**Heredity (2006) 97**, 111–118

www.nature.com/hdy

# Evolution of MHC class IIB in the genome of wild and ornamental guppies, *Poecilia reticulata*

C van Oosterhout, DA Joyce and SM Cummings
*School of Biological Sciences, University of Hull, Cottingham Rd, Hull HU6 7RX, UK*

This is the first study to quantify genomic sequence variation of the major histocompatibility complex (MHC) in wild and ornamental guppies, *Poecilia reticulata*. We sequenced 196–219 bp of exon 2 MHC class IIB (*DAB*) in 56 wild Trinidadian guppies and 14 ornamental strain guppies. Each of two natural populations possessed high allelic richness (15–16 alleles), whereas only three or fewer *DAB* alleles were amplified from ornamental guppies. The disparity in allelic richness between wild and ornamental fish cannot be fully explained by fixation of alleles by inbreeding, nor by the presence of non-amplified sequences (ie null alleles). Rather, we suggest that the same allele is fixed at duplicated MHC *DAB* loci owing to gene conversion. Alternatively, the number of loci in the ornamental strains has contracted during $>100$ generations in captivity, a hypothesis consistent with the accordion model of MHC evolution. We furthermore analysed the substitution patterns by making pairwise comparisons of sequence variation at the putative peptide binding region (PBR). The rate of non-synonymous substitutions (d*N*) only marginally exceeded synonymous substitutions (d*S*) in PBR codons. Highly diverged sequences showed no evidence for diversifying selection, possibly because synonymous substitutions have accumulated since their divergence. Also, the substitution pattern of similar alleles did not show evidence for diversifying selection, plausibly because advantageous non-synonymous substitutions have not yet accumulated. Intermediately diverged sequences showed the highest relative rate of non-synonymous substitutions, with d*N*/d*S*$>14$ in some pairwise comparisons. Consequently, a curvilinear relationship was observed between the d*N*/d*S* ratio and the level of sequence divergence.
*Heredity* (2006) **97,** 111–118. doi:10.1038/sj.hdy.6800843; published online 31 May 2006

**Keywords:** *Poecilia reticulata*; major histocompatibility complex; multigene evolution; gene duplication; gene conversion; teleost

## Introduction

The highly polymorphic genes of the major histocompatibility complex (MHC) encode cell-surface receptors that recognise and bind peptides for presentation to T cells. They have an important adaptive significance with respect to conferring parasite resistance (Hill *et al*, 1991; Hedrick *et al*, 1998), mate choice and preference (Penn, 2002; Milinski, 2003) and possibly maternal–fetal interactions (Ober and van der Ven, 1997).

Several theories have been put forward to explain the evolution of the MHC. The birth-and-death model proposes that MHC genes are produced in repeated tandem or block gene duplications, and some of the duplicated genes diverge functionally whereas others become pseudogenes owing to deleterious mutations or are deleted from the genome (Nei *et al*, 1997). In the accordion model (Klein *et al*, 1993), the number of MHC genes is assumed to be expanding or contracting depending on the need to protect the host from ever-changing parasite pressure. The high levels of MHC polymorphism are thought to be maintained by various forms of balancing selection, including overdominant selection, frequency-dependent selection and selection that varies in time and/or space (Hedrick and Thomson, 1983; Hedrick, 2002). The most commonly held view is that MHC variation is maintained primarily by balancing selection and point mutation in combination with gene duplication and translocation (Nei *et al*, 1997).

MHC polymorphism appears to be maintained even in populations that are monomorphic at other genetic loci. For example, the San Nicolas Island fox has no variation in other genetic markers (including allozymes, minisatellites and microsatellites) but exhibits remarkably high levels of MHC variation (Aguilar *et al*, 2004). Such a high level of MHC polymorphism suggests that strong balancing selection is acting on MHC alleles or genotypes (Aguilar *et al*, 2004), or that other stochastic processes are involved in maintaining MHC diversity (Hedrick, 2004).

Important insights into the evolution, maintenance and adaptive significance of MHC variation can be made by studying inbred captive populations in which the effects of drift and stochastic processes override the deterministic effects of balancing selection on MHC. Inbred strains of mice vary in the genomic organisation and gene copy number of MHC class IA loci owing to frequent gene duplication and deletions (Stroynowski, 1990). Alleles of MHC class IA also appear not to show locus specificity (Gu and Nei, 1999), observations that are consistent with inter-locus recombination and gene conversions (Yun *et al*, 1997). Phylogenetic analysis on MHC class II genes in mice does not show evidence,

*Correspondence: C van Oosterhout, School of Biological Sciences, University of Hull, Cottingham Rd, Hull HU6 7RX, UK.*
*E-mail: c.van-oosterhout@hull.ac.uk*
Received 27 September 2005; accepted 12 April 2006; published online 31 May 2006

however, for inter-locus recombination as alleles of loci form monophyletic groups (Gu and Nei, 1999). Interestingly, Hogstrand and Bohme (1999) found direct empirical evidence of gene conversion in MHC class II loci in mice sperm, the frequency of which is estimated to be as high as 1/40 000 between alleles. The discrepancy between the phylogenetic analysis and the direct measure of gene conversion could be explained by the fact that many gene conversions are not fixed in natural populations of mice because of selective disadvantage compared to wild-type alleles (Gu and Nei, 1999).

Inbred strains and selection lines of guppies (*Poecilia reticulata* Peters) could also be used to study the evolution of MHC. The genetics of colour pattern in the guppy has been studied over the last 80 years, and during this time, many selected and inbred strains have been established (Winge, 1922; van Oosterhout *et al*, 2003). The guppy is native to the freshwater streams, ponds and brackish water swamps of Trinidad, Tobago, Barbados, Venezuela, Guyana and northeastern Brazil (Haskins and Haskins, 1951) and made its first appearance in Europe as aquarium fish in 1905. Guppies were also introduced to many parts of Asia for mosquito control during the early twentieth century (Fernando and Phang, 1990). In Singapore, commercial culture of fancy tail or ornamental guppy strains began in the early 1950s, and around 30–40 different strains are currently reared in monoculture farms (Fernando and Phang, 1985; Khoo *et al*, 1999). Previous studies on the MHC of guppies focused on expressed loci and detected a very low copy number of MHC class I and II loci. Sato *et al* (1995) studied expressed MHC variation in guppies from various ornamental lines and descendants of wild-caught fish from the Aripo and El Cedro Rivers in Trinidad. Possibly because the wild fish were inbred over a few generations, Sato *et al* (1995) did not find differences in genetic variation between wild and ornamental strains and identified only one or two expressed MHC class I loci. They also detected one class II MHC locus, which showed 92% sequence similarity to the MHC class IIB *DAB* sequence of *Xiphophorus maculatus* (see McConnell *et al*, 1998a, b). McConnell *et al* (1998b) found evidence for another highly diverged MHC class IIB locus, *MhcPore-DXB*. The presence of just two expressed class II genes in the guppy is in sharp contrast with the high number of duplicated genes found in more advanced neoteleosts (cichlids, sticklebacks, cod) (Miller *et al*, 2002).

Here we investigate MHC class IIB (*DAB*) variation in the genomes of ornamental and wild-caught guppies using specific and degenerate MHC class II *DAB* locus primers. We analyse the effects of >100 generations of selective breeding and inbreeding on MHC variation in the genomes of these small teleosts. We further investigate the effects of selection on the substitution pattern of alleles by analysing the rate of synonymous (d*S*) and non-synonymous (d*N*) substitutions between pairwise sequences with varying levels of divergence.

## Materials and methods

### Guppy populations and strains

We analysed the MHC variation in two commercial ornamental strains from Singapore, that is, the Red tail (Red), the Blue variety (Blue), and one commercial English strain known as Green Snakeskin (Green). The date of domestication of Red in Singapore is not exactly known, but the first mention in the literature of selection of the red tail coloration was by Axelrod and Whitern (1965). The Blue variety, also known as the Istanbul strain, dates back to the 1950s and has been described by Dzwillo (1959). The Green strain is an ornamental guppy variety cultured in England in the late 1940s or early 1950s (Gordon, 1955). The generation time of guppies can be as little as 2–3 months with longevity of 1–2 years (van Oosterhout, personal observation). This means that the ornamental guppies have been in captivity for a minimum of 25 and a maximum of 300 generations, or approximately >100 generations. We also analysed MHC of wild-caught guppies from two natural Trinidadian guppy populations of the Upper Aripo (UA, grid reference PS693100E and 1181800N) and Lower Aripo (LA, PS6914000E and 1177700N), in October 2001.

### PCR procedures

Genomic DNA was extracted from caudal finclips using the HotShot method (Truett *et al*, 2000). We screened four guppies of the Green strain, five guppies of Red strain, five guppies of the Blue strain, 35 guppies from the Lower Aripo (LA) and 21 guppies from the Upper Aripo (UA).

A section of MHC IIB exon II between codons 11 and 83, containing 12 putative antigen recognition sites, was amplified using the reverse primer (Sato *et al*, 1995) and either a modified version of the forward primer (5′-GCA GGT GGA TTC AGA GAA TAT GAA GTG GAT CG-3′) or a degenerate forward primer *DAB*degF (5′-GTG TCT TYA RCT SNC TGA RC-3′). These primers spanned regions of 219 and 196 bp, respectively. A degenerate reverse primer, *DAB*degR (5′-TYR ACG KKG KTC AGA CAG-3′), failed to amplify any additional sequences when used in conjunction with either the specific or degenerate forward primers, and was therefore not used in these experiments.

Amplification took place in a total volume of 25 µl containing either 2.5 pmol of each specific primer or 12.5 pmol of the degenerate primer, 3 mM $MgCl_2$ and 0.2 mM each dNTP and 0.1 µl of Biotaq polymerase (Bioline). A touchdown PCR consisting of an initial step of 94°C for 3 min, followed by one cycle of 94°C: 45 s, 61°C: 1 min, 72°C: 1 min, then one cycle of 94°C: 45 s, 59°C: 1 min, 72°C: 1 min, then 28 cycles of 94°C: 45 s, 58°C: 1 min, 72°C: 1 min was performed. Products were resolved on 1.5% agarose gels.

We quantified the PCR error rate due to recombinant products originating from heteroduplex mismatch repair during cloning, as well as base nucleotide substitution during PCR (Kanagawa, 2003; Lukas and Vigilant, 2005). The probability of erroneous base substitution was estimated to be around $4.2 \times 10^{-5}$ nucleotide substitution per site per duplication, which compares to $7.3 \times 10^{-5}$ in a study by Kobayashi *et al* (1999). The expected proportion of correct sequences of 219 bp is 61.9%, and there are predicted to be approximately 27.6% with one, 7.6% with two and 2.1% with three artificial PCR base substitutions. To minimise the probability of erroneously identifying an artificial sequence as an actual *DAB* allele, we only recognised 'real' alleles as those that occurred in

two independent PCRs. In addition, sequences that differed by more than three base pairs from another sequence found within the same individual were also recognised as 'real' MHC alleles, as this level of sequence differentiation is unlikely to be caused by PCR misincorporations. This binning protocol thus minimised the risk of identifying PCR artefacts as real alleles and reduced the probability of ignoring accurate sequence information.

By comparing the results from different sets of primers, we assessed whether we have missed alleles due to mutations in primer annealing sites (see eg Lukas and Vigilant, 2005). The degenerate primers were designed by aligning all *P. reticulata* (guppy), *P. picta*, *P. sphenops* and *X. maculatus DAB* sequences published on GenBank, as well as those previously amplified in our lab (SM Cummings and C van Oosterhout, unpublished manuscript).

In addition to the degenerate primer set, we also developed a new reverse primer in exon 3, DABR4 (5′-TGC AGA CCA GCA TGG AGG GGT GC-3′). This allowed us to discriminate between genomic (gDNA) and expressed (cDNA) copies, as the size of the insert depends on the presence or absence of the intron sequence. DABR4 thus allows us to test whether we amplified all *DAB* alleles present and investigate which genomic copies are expressed. We analysed four guppies and found that of all of the eight gDNA alleles were also detected in the cDNA. This suggests that all sequences amplified from gDNA are expressed, and that we are not amplifying potential MHC pseudogenes.

### Cloning

PCR products were ligated into pGEM-T (Promega) and transformed into DH5α (Invitrogen) competent bacterial cells according to the manufacturer's instructions. Positive colonies were picked from plates and dipped directly into 25 μl PCRs prepared as described in the previous section. Positive colonies were subjected to allele identification using single-strand conformational polymorphism (SSCP).

### SSCP

A 5 μl measure of each PCR product was added to 7 μl of formamide:bromophenol:xylene cyanol solution and heated to 95°C for 5 min before being quenched in ice water. The samples were immediately loaded on a 12.5% MDE acrylamide gel containing 10% glycerol. These were run at 40 V in freshly prepared 0.6 × TBE buffer at 4°C and stained using Sybr-gold (10 μl per 100 ml 0.6 × TBE buffer).

Different conformations were verified as different alleles by sequencing Qiagen-cleaned PCR products from the initial colony screen using standard M13 primers and fluorescently labelled terminators on a Beckman Coulter CEQ8000. Approximately 25 identical conformations were also sequenced, which confirmed that the SSCP identified all unique sequences (data not shown).

We initially screened six clones per fish. However, because all ornamental strain guppies were genetically uniform and showed extremely low allelic variation (up to two SSCP patterns), we increased the number of clones per ornamental fish to check we had not missed any *DAB* alleles. We then analysed between 24 and 48 clones per ornamental strain fish and screened a total of 656 clones from four Green, five Red and five Blue guppies, and a total of 336 clones from 56 wild guppies.

### Sequence analysis

Chromatograms were edited and manually aligned using Proseq (Filatov, 2002) and Codon-Code Aligner (CodonCode Corp.). The $\omega$-ratio, that is, the ratio of non-synonymous (d$N$) to synonymous (d$S$) amino-acid substitution rate ($\omega = dN/dS$) was calculated following the method of Yang and Nielsen (2000) (ie equal weighting of pathways) in PAML 3.13d (Yang, 1997). Utilising the inferred structural molecular model of MHC (Hughes *et al*, 1996), we analysed the MHC sequence variation by partitioning it into two groups, that is, putative peptide binding region sites (PBR) and those outside (non-PBR). The PBR codons were identified by aligning the amino-acid sequence to that of an African cichlid (Figueroa *et al*, 2001), translated and labelled in MEGA 2.1 (Kumar *et al*, 2001). An alignment of all sequences in the study indicating the PBR codons is available as Supplementary Information. The PBR codons are expected to experience diversifying selection, as this promotes non-synonymous (replacement) substitutions, which enhances amino-acid diversity of transcribed proteins (Hughes, 2002). Consequently, PBR codons are expected to show $\omega$-ratio in excess of unity. By contrast, purifying selection is generally found to govern sequence variation outside the PBR, probably owing to functional constraints of the protein. We used pairwise comparisons between allelic sequences to analyse the substitution patterns both at the PBR and non-PBR codons. We calculated the polynomial regression between the $\omega$-ratio of PBR and the sequence divergence at non-PBR codons for all pairwise comparisons between alleles. This analysis was performed to investigate whether the substitution pattern changes with different levels of divergence (see eg Hughes, 1999; Hughes and Friedman, 2004).

To analyse the substitution patterns more comprehensively, we also calculated maximum likelihood estimation of d$S$ and d$N$ in pairwise comparisons in PAML 3.13d (Yang, 1997). We performed four tests to compare the relative rates of d$N$ and d$S$ between PBR and non-PBR sites. We first tested whether PBR codons display a significantly higher d$N$ than non-PBR codons. Secondly, we tested whether d$N$ significantly exceeds the d$S$ in the PBR. Both analyses test whether the substitution patterns are consistent with diversifying selection at the PBR. Thirdly, we analysed whether d$S$ exceeded d$N$ significantly in the non-PBR codons, which would be evidence for purifying selection. Finally, we tested whether d$S$ is not significantly different between PBR and non-PBR codons, which would be expected if synonymous sites evolve neutrally in the PBR as well as in the non-PBR. We utilised a simple Mann–Whitney $U$-test to examine the substitution patterns between all pairwise comparisons.

Sequences have been deposited on GenBank nucleotide database (accession numbers: wild guppies, AY770035, AY770040, AY770042, AY745517 and DQ396604–DQ396617; and ornamental guppies, AY747137, AY747142 and AY751542.

### Population genetic analysis

The MHC class IIB (*DAB*) allelic richness was compared between populations and strains using a bootstrap to calculate the mean and 95% CI of the numbers of unique alleles as a function of the number of clones screened (see Leberg, 2002). A given number of clones were therefore re-sampled 1000 times (without replacement) and the number of unique alleles was observed in each random subsample. A two-parameter exponential rise to maximum exploration curve ($y = a(1-e^{-bx})$) was fitted to estimate the mean number of unique sequences in each population or strain in SigmaPlot 7.0.

## Results

### Number of *DAB* alleles

We screened genomic copies of MHC amplified from 35 LA, 21 UA, five Red, five Blue and four Green guppies, and sequenced six clones per fish initially. The initial screen detected up to four distinct alleles in wild guppies, which suggests that our primers amplify a minimum of two *DAB* loci. Surprisingly, we detected only two *DAB* alleles in the sample of Green guppies, and one *DAB* allele in all Blue and Red strain samples. We subsequently re-amplified the DNA extractions with our degenerate primers to test whether we amplified all alleles in the ornamental strains. Analysis of 24–48 clones per ornamental fish detected only one more *DAB* allele, which was present in individuals of the Green strain. The discovery curves presented in Figure 1 clearly show that large differences exist in the number of *DAB* alleles detected when sequencing 100 random clones in ornamental strains and wild populations.

### MHC sequence variation

The PBR codons showed evidence for diversifying selection with a $\omega$-ratio significantly in excess of unity (mean($\pm$SE) = 2.28($\pm$0.16), one sample *t*-test, $t = 7.77$, $P < 0.0001$). Outside the PBR, the substitution pattern was consistent with purifying selection and characterised by a relative excess of synonymous substitutions, with a $\omega$-ratio significantly below unity (mean($\pm$ SE) = 0.58($\pm$0.017), one sample *t*-test, $t = -24.03$, $P < 0.0001$).

Figure 2 shows the relationship between the sequence divergence and the $\omega$-ratio of the PBR codons using pairwise comparisons between alleles. A significant curvilinear relationship exists between the level of polymorphism and $\omega$-ratio ($F_{2,222} = 17.24$; $P < 0.0001$). On average, the nucleotide substitution pattern of moderately diverged sequences (sequence divergence ranging between 8 and 25%) is characterised by a relative excess of non-synonymous substitutions ($\omega > 1$). Several pairwise comparisons between moderately diverged sequences show, however, surprisingly low $\omega$-ratios ($\omega \approx 1$). These comparisons are between *DAB* alleles of ornamental and wild guppies, and represent diverged *DAB* sequences with relatively similar PBRs.

Further inspection of Figure 2 shows that the $\omega$-ratio does not exceed unity when comparing highly divergent sequences (ie sequences diverged by more than 25%). These points represent pairwise comparisons that involve the Green *DAB* allele amplified with the degenerate primers. It tentatively suggests that during divergence over evolutionary timescales, the PBR accumulates silent substitutions that offset the initial excess in non-synonymous base changes.

Interestingly, the $\omega$-ratio is (close to) zero in five pairwise comparisons between very recently diverged sequences (ie percentage polymorphism <5%; see Figure 2). Because these sequences are separated by only a few polymorphisms, it is possible that insufficient mutations have accumulated to fuel diversifying selection, and consequently, the $\omega$-ratio does not show an excess of non-synonymous substitutions. As a result of these processes, the regression equations shows a significant quadratic term between sequence divergence of alleles and $\omega$-ratio of the PBR ($F_{1,222} = 133.05$, $P = 0.000001$).

Figure 3a shows the number of synonymous (d*S*) and non-synonymous (d*N*) substitutions for all 253 pairwise comparisons between alleles estimated using the method of Yang and Nielsen (2000). Base substitutions in PBR

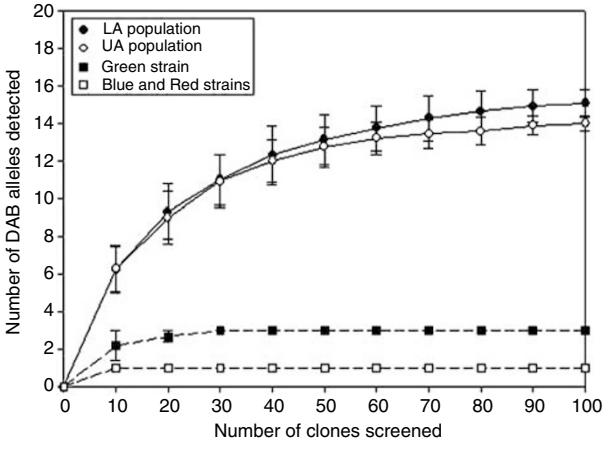


**Figure 1** Discovery curves of the mean ($\pm$95% CI) number of *DAB* alleles in bootstrapped samples of wild populations (solid lines) and ornamental strains (dashed lines). The LA (solid circles) and UA (open circles) show a very similar allelic richness and a much higher number of alleles than the Green (solid squares), and Red and Blue strains (open squares).

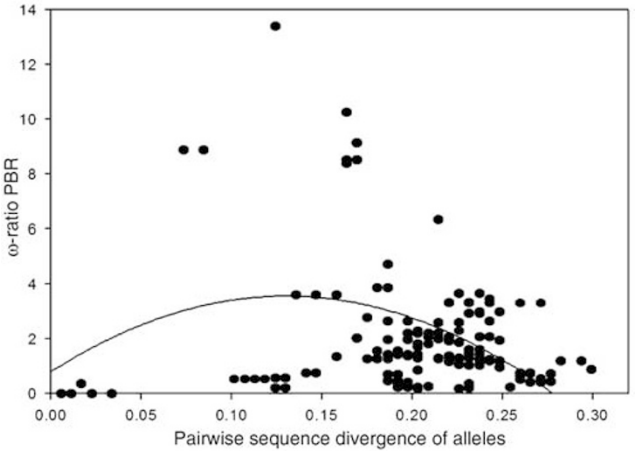


**Figure 2** Relationship between the total sequence divergence and the $\omega$-ratio of the PBR codons in pairwise comparisons between *DAB* alleles.

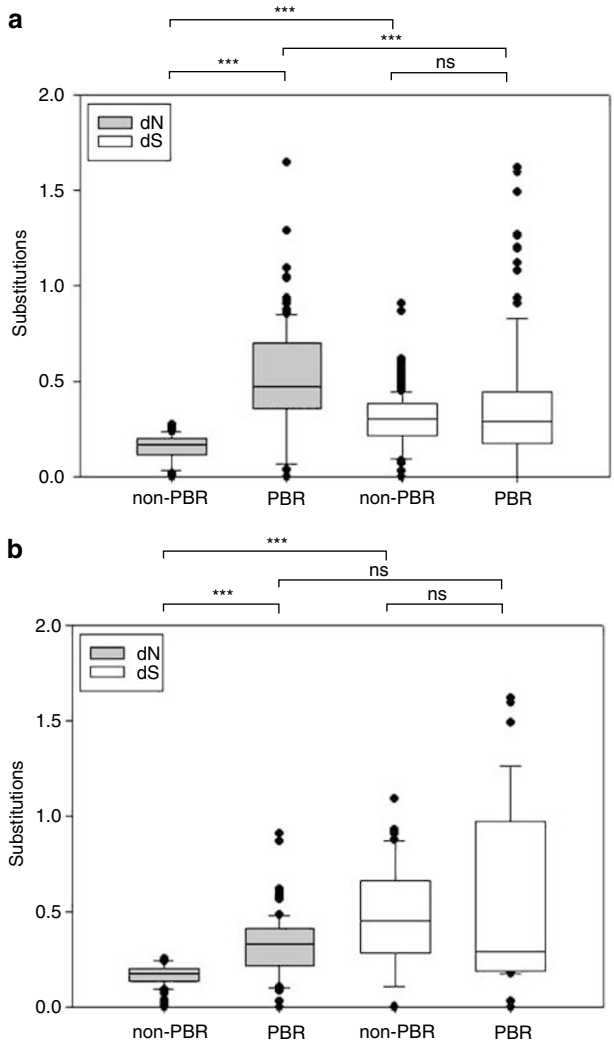


**Figure 3** (**a**) The rate of synonymous (d*S*) and non-synonymous (d*N*) substitutions at the PBR and non-PBR codons for all pairwise comparisons between alleles. See text for test results. (**b**) The rate of synonymous (d*S*) and non-synonymous (d*N*) substitutions at the PBR and non-PBR codons for pairwise comparisons between alleles of the ornamental lines and wild alleles. See text for test results.

and non-PBR codons are displayed separately. Four comparisons are particularly interesting: (1) PBR codons display a significantly higher d*N* than non-PBR codons (Mann–Whitney *U*-test: $W = 87420.0$, $P < 0.001$), which is evidence of diversifying selection. (2) The d*N* significantly exceeds the d*S* in the PBR (Mann–Whitney *U*-test: $W = 75377.0$, $P < 0.001$), which is also consistent with diversifying selection and evidence that non-synonymous substitutions are favoured in the PBR. (3) By contrast, d*S* exceeds d*N* significantly in the non-PBR codons (Mann–Whitney *U*-test: $W = 83903.0$, $P < 0.001$), which is consistent with purifying selection. (4) The d*S* is not significantly different between PBR and non-PBR codons (Mann–Whitney *U*-test: $W = 63797.0$, $P = 0.837$), which shows that the rates of synonymous substitutions are approximately equal between PBR and the non-PBR sites and suggests neutral evolution.

We then calculated the d*N* and d*S*, focusing on comparisons that include *DAB* alleles of ornamental strains *versus* wild strains only (see Figure 3b). As in the previous analysis, d*S* exceeds d*N* for the non-PBR codons significantly (Mann–Whitney *U*-test: $W = 8884.5$, $P < 0.001$), the d*N* is greater in the PBR than in the non-PBR (Mann–Whitney *U*-test: $W = 9227.5$, $P < 0.001$) and there is no significant difference in d*S* between PBR and non-PBR bases (Mann–Whitney *U*-test: $W = 6468.5$, $P = 0.305$). However, in contrast to the previous analysis, d*N* does not significantly exceed d*S* in the PBR (Mann–Whitney *U*-test: $W = 7049.5$, $P = 0.325$). This corroborates the regression analysis shown in Figure 2, in which the $\omega$-ratio was close to unity in pairwise comparisons between highly diverged *DAB* alleles of ornamental and wild guppies. Figure 3b shows that a large number of synonymous substitutions have accumulated since the divergence of these alleles, which offset the high rate of non-synonymous substitutions in the PBR, thereby rendering the $\omega$-ratio close to unity.

## Discussion

This is the first comparative analysis on partial sequences (196–219 bp) of genomic MHC class IIB (*DAB*) nucleotide variation between wild and ornamental guppies (*P. reticulata*). The study was undertaken with the aim of investigating the effects of domestication and selective breeding on the sequence variation of these immunological important loci. In samples of two natural Trinidadian guppy populations, we detected 15 or 16 MHC alleles, whereas only one to three *DAB* alleles were detected in the ornamental stocks. A previous study had shown that even though the wild populations differ substantially in their long-term effective populations sizes, random genetic drift in the small and genetically isolated headwater population of the Upper Aripo did not appear to have reduced the MHC variation noticeably (C van Oosterhout *et al*, unpublished manuscript). By contrast, domestications and selective breeding appear to have had a dramatic impact on the allelic richness of *DAB* loci in all ornamental lines examined. We will first discuss the sequence divergence in more detail and then examine three hypotheses that could explain differences in allelic richness of *DAB* loci between wild and ornamental guppies.

### MHC sequence variation

The birth-and-death model (Nei *et al*, 1997) and accordion model (Klein *et al*, 1993) are two important models of multigene evolution that can explain the generally large and variable number of MHC loci and high levels of polymorphism. According to these models, repeated tandem and block gene duplications generate the high level of diversity typical for MHC genes (Nei *et al*, 1997). Duplicated MHC sequences are thought to be governed by diversifying selection over a short evolutionary time, which may favour their specialisation (Hughes, 1999). Once the proteins of duplicated genes have become specialised for distinct functions, new amino-acid changes are no longer advantageous and purifying selection will again predominate (Hughes, 1999; Hughes and Friedman, 2004). Eventually, the number of synonymous substitutions will exceed the non-synonymous ones ($dS > dN$), resulting in an apparently atypical $\omega$-ratio at the PBR (ie $\omega$-ratio not in excess of unity).

Pairwise comparisons of sequence variation between MHC class IIB *DAB* genes of guppies in the current study support the hypothesis that the signal of diversifying selection is transient over evolutionary time. A significant curvilinear relationship was detected between the sequence divergence in pairwise comparisons of alleles and the $\omega$-ratio of their PBR codons. The $\omega$-ratio was close to unity in pairwise comparisons between highly diverged sequences. Interestingly, $\omega$-ratios were not significantly elevated in pairwise comparisons between very similar sequences, possibly because no beneficial non-synonymous substitutions had yet accumulated in the PBR. Moderately diverged sequences showed on average the highest $\omega$-ratios, with $\omega$-ratios in excess of 14 for some comparisons. As a consequence of these integrate effects of mutation and selection over evolutionary time, a curvilinear relationship exists between sequence divergence and signal of positive Darwinian selection as expressed in the $\omega$-ratio at the PBR. This relationship reflects that only when sufficient genetic variation is generated by mutations, evidence for divergent selection can be detected at the PBR. However, the signal of diversifying selection inevitably will be eroded over time because of the ongoing accumulation of synonymous substitutions.

The $dN/dS$ (or $\omega$-ratio) is a popular measure to infer positive selection. However, the current study suggests that the $\omega$-ratio may be close to unity (or zero) despite putative balancing selection. Furthermore, there are two distinct reasons for an atypical $\omega$-ratio; either no non-synonymous mutations have accumulated for diversifying selection to operate on, or the evidence for diversifying selection has become eroded over evolutionary time (Hughes and Friedman, 2004). By comparing $dN$ and $dS$ between PBR and non-PBR regions in conjunction, the evolution of substitution patterns can be understood more comprehensively. An additional limitation of the $\omega$-ratio is that its value is undetermined when no synonymous substitutions have accumulated ($dS = 0$). Hence, the analysis of $dN$ and $dS$ separately offers a further advantage over the use of $\omega$-ratio alone and we advocate its use, particularly in population genetic and phylogenetic studies on the MHC.

### MHC variation and models of multigene evolution

The most intriguing finding of our study is the large difference in the number of unique MHC class IIB *DAB* alleles between the wild and ornamental guppy populations. In population samples of the Blue, Red and Green strains, we detected one or three *DAB* alleles. Wild populations, on the other hand, possessed many *DAB* alleles, and we detected 15 and 16 alleles in our population samples of the Upper and Lower Aripo, respectively. We will discuss three hypotheses that could explain the differences in sequence numbers between wild populations and ornamental lines: (i) the presence of non-amplification (null alleles); (ii) the fixation of alleles during selective breeding; and (iii) multigene evolution reducing the allelic richness of *DAB*.

The low number of *DAB* alleles in the ornamental guppies could be caused by failure to amplify or detect all *DAB* alleles due to mutations at the primer sites. We tested for the presence of these so-called null alleles by designing degenerate primers, internal to the original primers. These new primers were highly conserved in all our sequences, as well in those of the related species (*X. maculatus*). In addition, these primers amplify *DAB* alleles in two closely related poecilids (*P. picta* and *P. sphenopsis*) (SM Cummings and C van Oosterhout, unpublished manuscript). However, these new primers amplified only one novel *DAB* sequence from Green fish (in addition to two previously amplified alleles in this line). No novel sequences were amplified from either the Red or the Blue strains using these degenerate primers. In addition, we did not detect new sequences in the expression study in which we amplified *DAB* sequences using a degenerative forward primer and reverse primer designed in the conserved exon 3. Although we cannot completely rule out the possibility that some *DAB* alleles remained undetected, the large differences between the wild and ornamental gene pools are unlikely to be explained by null alleles alone.

A second possible explanation involves the effects of inbreeding during $>100$ generations in captivity. Selective breeding of ornamental guppy strains is expected to result in extraordinary levels of random genetic drift as colour pattern mutations of single genes occur very sporadically and in few animals (Winge, 1922; Haskins and Haskins, 1951). Colour pattern strains are commonly started with a single founder male that carries the desirable mutation, whose offspring is subsequently artificially selected to improve the quality of the desired trait. Prolonged selection further increases inbreeding and level of homozygosity (van Oosterhout *et al*, 2003), which would ultimately result in the fixation of alleles at loci. However, we would expect to find three rather than just one *DAB* allele in each strain, corresponding to the number of *DAB* loci inferred in our previous study (van Oosterhout *et al*, unpublished MS). The present study also shows there are at least two *DAB* loci, as we detected four alleles in six randomly amplified clones of some wild guppies. Inbreeding cannot fully explain why we only detected a single *DAB* allele in two ornamental lines.

The third hypothesis postulated to explain differences in the number of *DAB* alleles between wild and ornamental guppies involves models of multigene evolution, in particular gene conversion and the accordion model of MHC evolution (Klein *et al*, 1993). In extant bony fishes, the number of duplicated MHC genes are thought to range from 1 to 42, with the more primitive euteleosts (eg carp, zebrafish, salmon) having a low number and the more advanced neoteleosts (cichlids, sticklebacks, cod) a large number ($\geq 17$) of duplicated MHC genes (Miller *et al*, 2002). Studies on cichlids, swordtails and sticklebacks show that the MHC class I and IIB copy number varies within individuals and populations of a single species (Malaga-Trillo *et al*, 1998; Figueroa *et al*, 2001; Reusch *et al*, 2001). According to the accordion model of multigene evolution, the MHC copy number expands and contracts depending on the diversity and selection by parasites (Klein *et al*, 1993). The parasite diversity and selection intensity by parasites is probably reduced considerably in a captive environment compared to the wild. Consequently, MHC genes may have been deleted from the genome of ornamental strain guppies, thereby reducing the number of *DAB* alleles below the number expected based on drift alone.

Alternatively, rather than a reduced number of loci, the same *DAB* allele might be fixed at different genetic loci. Multigene families whose member genes have similar functions are believed to undergo concerted evolution (for review see Martinsohn *et al*, 1999). During concerted evolution, gene conversion can change the number of copies of a particular allele within an individual owing to non-reciprocal exchange of DNA between two sequences at different genetic loci. Eventually, this will homogenise the DNA sequences of genes in the gene family, as the same sequence is fixed at different loci (Nei *et al*, 1997). Also recombination within a locus (intergenic gene conversion) will inevitably homogenise sequences of member genes. The MHC class II sequence variation in birds seems to support concerted evolution (Garrigan and Edwards, 1999; Wittzell *et al*, 1999; Hess and Edwards, 2002), and inter-locus recombination has recently also been reported in a teleost, *Gasterosteus aculeatus* (Reusch *et al*, 2004). In the Red and Blue strains, we detected only one *DAB* allele, an observation that appears to be inconsistent with one locus and the amplification of up to four *DAB* alleles in some wild Aripo guppies. Exhaustive screening with different primer combinations did not reveal more alleles. We therefore expect that the presence of only one allele in the Red and Blue strains is not an artefact of null alleles. Rather, we suggest that the same *DAB* allele is either fixed at multiple MHC loci owing to gene conversion, or that the number of loci in the ornamental strains has contracted to only one *DAB* locus, a hypothesis consistent with the accordion model of MHC evolution (Klein *et al*, 1993). Over 100 generations of selective breeding and inbreeding, ornamental guppies may have experienced the effects of multigene evolution, such as gene deletion and/or inter-locus recombination. These processes are likely to resemble those occurring in various strains of inbred mice, which are known to vary in the genomic organisation and gene copy number of MHC class IA loci (Stroynowski, 1990; Yun *et al*, 1997). Future MHC studies using backcrossed ornamental–wild hybrids will shed further light on the number of *DAB* loci, the locus-affiliation of their alleles and the evolutionary mechanism responsible for the apparent fast decline of MHC variation in ornamental guppy strains.

## Acknowledgements

We thank Joanne Cable for helpful comments on the manuscript and Alan Smith for help with fish husbandry. This work was funded by NERC Grants NER/M/S/2002/00101 and NER/I/S/2000/00885 and Leverhulme Trust Grant F/00 181/F awarded to CVO.

## References

Aguilar A, Roemer G, Debenham S, Binns M, Garcelon D, Wayne RK (2004). High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. *Proc Nat Acad Sci USA* **101**: 3490–3494.

Axelrod HR, Whitern W (1965). *Guppies*. TFH Publications: Jersey City, NJ. 64pp.

Dzwillo M (1959). Genetische Untersuchungen an Domestizierten Stammen von *Lebistes retilulatus* (Peters). *Mitt Hamburg Zool Mus Inst* **57**: 143–186.

Fernando AA, Phang VPE (1985). Culture of the guppy, *Poecilia reticulata*, in Singapore. *Aquaculture* **51**: 49–63.

Fernando AA, Phang VPE (1990). Inheritance of Red and Blue caudal fin colorations in two domesticated varieties of the guppy, *Poecilia reticulata*. *J Aqua Trop* **5**: 209–217.

Figueroa F, Mayer WE, Sultmann H, O'Huigin C, Tichy H, Satta Y *et al* (2001). Mhc class IIB gene evolution in East African cichlid fishes. *Immunogenetics* **51**: 556–575.

Filatov DA (2002). ProSeq: a software for preparation and evolutionary analysis of DNA sequence datasets. *Mol Ecol Notes* **2**: 621–624.

Garrigan D, Edwards SV (1999). Polymorphism across an exon–intron boundary in an avian Mhc class IIB gene. *Mol Biol Evol* **16**: 1599–1606.

Gordon M (1955). *Guppies as Pets*. TFH Publications: Neptune City, NJ.

Gu X, Nei M (1999). Locus specificity of polymorphic alleles and evolution by a birth-and-death process in mammalian MHC genes. *Mol Biol Evol* **16**: 147–156.

Haskins CP, Haskins EF (1951). The inheritance of certain colour patterns in wild populations of *Lebistes reticulates* in Trinidad. *Evolution* **5**: 216–225.

Hedrick PW (2002). Pathogen resistance and genetic variation at MHC loci. *Evolution* **56**: 1902–1908.

Hedrick PW (2004). Comment on 'Parasite selection for immunogenetic optimality'. *Science* **303**: U1.

Hedrick PW, Kim TJ, Parker KM (1998). Parasite resistance and genetic variation in the endangered Gila topminnow. *Animal Cons* **4**: 103–109.

Hedrick PW, Thomson G (1983). Evidence for balancing selection at HLA. *Genetics* **104**: 449–456.

Hess CM, Edwards SV (2002). The evolution of the major histocompatibility complex in birds. *Bioscience* **52**: 423–431.

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.

Hogstrand K, Bohme J (1999). Gene conversion can create new MHC alleles. *Immun Rev* **167**: 305–317.

Hughes AL (1999). *Adaptive Evolution of Genes and Genomes*. Oxford University Press: New York.

Hughes AL (2002). Natural selection and the diversification of vertebrate immune effectors. *Immunol Rev* **190**: 161–168.

Hughes AL, Friedman R (2004). Recent mammalian gene duplications: robust search for functionally divergent gene pairs. *J Mol Evol* **59**: 114–120.

Hughes AL, Yeager M, Carrington M (1996). Peptide binding function and the paradox of HLA disease associations. *Immunol Cell Biol* **74**: 444–448.

Kanagawa T (2003). Bias and artifacts in multitemplate polymerase chain reactions (PCR). *J Biosci Bioeng* **96**: 317–323.

Khoo G, Lim TM, Chan W-K, Phang VPE (1999). Genetic basis of the variegated tail pattern in the guppy, *Poecilia reticulata*. *Zool Sci* **16**: 431–437.

Klein J, Ono H, Klein D, O'hUigan C (1993). The accordion model of MHC evolution. *Prog Immunol* **8**: 137–143.

Kobayashi N, Tamura K, Aotsuka T (1999). PCR error and molecular population genetics. *Biochem Gen* **37**: 317–321.

Kumar S, Tamura K, Jakobsen IB, Nei M (2001). *MEGA2: Molecular Genetics Analysis Software*. Arizona State University: Tempe.

Leberg PL (2002). Estimating allelic richness: effects of sample size and bottlenecks. *Mol Ecol* **11**: 2445–2449.

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

Malaga-Trillo E, Zaleska-Rutczynska Z, McAndrew B, Vincek V, Figueroa F, Sultmann H *et al* (1998). Linkage relationships and haplotype polymorphism among cichlid Mhc class II B loci. *Genetics* **149**: 1527–1537.

Martinsohn JT, Sousa AB, Guethlein LA, Howard JC (1999). The gene conversion hypothesis of MHC evolution: a review. *Immunogenetics* **50**: 168–200.

McConnell TJ, Godwin UB, Cuthbertson BJ (1998a). Expressed major histocompatibility complex class II loci in fishes. *Immunol Rev* **166**: 294–300.

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

Milinski M (2003). The function of mate choice in sticklebacks: optimizing Mhc genetics. *J Fish Biol* **63**(Suppl A): 1–16.

Miller KM, Kaukinen KH, Schulze AD (2002). Expansion and contraction of major histocompatibility complex genes: a teleostean example. *Immunogenetics* **53**: 941–963.

Nei M, Gu X, Sitnikova T (1997). Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc Natl Acad Sci USA* **94**: 7799–7806.

Ober C, van der Ven K (1997). Immunogenetics of reproduction: an overview. *Reprod Immunol Curr Top Microbiol Immunol* **222**: 1–23.

Penn DJ (2002). The scent of genetic compatibility: sexual selection and the major histocompatibility complex. *Ethology* **108**: 1–21.

Reusch TBH, Haberli MA, Aeschlimann PB, Milinski M (2001). Female sticklebacks count alleles in a strategy of sexual selection explaining MHC polymorphism. *Nature* **414**: 300–302.

Reusch TBH, Schaschl H, Wegner KM (2004). Recent duplication and inter-locus gene conversion in major histocompatibility class II genes in a teleost, the three-spined stickleback. *Immunogenetics* **56**: 427–437.

Sato A, Figueroa F, O'hUigin C, Reznick DN, Klein J (1995). Identification of major histocompatibility complex genes in the guppy, *Poecilia reticulata*. *Immunogenetics* **43**: 38–49.

Stroynowski I (1990). Molecules related to class-I major histocompatibility complex antigens. *Annu Rev Immunol* **8**: 501–530.

Truett GE, Heeger P, Mynatt RL, Truett AA, Walker JA, Warman ML (2000). Preparation of PCR-quality mouse genomic DNA with hot sodium hydroxide and tris (HotSHOT). *Biotechniques* **29**: 52.

van Oosterhout C, Trigg RE, Carvalho GR, Magurran AE, Hauser L, Shaw PW (2003). Inbreeding depression and genetic load of sexually selected traits: how the guppy lost its spots. *J Evol Biol* **16**: 273–281.

Winge O (1922). One-sided masculine and sex-lined inheritance in *Lebistes reticulatus*. *J Genet* **12**: 145–162.

Wittzell H, Bernot A, Auffray C, Zoorob R (1999). Concerted evolution of two Mhc class II B loci in pheasants and domestic chickens. *Mol Biol Evol* **16**: 479–490.

Yang Z (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* **13**: 555–556.

Yang Z, Nielsen R (2000). Estimating synonymous and non-synonymous substitution rates under realistic evolutionary models. *Mol Biol Evol* **17**: 32–43.

Yun TJ, Melvold RW, Pease LR (1997). Complex major histocompatibility complex D locus variant generated by an unusual recombination mechanism in mice. *Proc Natl Acad Sci USA* **94**: 1384–1389.

Supplementary Information accompanies the paper on Heredity website (http://www.nature.com/hdy)