population differentiation through selection for migrants (Schierup *et al.*, 2000; Muirhead, 2001). Such selection for rare MHC alleles can be strong enough that alleles are retained long after speciation, which is termed transspecies polymorphism (Klein *et al.*, 1998). Retention of alleles from an ancient founding population, either from frequency-dependent or overdominant selection, may have led to the observed pattern of shared alleles across divergent drainages in the guppy. For example, in addition to finding the presence of the Pore_a132 allele across all populations, the Pore_a73 allele was found at relatively high frequencies in all populations and 13 other alleles were shared across at least two of the drainages. Although the populations from the different drainages likely diverged some 2.5 million years ago (Magurran, 2006), it is plausible that alleles have been retained since the time of colonization of the various drainages. For example, retention of MHC alleles has been found in such divergent taxa as humans and chimpanzees, and mice and rats (Klein *et al.*, 1998). Furthermore, selection for rare migrants is unlikely to fully account for the lack of population differentiation at the MHC because migration among all populations is unlikely (Magurran, 2006; Suk and Neff, 2009). However, we also found that the observed heterozygosity at the MHC was consistently lower than the expected heterozygosity in our populations. This latter result suggests that there is no direct selection for heterozygosity at the MHC, at least not in the current generation. The lower than expected heterozygosity may be due to underdominance at the MHC (Pitcher and Neff, 2006), selection for an optimal number of alleles (Wegner *et al.*, 2003) or methodological errors such as null alleles. A direct test of balancing selection requires the comparison of levels of expected heterozygosity to observed heterozygosity (Ewens, 1972; Watterson, 1978). We were unable to provide a meaningful comparison because of the duplication event at the MHC class IIB. Thus, we cannot yet confirm a role for balancing selection on maintaining genetic diversity at the MHC and reducing differentiation across the populations we studied.

We were also able to quantify long-term positive selection on individual codons of the MHC. We found that of the 43 alleles, 39 encoded for different amino acid sequences. Positive selection was detected at five codons, of which three were in the putative PBR. We also found that the six codons of the putative PBR had a higher average dN/dS ratio than the 66 non-PBR codons. Positive selection at the PBR is a well-documented property of the MHC (Hughes and Yeager, 1998) and similar results have been found in the majority of studies of MHC (see Supplementary Material in Bernatchez and

Landry, 2003). Interestingly, the other two codons that we detected positive selection on were adjacent to the putative PBR. These sites may also be involved in the affinity between the MHC and foreign antigens as amino acids that do not come into direct contact with the antigen can still affect binding through conformational changes (Foote and Winter, 1992). Positive selection in codons adjacent to the PBR has also been documented in the gila trout (*Oncorhynchus gilae gilae*; Peters and Turner, 2007), haplochromine cichlids (*Pseudotropheus fainziberi* and *P. emmiltos*; Blais *et al.*, 2007) and in humans (Yang and Swanson, 2002).

We also found purifying selection at two codons within the putative PBR. Through the examination of the crystal structure of human MHC DRB1 using Deepview/Swiss-PBDviewer v3.7 (Guet and Peitsch, 1997), we determined that both of these codons were located at an open end of the putative peptide recognition groove. Interestingly, the three codons within the putative PBR that were undergoing positive selection were close together and located at the center of the putative peptide-binding groove. These spatial results may be significant because the MHC class II does not anchor peptides at the end of the peptide-binding groove like the class I locus (Klein, 1990). Instead, our data suggest that stronger interactions with antigens may occur at the center of the groove. The codons at either end of the groove may be undergoing purifying selection for some other function. To our knowledge, purifying selection within the PBR has not previously been documented. However, some caution must be exercised when making conclusions on the structure of the MHC because X-ray crystallography has so far only been completed for human MHC genes. Nevertheless, by using a random sites model, we have been able to elucidate the type of selection acting at the MHC more clearly, and, consequently, we have been able to infer potential functional diversity among and within MHC alleles.

Overall, we found evidence that selection is acting on the MHC both at the sequence and population level in guppies. At the sequence level, we found evidence for both positive and purifying selection within the putative PBR that may relate to protein confirmation and antigen recognition. At the population level, directional or balancing selection may be acting to make populations more similar than expected by neutrality. These results are particularly interesting given the long divergence time between some of our populations. Further studies are required to elucidate the roles of each of these mechanisms and to identify the agent of selection.

166

References

Alexander HJ, Taylor JS, Wu SS, Breden F (2006). Parallel evolution and vicariance in the guppy (*Poecilia reticulata*) over multiple spatial and temporal scales. *Evolution* **60**: 2352–2369.

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.

Beaumont MA, Balding DJ (2004). Identifying adaptive genetic divergence among populations from genome scans. *Mol Ecol* **13**: 969–980.

Beaumont MA, Nichols RA (1996). Evaluating loci for use in the genetic analysis of population structure. *Proc R Soc London Ser B* **263**: 1619–1626.

Becher SA, Russel ST, Magurran AE (2002). Isolation and characterization of polymorphic microsatellites in the Trinidadian guppy (*Poecilia reticulata*). *Mol Ecol Notes* **2**: 456–458.

Bernatchez L, Landry C (2003). MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *J Evol Biol* **16**: 363–377.

Blais J, Rico C, van Oosterhout C, Cable J, Turner GF, Bernatchez L (2007). MHC adaptive divergence between closely related and sympatric African chichlids. *PloS One* **8**: e734.

Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL et al. (1993). Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* **364**: 33–39.

Cable J, van Oosterhout C (2007a). 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.

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

Carvalho GR, Shaw PW, Magurran AE, Seghers BH (1991). Marked genetic divergence revealed by allozymes among populations of the guppy *Poecilia reticulata* (Poeciliidae) in Trinidad. *Biol J Linn Soc* **42**: 389–405.

Clarke BC, Kirby DRS (1966). Maintenance of histocompatibility polymorphism. *Nature* **211**: 999–1000.

Consuegra S, Megans HJ, Leon K, Stet RJM, Jordan WC (2005). Patterns of variability at the major histocompatibility class II alpha locus in Atlantic salmon contrast with those at the class I locus. *Immunogenetics* **57**: 16–24.

Crispo E, Bentzen P, Reznick DN, Kinnison MT, Hendry AP (2006). The relative influence of natural selection and geography on gene flow in the guppies. *Mol Ecol* **15**: 49–62.

Dionne M, Miller K, Dobson JJ, Caron F, Bernatchez L (2007). Clinal variation in MHC diversity with temperature: evidence for the role of host-pathogen interaction on local adaption in Atlantic salmon. *Evolution* **61**: 2154–2164.

Doherty PC, Zinkernagel RM (1975). Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature* **256**: 50–52.

Ewens WJ (1972). The sampling theory of selectively neutral alleles. *Theor Pop Biol* **3**: 87–112.

Fajen A, Breden F (1992). Mitochondrial DNA sequence variation among natural populations of the Trinidad guppy *Poecilia reticulata*. *Evolution* **46**: 1457–1465.

Figueroa F, Mayer WE, Sultmann H, O'hUigin C, Tichy H, Satta Y et al. (2000). MHC class II B gene evolution in east african cichlid fishes. *Immunogenetics* **51**: 557–575.

Foote J, Winter G (1992). Antibody framework residues affecting the conformation of hypervariable loops. *J Mol Biol* **224**: 487–499.

Goudet J (2001). *FSTAT, a Program to Estimate and Test Gene Diversities and Fixation Indices, Version 2.9.3*. Institut d'Ecologie, Universite de Lausanne: Laussane, Switzerland.

Guet N, Peitsch MC (1997). Swiss-model and Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**: 2714–2723.

Hamilton WD, Zuk M (1982). Heritable true fitness and bright birds: a role for parasites? *Science* **218**: 384–387.

Hardy OJ, Vekemans X (2002). SPAGeDi: a versatile computer program to analyze spatial genetic structure at the individual or population levels. *Mol Ecol Notes* **2**: 618–620.

Haskins CP, Haskins EF (1954). Note on a 'permanent' experimental alteration of genetic constitution in a natural population. *Proc Nat Acad Sci USA* **40**: 627–635.

Hedrick PW, Kim TJ, Parker KM (2001). Parasite resistance and genetic variation in the endangered Gila topminnow. *Animal Cons* **4**: 103–109.

Hey J, Nielson R (2004). Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. permilis*. *Genetics* **167**: 747–760.

Hill AVS (1991). HLA associations with malaria in Africa: some implications for MHC evolution. In: Klein J, Klein D (eds). *Molecular Evolution of the Major Histocompatibility Complex*. Springer: Berlin and Heidelberg. pp 403–419.

Hill AVS, Allsopp CEM, Kwiatkowski D, Anstey NM, Twumasi P, Rowe PA et al. (1991). Common West African HLA antigens are associated with protection from severe malaria. *Nature* **352**: 595–600.

Hughes AL, Nei M (1988). Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335**: 167–170.

Hughes AL, Yeager M (1998). Natural selection at the major histocompatibility complex loci of vertebrates. *Annu Rev Genet* **32**: 415–435.

Jukes TH, Cantor CR (1969). Evolution of protein molecules. In: Munro HN (ed). *Mammalian Protein Metabolism*. Academic Press: New York.

Kimura M (1983). *The Neutral Theory of Molecular Evolution*. Cambridge University Press.

Klein J (1990). *The Natural History of the Major Histocompatibility Complex*. John Wiley & Sons: NW, USA.

Klein J, Sato A, Nagl S, O'hUigin C (1998). Molecular transspecies polymorphism. *Annu Rev Ecol and Syst* **29**: 1–21.

Landry C, Bernatchez L (2001). Comparative analysis of population structure across environments and geographical scales at major histocompatibility complex and microsatellite loci in Atlantic salmon (*Salmo salar*). *Mol Ecol* **10**: 2525–2539.

Langefors A, Lohm J, Grahn M, Andersen O, von Schantz T (2001). Association between major histocompatibility complex class IIB alleles and resistance to *Aeromonas salmonicida* in Atlantic salmon. *Proc R Soc London Ser B* **268**: 479–485.

Lohm J, Grahn M, Langefors A, Andersen O, Storset A, von Schantz T (2002). Experimental evidence for major histocompatibility complex-allele-specific resistance to a bacterial infection. *Proc R Soc London Ser B* **269**: 2029–2033.

Lukas D, Vigilant L (2005). Reply: facts, faeces and setting standards for the study of MHC genes using noninvasive samples. *Mol Ecol* **14**: 1601–1602.

Lyles AM (1990). Genetic variation and susceptibility to parasites: *Poecilia reticulata* infected with *Gyrodactylus turnbulli* PhD thesis, Princeton University.

Magurran AE (2006). *Evolutionary Ecology The Trinidadian Guppy*. Oxford University Press.

Martin CH, Johnsen S (2007). A field test of the Hamilton-Zuk hypothesis in the Trinidadian guppy (*Poecilia reticulata*). *Behav Ecol Sociobiol* **61**: 1897–1909.

McConnell TJ, Godwin UB, Norton SF, Nairn RS, Kazianis S, Morizot DC (1998). Identification and mapping of two divergent, unlinked major histocompatibillity complex class II *B* genes in xiphophorus fishes. *Genetics* **149**: 1921–1934.

Meyer-Lucht Y, Sommer S (2005). MHC diversity and the association to nematode parasitism in the yellow-necked mouse (*Apodemus flavicollis*). *Mol Ecol* **14**: 2233–2243.

Miller KM, Kaukinen KH, Beacham TD, Withler RE (2001). Geographic heterogeneity in natural selection on an MHC locus in sockeye salmon. *Genetica* **111**: 237–257.

Muirhead CA (2001). Consequences of population structure on genes under balancing selection. *Evolution* **55**: 1532–1541.

Oliver MK, Telfer S, Piertney SB (2009). Major histocompatibility complex (MHC) heterozygote superiority to natural multi-parasite infections in the water vole (*Arvicola terrestris*). *Proc R Soc B* **276**: 1119–1128.

Ono H, O'hUigin C, Vincek V, Klein J (1993). Exon-intron organization of fish major histocompatibility complex class II B genes. *Immunogenetics* **38**: 223–234.

Paterson IG, Crispo E, Kinnison MT, Hendry AP, Bentzen P (2005). Characterization of tetranucleotide microsatellite markers in guppy (*Poecilia reticulata*). *Mol Ecol Notes* **5**: 269–271.

Paterson S, Wilson K, Pemberton JM (1998). Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population (*Ovis aries L.*). *Proc Natl Acad Sci USA* **95**: 3714–3719.

Peters MB, Turner TF (2007). Genetic variation of the major histocompatibility complex (MHC class II B gene) in the threatened Gila trout *Oncorhynchus gilae gilae*. *Cons Genet* **9**: 257–270.

Piertney SB, Oliver MK (2006). The evolutionary ecology of the major histocompatibility complex. *Heredity* **96**: 7–21.

Pitcher TE, Neff BD (2006). MHC class IIB alleles contribute to both additive and non-additive genetic effects on survival in Chinook salmon. *Mol Ecol* **15**: 2357–2365.

Pons O, Petit RJ (1995). Estimation, variance and optimal sampling of gene diversity I. Haploid locus. *Theor Appl Genet* **90**: 462–470.

Richman A (2000). Evolution of balanced genetic polymorphism. *Mol Ecol* **9**: 1953–1963.

Ronfort J, Janczewski E, Bataillon T, Rousset F (1998). Analysis of population structure in autotetraploid species. *Genetics* **150**: 921–930.

Rozas J, Sanchez-Delbarrio JC, Messeguer X, Rozas R (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**: 2496–2497.

Saitou N, Nei M (1987). The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406–425.

Schad J, Ganzhorn JU, Sommer S (2005). Parasite burden and constitution of major histocompatibility complex in the malagasy mouse lemus, *Microcebus murinus*. *Evolution* **59**: 439–450.

Schaschl H, Suchentrunk F, Hammer S, Goodman SJ (2005). Recombination and the origin of sequence diversity in the *DRB* MHC class II locus in chamois (*Rupicapra* spp.). *Immunogenetics* **57**: 108–115.

Schierup M, Vekemans X, Charlesworth D (2000). The effect of subdivision on variation at multi-allelic loci under balancing selection. *Genet Res* **76**: 51–62.

Schneider S, Roessli D, Excoffier L (2000). *Arlequin V. 2.000*. Genetics and Biometry Laboratory, University of Geneva: Switzerland.

Sommer S (2003). Effects of habitat fragmentation and changes of dispersal behaviour after a recent population decline on the genetic variability of noncoding and coding DNA of a monogamous Malagasy rodent. *Mol Ecol* **12**: 2845–2851.

Suk HY, Neff BD (2009). Microsatellite genetic differentiation among populations of the Trinidadian guppy. *Heredity* **102**: 425–434.

Takahata N, Nei M (1990). Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics* **124**: 967–978.

Tamura K, Dudley J, Nei M, Kumar S (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**: 1596–1599.

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

van Oosterhout C, Joyce DA, Cummings SM (2006a). Evolution of MHC class IIB in the genome of wild and ornamental guppies, *Poecilia reticulata*. *Heredity* **97**: 111–118.

van Oosterhout C, Joyce DA, Cummings SM, Blais J, Barson NJ, Ramnarine IW *et al.* (2006b). Balancing selection, random genetic drift, and genetic variation at the major histocompatibility complex in two wild populations of guppies (*Poecilia reticulata*). *Evolution* **60**: 2562–2574.

van Oosterhout C, Mohammed RS, Hansen H, Archard GA, McMullan M, Weese DJ *et al.* (2007). Selection by parasites in spate conditions in wild Trinidadian guppies (*Poecilia reticulata*). *Int J Parasitol* **37**: 805–812.

Watterson GA (1978). The homozygosity test of neutrality. *Genetics* **88**: 405–417.

Wegner KM, Kalbe M, Kurtz J, Reusch TBH, Milinski M (2003). Parasite selection for immunogenetic optimality. *Science* **301**: 1343.

Weir BS, Cockerham CC (1984). Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.

Westerdahl H, Hansson B, Bensch S, Hasselquist D (2004). Between-year variation of MHC allele frequencies in great reed warblers: selection or drift? *J Evol Biol* **17**: 485–492.

Yang Z, Bielawski JP (2000). Statistical methods for detecting molecular adaptation. *Trends Ecol Evol* **15**: 496–503.

Yang Z, Nielson R, Goldman N, Krabbe Pedersen A (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**: 431–449.

Yang Z, Swanson WJ (2002). Codon-substitution models to detect adaptive evolution that account for heterogenous selective pressures among site classes. *Mol Biol Evol* **19**: 49–57.

Yang ZH (2007). PAML 4; a program package for phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**: 1586–1591.

Yang ZH, Wong WSW, Nielson R (2005). Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol Biol Evol* **22**: 1107–1118.