Host-parasite co-evolution and genetic variation at the Major Histocompatibility Complex in the Trinidadian guppy (*Poecilia reticulata*)

A Thesis submitted for the Degree of PhD.

in the University of Hull

by

Mark M^{c}Mullan, Biology BSc (Hons)

August, 2010

Acknowledgments

This study was funded by the NERC Ph.D. studentship to M.M. (NER/S/A/2005/13362A). Ryan Mohammed assisted with all aspects of the work conducted in Trinidad including the collection of guppies, collection of site data, provision of food shelter and almost all other basic requirements in Trinidad and Tobago, Bettina helped. Alan Smith maintained live stocks in the laboratory in Hull. Dr. Nicola Barson allowed the adaption of her map for use in this thesis. Dr. Jo Cable parasite screened all guppies and along with Dr. Violaine Llaurens Dr. Domino Joyce and Dr. Cock van Oosterhout provided critical advice on the structure and content of this Thesis.

I owe a personal thank you to John, for buying the beers, and to Rich for drinking them.

Thank you Angela for your love and patience.

I would also like to thank Cock, for taking the time.

Thank you Dr Bernd Hänfling and Dr Stuart Piertney for reviewing this thesis.

Contents

Abstract

The Major Histocompatibility Complex (MHC) is a region of the vertebrate genome believed to be responsible for an individual's ability to detect and recognise invading parasites. The MHC molecule has been shown to bind to short fragments of parasite and present these to the adaptive immune system. Theory developed to describe the maintenance of polymorphism within the MHC has focussed on the principle that different MHC alleles recognise different groups of parasites and that parasite diversity maintains diversity in the MHC (i.e. Red Queen dynamics). Nevertheless, after 50 years of research, the precise mechanism for the maintenance of extraordinary levels of polymorphism in the MHC remains yet to be resolved.

In the present thesis, I use guppies (*Poecilia reticulata*) as a model to investigate the role of parasites in the maintenance of polymorphism in the MHC. In a study of spatial variation of both MHC and microsatellite variation, I find evidence to suggest that upstream populations of guppies have small population sizes and reduced gene flow into the population. However, these populations of guppies maintain similar levels of MHC polymorphism to that of larger populations of guppies further downstream. The maintenance of MHC in small upstream populations provides evidence for selection for the maintenance of MHC polymorphism despite the effect of random genetic drift. This finding is particularly interesting given that I show that this small upstream population has a significantly reduced parasite fauna. These data therefore provide evidence for other sources of selection on the MHC such as sexual selection or Associative Balancing Complex (ABC) evolution.

Balancing selection maintains polymorphism above that expected under a neutral model of evolution. However, balancing selection is a general term that encompasses several distinct mechanisms, including negative frequency dependent selection (rare allele advantage), overdominance (heterozygote advantage) and selection that favours distinct alleles in different times and/or places (fluctuating selection). I use a temporal dataset of MHC and microsatellite variation to distinguish between these different models of balancing selection. In particular, I evaluate the ability of the different models of balancing selection to explain the empirical data of two guppy populations sampled in 2001 and 2007. I conclude that the relatively low level of spatial genetic divergence of the MHC in 2001 is most consistent with the overdominance and negative frequency dependent selection models of balancing selection. By contrast, the data in 2007 suggest that MHC is subject to fluctuating selection, showing a higher level of spatial genetic differentiation than the microsatellites. Over time, the MHC appears to change more rapidly than neutral microsatellite loci. Using a verbal model, I argue that this pattern of temporal genetic divergence and the rapid turnover of MHC alleles is consistent with all types of balancing selection. Balancing selection increases the effective migration rate, resulting in a rapid differentiation of the MHC gene pool over generations. Such allelic turnover can however only be realised in a metapopulation with a large gene pool of MHC alleles. Both these conditions (i.e. a source-sink metapopulation with circa 85 MHC alleles) have recently been demonstrated for guppy populations in the Caroni Drainage in Trinidad by different authors. Importantly, these findings demonstrate that in an open metapopulation, the impact of migration can differ dramatically between neutral genes and genes under balancing selection.

Using a simulation model, I further explore whether the combination of balancing selection (in particular, overdominance) and gene flow in a metapopulation system can explain the large observed spatio-temporal differentiation of the MHC. Traditionally, authors have interpreted large temporal fluctuations in MHC allele frequencies as evidence for a coevolutionary arms race between host immune genes and parasite virulence genes (i.e. Red Queen dynamics). In this theoretical chapter, I explore whether such data can also be explained by simple overdominant selection in combination with migration from a source population with many distinct MHC alleles. I find that the commonly held assumption that balancing selection homogenises gene frequencies and reduces the level of genetic differentiation (G'_{ST}) is not always correct, and it depends on the interaction between evolutionary forces and population demography. Furthermore, I demonstrate that balancing selection (overdominance) can explain the rapid turnover in MHC alleles, and that this observation should not be taken as evidence of Red Queen dynamics through host-parasite co-evolution.

Altogether, this thesis highlights two main considerations that should be made in future studies of the MHC. Firstly, other sources of balancing selection should be considered in addition to parasite selection, particularly where no causal relationship between parasites and MHC alleles has been identified. Secondly, population demography can have a different impact on the population genetics of the MHC compared to that of neutral loci. The effect of a higher effective migration rate on MHC alleles is all too often interpreted as evidence for changes in the direction of parasite-mediated selection (i.e. Red Queen dynamics). However, theoretically, other forms of balancing selection, including overdominance, negative frequency dependent selection, fluctuating selection and ABC evolution, can also drive the temporal dynamics of the MHC.

General Introduction

Chapter 1

Evolution by natural selection is a slow process (Darwin 1859) and documenting natural selection is made difficult because changes in the environment are also slow. Furthermore, it is difficult to predict the source of natural selection and a phenotypic response to a change in an organism's ecology. Nevertheless, selection has been documented in the lab, in bacteria (Barrick et al. 2009; Buckling et al. 2009; Paterson et al. 2010). Because the length of time between one generation and the next is short, selection acting over many thousands of generations can be measured over a relatively short period (see Barrick et al. 2009).

Longer generation times reduce the amount of change one might expect to observe in any period. One vertebrate example where predation has been shown to drive variation in life history traits is the Trinidadian guppy, *Poecilia reticulata* (Peters) (Reznick and Endler 1982; Reznick and Bryga 1987; Reznick et al. 1997). The diversity of habitats from the uplands to the lowlands made this fish species an ideal candidate to explore the effects natural selection (see Houde 1997; Magurran 2005). The effects of other abiotic factors such as canopy cover and water temperature have also been implicated in the evolution of guppy life history traits (reviewed by Endler 1995; see also Bassar et al. 2010). However, until recently, the role of parasites (defined here in the broadest sense to include all micro- or macro-parasitic infection) as a source of selection on guppies had been ignored (Cable In press).

Antagonistic co-evolution between host and parasite has been shown in invertebrates to lead to selection over relatively short timescales (see Dybdahl and Lively 1998; Lively and Dybdahl 2000; Decaestecker et al. 2007). Genetic changes within or among populations are the best source of evidence for adaptation in response to selection as oppose to phenotypic plasticity. However, the effect of genetic change on morphology can be difficult to determine (see Ford 2002). In this thesis, I contrast the effects of selection on a gene complex involved in host parasite interactions, with the neutral effects of population dynamics. Nucleotide changes within the coding region this gene complex are believed to be directly related to individual immunocompetence (see review by Hughes and Yeager 1998). Therefore, maintenance of polymorphism within these genes above that observed at neutral loci can be taken as evidence for the effects of natural selection acting in wild populations.

The Major Histocompatibility Complex (MHC) is a large multigene family present in all jawed vertebrates (Klein et al. 2000; see also Danchin et al. 2004; Kelley et al. 2005). It has a central role in immunocompetence and is implicated in mate choice and sexual selection (Apanius et al. 1997; Edwards and Hedrick 1998; Hughes and Yeager 1998; Penn and Potts 1999; Bernatchez and Landry 2003; Wegner et al. 2004; Sommer 2005; Piertney and Oliver 2006). The histocompatibility genes were first studied in relation to transplantation compatibility in humans and mice (Dausset 1958). The MHC is implicated with the detection of invading pathogens by recognition of small protein fragments known as peptides, or antigens.

Identification of the three dimensional structure of MHC class I and class II molecules revealed its role in recognition and binding of antigenic peptides and their presentation to T-cells to initiate a specific adaptive immune response (Bjorkman et al. 1987a, b; Brown et al. 1993;

Stern et al. 1994; see also Blais et al. 2007, for comparison in cichlid fish). The number of genes in the MHC varies enormously between species. For instance, there are over 200 MHC genes (including pseudogenes) within the human MHC (the human leukocyte antigen or HLA (The-MHC-sequencing-consortium 1999; Shiina et al. 2004) and just 19 within the chicken MHC (B locus) (Kaufman et al. 1999). The number of MHC genes can also vary between individuals of the same species (Malaga-Trillo et al. 1998; Figueroa et al. 2001). Variation in the number of genetic loci within the MHC is predicted by various multigene models, including the accordion model of MHC evolution (Klein et al. 1993), the Birth-and-death model and concerted evolution (see review by Nei and Rooney 2005).

The MHC is characterised by extraordinary levels of polymorphism, which is illustrated by the human MHC HLA-B locus, which has over 1000 alleles (Robinson et al. 2003) (IMGT/HLA Database, 2008, http://www.ebi.ac.uk/imgt/hla/stats.html). This level of polymorphism varies at different MHC genes (Robinson et al. 2003) (IMGT/HLA Database, 2008), as well as within and among populations (Wegner et al. 2003b; van Oosterhout et al. 2006b; Dionne et al. 2007). The highest level of polymorphism is observed within a specific region of the MHC known as the Peptide Binding Region (PBR) (Hughes and Yeager 1998). This region is responsible for the specificity of the MHC molecule to bind particular peptides from different sources, both micro- and macroparasites (Bjorkman et al. 1987a; Stern et al. 1994)

The abundance and composition of pathogens within populations is a mechanism proposed to maintain MHC polymorphism, and much of the work on this subject has focused on humans (Shiina et al. 2004) and mice (Penn and Potts 1999). However, studies on the MHC of non-model organisms increase understanding of the environmental factors required for the maintenance of MHC polymorphism (Bernatchez and Landry 2003; Sommer 2005; Piertney and Oliver 2006). Sexual selection for "good genes" or increased genetic diversity has also been proposed to maintain MHC polymorphism in many species (Penn and Potts 1999; Milinski 2006).

The research presented here focuses on MHC variation in the Trinidadian guppy. I measure genetic variation at MHC and neutral loci and quantify parasite fauna at three locations of the Aripo River in the Northern Mountain range of Trinidad (Fig. 1). In this first Chapter, I introduce, briefly the structure and function of the MHC. I consider the evolution of the MHC and the mechanisms of balancing selection proposed to maintain MHC polymorphism. I introduce the MHC of poeciliids and then present relevant work on the MHC of the guppy. This text will shortly be published as a book chapter (McMullan and van Oosterhout in press see also Appendix 14). Finally, I outline the research conducted and the structure of this thesis.

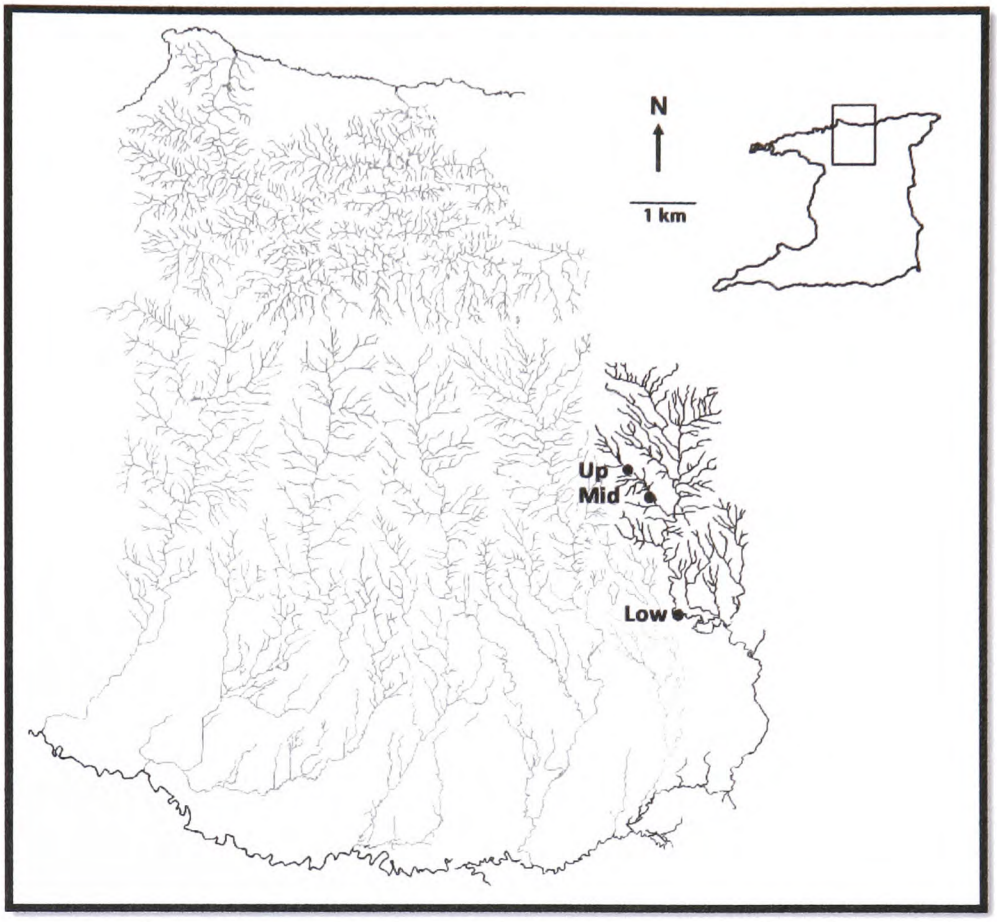


Figure 1. Map of Northern Trinidad showing the Upper Naranjo (Up), Mid Naranjo (Mid) and Lower Aripo (Low) populations sampled in the present study. The inset shows the location of the detailed map within Trinidad. Map adapted from Barson et al. (2009) with the permission of Dr. N. J. Barson.

The Major Histocompatibility Complex (MHC)

The MHC multigene family encodes a number of proteins responsible for the preparation, loading and specific binding of antigen fragments (reviewed by Ting and Trowsdale 2002; Boss and Jensen 2003; Danchin et al. 2004). Divided into classical and non-classical categories the classical MHC genes are typically highly polymorphic and separated into the MHC class I and MHC class II subfamilies. Both class I and class II genes are responsible for binding antigen (peptide) fragments (Hughes and Yeager 1998; Klein and Sato 1998). The non-classical MHC genes display low levels of polymorphism and gene expression, and the DNA sequences of these genes generally show little evidence of positive selection (Summers et al. 2008). The class III region contains a group of genes genetically and evolutionarily unrelated to class I or class II, although some of these genes have immune function, predominantly in the innate immune response (Hughes and Yeager 1998; Klein and Sato 1998). MHC class I and class II are linked in mammals and birds, but they occur unlinked on different chromosomes in teleost fishes (Sato et al. 2000; Stet et al. 2003; Kelley et al. 2005). Consequently, some authors refer to teleost MHC as the MH (see Ellis et al. 2006; Dixon 2008), although the more traditional abbreviation, MHC, is used here.

Class I and class II molecules present different types of antigens to different types of T-cells. Class I molecules are expressed on the surface of all nucleated somatic cells and present

peptides originating from within the cell, such as viral peptides. Class II molecules are expressed primarily on antigen presenting cells of the immune system, and they specialise in the presentation of extracellular pathogen peptides (e.g. bacteria and microparasites) (see review by Hughes and Yeager 1998). MHC class I and class II subfamilies rely on different antigen processing pathways to load antigen into the MHC molecule. Once degraded, antigen fragments are bound to the MHC glycoprotein and presented to T-cells. T-cells initiate the adaptive immune response which acts to target the pathogen specifically. T-cells need to respond only to foreign antigens and not to the host's own tissue. In order to recognise "self" from "non-self", autoreactive thymocytes with T-cell receptors are eliminated during thymic selection in the developing embryo (Jameson et al. 1994; Jordan et al. 2001).

Much population and evolutionary genetic research in non-model organisms seem biased toward the class II over class I MHC. This could be because of the relative ease of quantifying microparasite fauna over viruses. The class II molecule is a heterodimer consisting of an α chain and a β chain (class II*A* and class II*B,* respectively). Each of the monomers encodes part of the PBR (see Hughes and Yeager 1998). The PBR region of the α chain has been shown to have low levels of polymorphism in the chicken, a model for the MHC (Kaufman et al. 1999; Salomonsen et al. 2003). This may be why the majority of the research in non-model organisms has focussed on the polymorphic exon 2 of class II*B* region (but see Figueroa et al. 2001; de Eyto et al. 2007; Gomez et al. 2010). Within the β monomer positive selection is only evident in exon 2, and consequently, this exon shows the highest level of polymorphism of the MHC. Other exons of the MHC class II*B* are relatively conserved, and the sequence variation in these exons is governed mainly by purifying selection and neutral evolutionary forces.

Evolutionary genetics and molecular ecology of the MHC

Evolution of multigene families

Three main hypotheses have been proposed for the variation observed in the number of classical MHC genes, these are divergent evolution, concerted evolution and birth-and-death evolution (reviewed by Nei & Rooney, 2005). These mechanisms rely on gene duplication, mutation, recombination and gene conversion for the generation of new genes (Nei and Rooney 2005). The precise mechanisms of gene duplication are unknown although whole genome duplications, unequal crossing over, retroposition and segmental duplication are all proposed to be responsible for different duplication events (Zhang 2003).

Duplicated loci may be maintained as functional genes or they may be lost as pseudogenes (see reviews by Balakirev and Ayala 2003; Zhang 2003). The resulting relationship between genes maintained under divergent selection is a reduced divergence between orthologous genes, concerted evolution maintains genes with reduced divergence between paralogous genes, and birth-and-death evolution is a combination of these previous two mechanisms (see review by Nei and Rooney 2005). Birth-and-death evolution is the process by which genes created by duplication are maintained or lost from the multigene family (Nei et al. 1997). Gene conversion (and recombination) may act to increase or decrease genetic diversity in duplicated genes (Nei et al. 1997; Nei and Rooney 2005). Evidence for duplication and loss (deletion) of expressed genes is observed in the MHC of mammals, where some orthologous genes can be shared between species, known as trans-species polymorphism (TSP) whilst others are not (Klein et al. 1993; Klein and Sato 1998; Nei and Rooney 2005). Klein et al. (1993) developed this

idea as an expansion and contraction of MHC genes called the Accordion model. The contraction of genes is thought to be facilitated by pseudogenes reducing the number of translated genes within the multigene family.

Nowak et al. (1992) suggested that too many MHC genes can reduce the number of T-cells as a result of thymic selection (see reviews by Penn and Potts 1999; Milinski 2006). During thymic selection, T-cells that bind with too high an affinity to an individual's MHC molecules are eliminated before birth. Hence, increased MHC diversity may lead to the elimination of more T-cells during thymic selection and reduce an individual's immunocompetence. Reduction of immunocompetence through negative (thymic) selection of T-cells is the bases for the hypothesis for optimum MHC variability (see Milinski 2006)

Balancing selection and the MHC

In the last three decades, molecular ecologists have studied neutral evolution using allozymes, minisatellites, microsatellites and SNPs (Hedrick 2005a). Neutral theory asserts that the majority of new mutations are either deleterious and eliminated from a population by negative selection, or (nearly) neutral and subject to random genetic drift (see reviews byFord 2002; Garrigan and Hedrick 2003; Nielsen 2005). Neutral polymorphism is maintained in a mutation-drift equilibrium, and this variation can be used to analyse genealogies, establish paternities, track population migration, measure gene flow and estimate effective population size.

More recently, with the advent of economically viable sequencing and the identification of genes under Balancing selection, research begun to focus on adaptive evolution. Adaptive evolution is driven by positive selection promoting novel (favourable) sequence variants. These novel mutations may be fixed in the population or, under balancing selection maintained as a polymorphism. Genes that are evolving under balancing selection are distinct from neutral loci in that they show high levels of heterozygosity and a large numbers of alleles at similar frequencies (hence the term *balancing* selection). Some alleles can be maintained over great lengths of time, possibly exceeding speciation events, a phenomenon called trans-species polymorphism (reviewed by Klein et al. 1998). Furthermore, balancing selection can also affect the DNA substitution pattern, resulting in an elevated d_N/d_S ratio in the PBR and a reduced nucleotide divergence in the introns (see Apanius et al. 1997; Hughes 2002 and references therein).

Balancing selection can be inferred using several types of analyses of sequence data and the d_N/d_S ratio is one commonly used method (see reviews by Ford 2002; Garrigan and Hedrick 2003). The d_N/d_S value is essentially a ratio of the number of nonsynonymous mutations (i.e. amino acid changing mutations) to synonymous ones (also known as silent substitutions). Synonymous mutations do not change the protein produced by the gene and are predicted to be selectively neutral. Consequently, synonymous mutations are accumulated and lost in a neutral fashion and these mutations will occur in drift-mutation equilibrium (Crow and Kimura 1970; Hedrick 2005a). In contrast, nonsynonymous mutations can alter protein function and such mutations are thus usually subject to natural selection. Mutations that change the fitness of the organism can be either favourable (positively selected), or deleterious (negatively selected). In conserved genomic regions, nonsynonymous mutations tend to be purged by negative selection, as changes to the protein structure are generally deleterious. Synonymous mutations, on the other hand, can persist at a relatively high frequency, and consequently, in

conserved gene regions the d_N/d_S ratio will be below unity ($d_N/d_S<1$). However, in gene regions where evolutionary change is favourable, positive selection is predicted to maintain nonsynonymous mutations, resulting in an elevated d_N/d_S ratio ($d_N/d_S\geq1$). In the MHC, the protein region interacting with foreign peptides is transcribed from the codons of the PBR. The codons in this region are thought to accumulate nonsynonymous mutations as these allow the recognition of novel parasites. By contrast, the areas outside the PBR are conserved, and nonsynonymous mutations are purged from these gene regions. Studies on the MHC have been facilitated by the fact that the PBR codons of the human MHC have been identified through X-ray crystallography (Bjorkman et al. 1987b, a; Brown et al. 1993; Stern et al. 1994), and that the MHC is highly conserved across vertebrates (but see Blais et al. 2007 for differences in PBR codons between humans and cichlid fish). This allows researchers working on the MHC of other vertebrate taxa *a priori* to identify the codons under positive and negative selection.

Balancing selection plays a pivotal role in the evolution of the MHC and is thought to maintain polymorphism at a locus by one (or combination) of the following four processes: (1) heterozygote advantage (overdominant selection) (Doherty and Zinkernagel 1975a, b; Penn and Potts 1999), (2) rare allele advantage (negative frequency dependent selection) (Kojima 1971; Takahata and Nei 1990), (3) selection varying in time or space (spatial or temporally variable selection, here fluctuating selection) (Hedrick et al. 1976; Hedrick 2002), or (4) selection against recessive deleterious load (ABC evolution) (van Oosterhout 2009a, b).

1. Overdominant selection: Heterozygous individuals are thought to be able to recognise more pathogens than homozygous individuals, and therefore have a superior immune response and higher fitness (Doherty and Zinkernagel 1975b). Landry et al. (2001) demonstrated selection for increased proportions of Atlantic salmon offspring heterozygous at the MHC class II*B* locus. Similarly, Arkush et al. (2002) demonstrated increased survival of heterozygous MHC class II*B* Chinook salmon to infectious hematopoietic necrosis virus (IHNV). The level of inbreeding (inbred or out-bred) was not found to correlate with IHNV survival but was associated with increased resistance to another myxozoan pathogen. These findings highlight the importance of heterozygosity at MHC genes and throughout the genome as a means of pathogen defence (Spielman et al. 2004; Hale and Briskie 2007; van Oosterhout et al. 2007b; Oliver et al. 2009b). Penn (2002) performed mesocosm experiments and found no clear evidence of increased resistance through MHC heterozygote advantage. Evidence is accumulating that an optimal, rather than a maximal heterozygosity might provide the highest resistance (see Nowak et al. 1992; Wegner et al. 2003a; Milinski 2006). In this thesis overdominant selection is taken to indicate symmetric overdominance as these have been shown to maintain long term polymorphism (De Boer et al. 2004)

2. Negative frequency dependent selection (NFDS): NFDS is the maintenance of polymorphism through rare allele advantage (Clarke and Kirby 1966; Kojima 1971). Whereas overdominant selection acts on MHC genotypes, the model of frequency dependent selection assumes that each MHC allele recognises a particular set of parasites. Temporal surveys of MHC variation are required to demonstrate a cyclical pattern in MHC allele frequency predicted by NFDS (see Spurgin and Richardson 2010) although the identification of resistant and susceptible MHC alleles is considered consistent with the predictions of NFDS (Langefors et al. 2001; Lohm et al. 2002; Froeschke and Sommer 2005). Langefors et al. (2001) infected 4800 *Salmo salar* with a bacterium (*Aeromonas salmonicida*) and identified MHC alleles associated with both resistance

and susceptibility. This shows that besides selection for heterozygous genotypes, parasite selection can act on single MHC alleles, which may result in a dynamic coevolutionary arms race between the host immune system and parasite virulence genes (Lively and Dybdahl 2000; Carius et al. 2001). Parasites are expected to evolve to avoid detection by common MHC alleles. By contrast, the selection pressure on parasites for avoiding recognition by rare MHC alleles is much lower (Slade and McCallum 1992). Consequently, rare MHC alleles are more likely to detect pathogens, and hence such alleles are predicted to confer a higher fitness to the host than common alleles (Trachtenberg et al. 2003; Froeschke and Sommer 2005). Over time, parasite selection is expected to increase the frequency of rare alleles as hosts carrying these rare alleles have a higher probability of avoiding parasitism. The coevolutionary arms race between the host immune system and parasite virulence results in a dynamic and balanced polymorphism in both host MHC and parasite virulence genes (Sommer 2005). In this thesis NFDS is taken to act on all rare alleles (that confer resistance) regardless of whether they are old or new alleles (see Spurgin and Richardson 2010)

3. Fluctuating selection: Spatial and/or temporal heterogeneity in selection pressure or fluctuating selection can result in a balanced polymorphism at the MHC (Hedrick et al. 1976; Hedrick 2002). This mechanism assumes firstly, that MHC alleles confer resistance to different parasites (as in overdominance) and secondly, that parasite presence varies over time. A study of the MHC class I variation over nine years of great red warblers found variation in allele frequencies over time (Westerdahl et al. 2004). The authors suggest that these fluctuations in MHC allele frequencies are due to temporal fluctuations in pathogen fauna. However, they are also consistent with frequency dependent selection (see also Charbonnel and Pemberton 2005). A parasite of the Trinidadian guppy has also been implicated with temporal fluctuation in MHC allele frequency (Fraser et al. 2010b). Fraser et al. (2010b) found that MHC genetic divergence between guppy populations was lower than that of neutral markers but that this pattern had changed in a number of populations in the subsequent year. The authors argue this is evidence for a change in the mode of selection, from homogenising to diversifying (Fraser et al. 2010b).

4. Associative Balancing Complex (ABC) evolution: ABC evolution proposes the maintenance of MHC polymorphism through selection against homozygotes because they express recessive deleterious mutations that accumulate nearby MHC genes (van Oosterhout 2009a). The recessive deleterious mutations can accumulate in the MHC because the effective rate of recombination in some areas of the MHC is very low (see e.g. Stenzel et al. 2004; Gregersen et al. 2006). This reduces the efficiency of purifying selection to remove recessive deleterious mutations (e.g. Haddrill et al. 2007; see also review by Comeron et al. 2008). Some of those mutations become fixed in all copies of a particular haploblock in a process analogous to Muller's ratchet (Muller 1932).

These three models of balancing selection are not mutually exclusive, but rather, they are likely to act complementarily. De Boer et al. (2004) and Borghans et al. (2004) both modelled NFDS and overdominant selection, finding that overdominance was unlikely to be able to maintain the level of polymorphism observed at the MHC, unless each MHC allele conferred a very similar fitness (e.g. symmetric overdominance). However, it is difficult to discriminate between different types of balancing selection in wild systems. For example, rare alleles that are predicted to be maintained under NFDS are also more likely to be presented in heterozygote condition (Apanius et al. 1997).

Selection for specific MHC alleles can increase their effective migration rate compared to neutral markers, such as microsatellite alleles (Schierup 1998). Random genetic drift drives changes in microsatellite allele frequency and therefore, novel alleles introduced to a new population may be lost by chance (see Charlesworth 2009). However, if selection favours the introduction of an MHC allele, it is more likely to be maintained in the population. This is of particular consequence for NFDS, overdominant selection and ABC evolution because they predict that selection favours rare alleles (Apanius et al. 1997; van Oosterhout 2009a). Consequently, genetic differentiation between populations (with gene flow) is expected to be reduced under these modes of balancing selection (Schierup 1998; Schierup et al. 2000; Muirhead 2001). Fluctuating selection is expected to increase genetic differentiation between populations because alternate alleles are favoured.

Sexual selection, parasite selection and inbreeding avoidance

According to the Red Queen Hypotheses, the hosts' immune system evolves quickly in response to the rapid evolution of their parasites (see also review by Milinski 2006; e.g. Paterson et al. 2010), and balancing selection by parasites can maintain MHC polymorphism. However, mate choice can also drive rapid MHC evolution, or it may aid maintenance of polymorphism at these genes (Milinski 2006). The MHC is implicated in sexual selection at both pre- and post zygotic levels, for example, by means of mate choice for mates with a specific MHC genotype (Reusch et al. 2001), differential fertilisation success (Skarstein et al. 2005), and the selective abortion of MHC-homozygous embryos (Dorak et al. 2002). Studies on a wide variety of animals support the role of MHC in pre-copulatory sexual selection, including research on humans and rodents (reviewed by Penn and Potts 1999; Kavaliers et al. 2004), salmon (Landry et al. 2001; Neff et al. 2008) and sticklebacks (reviewed by Milinski 2003). This work has focused on the role of olfaction in particular, which is thought to enable individuals to determine the compatibility of the MHC of prospective mates. For example, research on sticklebacks showed that females use evolutionarily conserved structural features of MHC molecules to evaluate diversity of the MHC of males (Milinski et al. 2005).

There has been considerable debate on the relative roles of MHC variation and genome-wide heterozygosity and their consequence on fitness (Hedrick et al. 2001b; Landry et al. 2001; e.g. Arkush et al. 2002; see also review by Sommer 2005; van Oosterhout et al. 2007b). Landry et al. (2001) were able to separate the effects of MHC heterozygosity and inbreeding avoidance (i.e. the relatedness, measured using similarity at five neutral microsatellite loci). By sampling spawning Atlantic salmon and then juvenile salmon after hatching the authors found that mates selected in order to increase genotypic differences in the MHC of juveniles, and not to offer increased heterozygosity in the genome. This supports the role of the MHC in sexual selection for "good genes" (or genetic compatibility), although the authors acknowledge the experiment cannot differentiate between active mate choice, differential fertilisation success, and the higher survival of juveniles of heterozygous MHC genotypes (see Skarstein et al. 2005; de Eyto et al. 2007; Neff et al. 2008). Disentangling the relative importance of inbreeding avoidance, "good gene" selection and parasites resistance in MHC evolution is an important future challenge in MHC research. A particularly promising new avenue of research is offered by the comparison of MHC evolution in a pair of closely-related poeciliids, one species being entirely clonal and the other sexually reproducing (see Tobler and Schlupp 2005; Schaschl et al. 2008).

Poeciliid MHC

The MHC class I and II genes of the teleosts are unlinked and believed to reside on separate chromosomes, an arrangement that is markedly different from that of tetrapods (Sato et al. 2000; Stet et al. 2003; see also review by Kelley et al. 2005; Dixon 2008). Free recombination between class I and II MHC regions in teleost fish is believed to be a derived characteristic because the class I and II genes are linked in cartilaginous fishes, the oldest class of extant vertebrates (Ohta et al. 2000). Class III genes are divided over several different chromosomes in teleosts (Flajnik et al. 1999; Sambrook et al. 2002). The occurrence of separate linkage groups may have important implications for the evolution of teleost MHC, as with free recombination, selection can operate independently on the different gene classes. This suggests that teleost MHC is more versatile, and perhaps less prone to accumulate mutations by genetic hitchhiking. Relatively little work has been done on MHC class I and III in poeciliids (but see Sato et al. 1995; Sato et al. 2000; Schaschl et al. 2008), possibly because class I genes are involved in the antigen presentation of more elusive intracellular parasites such as viruses. However, Figueroa et al. (2001) conduct a comprehensive survey of the MHC class I of 22 swordtail fishes. Sequences were found to fall within one of two ancient lineages and the pattern of sequence similarity was consistent with the birth-and-death theory of multigene evolution (Nei and Rooney 2005). In this thesis I focus on research on MHC class II loci.

Poeciliid MHC DAB genes

The MHC multigene family and its large number of genes necessitate a clear classification of alleles and genes. In the present thesis I follow the nomenclature first proposed by Klein et al. (1990). The *DAB* loci, in poeciliids (and many other fish species), are possibly the best studied genes. *DAB* specifies the region of the MHC, in which *D* is an abbreviation of Duo, in reference to class II, is *A* represents the family of genes and *B* specifies the β monomer of the MHC heterodimer (class II*B*). The species from which the MHC sequence was amplified is denoted by abbreviation of genus and species names, for example, *Mhc-Pore* specifies that the MHC sequence was obtained from the organism *Poecilia reticulata* (see Klein et al. 1990). Presently, in guppies, MHC allele locus affiliations are unknown, consequently all alleles are given a *DAB* prefix.

Sato et al. (1995) first described the MHC of the guppy using ornamental stocks and inbred wild fish to define the class I and class II regions. The authors observed low variability at both classes of genes and predicted that guppies possess one MHC class II gene and one or two MHC class I genes. A subsequent study by van Oosterhout et al. (2006a) confirmed this finding, showing that ornamental guppy lines possessed between one and three MHC class II*B* alleles. However, the authors detected 15 to 16 MHC alleles within wild guppy populations, demonstrating these populations had a much higher level of MHC polymorphism than ornamental guppy stocks. Within individual wild-caught guppies they detected up to four alleles, which implies there are at least two *DAB* genes (see also Fraser et al. 2010a). The disparity in allelic richness between wild and ornamental fish could not be explained by fixation of alleles by inbreeding in captivity, nor by the presence of non-amplified sequences (i.e. null alleles). Rather, van Oosterhout et al. (2006a) suggested that during many generations in captivity, multigene evolution may have played an important role shaping the organisation and polymorphism of the MHC. In the ornamental lines, gene conversion may have fixed the same allele at duplicated MHC *DAB* genes (see e.g. Reusch et al. 2004, for the role of gene conversion in teleost MHC; Reusch and Langefors 2005). van Oosterhout et al. (2006a)

furthermore hypothesised that a reduced level of parasitism during >100 generations in captivity may have reduced the number of duplicated *DAB* genes in the ornamental strains, a hypothesis consistent with the accordion model of MHC evolution (see Klein et al. 1993; Gu and Nei 1999).

Schaschl *et al.* (2008) analysed the MHC class I and class II*B* diversity of two closely related mollies. The Amazon molly (*Poecilia formosa*) is an all-female clonal Poeciliid that lives in sympatry with the sexually reproducing Sailfin molly (*Poecilia latipinna*) (Schaschl et al. 2008). Genetic diversity was lower in the asexual molly in comparison to its asexual counterpart at both neutral genetic and MHC class I and II loci, though the asexual molly still maintained a considerable level of MHC diversity (Schaschl et al. 2008). The Red Queen hypothesis suggests that the asexual molly should be unable to provide a "moving target" against rapidly evolving pathogens, as (meiotic) recombination cannot provide novel gene combinations in clonally reproducing organisms. However, previous research showed no difference in parasite prevalence between both sympatric species (Tobler and Schlupp 2005; Tobler et al. 2005). Comparison of MHC diversity between the asexually and sexually reproducing species offers a unique opportunity to disentangle the relative roles of parasites and sexual selection in the maintenance of MHC polymorphism.

Poeciliid MHC DXB gene

In 1998, McConnell et al. (1998b) reported a novel MHC class II*B*-like gene in *Xiphophorus maculatus* and *X. helleri*. This gene was labelled *DXB,* because its family designation remains unknown (McConnell et al. 1998b). This locus is not linked to the *DAB* locus and is expressed in *X. maculatus*, *X. multilineatus* and *X. pygmaeus* (McConnell et al. 1998a; Roney et al. 2004). In addition, genomic copies of this gene have been identified in *X. helleri* and *P. reticulata*, with features consistent with functional genes (McConnell et al. 1998a; McConnell et al. 1998b). *Xima-DXB* and *Pore-DXB* are thought to be orthologous, they share 98% sequence similarity in exon 3 (i.e. *Xima-DXB*01* and *Pore-DXB*01*). *DXB* has a sequence similarity of only 61-63% in exon 3 with *DAB* alleles, and both genes are hypothesised to be the result of an ancient gene duplication event. However, unlike *DAB*, the *DXB* genes show very little polymorphism and only a weak signal of positive selection (Summers et al. 2008).

The exon-intron structure of the *DXB* is similar to that of amphibian, chicken and human class II*B* structure, suggesting a *DXB*-like lineage is phylogenetically related to tetrapod MHC class II*B* genes (McConnell et al. 1998a; McConnell et al. 1998b). Interestingly, Roney et al. (2004) identified two alternative splice patterns of the *DXB* gene and an additional truncated version from a sample of *X. multilineatus* and *X. pygmaeus*. Alternative splicing is the transcription of only certain exons of a gene, which allows a single gene to encode many distinct proteins (reviewed by Stetefeld and Ruegg 2005; Artamonova and Gelfand 2007). Alternative splicing may be an important feature of the MHC, potentially increasing its versatility. In one human non-classical class II chain molecule (*HLA-DM*) four alternative splice patterns have been detected (Modrek et al. 2001). Gene expression and the population genetics of *DXB* genes in poeciliids has received relatively little attention. However, the interactions between *DXB* gene products and other MHC class II molecules (see Ting and Trowsdale 2002; Danchin et al. 2004), and the relative conservation of this gene across phylogenetically distinct taxa suggest it plays an important role in MHC evolution (Jensen et al. 2008). As recently highlighted by Summers et

al. (2008), research on fish, and in particular poeciliids, may help to unravel the enigmatic role of (non-classical) MHC genes in other vertebrates.

The genomic organisation of Poeciliid MHC is relatively simple, more similar to that of salmonids than that of many other important evolutionary and ecological model species. Salmonids are believed to contain just one *DAB* locus (Langefors et al. 1998; Langefors et al. 2000; Landry and Bernatchez 2001; Stet et al. 2002), which makes detecting putative associations between resistance and MHC alleles or genotypes easier than for species with many duplicated MHC genes (see also Kaufman et al. 1999). For example, the three-spined stickleback is thought to have four *DB* gene loci (*Gaac-DAB*, *-DBB*, *-DCB* & *-DDB*) (Reusch and Langefors 2005), and cichlids may have up to 17 loci (Malaga-Trillo et al. 1998). Consequently, in these species associations between resistance and MHC can only be detected when analysing multilocus genotypes. PCR amplifying alleles of a large number of loci is a technically challenging task. Indeed, the relative simplicity of Poeciliid MHC makes this family an excellent model for future immunological, parasitological and evolutionary studies. Furthermore, the apparent difference in genomic organisation between ornamental and wild guppies and their global distribution through release from ornamental stocks (see Lindholm et al. 2005) makes this an interesting species to examine the role of different models of multigene evolution.

Poecilia reticulata, MHC and molecular evolution and ecology

Parasite selection and MHC

In natural populations, all genetic variation is effected by mutation and random genetic drift (Hedrick 2011). In addition to these evolutionary forces, MHC genes are also affected by selection (Ford 2002; Garrigan and Hedrick 2003; Nielsen 2005). To study the effects of selection, researchers tend to compare the genetic variation at MHC loci to that of neutral genetic markers (e.g. microsatellites, see van Oosterhout et al. 2006b). The signal of natural selection should be detectable only in the MHC, while population demographic forces (such as genetic drift and migration) will be evident at both neutral and immunogenetic loci (Garrigan and Hedrick 2003). This allows one to disentangle these demographic effects from the effects of selection on the MHC (e.g. van Oosterhout et al. 2006b).

van Oosterhout et al. (2006b) assessed MHC class II*B* and neutral microsatellite diversity in two wild populations of guppies. The authors also recorded differences in parasite fauna. Guppy populations inhabited an upland ($N_e \approx 100$) and lowland ($N_e \approx 2400$) region of the Aripo River in Trinidad. The sites are characterised by a distinctly different parasite fauna (van Oosterhout et al. 2006b). The upland population, was more frequently infected with a virulent species of gyrodactylid parasite, and was shown to maintain high levels of MHC diversity despite the effects of random genetic drift on this small population. Using computer simulations, the authors estimated selection coefficients of $S \geq 0.2$ acting on the MHC in the most parasitized population.

In a subsequent study, we estimated the potential for parasite selection directly, using a mark-recapture experiment on the same guppy population in Trinidad (van Oosterhout et al. 2007a). Almost 200 guppies were individually marked and their gyrodactylid parasite load counted. Individuals (males but not females) with increased parasite burdens were recaptured with a significantly reduced rate, confirming the role of selection imposed through *Gyrodactylus*

infection. However, the selection coefficient estimated from reduced recapture rate of infected fish was lower ($S \leq 0.14$) than that required to explain the MHC variation observed in the previous study (van Oosterhout et al. 2006b). This suggests that other selective forces are operating on the MHC, such as sexual selection through female mate choice, or selection on linked mutations that hitch-hike with the MHC alleles.

Fraser and colleagues have also studied guppy MHC class II*B* variation in some detail (see Fraser and Neff 2009, 2010; Fraser et al. 2010a, b). In a laboratory study, Fraser and Neff (2009) found an association of one MHC allele (*Pore a132*) with reduced individual *Gyrodactylus* burden, a relationship also observed in wild populations (Fraser and Neff 2010). Fraser et al. (2010b) conducted a temporal investigation of nine populations of guppies in the Northern Mountain range of Trinidad. In 2006, the authors found evidence of selection acting to reduce MHC to neutral genetic differentiation between populations and in 2007 selection was found to increase MHC differentiation to microsatellite variation. Fraser et al. (2010b) suggest that there was a change from homogenising to diversifying selection over this period. Fluctuating selection has been suggested in a number of temporal studies of MHC variation (e.g. Westerdahl et al. 2004; Charbonnel and Pemberton 2005).

Sexual selection and MHC

Female mate choice based on male colour pattern and ornamentation has been studied extensively in guppies (reviewed by Houde 1997). However, sexual selection based on other senses, such as olfaction, has received relatively little attention (Crow and Liley 1979; Meyer and Liley 1982; Griffiths and Magurran 1999; Shohet and Watt 2004; Archard et al. 2008). Nevertheless, studies on other fish species have indicated that multimodal interactions including visual and chemosensory cues potentially play a role in mate choice and sexual selection (e.g., three-spined sticklebacks, Gasterosteus aculeatus, see McLennan 2003). Archard et al. (2008) showed that high concentrations of chemosensory stimulus were required to elicit a behavioural response in guppies. Such high concentrations are only attained when the fish are in close proximity. This may arise when guppies shoal, or when males display to females, which occurs within distances of a few centimetres (Endler 1995; Long and Rosenqvist 1998). Shohet & Watt (2004) used an experimental flume to study the attraction of female guppies to conspecific chemosensory cues. The authors found that females were attracted to cues from other females, and that visually attractive males were chemically unattractive, and vice versa. The potential role of the MHC in this model has however not been investigated. Future research into MHC and sexual selection is possibly best studied in poeciliid species with little variation in male colour patterns, since in these species olfaction may play a more prominent role than visual communication.

Ecology of *Poecilia reticulata*

The guppy is a small live bearing fish native to freshwater streams of Trinidad and Tobago and South America (see Houde 1997). Initially noted for the conspicuous colouration of males, guppies are easily studied in their natural habitat and with generation a time of approximately three months they are also well suited to laboratory research (see Endler 1980; Reznick and Endler 1982; Houde 1997; Magurran 2005). Guppies are a model organism for studies on the effects of natural selection in the wild (see review by Endler 1995). Trinidadian guppy populations show marked phenotypic and behavioural differences between habitats that are correlated with predator fauna (Endler 1995). Guppies in Trinidad occupy habitats that differ

widely in parasite fauna on a spatial (van Oosterhout et al. 2006b) and temporal scale (van Oosterhout et al. 2007a; Fraser et al. 2010b). The importance of parasite-mediated selection in adaptive evolution and the effects on the ecology and evolution of guppies have been largely ignored (but see van Oosterhout et al. 2003; van Oosterhout et al. 2006b; Cable and van Oosterhout 2007b; van Oosterhout et al. 2007a; Fraser and Neff 2009, 2010). In the present study, I measure parasite diversity at three sites of the Aripo River in Trinidad's Northern Mountain range. The aim of this study is to disentangle the effects of parasitism on the maintenance of polymorphism in small bottlenecked populations.

Guppies in the present study are from three locations in Trinidad, the Upper and Mid Naranjo River (Up and Mid, respectively) and the Lower Aripo (Low) River. The Naranjo is a tributary of the Aripo River, which feeds the Caroni River within the Caroni drainage (Fig. 1). Predators in the Low population include two cichlid species (Crenicichla *alta* and *Aequidens pulcher*) and two shrimp species (*Macrobrachium crenulatum* and *M. carcinus*), which predate guppies of any size (see Magurran 2005 for full list of predatory species in the Aripo). The population also has a rich biodiversity of parasites (monogeneans, cestodes, digeneans and nematodes, see review by Cable In press). The Haskins waterfall separates the Low from the Mid population. The Haskins waterfall is a barrier to the majority of guppy predators (Endler 1980), but not to parasites. *R. hartii* is the main guppy predator found above the Haskins waterfall (see Magurran 2005). This killifish is found in the Low, Mid and the Up population. *R. hartii* are gape limited and prey mainly on small guppies. Consequently, the Mid and Up guppy populations have males that are brightly coloured in comparison to the Low population (Endler 1980). The Up, the uppermost population is located approximately 0.5 km above the Mid. These populations are separated by a 2-3 m manmade damn. This dam may also represent a cut off point for the majority of guppy parasites. Overall, these three guppy populations represent a hierarchical variation in selection pressure from upland to lowland. The Up population has few predators and few parasites, the Mid population has few predators and parasites and the Low population contains large predators and parasites.

Gyrodactylus

Gyrodactylus parasites are just one genera of parasites than are known to infect guppies, however they are mentioned here because they are the most prevalent guppy parasites (Martin and Johnsen 2007). Resistance of guppies to gyrodactylid infections has been shown to be population specific (van Oosterhout et al. 2003) and have a genetic basis (Cable and van Oosterhout 2007b; Fraser and Neff 2009). Increased G*yrodactylus* spp. load has been shown to significantly reduce guppy survival in an upstream site (van Oosterhout et al. 2007a). *Gyrodactylus turnbulli* (Harris, 1986) and *G. bullatarudis* (Turnbull, 1956) ectoparasites graze on the mucus and epithelial cells of guppies (Bakke et al. 2007) and can have a dramatic affect on their guppy host, resulting in increased mortality (Cable and Harris 2002; van Oosterhout et al. 2007a), reduced feeding response (van Oosterhout et al. 2003), reduced male mating success (Houde and Torio 1992; Lopez 1998) and impaired female mate choice (Lopez 1999). These viviparous parasitic worms can reproduce at exponential rates with a doubling of parasite population every 24-48 hours under ideal conditions (Scott 1982). There is no specific transmission stage and they can spend their entire life cycle on a single guppy host (see reviews by Cable and Harris 2002; Bakke et al. 2007).

Thesis outline

In **Chapter 2** I use neutral variation at 13 microsatellite loci to establish population genetic relationships between the Upper Naranjo (Up), Mid Naranjo (Mid) and Lower Aripo (Low) guppy populations. I measure genetic differentiation between populations and estimate effective population size and migration rates in order to test the hypothesis that (1) population size is reduced in upstream localities and that (2) migration is biased in the downstream direction. This chapter is intended as a basis for later chapters and is to be published only in part as an addition to data in subsequent chapters.

In **Chapter 3** I measure MHC diversity in the Up, Mid and Low guppy populations. I quantify parasite fauna in these populations and use neutral microsatellite data from chapter 2 to demonstrate the effects of selection despite random genetic drift on small guppy populations. Random genetic drift is expected to reduce genetic variation at both MHC and microsatellite loci. However, selection by parasites is predicted to maintain MHC but not microsatellite variation. MHC and microsatellite data is used here along with parasite prevalence data in order to test the hypothesis that parasite mediated selection maintains MHC polymorphism in wild guppy populations. This chapter is intended for publication in addition to Cummings et al. (2010 see Appendix 13) which was generated using part of the work from this chapter.

In **Chapter 4** I reanalyse MHC and a subset of the microsatellite data from chapter 3 combined with MHC and microsatellite data from a study by van Oosterhout et al. (2006b). This data is used in a temporal analysis with the aim of distinguishing the mode of balancing selection acting at the MHC in the Mid and Low populations sampled in 2001 and 2007. MHC genetic differentiation that is reduced in comparison to that of microsatellites (observed in 2001 data) is consistent with a higher effective migration rate of MHC alleles and indicates maintenance of MHC polymorphism by heterozygote advantage or negative frequency dependent selection. However, this pattern of genetic differentiation of MHC and microsatellites is not observed in the 2007 data. This chapter explores the predictions of a model of selection that fluctuates over space and/or time and considers 2007 data in light of this model of balancing selection. This chapter is intended for publication in combination with a simulation model (see Chapter 5).

In **Chapter 5** I use a simulation model to demonstrate the combined effects of overdominant selection and migration on spatio-temporal genetic differentiation on MHC and neutral microsatellite loci. Chapter 4 highlights a change in the mode of selection in 2001 and 2007 and that this change can account for temporal variation in MHC alleles. In the simulation used in this chapter, only one mode of selection (overdominance) is used under the prediction that it can account for all the observations of the temporal dataset. This chapter is intended for publication in combination with Chapter 4.

In **Chapter 6** I summarise the results from previous chapters, discuss their implications and avenues for future research. This chapter is not intended for publication

Neutral variation in Poecilia reticulata of the Aripo River

Chapter 2

Abstract

The population genetics of guppies (*Poecilia reticulata*) has been studied extensively and is well understood. The effects of selection however, may, or may not be evident in population division as measured at neutral markers. The present study measures genetic differentiation of three populations of guppies, inhabiting the Aripo River and an upstream tributary (Naranjo). A finding of reduced population size with increasing altitude and downstream-biased migration is consistent with previous estimates. This work provides an understanding of the demography of these populations and forms a basis for future work on selection acting within and between these guppy populations.

Introduction

The population genetics of guppies (*Poecilia reticulata*) of Trinidad and Tobago has been intensively studied using a variety of molecular markers, including allozymes (Carvalho et al. 1991; Shaw et al. 1991; Carvalho et al. 1996), mitochondrial sequence variation (Fajen and Breden 1992), microsatellites (Crispo et al. 2006; Barson et al. 2009; Suk and Neff 2009) as well as nuclear genes (Alexander et al. 2006) and single nucleotide polymorphisms (SNPs) (Willing et al. 2010). Studies on adaptive evolution of life history traits of guppies (reviewed by Endler 1995) required a detailed understanding of the colonisation history of Trinidad on two temporal scales. Guppies that had been translocated for selection experiments between 10 and 40 years prior allowed Shaw et al. (1991) and Carvalho et al. (1996) to test expectations of bottlenecked populations where the precise details of founder events were known. This work established ideas of increased genetic diversity in the lowlands and developed understanding of temporal dynamics in guppy populations (Carvalho et al. 1991; Shaw et al. 1991). Allozyme and mitochondrial data as well as nuclear markers provided evidence of long term colonisation history known as the two arcs hypothesis. According to this hypothesis, guppies of eastern and western Trinidad have distinct colonisation histories (Fajen and Breden 1992; Carvalho et al. 1996; Alexander et al. 2006; but see Suk and Neff 2009; Willing et al. 2010).

Compared to other genetic markers including mtDNA and SNPs, microsatellite data suffers from a relatively low resolution at a phylogenetic timescale (millions of years) (but see Galindo et al. 2009), but it has an excellent resolution at a micro-evolutionary timescale (100s to 1000s of years). The relatively high mutation rates caused by slippage and proofreading errors tend to generate alleles more quickly than single base pair mutations (reviewed by Selkoe and Toonen 2006). Barson et al. (2009) used microsatellites to reveal the downwards-biased migration both between and within rivers in the Caroni Drainage. Crispo et al. (2006) used microsatellite variation to account for the source of selection for upstream traits. By separating the effects of waterfalls and predation on genetic differentiation of guppy populations, the authors showed that waterfalls were effective barriers to upstream gene flow

but predators did not seem to influence long term gene flow despite differences in morphology.

Population demography can be inferred by studying the genetic variation at multiple microsatellite loci. Assuming that these loci are selectively neutral, the effects of random genetic drift, mutation and migration can be quantified. The information can subsequently be used to study the effects of selection on specific loci, such as genes of the MHC (van Oosterhout et al. 2006b), male colour pattern genes (Willing et al. 2010) and other genes under directional selection. The study by Crispo et al. (2006) highlights the importance of understanding demography where, in that case, waterfalls present a barrier to one set of loci but not another (but see Shikano and Taniguchi 2002; Shikano et al. 2005; van Oosterhout et al. 2007b, for examples of relationships between neutral genetic variation and fitness related traits).

In the present study, I measure neutral variation at 13 microsatellite loci in guppies of the Aripo and Naranjo Rivers in Trinidad. Earlier work on a subset of these loci conducted on a number of other guppy populations in Trinidad demonstrated downstream-biased migration, as well as a progressive reduction in effective population size with increased altitude. In the present study, neutral genetic diversity is measured to estimate demographic parameters, which allows subsequent inference of selection acting within the major histocompatibility complex (see Chapter 3). I therefore focus on three populations of the Aripo River system: the Lower Aripo (Low, previously LA, see van Oosterhout et al. 2006b), the Mid Naranjo (Mid, previously UA, see van Oosterhout et al. 2006b) and the Upper Naranjo (Up). I measure neutral variation in order to establish effective population size (N_e) and migration (Nm) between populations.

Materials and Methods

Guppy collection

Wild caught guppies were captured (n = 150) (June 2007) using a seine net, given a lethal dose of MS222 and stored separately in molecular grade ethanol. In order to minimise allele frequency biases associated with sampling families or cryptic social structure, each site was sampled at multiple locations in order to maximise the chance of catching guppies from separate shoals. Three sampled guppy populations are found along the Aripo River, which is part of the Caroni Drainage in the Northern Mountain Range of Trinidad (grid reference: Upper Naranjo (Up), 20P 692498.44 E 118257.53 N Mid Naranjo (Mid), 20P 693100 E 1181800 N; Lower Aripo (Low), 20P 6914000 E 1177700 N).

Molecular analysis

Genomic DNA was extracted from the caudal fin using the HotSHOT protocol of Truett et al. (2000). Extracted DNA was PCR amplified at thirteen microsatellite loci, including two interrupted repeats: *Pr39*, *Pr92* (Becher et al. 2002); ten perfect dinucleotide repeats: *Pret-32*, *Pret-46*, *Pret-69*, *Pret-77* (Watanabe et al. 2003), *G72*, *G82*, *G211*, *G289*, *G350* (Shen et al. 2006), *Hull 9-1*; and a perfect tetranucleotide repeat, *Hull 70-2* (van Oosterhout et al. 2006). Forward primers were labelled with Cy5, Cy5.5 (Eurofins MWG Operon, Germany) and WellRED D2 (Sigma-Aldrich) dyes. Microsatellites were amplified using Qiagen multiplex PCR Kit according to the manufacturer's instructions with 30 PCR cycles and annealing

temperatures of 53°C (*Pr39*, *Pret-77*, *Hull 70-2*, *Pret-46*), 56°C (Pr92, Pret-69, G72 Hull 9-1, G350) 58°C (*Pret*-32, G82, *G289* & *G211*). Loci of similar annealing temperatures were multiplexed together in 10µl PCR (there were no more than five loci in a single reaction). PCR products were resolved on a Beckman Coulter CEQ 8000 sequencer using CEQ size standard kit (400 bp). Individuals for which information on one or more locus was not retrieved were not included in the final analysis (13 individuals discarded, final sample size Up, n = 47; Mid, n = 45; Low, n = 45).

Population genetic analyses

Microsatellite genotype frequencies were checked for the presence of null alleles, amplification bias and scoring errors using Micro-Checker 2.2.3 (van Oosterhout et al., 2004). Observed and expected heterozygosity were calculated. Genotypic disequilibrium was calculated using FSTAT version 2.9.3 (Goudet 1995, 2001) with a Bonferroni adjusted P-value for multiple testing at the 5% nominal level ($\alpha = 0.000641$).

Genetic differentiation, corrected for highly polymorphic loci (G'_{ST}) (Hedrick 2005b) was used to measure differentiation at microsatellite loci among the Up, Mid and Low populations. G'_{ST} values with 5-95% confidence intervals (CI) were calculated by bootstrapping (with replacement) over individuals using 1000 runs. Microsatellite bootstrap calculations were checked using the Nei's (1987) method within FSTAT 2.9.3 (Goudet 1995, 2001) (see Appendix 1). Microsatellite genetic differentiation was calculated for each locus and the mean values are calculated by averaging across 13 loci. Allelic richness (*A*, 5-95% CI) of microsatellites was bootstrapped with replacement over individuals, again using 1000 runs per locus per population.

STRUCTURE 2.3.1 (Pritchard et al. 2000) and NEWHYBRIDS 1.1 (Anderson and Thompson 2002) were used to further explore the differentiation between the Up, Mid and Low populations. STRUCTURE software uses allele frequency data and a Bayesian clustering system in order to assign individuals to distinct genetic clusters. This software was used to determine the number of clusters, *k* (*k* = 1-6) and the presence of upstream migration into the Upper Naranjo population. Upstream migration would be confirmed by the presence of a Mid cluster in the Up (or Low in Mid) population. Burn-in and run length of simulation were 500,000 and 1,000,000 repetitions, respectively. Six iterations were run and values are reported as means (± var[LnP(D)). The correlated allele frequency model was used based on observed allele frequencies (see Appendix 2). The model used a population location prior (Hubisz et al. 2009) and was run with an admixture prior. However, separate inference of alpha (initial alpha = 1.0) for each cluster was used because admixture is believed to be asymmetric due to biased downstream migration (Barson et al. 2009).

The Evanno method (Evanno et al. 2005) for estimating the appropriate value of the number of clusters (*k*) was calculated using STRUCTURE HARVESTER (Earl 2011). Rather than using the most likely value of *k* the Evanno method uses a calculation that incorporates the rate of change of *k* (Delta *k*) (see Evanno et al. 2005). Evanno et al. (2005) argue that because the *k* value asymptotes the highest likelihood does not represent the most likely number of clusters.

NEWHYBRIDS software was used as a separate means by which to confirm upstream migration into the Upper Naranjo population. This software uses genotype frequency data to classify F1, F2 and backcross (Bx) individuals (where $n = 2$, i.e. a maximum of two generations of

backcrosses) (Anderson and Thompson 2002). NEWHYBRIDS is primarily used to test for hybrid individuals of sympatric populations of separate species or subspecies. The program assumes that some proportion of the population is not admixed. NEWHYBRIDS was used without a genotype frequency category prior (z) and was run multiple times for a burn-in of 500,000 and runtime of 1,000,000 steps. The number of genotype frequency classes was set to two generations of crossing (n = 2, equivalent to six frequency classes), however a reduced period of hybridisation (n = 1) gave a qualitatively similar result. The presence of hybrid individuals in an upstream site can be taken as evidence of upstream migration.

MIGRATE 2.4 (Beerli and Felsenstein 1999, 2001) was used in order to get a further estimate of population size and migration rate. Softwares that estimate such parameters may be prone to error but are used here in order to compare population sizes and directional biases in upstream or downstream migration that are indicated by the STRUCTURE analysis. MIGRATE uses a maximum likelihood coalescent approach to estimate theta (Θ), which is equal to four times the effective population size, N_e, multiplied by the mutation rate (per generation), μ ($\Theta = 4N_e\mu$), and M, the migration rate parameter which is the migration rate, m, divided by the mutation rate (m/μ). The migration rate (M) was converted into number of migrants per generation (Nm) by multiplying M by Θ and then dividing by four (nuclear inheritance scalar). Effective population size and migration rate were calculated based on a microsatellite mutation rate of $\mu = 2x10^{-4}$ (Ellegren 2000).

MIGRATE was run four times, using F_{ST} estimates to start the first run. Subsequent runs were started from estimates (Θ and M) of previous runs. The Brownian motion model was used as an approximation of the stepwise mutation model. The MCMC search criteria used 200 short chains of 10,000 steps and 10 long chains of 400,000 steps with heating scheme of four temperatures (1.0, 1.2, 1.5, and 3.0). The burn-in was set to 100,000. Runs were repeated until Θ and M estimates were consistent between runs, either reaching asymptote or having overlapping 95% confidence intervals between runs. Maximum likelihood estimates (with 5-95% CI) were used to compare effective population size (N_e) and migration rate (Nm) between populations.

Likelihood ratio tests of reduced migration models were tested against the full migration model in order to establish the likelihood of a barrier to migration between populations. Tests were conducted in a replicate fourth run and models were assessed by comparison of log likelihood of the test model in comparison to the full model and using Akaike's Information Criterion in the MIGRATE output. Reduced migration models included: (1) no upstream migration, (2) barrier to upstream migration between the Mid and Up populations, (3) barrier to upstream migration between the Low and Mid populations, and (4) migration only between adjacent populations (direct Up to Low migration blocked) (Table 5).

Results

Micro-Checker 2.2.3 (van Oosterhout et al., 2004) analysis indicated Hardy Weinberg equilibrium at all 13 microsatellite loci. Evidence for null alleles was found in *Hull 70-2* and *Hull 9-1*, however, in both cases, the combined probability for all classes was not significant. Therefore, these frequencies were used uncorrected in future analyses. From 234 tests of genotypic disequilibrium, 9 exhibited evidence of genotypic disequilibrium ($P < 0.05$). This is fewer than expected by chance as type I errors (11). Indeed, there was no genotypic disequilibrium at microsatellite loci as measured at the Bonferroni corrected significance level ($P > 0.0006$, adjusted P-value corresponding to 5% nominal level after Bonferroni correction). Consistent with Micro-Checker there was no significant difference between observed and expected heterozygosity over all microsatellite loci (see Table 1). The level of observed heterozygosity (Ho) was significantly different between populations (Kruskal-Wallis test (adjusted for ties): $P = 0.003$). Moreover Ho was significantly reduced from the Mid to the Up populations which are found separated by a man made dam (see Table 1). This finding is consistent with a population bottleneck caused by reduced population size and reduced upstream migration from the Mid to the Up population.

Table 1. Number of alleles, observed, and expected heterozygosity at 13 neutral microsatellite loci in three guppy populations of the Aripo River. There is no significant difference between observed and expected heterozygosity over all microsatellite loci (Kruskal-Wallis test (adjusted for ties): Up $P = 0.796$; Mid $P = 0.898$; Low $P = 0.980$). Ho was significantly reduced from the Mid to the Up populations (Mann-Whitney test (adjusted for ties): $P = 0.040$)

	Up (n = 47)			**Mid (n = 45)**			**Low (n = 45)**		
Locus	**A**	**Ho**	**He**	**A**	**Ho**	**He**	**A**	**Ho**	**He**
Pret-32	3	0.319	0.289	6	0.511	0.473	16	0.667	0.756
G82	1	0.000	0.000	2	0.133	0.124	2	0.200	0.180
G289	5	0.362	0.435	7	0.644	0.645	14	0.933	0.839
G211	2	0.043	0.042	3	0.156	0.166	5	0.244	0.280
Pret-77	2	0.064	0.062	3	0.289	0.271	6	0.556	0.458
Pr39	4	0.085	0.083	5	0.533	0.543	14	0.711	0.816
Hull 70-2	7	0.745	0.645	11	0.578	0.745	16	0.889	0.924
Pret-69	2	0.149	0.173	4	0.467	0.434	10	0.578	0.517
Pr92	2	0.043	0.042	2	0.133	0.162	8	0.556	0.560
Hull 9-1	4	0.128	0.122	7	0.400	0.444	12	0.644	0.768
G72	1	0.000	0.000	1	0.000	0.000	3	0.089	0.086
G350	1	0.000	0.000	1	0.000	0.000	2	0.044	0.043
Pret-46	3	0.085	0.082	7	0.422	0.383	14	0.733	0.656
Mean	**2.846**	**0.156**	**0.152**	**4.538**	**0.328**	**0.338**	**9.385**	**0.526**	**0.529**

Table 2. Allelic richness (*A*) and genetic divergence (G'_{ST}) (mean (5-95% CI) estimated within and between the Upper (Up) and Mid Naranjo (Mid) and the Lower Aripo (Low) River guppy populations (see also Fig. 1).

Population	*A*	Divergence	G'_{ST}
Up	34 (31-36)	**Up-Mid**	0.069 (0.027-0.134)
Mid	55 (52-58)	**Mid-Low**	0.238 (0.170-0.320)
Low	105 (97-112)	**Up-Low**	0.328 (0.253-0.40)

Genetic differentiation (mean G'_{ST}, 5-95% CI) among populations significantly increases with distance between population (Randomisation test: $P < 0.001$ for all tests) (Table 2 and Fig. 1). There is a large and significant difference in microsatellite allelic richness between the Upper (Up) and Mid (Mid) Naranjo and Lower Aripo (Low) populations (Randomisation test: $P < 0.001$ for all tests) (Table 2 and Fig. 1). This progressive loss in microsatellite allelic richness and heterozygosity with increasing upstream location is indicative of a bottleneck and a reduction in effective population size.

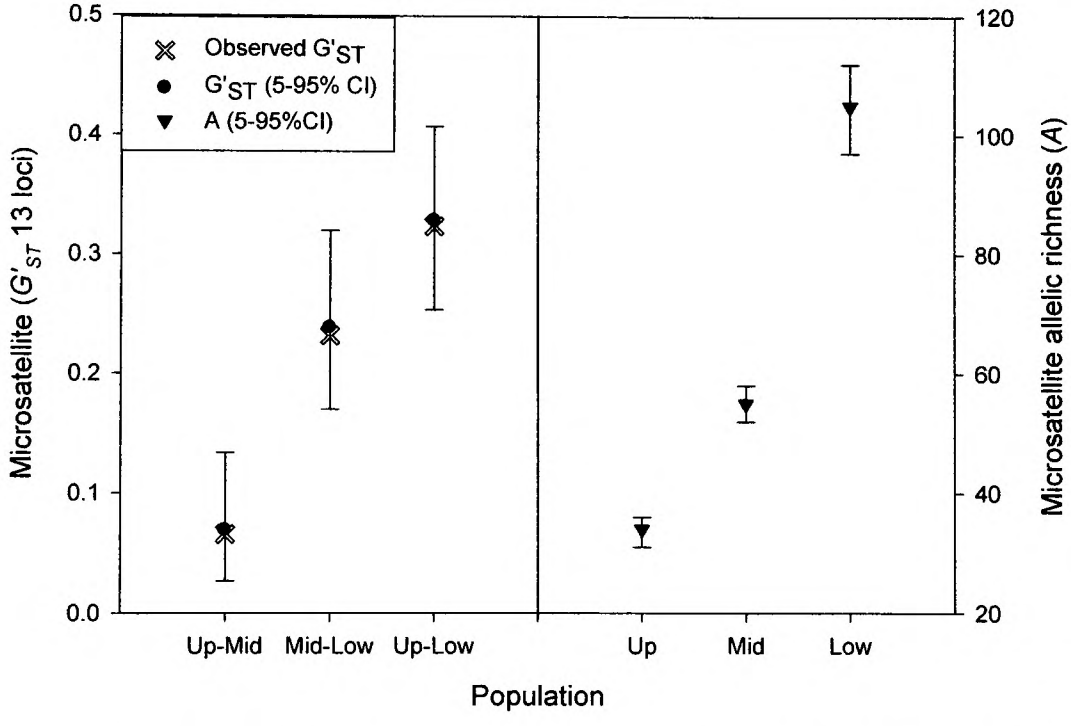


Figure 1. Genetic differentiation (G'_{ST}, ± 5-95% CI) at 13 microsatellite loci among pairwise comparisons of the guppies of the Upper and Mid Naranjo, and Lower Aripo population (left panel). Genetic differentiation is significantly different between all three comparisons (Randomisation test: $P < 0.001$ for all tests). Allelic richness (*A*, 5-95% CI) of microsatellite loci (right panel). Populations differ significantly in allelic richness (Randomisation test: $P < 0.001$ for all tests) indicating reduced population size with increasing upstream locality.

Structure 2.3.1 (Pritchard et al. 2000) was used to assess population structure in the Up, Mid, and Low populations (Fig. 2). Figure 2 shows cluster assignment in the Up, Mid, and Low based on three clusters. Clusters are not confined within populations and Figure 2 shows a clear distinction between the Low versus the Mid and Up populations based on the assignment to cluster 3 (Lower: 96.1%; Mid: 0.9% and Upper 0.0%). Assignment of the Up population to cluster 1 is high (cluster 1 = 95.1%, cluster 2 = 4.9%) but Mid population assignment is split between clusters 1 and 2 (cluster 1 = 67.8%, cluster 2 = 31.3%) (Fig. 2). The pattern of assignment of the Mid population to both clusters 1 and 2 suggests that migration is downstream biased from Up into Mid Naranjo (Fig. 2). The highest likelihood for the number of clusters (mean LnP(D)) for the Up, Mid and Low populations was $k = 3$ (Fig. 3). However interpretation of the most probable value of k using STRUCTURE HARVESTER (Earl 2011) was $k = 2$ (Fig. 3). This reduced k would put the Up and Mid into a single cluster despite these populations having a genetic differentiation significantly greater than zero (Fig. 1) and an observed heterozygosity and allelic richness that is significantly lower in the Up compared to the Mid population (Fig. 1 and Table 1).

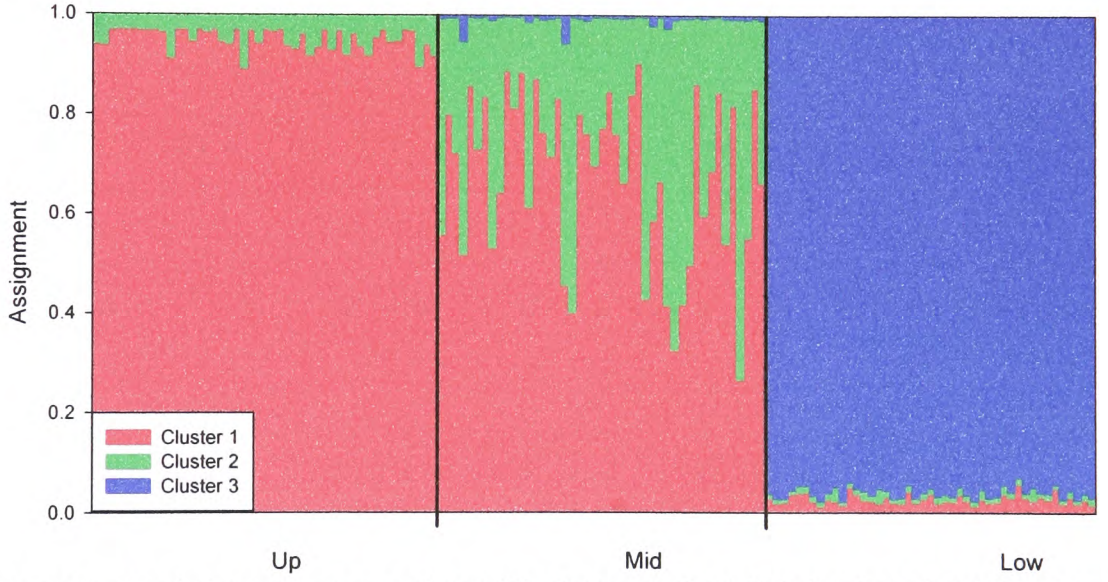


Figure 2. Population assignment estimated by STRUCTURE 2.3.1 (Pritchard et al. 2000), using a location prior (Hubisz et al. 2009). Guppies were sampled from the Upper (Up, $n = 47$; 1 = 95.1%, 2 = 4.9%) and Mid Naranjo (Mid, $n = 45$; 1 = 0.678, 2 = 31.3%, 3 = 0.9%) and Lower Aripo (Low, n = 45; 1 = 2.7%, 2 = 1.2%, 3 = 96.1%) populations. Each individual is represented by a single vertical bar with proportions of individuals' assignment to cluster 1 (red), cluster 2 (green) or cluster 3 (blue). The number of estimated clusters (k) was three. An admixture prior is used with separate inference of alpha.

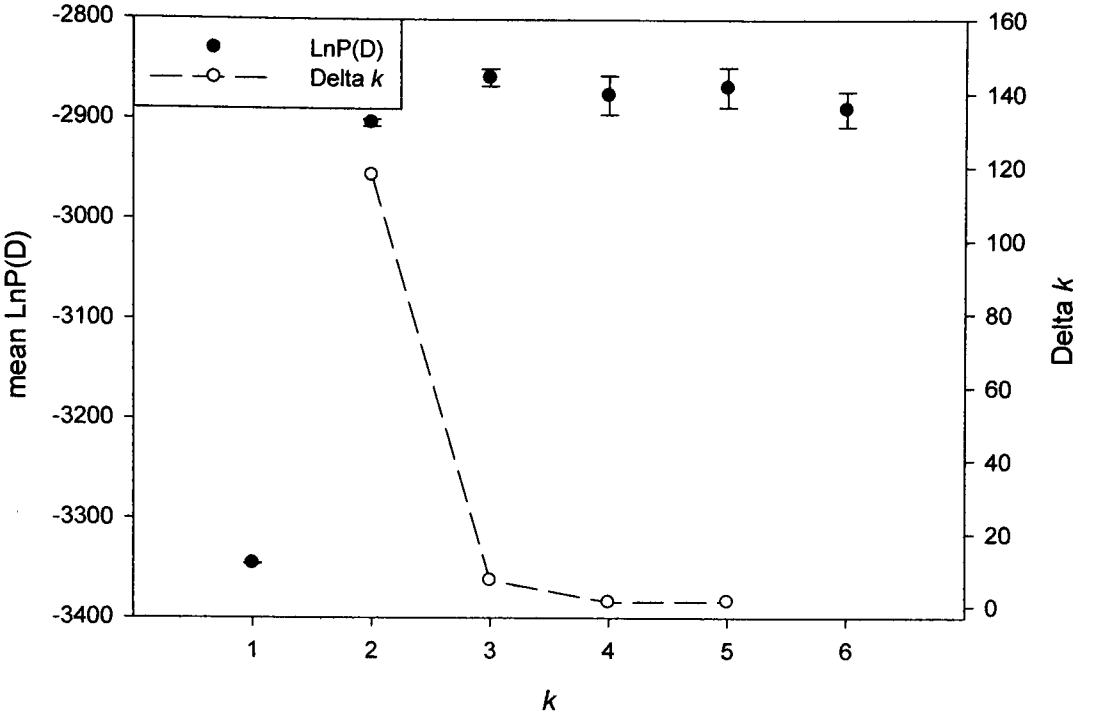


Figure 3. Mean LnP(D) (± stDev) for the number of clusters, *k* and Delta *k* (filled and open circles, respectively) of the Up, Mid and Low populations of guppies. Mean LnP(D) was estimated in STRUCTURE 2.3.1 (Pritchard et al. 2000), using a location prior (Hubisz et al. 2009). Delta *k* was estimated using the Evanno method (Evanno et al. 2005) in STRUCTURE HARVESTER (Earl 2011). Using mean LnP(D) from STRUCTURE the most likely estimate of *k* is 3 (2857.85 ± 8.63), the delta *k* estimate however, indicates *k* = 2 is the most likely value for the number of clusters.

In order to further explore the divergence between the Up and Mid populations and confirm the number of clusters in the UN and MN Structure (Pritchard et al. 2000) software was used to estimate admixture of the Up and Mid populations (without the Low population). The number of clusters observed between the UN and MN populations was two (*k* = 2). This was confirmed using the Evanno method (Evanno et al. 2005) (see Fig. 4).

NEWHYBRIDS 1.1 (Anderson and Thompson 2002), used to further explore the relationship between the Up and Mid population showed 95.7% of Up individuals received pure posterior probability assignment to the Up population (pure = P (Up) ≥ 0.95) (Fig. 5). However, assignment of Mid individuals to a pure Mid category was not observed. Almost half (44.4%) of the Mid individuals received pure Up assignment and the remainder were assigned to one of four hybrid classes (F1, F2, Up Bx, Mid Bx). Altogether, these results indicate that there is little evidence for upstream migration from the Mid to the Up population. The rate of migration in the downstream direction (Up to Mid), appears considerably higher. The pure assignment in the Mid population may represent a more distant population, such as the Low (see Fig. 5).

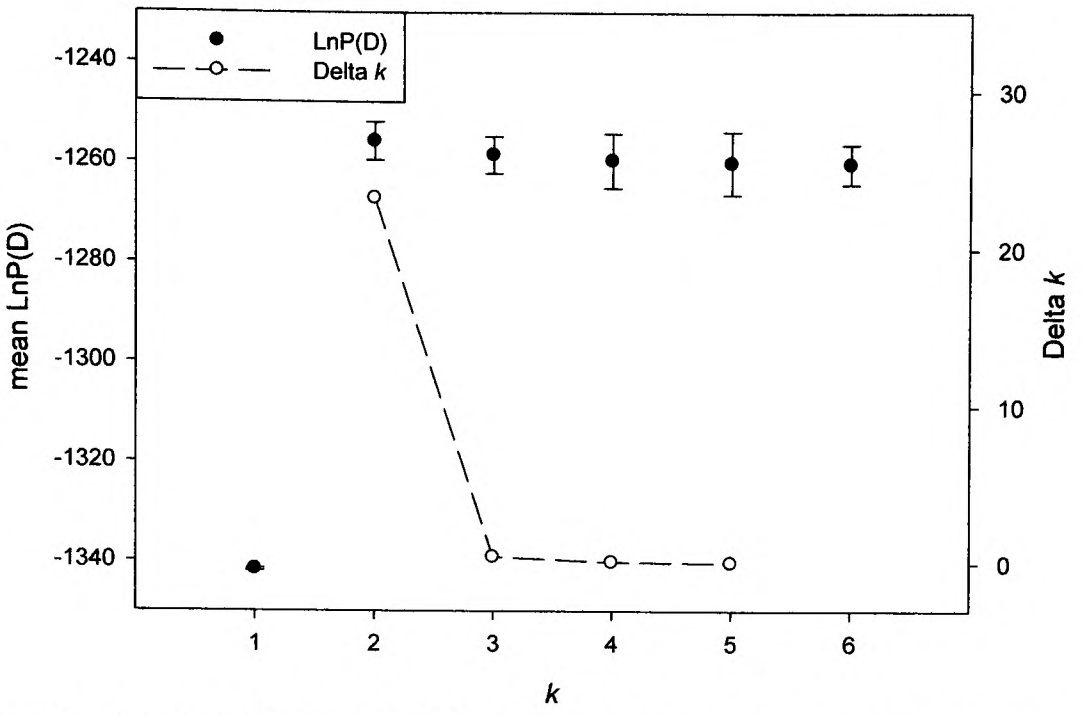


Figure 4. Mean LnP(D) (± stDev) and Delta *k* (filled and open circles, respectively) of the Up and Mid populations of guppies. Mean LnP(D) was estimated in STRUCTURE 2.3.1 (Pritchard et al. 2000), using a location prior (Hubisz et al. 2009). Delta *k* was estimated using the Evanno method (Evanno et al. 2005) in STRUCTURE HARVESTER (Earl 2011). Using mean LnP(D) from STRUCTURE the most likely estimate of *k* is 2 (-1255.65 ± 3.82), the delta *k* estimate also indicates *k* = 2 is the most likely value for the number of clusters.

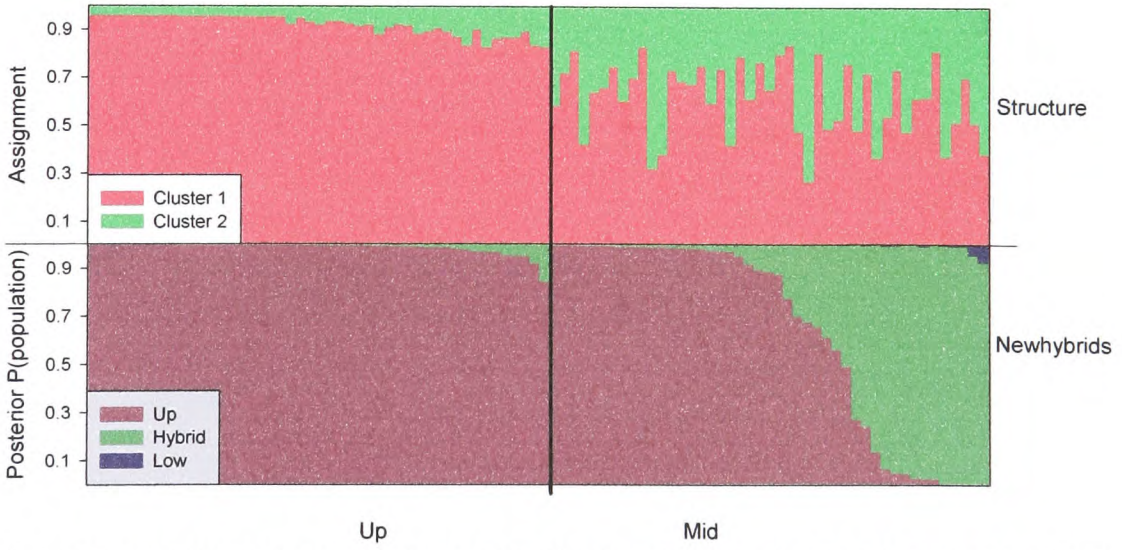


Figure 5. Population assignment estimated by Structure 2.3.1 (Top) and posterior probability of individual assignment to the pure Up, Hybrid, and Low category estimated using NEWHYBRIDS (Bottom). Each individual is represented by a single vertical bar. Guppies were sampled from the Upper (Up, n = 47; assignment, 1 = 92.6%) and Mid Naranjo (Mid, n = 45; assignment, 2 = 36.8%) populations. The number of estimated clusters (*k*) was two which was confirmed using the Evanno method (Evanno et al. 2005) (see Fig. 4).

Effective population size, N_e and migration rates, Nm, were estimated using MIGRATE 2.4 (Beerli and Felsenstein 1999, 2001). Theta (Θ) declines increasing altitude (Θ (5-95% CI): Low = 1.856 (1.750-1.971); Mid = 1.001 (0.937-1.073); Up = 0.868 (0.827-0.911)) (see Fig. 6 and Table 3 for estimates of N_e). Migration rates are greater downstream than upstream (Table 3 and Fig. 7). Direct migration between the Up and Low populations was low in comparison to migration rates among adjacent populations (Table 3 and Fig. 7). However, the full migration model showed the best fit with the data (Table 5). This indicates that while the migration rates are low, upstream and downstream migration occurs across the three populations.

Table 3. Estimates of population size ($N_e \pm$ 5-95% CI) (on the diagonal) and effective migration ($Nm \pm$ 5-95% CI) (migration from row to column, upstream and downstream migration above and below the diagonal, respectively). Effective population size is estimated based on a microsatellite mutation rate of $2x10^{-4}$ (see Ellegren 2000) (see Table 4 for comparison of Mid and Low to estimates from other studies).

		Migration from		
Migration to		Up	Mid	Low
	Up	1085 (1034-1139)	0.55 (0.49-0.61)	0.03 (0.02-0.06)
	Mid	1.90 (1.77-2.03)	1251 (1171-1341)	0.61 (0.53-0.69)
	Low	0.14 (0.11-0.18)	1.20 (1.09-1.34)	2320 (2188-2464)

Table 4. Estimates, from van Oosterhout et al. (2006b) and Barson et al. (2009), of population size (N_e) (on the diagonal) and effective migration (Nm) (migration from row to column, upstream and downstream migration above and below the diagonal, respectively) between the Mid and Low populations. Effective population size is estimated based on a microsatellite mutation rate of $2x10^{-4}$. Barson et al. (2009) estimates are presented twice as estimated from Aripo populations alone and from a whole drainage approach, respectively.

		Migration from					
Migration to		van Oosterhout				Barson	
		Mid	Low			Mid	Low
	Mid	52	0.06		Mid	381, 545	0.10, 0.14
	Low	0.27	1197		Low	0.82, 0.14	2118, 350

Table 5. Results of hierarchical reduction in migration models between the Upper and Mid Naranjo and the Lower Aripo populations calculated in MIGRATE 2.4 (Beerli and Felsenstein 1999, 2001). The full open model has the highest Log(likelihood) and the lowest Akaike's Information Criterion (AIC) indicating it has the greatest applicability to the data (Beerli 2002).

	Model	**LnL(test)**	**AIC**
	Full open migration model	-824.72	1667.44
1	All upstream migration blocked	-12145.48	24302.97
2	Effective barrier to upstream migration between the Mid and Up populations	-8184.26	16382.52
3	Effective barrier to upstream migration between the Low and Mid populations	-4291.44	8596.87
4	Migration only between adjacent populations (direct Up Low migration blocked)	-1355.43	2724.85

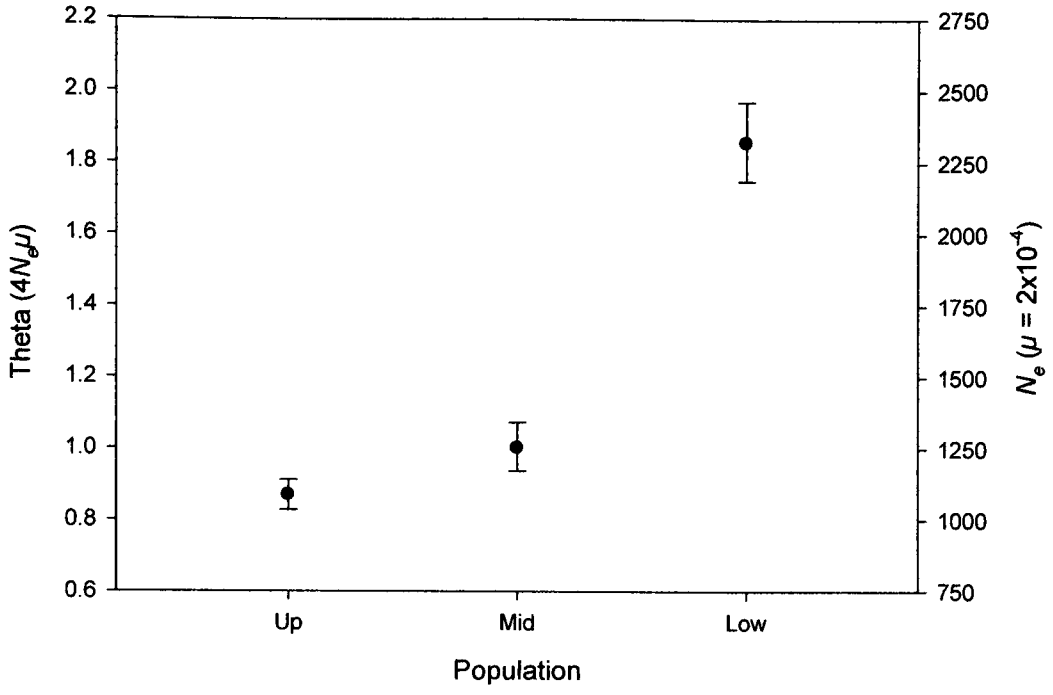


Figure 6. Estimates of theta and effective population size, N_e ($\mu = 2x10^{-4}$), using MIGRATE 2.4 software and 13 microsatellite loci. Population size decreases with increasing altitude as expected when rivers become smaller in the upland sites (see also Table 3).

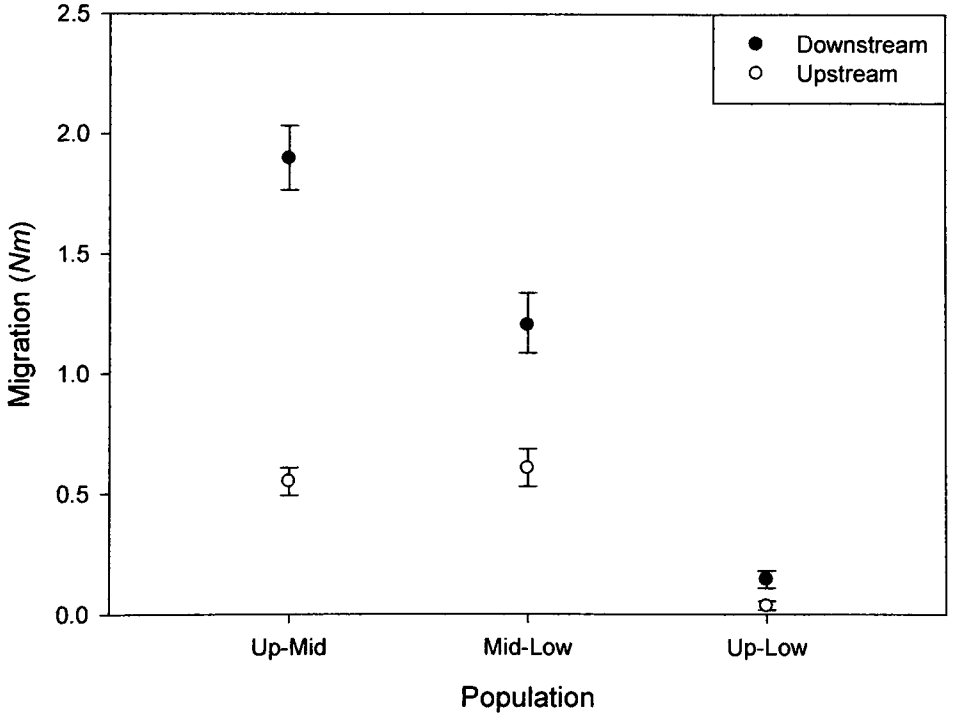


Figure 7. Estimates of downstream (solid symbols) and upstream (open symbols) migration (*Nm*) amongst populations of guppies from the Upper and Mid Naranjo, and the Lower Aripo River. Migration is downstream biased and higher between neighbouring populations (see also Table 3).

Discussion

The present study assesses the genetic variation at (neutral) microsatellite loci in three guppy (*Poecilia reticulata*) populations in the Aripo River in Trinidad. Genetic diversity and heterozygosity at 13 microsatellite loci was significantly higher in the lowland population, the Lower Aripo (Low) than in the upstream Mid Naranjo (Mid) population, while the most upstream, Upper Naranjo (Up) population showed the lowest level of microsatellite variation. The progressive loss of neutral genetic variation in upstream guppy populations is consistent with the small effective population size and reduced river volume in the upland habitats compared to the lowlands (Carvalho et al. 1991; Shaw et al. 1991; Shaw et al. 1994; Alexander et al. 2006; Barson et al. 2009). Neutral genetic variation in small upstream populations is further reduced by limited contemporary upstream migration due to waterfalls (Crispo et al. 2006). Concordantly, in the present study estimated migration rates are downstream biased.

When considering the Up, Mid, and Low guppy populations, the number of distinct genetic clusters (k) as estimated by the software STRUCTURE 2.3.1 (Pritchard et al. 2000) was three (highest log likelihood). However, the Evanno et al. (2005) method of estimation of k indicated just two distinct genetic clusters among these three populations. Further investigation into the number of clusters showed that when considering just the Up and Mid populations the number of clusters estimated by STRUCTURE and the Evanno et al. (2005) method was two. Evanno et al (2005) explored how STRUCTURE partitions clusters in a range of migration models (Island, hierarchical island and contact zone) where populations are equally different from one another. In the present study genetic divergence is not equally partitioned between the Up, Mid, and Low. Kalinowski (2011) argues that equality of divergence is a biologically simplistic assumption and that more realistic situations may affect the manner in which Structure assigns individuals to clusters. Kalinowski (2011) uses a simulated dataset to show that STRUCTURE can cluster populations into too few clusters, particularly when divergence is not equally partitioned between populations.

Estimates of upstream migration rates were significantly greater than zero and higher than that estimated in previous studies (van Oosterhout et al. 2006b; Barson et al. 2009) (compare Tables 3 and 4). Similarly, the effective population size estimates differ between these studies (compare Tables 3 and 4). The discrepancies in migration rate and effective population size estimates can be explained by: (1) differences in the microsatellite loci and number of loci analysed in the three studies, and (2) differences in the number of populations analysed and hence, the migration model employed in the software.

Barson et al. (2009) used variation at eight microsatellite loci in eight (four upland and four lowland) guppy populations to test for recent bottlenecks. The authors found that the Mid population did show signs of fluctuations in population size. Fluctuations in population size in the Mid population have been observed in a mark recapture study that showed a decline in the numbers of guppies in the Mid at the start of the wet season rains (van Oosterhout et al. 2007a). The resulting in a population-dynamics may explain the differences in N_e and Nm estimates between studies across years. However, there may also be methodological differences among these studies that could explain the discrepancies in estimates. For example, the present study analyses the genetic variation of 13 microsatellite loci, which is a considerably larger panel compared to previous studies (eight microsatellite loci in van Oosterhout et al. 2006b; and eight used in Barson et al. 2009). Stochastic differences in gene

diversity between markers may affect the demographic estimates, particularly when only a small number of loci are analysed (see Jorde et al. 1997 for type II errors as a function of number of microsatellite loci). Barson et al. (2009) demonstrated the effects of unknown populations (or "ghost-populations") on the estimates of the Ne and *Nm* of guppy populations of the Caroni Drainage in Trinidad. Their analysis confirms the study by Beerli (2004) who showed that the estimates of effective population sizes become inflated with an increased number of immigrants from ghost populations. The potential bias caused by such ghost populations is particularly pronounced with high migration rate among populations (Slatkin 2005). Concordantly, the study by Barson et al. (2009) showed that the effects of ghost populations is considerably reduced by considering between-river migration, resulting in a marked decrease in effective population size estimates. With respect to the present study, the estimates of N_e (and *Nm*) may be inflated because the tributary furthest downstream that acts as a "super-sink population" (see Barson et al. 2009) was not included in the present analysis.

Guppies are found throughout the rivers of Trinidad and Tobago (see Carvalho et al. 1996; Barson et al. 2009; Willing et al. 2010) and are renowned for their presence above larger barriers such as waterfalls (see Crispo et al. 2006). These barriers to gene flow prevent the migration of the majority of their predators (reviewed by Endler 1995; Houde 1997; Magurran 2005) , but they do not completely prevent the upstream migration of guppies. In fact, while estimates of upstream migration are low they are still significantly higher than zero. One study provides evidence of admixture between remote populations separated by land and/or the sea (Willing et al. 2010). While artificial translocations cannot be ruled out, the considerable morphological differences between guppies of Trinidad should not be taken as evidence of barriers to migration. Indeed neutral genetic differentiation may be influenced by barriers to gene flow that do not influence gene flow at other genetic loci (see Schierup 1998).

Concluding remarks

Evidence from this and previous analyses using microsatellite and mark-release-recapture techniques suggest that the Naranjo River populations could be subject to annual fluctuations in population size (van Oosterhout et al. 2007a; Barson et al. 2009). Annual flooding events may contribute to downstream migration and reduce genetic diversity within upstream sites. However, population bottlenecks are expected to be less severe for loci under the effects of balancing selection (e.g. Aguilar et al. 2004). In the subsequent chapters, I investigate the role of migration, effective population size, and parasite selection on the maintenance of the genes responsible for recognition of parasites and initiation of the specific immune response, the Major Histocompatibility Complex (MHC).

Genetic variation at MHC class II*B* (*DAB*) genes, microsatellite variation and parasites in three wild guppy (*Poecilia reticulata*) populations in Trinidad

Chapter 3

Abstract

The present study compares the genetic variation at (neutral) microsatellite loci to that of (non-neutral) MHC genes among three guppy (*Poecilia reticulata*) populations in the Aripo River in Trinidad. In addition, a measure of parasite fauna is used to ascertain the source of selection on the MHC. A small upland guppy population appears to maintain a high MHC diversity despite little or no selection by microparasites in combination with considerable genetic drift. Population genetic analysis of neutral microsatellite variation rules out the possibility that migration can explain the high level of MHC diversity in this population. These findings suggest that either sexual selection and/or Associative Balancing Complex (ABC) evolution may be responsible for the exceptionally high MHC diversity detected in this population.

Introduction

The Major Histocompatibility Complex (MHC) is one of the most polymorphic loci in the vertebrate genome (Klein et al. 2000), and this level of polymorphism is greater than expected from the neutral model of evolution (see Kimura et al. 1963; Ford 2002; Garrigan and Hedrick 2003). Balancing selection is generally considered to account for the maintenance of such polymorphism (Edwards and Hedrick 1998; Hughes and Yeager 1998; Penn and Potts 1999; Piertney and Oliver 2006). The MHC heterodimer binds to self and non-self peptides or antigens, and these antigens are bound within a region of the molecule known as the peptide binding region (PBR). The present study focuses on class II (*B*) MHC, responsible for binding intracellular parasites (reviewed by Hughes and Yeager 1998). MHC binding foreign peptide initiates the adaptive immune response through T-cell recognition and activation (see reviews by Hughes and Yeager 1998; Penn and Potts 1999; Klein et al. 2000), and hence, parasites that can avoid detection by the MHC have a significant selective advantage.

Selection and drift

DNA polymorphism within the PBR of the MHC is believed to be maintained by selection for recognition of the numerous parasites an individual may encounter throughout its life. Natural selection acting on parasites to avoid detection by the MHC leads to continuous adaptation and counter-adaptation in both the parasites and hosts (coevolutionary arms race, or Red Queen dynamics, see Lively and Dybdahl 2000; Decaestecker et al. 2007; Paterson et al. 2010).

The MHC is therefore under balancing selection due to parasites that enhance the allelic diversity in MHC locus. Four (non-mutually exclusive) hypotheses of balancing selection include (1) the heterozygote advantage hypothesis by overdominance (Doherty and Zinkernagel 1975a; Penn and Potts 1999), (2) negative frequency dependent selection (Kojima 1971; Takahata and Nei 1990), (3) heterogeneous or fluctuating selection (Hedrick et al. 1976; Hedrick 2002), and (4) Associative Balancing Complex (ABC) evolution (van Oosterhout 2009a, b). The first three models describe how parasites act to maintain polymorphism at the MHC by increasing the fitness of individuals with a range of MHC alleles. In contrast, ABC evolution, postulates that selection operates against the mutations that have accumulated in or nearby the MHC genes. Recessive deleterious mutations located in the genomic region surrounding MHC do not often become expressed in large outbreeding populations, because MHC alleles are generally found in heterozygous genotypes. However, if a population goes through a bottleneck, the level of homozygosity increases and selection can act against MHC alleles that have increased in relative frequency by random genetic drift. This reduces the frequency of common alleles and maintains the MHC diversity despite drift during a bottleneck. Importantly, selection due to ABC evolution acts independently of selection due to parasites.

Random genetic drift tends to reduce genetic polymorphism across loci in the entire genome. Drift increases with decreasing effective population size (N_e), however, a number of studies have demonstrated the maintenance of MHC polymorphism in populations with small effective population size (Aguilar et al. 2004; van Oosterhout et al. 2006b, but see Hedrick et al. 2001; Bollmer et al. 2007). van Oosterhout et al. (2006b) measured MHC variation in two populations of guppies (*Poecilia reticulata* Peters), the Mid and Lower Aripo populations in Trinidad. The small upstream population (Mid) showed significant reduction in allelic richness at microsatellite loci, but not at the MHC (class II*B*) loci. Simulations showed that strong selection ($s \geq 0.2$) was required to explain the level of polymorphism observed in the small Mid population. The authors also measured parasite prevalence in the two populations and suggested that selection by parasites was primarily responsible for maintaining the high level of MHC polymorphism in the Mid. However, sexual selection could not be ruled out as an alternative mechanism maintaining MHC diversity (van Oosterhout et al. 2006b). Similar conclusions were drawn in an earlier study by Aguilar et al. (2004) who demonstrate that strong selection ($s > 0.5$) was required to maintain the high level of polymorphism at the MHC in small bottlenecked population of the San Nicolas Island fox (*Urocyon littoralis dickeyi*).

Parasites

Gyrodactylus species (particularly *G. turnbulli* and *G. bullatarudis*) are the most abundant guppy ectoparasites in the Northern Mountain range of Trinidad (Martin and Johnsen 2007; Cable In press). These parasites are born already pregnant with a single almost fully developed embryo. Subsequent generations are enveloped like Russian dolls. These parasites do not require an intermediate host, and consequently, have short generation times (as little as 24h at 25°C) (see Cable and Harris 2002; Bakke et al. 2007). They graze on the mucus and epithelial layer of guppies, and can reach intensities in excess of 100 worms on an individual guppy. Field experiments have shown that these parasites can significantly reduce the survival (or recapture rate) of their guppy hosts (van Oosterhout et al. 2007a). Damage caused by these parasites has furthermore been shown to increase susceptibility of the cichlid fish, Nile tilapia (*Oreochromis niloticus* L.) to bacteria through secondary infections (Xu et al. 2007). In addition, evidence for the maintenance of a particular MHC class II*B* allele (a-type) by gyrodactylids has

been shown in guppies (Fraser and Neff 2009, 2010). Altogether, these parasites represent a significant selective force in many wild guppy populations in Trinidad.

Digenean metacercariae are more diverse than gyrodactylids but less numerous in their infection of guppies (see review by Cable In press). These larval stage digeneans will encyst on almost any host tissue including the skin, musculature, gills, blood vessels, and internal organs, such as the heart and brain. Infections by adult digeneans are less common. In other hosts, digeneans have been shown to alter host behaviour to increase predation risk and transmission to the terminal host (e.g. Lafferty and Morris 1996; also see review by Barber et al. 2000).

Putative fungal infections observed on guppies consist of white fibrous growths on the gills, fins and/or body. The reason individuals succumb to such infections is unknown but it is generally assumed that many microparasitic diseases are related to reduced immune competence, sometimes triggered by environmental stressors (e.g. Bly et al. 1997)

Maintenance of MHC polymorphism in three populations of guppies

MHC and neutral microsatellite variation may be lost due to the fixation of alleles in small or bottlenecked populations (i.e. genetic drift) (Hedrick et al. 2001a; Miller and Lambert 2004; Bollmer et al. 2007). However, MHC variation may also be reduced as a consequence of a relaxed parasite selection (Slade 1992). In the present study, I aim to demonstrate the role of parasite selection in the maintenance of MHC polymorphism. I compare MHC variation between three guppy populations: 1) Upper Naranjo (Up) with reduced parasite fauna and small effective population size, 2) Mid Naranjo (Mid), 0.5km further downstream, a small population characterised by a diverse parasite fauna (see van Oosterhout et al. 2006b; van Oosterhout et al. 2007a), and 3) Lower Aripo (Low), furthest downstream, a large population with diverse parasite fauna and high predation (Magurran 2005; van Oosterhout et al. 2006b). I predict that neutral genetic variation (at 13 microsatellite loci) will decrease with population size, and also because of limited upstream migration (Shaw et al. 1991; Carvalho et al. 1996; Crispo et al. 2006, see also Chapter 2; Barson et al. 2009). Similarly, variation at the MHC is predicted to decrease in the upstream direction with population size and parasitism. The comparison of allelic richness between both types of genetic loci allows me to quantify the effect of balancing selection which maintains the number of MHC alleles above that predicted from a (neutral) drift-mutation-migration equilibrium

Materials and Methods

Guppy collection

Wild caught guppies were analysed for MHC ($n = 60$) and neutral variation at thirteen microsatellite loci ($n = 150$). Guppies were captured (June 2007) using a seine net, given a lethal dose of MS222 and stored separately in molecular grade ethanol. In order to minimise allele frequency biases associated with sampling families or cryptic social structure, each site was sampled at multiple locations in order to maximise the chance of catching guppies from separate shoals. Three sampled guppy populations are found along the Aripo River, which is part of the Caroni Drainage in the Northern Mountain Range of Trinidad (grid reference: Up, 20P 692498.44 E 118257.53 N; Mid, 20P 693100 E 1181800 N; Low, 20P 6914000 E 1177700 N).

Molecular analysis

Genomic DNA was extracted from the caudal fin using the HotSHOT protocol of Truett et al. (2000). Extracted DNA was PCR amplified at 13 microsatellite loci (see Chapter 2). MHC class II*B* variation from guppies was cloned and sequenced from HotSHOT extracted DNA (Truett et al. 2000). The first PCR amplification took place using a degenerate forward primer (*DABdegfb*: 5'-GTG TCT TTA RCT CSH CTG ARC-3'), situated near the 5' end of exon 2 of the MHC class II*B* loci and a reverse primer (*DABR6b*:5'-TTA GGG TAG AAA TCA TAA ACT CTG CA-3'), situated near the 5' end of exon 3. These primers amplify approximately 750-850 bp of genomic DNA, including 222 bp of exon 2, from codon 22, through intron 2, up to and including the first 68 bp of exon 3. The forward primer (*DABdegfb*) is known to amplify all previous guppy MHC class II*B* alleles (see van Oosterhout et al. 2006a; van Oosterhout et al. 2006b). This primer is based on *DABdegF*, a degenerate primer designed by aligning all GenBank published *P. reticulata* and *Xiphophorus maculatus* (Southern platy) *DAB* sequences, as well as *P. reticulata*, *P. picta* (swamp guppy), and *P. sphenops* (short-finned molly) *DAB* sequences previously amplified in our laboratory. The forward primer was modified here to reduce degeneracy and avoid the polymorphism that was formally introduced by PCR errors (see Cummings et al. 2010). *DABR6b* was designed using all previous amplified guppy alleles from van Oosterhout et al. (2006a; 2006b).

The first PCR of 25µl contained 2.5 pmoles of the specific reverse primer and 12.5 pmoles of the degenerate forward primer, 2.5 mM $MgCl_2$, and 0.2 mM each dNTP and 0.5 U *Taq* polymerase (Bioline Ltd., London). The touchdown PCR reaction consisted of an initial step of 95°C for 3 min followed by one cycle of 94°C for 1 min, 61°C for 1:30 min, 72°C for 1:30 min; then one cycle of 94°C for 1 min, 59°C for 1:30 min, 72°C for 1:30 min; and then 28 cycles of 94°C for 1 min, 58°C for 1:30 min, 72°C for 1:30 min; with a final step of 72°C for 30 min. Products were resolved on a 1% agarose gel.

Cloning took place using DH5α (Invitrogen) competent bacterial cells, using pGEM-T Easy Vector (Promega Ltd.) according to the manufacturer's instructions. For each individual sample, between 12 and 18 (mean 16.5) colonies were picked from plates and dipped directly into the second PCR of 35µl, containing 7 pmoles of each M13 primer, 2.5 mM $MgCl_2$, 0.2 mM of each dNTP, 0.7 U *Taq* polymerase. Additional (30 in total) colonies were picked for one individual because none of the MHC alleles were confirmed by alleles from other individuals so an additional PCR, cloning and sequencing step was run to confirm those sequences. The PCR reaction consisted of an initial step of 95°C for 5 min followed by 31 cycles of 94°C for 1 min, 54°C for 1:30 min and 72°C for 1:30 min; with a final step of 72°C for 30 min. The products were resolved on an agarose gel. ExoSAP cleanup and sequencing was performed by SymBio Corporation USA on an ABI3730xl.

Cloning effort was originally 18 clones per individual. The binomial distribution calculation performed in MS Excel shows that 18 clones gives greater than 95% chance of observing all an individual's alleles (assuming three *DAB* loci). However, very few individuals had the maximum number of six alleles. Of the 41 guppies that were sequenced for 18 clones, 1 individual had 6 alleles, 2 had 5, and 38 had 4 for fewer alleles; mean = 2.54). Consequently, in order to reduce redundancy, 12 clones were sequenced per individual in order to detect 4 alleles with more than 95% probability. The variation in cloning effort differed between populations (total number of clones sequenced per 20 individuals per population: Up = 342, Mid = 336, Low =

288). For this reason, measures of allelic richness and genetic differentiation were calculated using all clones per individual using a re-sampling technique, equalising the cloning effort to 12 clones per individual. The reduction in the number of clones per individual from 18 to 12 did not significantly change allelic richness (A) or genetic differentiation (G'_{ST}) in any of the populations (Randomisation test: A, Up P = 0.297, Mid P = 0.159, Low P = 0.238; G'_{ST}, Up-Mid P = 0.446, Mid-Low P = 0.525, Up-Low P = 0.481) (see Appendix 3).

Sequences were checked and aligned using CODONCODE ALIGNER version 2.0 and MEGA 4.1 (Tamura et al. 2007). The forward primer and the intron and exon 3 regions were removed and exon 2 sequences were aligned to other guppy MHC class II*B* sequences (Sato et al. 1995; van Oosterhout et al. 2006b) and to cichlid class II*B* sequences (Figueroa et al. 2000). Errors can be introduced to sequences by heteroduplex mismatch repair or sporadic substitutions caused by *Taq* polymerase misincorporations (Keohavong and Thilly 1989; Kobayashi et al. 1999; Kanagawa 2003; Cummings et al. 2010). MHC alleles were confirmed only when they were observed in at least two independent PCRs (Lukas et al. 2004; Lukas and Vigilant 2005). Two individuals had MHC alleles that were not confirmed from other PCRs. Cloning and sequencing was repeated for one of these individuals and all four alleles were confirmed. However, the second individual was not cloned and sequenced again, and therefore was removed from the study (Low, n = 19).

Sequence analysis

OMEGAMAP version 0.5 (Wilson and McVean 2006) is a program that uses Bayesian inference to calculate the selection parameter, ω (d_N/d_S), in the presence of recombination (ρ). This software was used to estimate the location of the PBR in the guppy MHC class II*B*. Divergent selection is expected to act on codons of the PBR in order to allow recognition of divergent parasite antigen by separate MHC alleles. Therefore, a significant ω value for a codon is an indication that it is part of the PBR. This firstly, confirms the presence of balancing selection acting to maintain polymorphism in exon 2. Secondly, it allows phylogenetic trees to be built that consider variation that is inside and outside the PBR. Each simulation was run 600,000 Markov-chain Monte Carlo iterations and the first 100,000 were discarded as a burn-in. Iterations were recorded at intervals of 100 with 10 orderings of sequences. As well as ω and ρ, the software estimates the transition transversion rate ratio (κ), a mutation rate (μ) and the rate of insertion/deletions (φ). These parameters were estimated independently for each codon because the roles of adjacent codons may be different in the MHC class II molecule (Brown et al. 1993; Stern et al. 1994). No codon bias was assumed. priors used were as suggested by software documentation in the case where no information is available on parameter estimates: improper inverse distributions were thus applied to μ, κ and φ with starting values of 0.1 for μ and κ and 3.0 for φ, inverse distributions were applied to ω and ρ ranging between 0.01 and 100. The model was run twice for the MHC allele frequencies from the Aripo River which was treated as a single population. Convergence was confirmed between separate runs and they were combined to obtain the final estimates of ω.

MEGA was used to construct unrooted neighbour joining (NJ) trees of nucleotides and of amino acids (No. of differences). The NJ tree is used because it does not require equivalent amounts of divergence between lineages (Saitou and Nei 1987; Kalinowski 2009). The tree is used primarily to observe allele locus specificity. Increased intra-locus (to inter-locus) recombination will be evident in the tree as locus specific clades (Reusch and Langefors 2005). Such a result

would be confirmed where individuals have no more than two alleles from each clade. Amino acid trees are used to compare the relative level of functional (amino acid changing) divergence with that of nucleotide divergence. MEGA was also used to test sequences for deviation from neutrality (Tajima's *D*). Codons that are within the PBR of the MHC molecule are expected to be under balancing selection and are therefore expected to have a ratio of nonsynonymous (d_N) to synonymous (d_S) nucleotide substitutions in excess of 1 (see Ford 2002; Garrigan and Hedrick 2003). The d_N/d_S ratio was calculated on all alleles in the program MEGA 4. A pairwise analysis (and overall average) was calculated for whole sequences and for regions of sequence pertaining to inside and outside the PBR, using the method of Nei and Gojobori (Nei and Gojobori 1986). DNASP (Librado and Rozas 2009) was used to calculate synonymous and nonsynonymous nucleotide differentiation (π). Mean population nucleotide differentiation (5-95% CI) was calculated in Minitab 12.1 using a bootstrap procedure written in a global macro, and using 100,000 runs to estimate the parameter values.

Parasite analysis

Selection from an array of parasites is predicted to maintain polymorphism at the MHC (Hughes and Yeager 1998). Therefore, MHC allelic richness is expected to correlate with parasite diversity. Guppies sampled from 2003, 2006 and 2007 were screened for ectoparasites, endoparasites and fungal growth. These data contain two (known) species of gyrodactylid ectoparasite (*Gyrodactylus turnbulli* and *G. bullatarudis*) and nine species of adult digenean with additional larval (cyst) stages (see review by Cable In press). No distinction is made between particular species of gyrodactylids or digeneans and fungal infection is scored in three classes (0 = no detectable fungus, 1 = fungal growth, 2 = heavy growth). Guppies were caught, anesthetised and preserved using the method described above. Guppy dissection and parasite counts were conducted in the laboratory of Dr Cable (University of Cardiff). Briefly, the preserved fish was transferred to a Petri dish and completely immersed in ethanol. Using a dissecting microscope the surface of the fish was scanned and any parasites removed. The ethanol in which the fish had been originally fixed was also screened for any dislodged parasites. The fish was then dissected by removing all major organs in turn (each gill arch, brain, eyes, intestine, liver, heart, spleen, swim bladder) and recording any visible parasites. Gyrodactylid prevalence was analysed here using a bootstrap method of 1000 replicates and a sample size of 16, based on the smallest non-zero sample. Prevalence data is compared across years among populations. Populations with significantly reduced parasite fauna over three sampling years are predicted to have reduced MHC variation due to reduced parasite selection combined with the effects of genetic drift on a small population.

Population genetic analyses

Microsatellite genotype frequencies were checked for the presence of null alleles, amplification bias and scoring errors using Micro-Checker 2.2.3 (van Oosterhout et al., 2004). Observed and expected heterozygosity (Ho and He) were calculated in Minitab 12.1. Genotypic disequilibrium was calculated using FSTAT version 2.9.3 (Goudet 1995, 2001) with a Bonferroni adjusted P-value for multiple testing at the 5% nominal level ($\alpha = 0.000641$).

Allelic richness (*A*, 5-95% CI) of both microsatellites and MHC was bootstrapped with replacement over individuals, using 1000 runs, using a global macro in Minitab 12.1. Given that the locus affiliation of MHC sequences remain unknown, it was not possible to calculate heterozygosity. Hence, allelic richness is used as proxy of gene diversity to measure the

amount of selection acting to maintain alleles within a population. Both neutral and MHC variation are subject to random genetic drift which reduces the number of alleles in a closed population. However, MHC loci are thought to be under balancing selection which acts to maintain polymorphism. Comparing allelic richness between both types of genetic loci provides a measure of balancing selection which maintains the number of MHC alleles above that predicted from a (neutral) drift-mutation-migration equilibrium. A reduction in the level of selection caused by a reduced parasite fauna is also expected to reduce the MHC allelic richness in that population.

Genetic differentiation (G'_{ST}), corrected for highly polymorphic loci, (G'_{ST}) (Hedrick 2005b) was used to measure differentiation at microsatellite and MHC loci among the Up, Mid and Low populations. Genetic differentiation may be different for MHC and microsatellite loci (Schierup 1998; Muirhead 2001). The expectation is that for loci under the effects of balancing selection (negative frequency dependent selection or overdominance) rare alleles are favored. These MHC alleles, introduced through migration between populations are maintained where microsatellite alleles are not. This reduces genetic differentiation in MHC but not microsatellite loci (e.g. in guppies see van Oosterhout et al. 2006b). A finding to the contrary would be indicative of fluctuating selection through a spatial variation in selection pressure. G'_{ST} values with 5-95% confidence intervals (CI) were calculated by bootstrapping (with replacement) over individuals, using 1000 runs. Microsatellite genetic differentiation between populations was calculated across 13 loci (for genetic differentiation at each locus see Appendix 4). MHC genetic differentiation was bootstrapped over individuals. This incorporated the variation in MHC alleles per individual (minimum 1, maximum 6; mean: Up = 2.75, Mid = 2.25 Low= 2.52). MHC genetic differentiation is calculated for the alleles combined from a number of loci. Therefore, MHC G'_{ST} is influenced by: (1) differences in the level of genetic diversity (i.e. heterozygosity), (2) variation in the number of MHC loci (see e.g. Malaga-Trillo et al. 1998) and (3) biases introduced by preferential amplification of specific alleles during PCR. Differences in heterozygosity and variations in the number of loci represent a contribution to G'_{ST} caused by population differentiation driven by a combination of demography and selection on the MHC. PCR amplification bias is an artefact that can also affect G'_{ST}. In the present study it is not possible to make a distinction between variations in the number of loci due to fixation of alleles or increased representation of specific alleles due to PCR bias (see Appendix 5).

Results

Sequence analysis

In total, 33 MHC class II*B* alleles (Fig. 1) were found in 59 individuals distributed throughout the Aripo River and Naranjo tributary (Fig. 2). There did not seem to be any connection between the alleles found in an individual and their position on the NJ tree, thus rejecting the possibility that clades are representative of loci (see Fig. 1). Between one and six alleles were found in a single individual (mean alleles per individual: Up = 2.75; Mid = 2.25; Low = 2.52) indicating a minimum of three loci in at least some individuals (but see Malaga-Trillo et al. 1998 for evidence of gene copy number variation among individual cichlid fish). There was no significant difference between the number of alleles per individual among populations (Kruskal-Wallis test (adjusted for ties): P = 0.355) and no difference between the number of MHC alleles (Allelic richness, *A*) found in each population (Randomisation test: Up-Mid P = 0.539; Up-Low P = 0.153; Mid-Low P = 0.098) (see Fig. 2 and 7).

All MHC class II*B* sequences were observed in a minimum of two separate cloning procedures, eliminating the chance for identification of PCR erroneous sequences (Keohavong and Thilly 1989; Kobayashi et al. 1999; Kanagawa 2003; Cummings et al. 2010). In addition, none of these sequences (exon 2 to exon 3) contains indels within exons or premature stop codons, suggesting they are functional. (See van Oosterhout et al. 2006b on the analysis of cDNA and expression of MHC alleles in guppies). Exon 2 is 222 bp, starting from the second nucleotide of codon 22 to the 3' end of the exon. There are 126 variable sites (56%). The mean ($\pi \pm$ 5-95% CI) pairwise nucleotide diversity per site is 0.183 (0.009-0.270) (π: Up = 0.155 (0.004-0.270); Mid = 0.174 (0.005-0.275); Low = 0.201 (0.113-0.275), and this did not differ significantly between populations (Randomisation test: $P \geq 0.37$ for all tests. Of the 33 alleles, there were 30 different amino acid sequences containing 50 polymorphic codons out of 73 (68.5%). Nine MHC class II*B* sequences are identical to those found in previous studies (see van Oosterhout et al. 2006b; Fraser et al. 2010a) (see Table 1).

Tajima's *D*, a test of neutrality, is significantly larger than zero for the MHC class II*B* alleles found in the Aripo River and Naranjo tributary (*Tajima D* = 2.350; P < 0.05), indicating a deviation from the expectation of sequence evolution governed by neutral processes (Tajima 1989). Evidence for positive selection over exon 2 (d_N/d_S; Nei-Gojobori Method, p-distance model) was apparent in just three (out of 528) pairwise comparisons and was not significant when considering the overall average (Z = -1.840, P = 1) (Nei and Gojobori 1986). Two reasons why positive selection may not be apparent are that (1) there is no positive selection and nonsynonymous (d_N) mutations are selected against or, (2) divergent alleles are maintained over long evolutionary timescales (see review by Klein et al. 1998) and synonymous (d_S) mutations have accumulated and eroded the signal of positive selection (van Oosterhout et al. 2006a). Indeed, nucleotide diversity (π) for both synonymous and nonsynonymous mutations is greater than zero (π Jukes Cantor corrected: d_S = 0.293; d_N = 0.195), which suggests that the relative excess of nonsynonymous substitutions may have been eroded over time.

A signal of positive selection is also not significant when considering the putative human PBR (HLA PBR) as estimated by Brown et al. (1993) (within PBR: Z = 0.76, P = 0.223; outside PBR: Z = -2.57, P = 1.000). However, estimation of the guppy PBR using omegaMap software (Wilson and McVean 2006) revealed codons with significant posterior probability of positive selection

($\omega > 1$; i.e. $d_N > d_S$). Figure 3 highlights 12 of 73 codons (29, 31, 38, 39, 61, 62, 68, 72, 73, 78, 86 and 91) of exon 2 of the guppy MHC class II*B* identified as under significant effects of positive selection. Such effects are predicted for polymorphic codons of the PBR. Of the 12 identified codons, 7 (58%) are located within the region of the putative HLA PBR region (Brown et al. 1993; Stern et al. 1994) and of the remaining 5 codons 3 are immediately adjacent to polymorphic codons of the putative HLA PBR (Fig. 3). Overall, there was a significant signal of positive selection on codons of the PBR identified using omegaMap (within guppy PBR: $Z = 4.03$, $P < 0.001$). In contrast, codons outside the guppy PBR showed no sign of positive selection ($Z = -2.97$, $P = 1.000$). Moreover, these non-PBR codons were found to have a significant signal of purifying selection ($Z = 2.95$, $P = 0.002$).

PBR codons identified using OMEGAMAP were used to construct two amino acid neighbour joining trees of exon 2 of the MHC class II*B* (Fig 4). The first tree was made up of codons outside the PBR and the second of codons within the PBR. Firstly, these phylogenies represent functionally distinct clades. Much of the nucleotide polymorphism (Fig. 1) is preserved in both the outside PBR and inside the PBR trees (Fig. 4). However, some of the alleles are functionally equivalent at the PBR which suggests maintenance through selection acting at other regions of the MHC (see van Oosterhout 2009a). Tree topology is different for each region of exon 2. This discordance between the evolutionary histories of amino acids suggests gene conversion or recombination acting to reduce evolutionary distance in one region but not the other (Posada and Crandall 2002; von Salomé et al. 2007).

Table 1. MHC class II*B* alleles with 100% identity to those from previous studies are shown. Alleles may be matched with more than one other in cases where query coverage is not 100% due to variation in sequence length (see van Oosterhout et al. 2006b; Fraser et al. 2010b for alleles identified in previous studies)

Allele (present study)	Allele (previous study)	Accession No.
Pore 107	*Pore bb*	-
Pore 131		
Pore 119	*Pore h54*	DQ396613
	Pore h61	DQ396614
	Pore p	
Pore 129	*Pore hh*	-
Pore 126	*Pore DBLA1-18*	AY745517
	Pore aa	
Pore 124	*Pore h5*	AY770035
Pore 106	*Pore h23*	DQ396608
Pore 138	*Pore h6*	DQ396605
Pore 227	*Pore h38*	AY770040

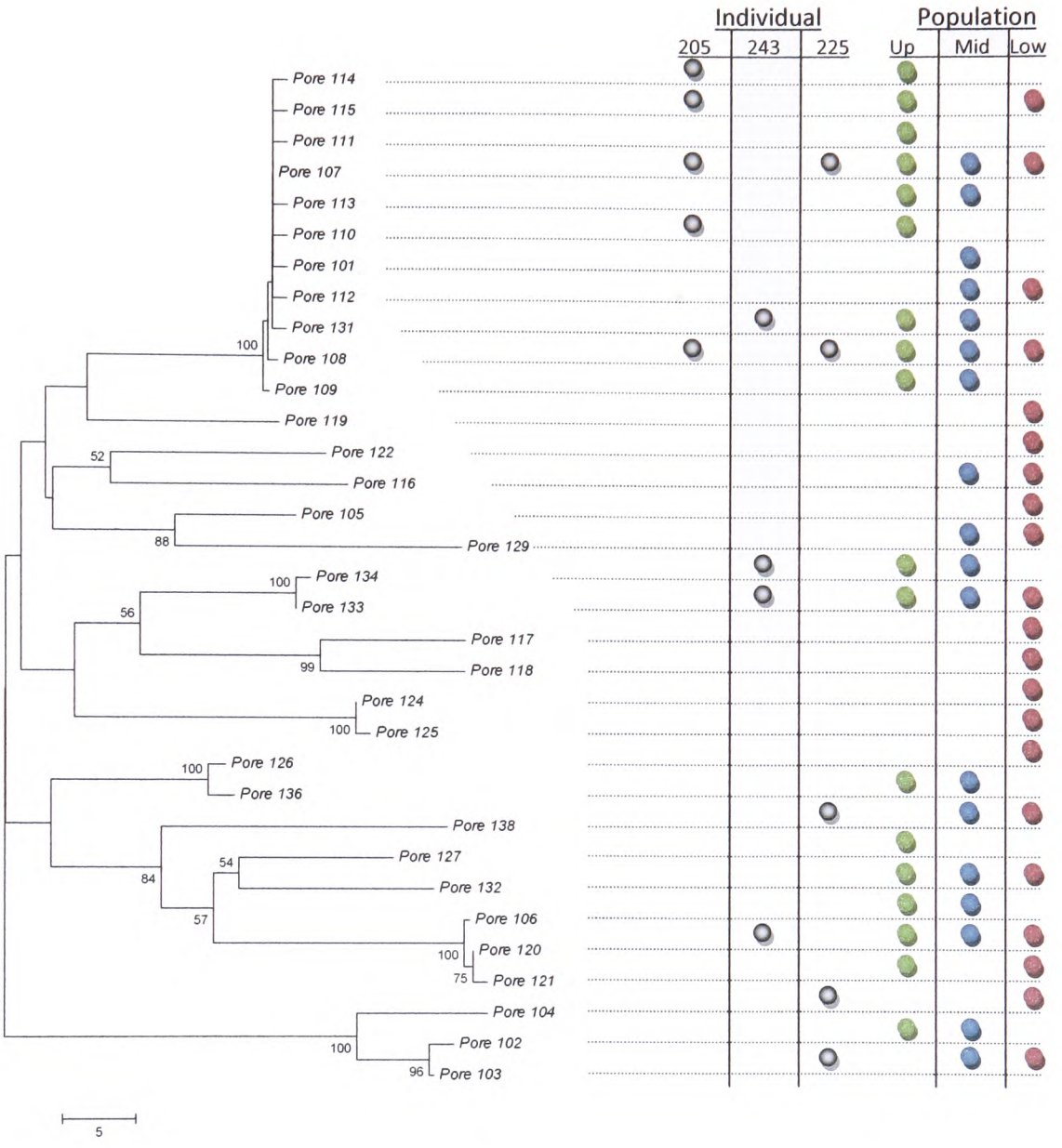


Figure 1. Neighbour-joining tree of MHC class II*B* alleles from guppies (Poecilia reticulata = *Pore*) of the Aripo River, Trinidad. The tree was computed using No of nucleotide differences with 1000 simulations. Bootstrap support values ≥ 50 are shown. Alleles from three individuals from the Up, Mid, and Low populations (205, 243, and 225, respectively) are shown to the right of the tree. Clades are not locus specific as demonstrated in individual 205 which has five alleles from the top clade. The presence of alleles in each population is also indicated to the right of the tree. The scale refers to base pair differences among alleles. Nomenclature does not include *DAB* locus affiliation because affiliation is unknown (see Table 1 for list of alleles identified in previous studies).

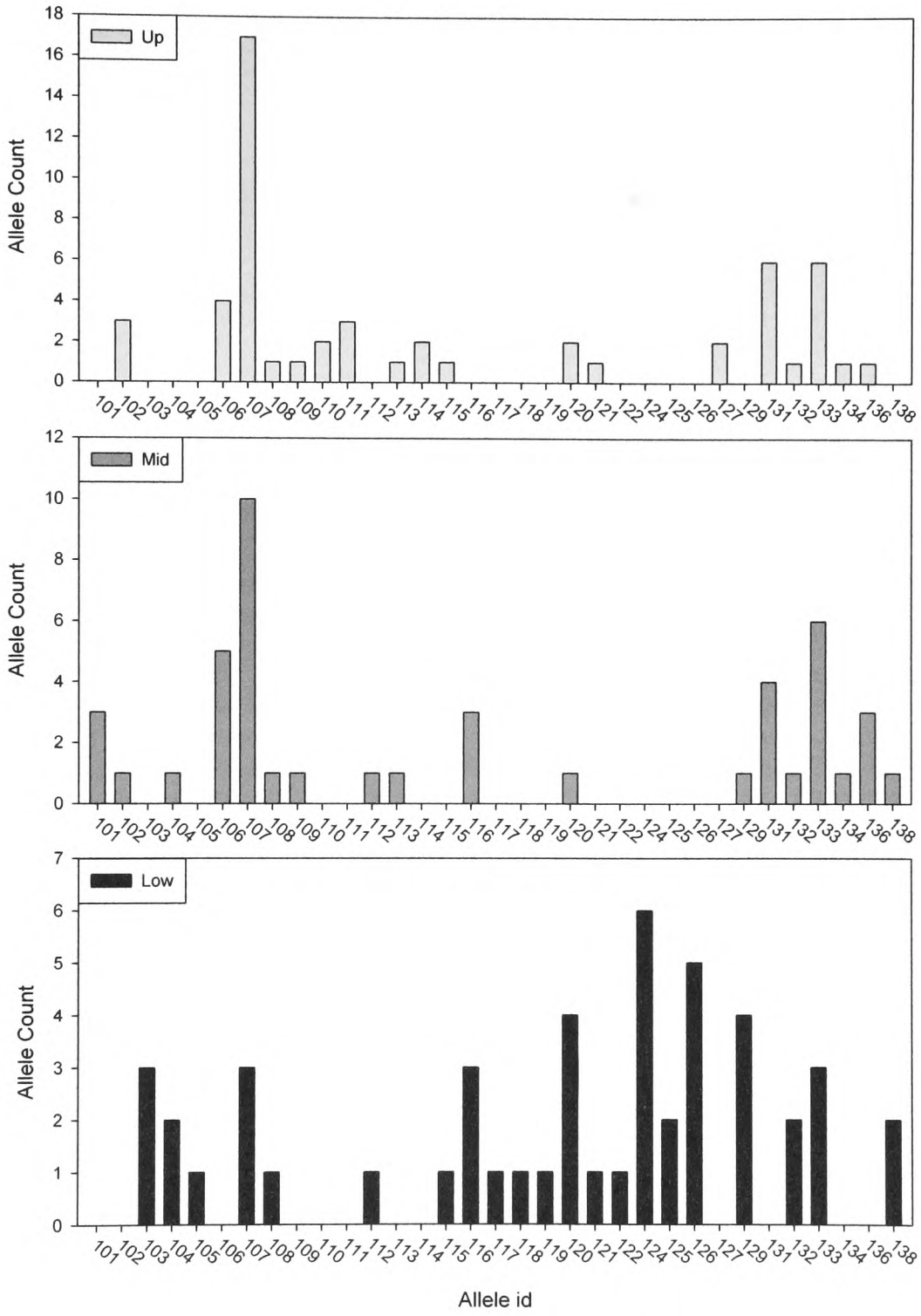


Figure 2. MHC class II*B* allele counts for Upper (n = 20) and Mid Naranjo (n = 20) and Lower Aripo (n = 19) guppy populations. Symmetry between the Up and Mid populations is not reflected in the comparison between the Mid and Low populations (G'_{ST} (5-95 CI): Up-Mid = 0.139 (0.135 - 0.386); Mid-Low = 0.605 (0.501 - 0.835) (see Fig. 8).

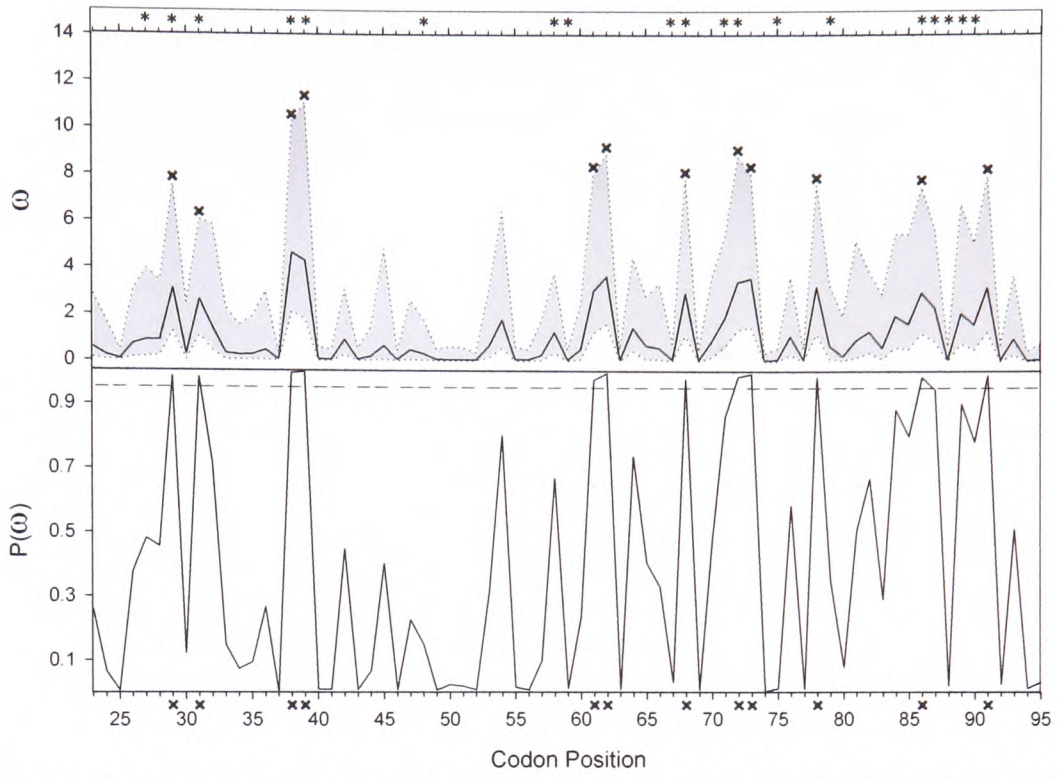


Figure 3. Analysis of selection on MHC class II*B* sequences of guppies (*Poecilia reticulata*) of the Aripo River, Trinidad. The upper plot shows the amount of positive selection (ω) at each codon (sliding window of 1 codon). Asterisks highlight the polymorphic sites of the HLA DRB1 PBR region (see Brown et al. 1993), shifted up by one codon to account for an insertion near the amino terminus of the fish MHC class II*B* molecule (Figueroa et al. 2000). Crosses highlight codons of the guppy MHC with a significant posterior probability (> 0.95) of positive selection. Posterior probability of positive selection features in the lower plot, where the dashed (short-short) line indicates the 0.95 acceptance criteria for positive selection. Crosses indicating significant probability of positive selection are repeated in the second plot at codons in the x-axis. Estimates are based on two combined runs of OMEGAMAP software using standard priors (see materials and methods). The Markov chain Monte Carlo was run for 600,000 iterations in each run, with a burn-in of 100,000 iterations.

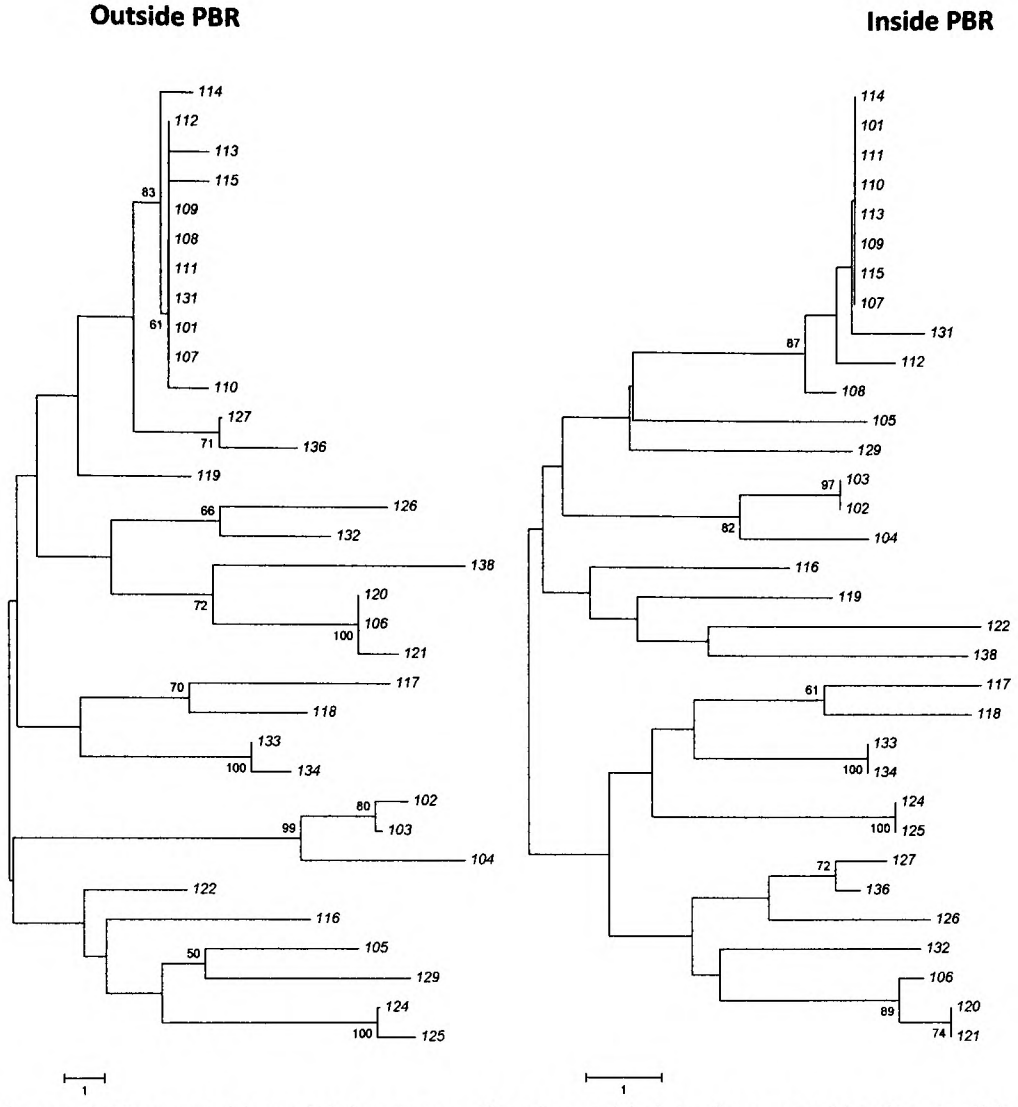


Figure 4. Two neighbour-joining trees of amino acid data from outside the peptide binding region (PBR) and inside the PBR of the MHC class II*B* alleles from guppies (*Poecilia reticulata*) of the Aripo River, Trinidad. Trees were computed using number of amino acid differences with 1000 simulations. Bootstrap support ≥ 50 is shown. Tree topology is different for different regions of exon 2 suggesting separate evolutionary histories of regions inside and outside the PBR.

Parasite analysis

The prevalence of a particular group of parasites is taken as evidence for selection on the MHC acting within a population. Where a group of parasites is found in all populations, parasite burden is used as a proxy for the strength of selection acting on the individuals of that population (see e.g. van Oosterhout et al. 2007a). Of the 201 guppies sampled in the Up population in 2003 (wet and dry season), 2006 and 2007 there were no digenean or fungal infections (Fig. 5). Only in 2007 were gyrodactylids found in the Up population where six of 153 (3.9%) individuals were infected, compared to 30 of 138 (21.7%) in the Mid and 48 of 125 (38.4%) in the Low the same year (Fig. 5 and 6). There was a significant difference in gyrodactylid burden between Up, Mid and Low population in 2007 (Kruskal-Wallis test (adjusted for ties): $H = 51.52$, $d.f. = 2$, $P < 0.001$) and the Up population had a significantly lower gyrodactylid burden than the Mid population (Mann-Whitney test (adjusted for ties): $P < 0.001$) (Fig. 6). In addition, the Up 2007 population had significantly a lower gyrodactylid

burden and prevalence compared with all other samples (Mann-Whitney test (adjusted for ties): burden $P < 0.036$, all pairwise comparisons; Chi-square tests for prevalence: $\chi^2 \geq 21.248$, *d.f.* = 1, $P < 0.001$, all pairwise comparisons). The year 2007 was the first recorded instance of any gyrodactylids in the Up population, but anecdotal evidence suggests that these gyrodactylids may be a new genus introduced by a guppy predator, *Rivulus hartii* (Dr Cable, pers. Com.).

Gyrodactylid burden varied significantly between populations over all years (Mann-Whitney test (adjusted for ties): $P < 0.001$). It did not vary significantly in the Mid population over time (Kruskal-Wallis test (adjusted for ties): $H = 6.72$, *d.f.* = 3, $P = 0.081$). However, there was a significant variation in gyrodactylid burden in the temporal samples of the Low population (Kruskal-Wallis test (adjusted for ties): $H = 67.13$, *d.f.* = 3, $P < 0.001$) (Fig. 6). Digenean load was significantly different between populations sampled over all years (Mann-Whitney test (adjusted for ties): $P < 0.001$). Variation in digenean burden is significant over time within the Mid but not the Low (Kruskal-Wallis test (adjusted for ties): Mid, $H = 11.41$, *d.f.* = 2, $P = 0.003$; Low, $H = 2.00$, *d.f.* = 2, $P = 0.368$).

There was no significant difference between the Mid and Low populations sampled for fungi over all years (Mann-Whitney test (adjusted for ties): $P < 0.140$). Variation in fungal burden is significant over time within the Mid and Low Kruskal-Wallis test (adjusted for ties): Mid, $H > 117.08$, *d.f.* = 2, $P < 0.001$ for both tests).

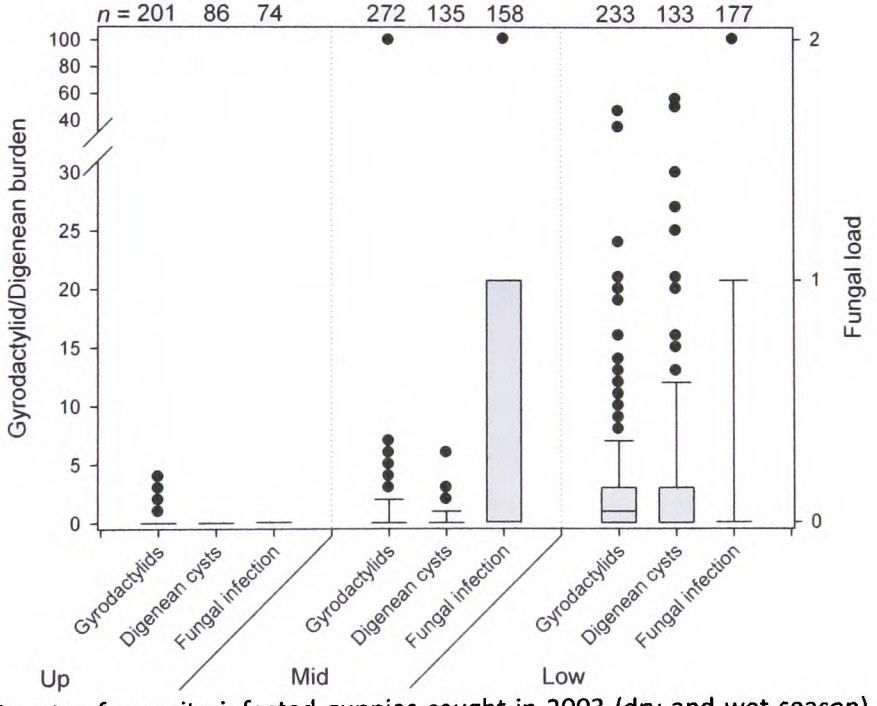


Figure 5. Counts of parasite infected guppies caught in 2003 (dry and wet season), 2006 and 2007 from the Upper Naranjo, Mid Naranjo and Lower Aripo populations (Up, Mid and Low). Parasite infections (burden) include numbers of Gyrodactylid, (internal) Digenean metacercariae and fungal infected fish (see *n* = number of fish screened). Fungal load is scored as 0 (no fungus), 1 (slight fungal growth) and 2 (moribund fungal growth), and is measured on the secondary y-axis. Although parasite burden shows significant temporal variation, the data is grouped (across years) into population to illustrate that, the Up population has significantly fewer parasites (see Fig. 6) than the Mid and Low populations.

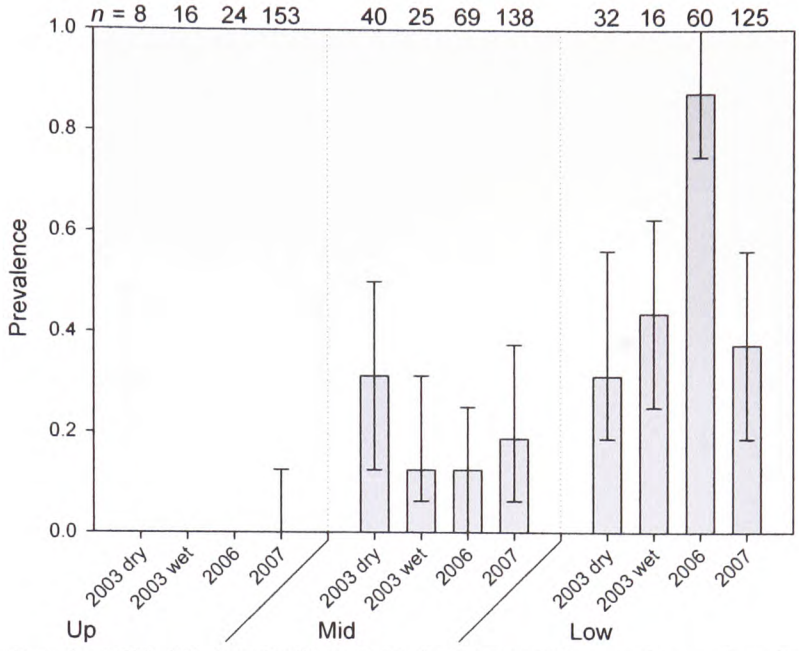


Figure 6. Gyrodactylid (*G. turnbulli* and *G. bullatarudis*) prevalence (median, 5-95% CI) on guppies from the Upper Naranjo, Mid Naranjo and Lower Aripo (Up, Mid and Low) sampled over three years (2003 dry and wet season, 2006 dry season and 2007 dry season, *n* = number of fish screened, on the top line of the histogram).

Population genetic analysis

There is no difference in MHC allelic richness (*A*) between the Up, Mid or Low populations (Randomisation test: P > 0.098 for all tests) (see Fig. 7). However, there is a large and significant difference in microsatellite allelic richness between the Up, Mid and Low populations (Randomisation test: P < 0.001 for all tests). This reduction in microsatellite allelic richness with upstream locality is indicative of a bottleneck possibly because of a reduction in population size.

Table 2. Allelic richness (*A*, median (5-95% CI)) at MHC and microsatellite loci among populations of guppies of the Upper Naranjo, Mid Naranjo and Lower Aripo Rivers. Allelic richness differs significantly between populations for microsatellite loci, while in contrast, the allelic richness of the MHC is similar in all three populations.

	MHC	Microsatellites
		*
Up	15 (11-17)	34 (31-36)
Mid	14 (11-17)	55 (52-58)
Low	18 (13-21)	105 (97-112)

* Values are significant at P < 0.001.

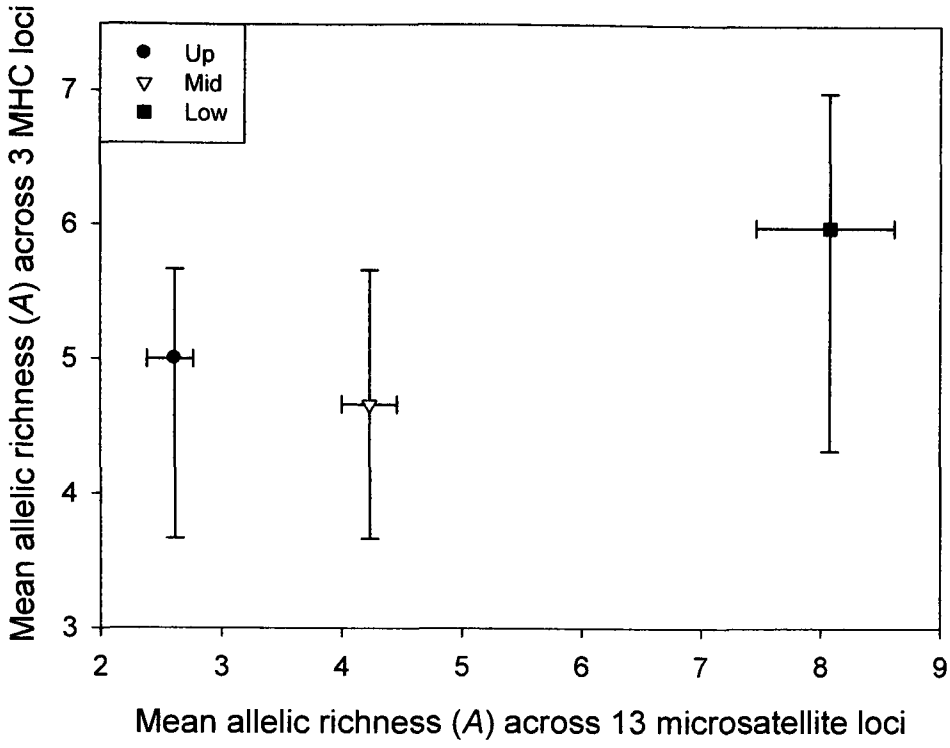


Figure 7. Mean allelic richness (*A*, 5-95% CI) of microsatellite and MHC loci of guppies from the Upper Naranjo, Mid Naranjo and Lower Aripo guppy populations (see also Table 2 for total *A*). The mean *A* for the MHC was calculated based on an estimated 3 loci. Populations differ significantly for their microsatellite variation (Randomisation test: $P < 0.001$ for all tests), but not for the variation at their MHC (Randomisation test: $P > 0.098$ for all tests).

Genetic differentiation (mean G'_{ST}, 5-95% CI) among populations is significantly higher for MHC loci than neutral microsatellite loci (Randomisation test: $P < 0.001$; see Table 3 and Fig. 8). This finding is inconsistent with the signal expected from NFDS and overdominance (Schierup 1998; Schierup et al. 2000; Muirhead 2001), but rather, supports the hypothesis of fluctuating selection, with spatial variation in the direction of selection pressures (see Hedrick et al. 1976; Hedrick 2002).

MHC genetic differentiation is significantly higher between the Mid and Low and between the Up and Low, than between the Up and Mid (randomisation tests: $P < 0.001$; see Table 3 and Fig. 8). For the MHC, the Mid population is genetically as distinct from the Low population, as is the Up population from the Low population (Randomisation test: $P = 0.177$; see Table 3). Hence, genetic differentiation between populations increases for comparisons over the Haskins waterfall.

Genetic differentiation at microsatellite loci is significantly smaller than for the MHC, albeit still significantly larger than zero (see Table 3 and Fig. 8). In addition, the level of genetic differentiation at microsatellites differs significantly between all pairwise comparisons (Randomisation tests: $P < 0.001$ for all tests; see Table 3 and Fig. 8). Hence, microsatellite genetic differentiation increases for comparisons over greater distance.

Table 3. Genetic differentiation (G'_{ST}, median (5-95% CI) at MHC and microsatellite loci among populations of guppies of the Upper and Mid Naranjo and Aripo Rivers, Trinidad.

	MHC		Microsatellites	
	*			
Up-Mid	0.211 (0.120-0.370)	*	0.069 (0.027-0.134)	*
Mid-Low	0.690 (0.492-0.850)	*	0.238 (0.170-0.320)	*
Up-Low	0.733 (0.531-0.904)	*	0.328 (0.253-0.407)	*

* Values are significant at $P \leq 0.001$.

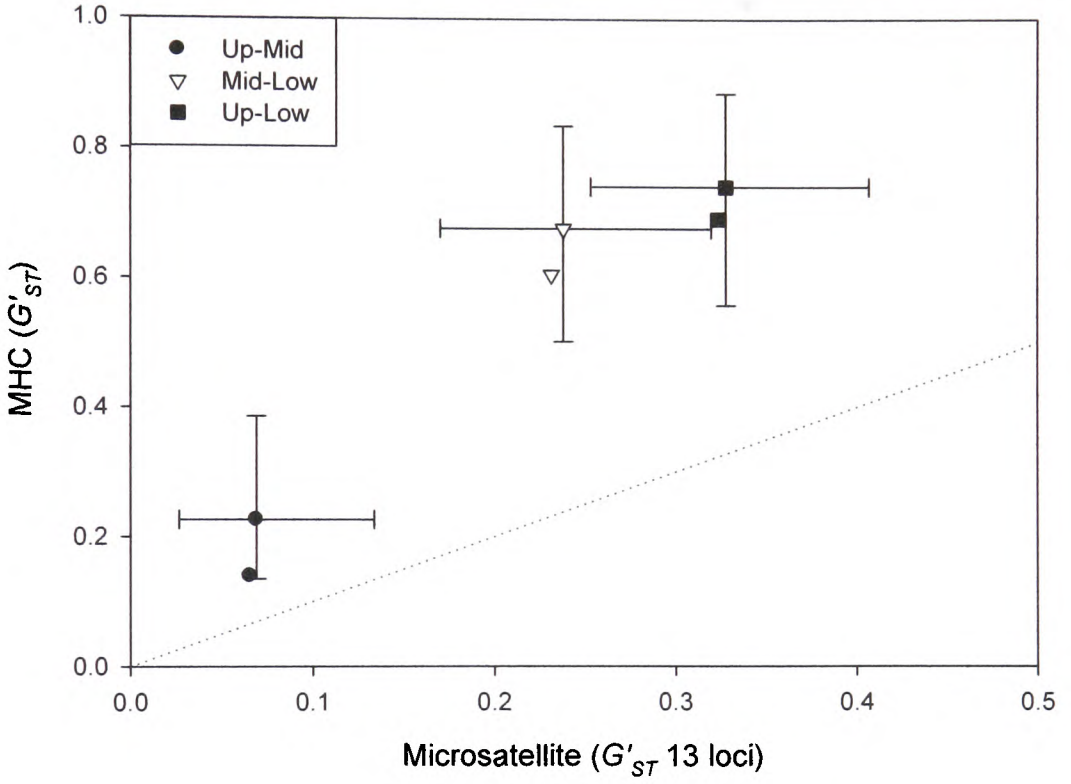


Figure 8. Genetic differentiation (mean G'_{ST} (5-95% CI)) at 13 microsatellite loci and at MHC loci among pairwise population comparisons between the Upper Naranjo (Up), Mid Naranjo (Mid) and Lower Aripo (Low). The dotted line (G'_{ST} microsatellites = G'_{ST} MHC) show the values where the neutral microsatellite and selected MHC loci are equally diverged between populations. Observed values are shown with respective symbols. Genetic divergence is greater between populations for MHC than for microsatellite loci (see also Table 2). Such a finding is inconsistent with predicted higher effective migration rate of MHC alleles (Schierup 1998; Schierup et al. 2000; Muirhead 2001), which predicts that the observed values should fall below the dotted line. This finding is, however, consistent with divergent selection pressures, for example by locally distinct parasite faunas (see Hedrick et al. 1976; Hedrick 2002).

Discussion

The present study compares the genetic variation at (neutral) microsatellite loci to that of (non-neutral) MHC loci in three guppy (*Poecilia reticulata*) populations of the Aripo River in Trinidad. This genetic data is also used to calculate population divergence at a spatial scale. In addition, a measure of parasite fauna is used to ascertain the source of selection on the MHC. The Upper Naranjo (Up) population has significantly reduced parasite fauna compared to the Mid Naranjo (Mid) and Lower Aripo (Low) populations. However, MHC variation in the Up population is not reduced according to the neutral expectation indicated by variation at microsatellite loci. In addition, genetic differentiation between populations is higher for MHC than for microsatellite loci, a finding opposite to the expectation of Negative frequency dependent selection and Overdominance (Schierup 1998; Schierup et al. 2000; Muirhead 2001). Altogether, these data suggest MHC variation is maintained by fluctuating selection pressure, though maintenance of MHC variation in the Up suggests that parasite mediated selection is not the source.

Positive selection in the MHC

There was no overall signal of positive selection on the exon 2 sequences of the MHC class II*B* alleles. Over the whole of exon 2, the number of synonymous substitutions exceeds that of the nonsynonymous mutations, reflecting that purifying selection dominates the signal of positive selection (Hughes and Yeager 1998). This finding is consistent with the conservation of a conserved and functional protein, i.e. the MHC. However, long term positive selection was identified in 12 out of the 73 codons of exon 2. This number is lower than previous estimates of the putative PBR for the guppy (van Oosterhout et al. 2006a; Fraser et al. 2010a). The present study employed OMEGAMAP software which estimates positive selection in the presence of recombination which can otherwise overinflate estimates of ω (Anisimova et al. 2003; Nei 2005; Wilson and McVean 2006). This may explain the lower estimate of the present study. The majority (83%) of codons identified using the method of Wilson and McVean (2006) corresponded, or were immediately adjacent to polymorphic sites within the HLA (Brown et al. 1993). Altogether, evidence from the nucleotide substitution pattern of exon 2 indicates that this region is, or has been (see Garrigan and Hedrick 2003), subject to balancing selection, specifically at the region responsible for recognition of parasites antigens.

High MHC variation in a small population with low parasite load

Genetic diversity and allelic richness at 13 microsatellite loci was significantly higher in the Lower Aripo (Low) than in the upstream Mid Naranjo (Mid) population, while the most upstream Upper Naranjo (Up) population showed the lowest level of microsatellite variation. This reduction of neutral genetic variation with upstream locality is consistent with strong genetic drift in small upstream populations with limited upstream migration from the lowlands (Shaw et al. 1991; Shaw et al. 1994, see Chapter 2). In the lowlands, the microsatellite genetic diversity is comparatively high due to their larger effective population sizes, downstream-biased migration, and source-sink metapopulation structure (Barson et al. 2009). In contrast to microsatellite variation, MHC class II*B* variation was not significantly reduced in populations further upstream.

Discordance between MHC and microsatellite variation is consistent with maintenance of MHC variation though balancing selection. Parasite-mediated balancing selection is commonly

proposed as the mechanism for the maintenance of MHC polymorphism despite the effects of drift in small populations. A number of studies have suggested parasites as a source of maintenance of MHC polymorphism (see reviews Apanius et al. 1997; Hughes and Yeager 1998; Bernatchez and Landry 2003; Piertney and Oliver 2006), including three studies on guppies in the Aripo River (van Oosterhout et al. 2006b; Fraser and Neff 2010; Fraser et al. 2010b). These latter studies implicate gyrodactylid ectoparasites, showing that these parasites significantly reduce the survival of guppies in the Mid Naranjo River (Mid) in Trinidad (Fraser and Neff 2010; Fraser et al. 2010b, see also van Oosterhout et al. 2007). In the present study, we observe maintenance of MHC polymorphism in the Up population which has no microparasites or fungal infections in 2003 and 2006 and, in 2007 has no internal digenean parasites or fungal infections. In addition, the introduction of gyrodactylid parasites in 2007 was marginal, with only six out of 153 individuals being infected, an infection prevalence that is significantly lower than in the Mid and Low in any previous year. Furthermore, the individual gyrodactylid burden in the Up population was significantly reduced compared to populations inhabiting downstream localities. The selection pressure due to parasites thus seems relaxed in the Up population as compared to Mid and Low populations whereas the MHC diversity is similar in the three populations. I conclude that parasite mediated selection might not be the only evolutionary force responsible for maintenance of MHC polymorphism in these guppy populations.

Migration is a well-known evolutionary process that could maintain allelic diversity. However, migration alone cannot account for the maintenance of MHC diversity in the Up population, given that the level of neutral microsatellite variation differs markedly between the Up and Mid population.

Maintenance of MHC polymorphism through sexual selection via mate choice has been shown in a number of fish species (Landry et al. 2001; Milinski et al. 2005; Neff et al. 2008) and might play a role in the variation of MHC observed in the guppy populations studied here. Identification of MHC compatible males is shown in many cases to be mediated through olfaction (reviewed by Penn and Potts 1999; Milinski 2006). While visual mediated sexual selection is a well established sensory modality for mate choice in guppies (reviewed by Houde 1997; Magurran 2005), olfactory sexual signalling appears secondary to the visual sense (but see Lopez 1998 for evidence of a visual signalling system; Shohet and Watt 2004; Archard et al. 2008). In addition sneaky mating often denies the choice of a female meaning that maintenance of MHC variation through sexual selection would have to be post-copulatory (Birkhead and Pizzari 2002; and see e.g. in Arctic charr Skarstein et al. 2005).

The MHC variation observed here could also be influenced by Associative Balancing Complex (ABC) evolution, which is a mechanism that has been theoretically shown to retain polymorphism in the MHC by reducing the fitness of MHC homozygous individuals (van Oosterhout 2009a). This model proposes that recessive deleterious mutations have built up as a "sheltered genetic load" in the MHC and in the extended genomic region. This sheltered load is expressed in homozygous individuals, reducing their fitness and maintaining polymorphism through associative overdominance (van Oosterhout 2009a). According to ABC evolution, the fitness of rare MHC alleles/haplotypes is relatively high, because they are less often found in a homozygous state. This results in a type of NFDS that can maintain polymorphism over long evolutionary timescales, potentially resulting in trans species polymorphisms. This type of balancing selection in ABC evolution can act continuously, both in the presence and absence of

parasite selection (van Oosterhout 2009a). Evidence for the maintenance of MHC alleles by mechanisms proposed by ABC evolution comes from the observation that some of the alleles are functionally equivalent at the PBR region. I propose that balancing selection during ABC evolution could maintain a high level of MHC diversity in the relaxed parasite-environment of the Up population.

Genetic differentiation at the MHC

Balancing selection through NFDS or overdominance acts to reduce genetic differentiation between populations (Schierup 1998; Schierup et al. 2000; Muirhead 2001). Genetic differentiation of the three populations in the present study is greater in the MHC than microsatellites. This is contrary to the expectation of MHC under NFDS and overdominant forms of balancing selection and, it is in disparity to previous observations on the MHC of guppies (van Oosterhout et al. 2006b; Fraser et al. 2010a; but see Fraser et al. 2010b). However, there are number of studies that have found a higher genetic differentiation in the MHC compared to neutral genetic loci (Hedrick and Parker 1998; Landry and Bernatchez 2001; Miller et al. 2001; Ekblom et al. 2007; Fraser et al. 2010b). The observations in the latter studies are consistent with selection varying in time and space (see Hedrick et al. 1976; Hedrick 2002).

Maintenance of MHC polymorphism in the Up population was observed despite the effects of drift and a reduced parasite fauna. ABC evolution has been theoretically shown to be able to maintain MHC polymorphism without continual parasite selection (van Oosterhout 2009a). Genetic divergence between populations was higher for MHC than microsatellite loci. This finding is inconsistent with NFDS and overdominant selection that act to reduce population genetic divergence at loci under balancing selection (Schierup et al. 2000; Muirhead 2001). ABC evolution however, may produce higher or lower genetic divergence of MHC (measured at exon 2) to neutral loci (see van Oosterhout 2009a). ABC evolution is again consistent with these observations but it is not possible to conclude it is operating without testing for sheltered genetic load directly.

Parasite fauna

Interestingly sharp differences in parasite fauna (*Gyrodactylus* spp) have been observed between the Mid Naranjo and Lower Aripo populations (van Oosterhout et al. 2006b). van Oosterhout et al. (2006b) quantified the relative abundances of gyrodactylid species (*Gyrodactylus turnbulli* and *G. bullatarudis*) and found significant variation between the Mid and Low populations. Such variation would not be evident in the present study where all gyrodactylid parasites were counted as one group. However, the species variation is likely to be an important factor in the selection of MHC alleles particularly since the authors also note that the virulence of *G. bullatarudis* is significantly higher than that of *G. turnbulli* (van Oosterhout et al. 2006b; see also Cable and van Oosterhout 2007a). In the present study, total gyrodactylid and digenean infection levels varied significantly between the Mid and Low populations (combined over all years), an indication that differences between these populations may be driving population specific selection and local adaptation of the MHC.

A temporal variation in parasite fauna was also observed in the Mid (for digeneans and fungi) and Low (for gyrodactylids and fungi) populations. These data demonstrate the dynamic nature of parasite infections of guppies in the Aripo River and Naranjo Tributary. Temporal variation in parasite fauna is predicted to drive adaptations of the MHC resulting in a

temporally dynamic MHC polymorphism. A change in the mode of selection from homogenising to diversifying was been observed in guppy populations of Trinidad by Fraser et al. (2010b). The authors also quantified gyrodactylid infections but as in the present study, they made no distinction between species. Notwithstanding, the authors did find a correlation between the change (over time) in the absolute number of infected individuals and temporal genetic differentiation of the MHC. This finding could be interpreted as evidence that MHC alleles vary over time with changes in gyrodactylid infection intensity. However, the authors do not establish a causal link between the change in parasite-fauna and MHC diversity. Indeed, a study presented in this thesis shows that a large turn-over of MHC alleles may also be caused by demographic events (see Chapter 5). The inflow of immigrants with novel MHC alleles and parasites may result in an autocorrelation between parasite and MHC diversity, which may be interpreted as evidence for Red Queen Dynamics.

Quantifying the variation in parasite fauna between populations in order to establish links with selection at the MHC is extremely challenging for two reasons. First, visible parasites are only a part of the total array of pathogens that challenge individual hosts. Bacterial species diversity has not been considered here, but even so, many bacterial species are opportunistic pathogens (see e.g. Tenaillon et al. 2010), which implies that noting only the presence or absence of such potential pathogens is of limited use. Secondly, quantifying species diversity of potential pathogens may not have sufficient resolution as within-species diversity could lead to balancing selection at the MHC (but see Dionne et al. 2007). Possibly, in the nearby future, high throughput genotyping analysis of the entire parasite community could be employed to study the Red Queen Dynamics and the coevolutionary arms race between a parasite fauna and the host MHC.

Concluding remarks

The present study draws attention to a population that maintains high MHC diversity despite little or no selection by microparasites and fungi on the MHC. Population genetic analysis of neutral microsatellite variation rules out the possibility that migration can explain the high level of MHC diversity in this population. This finding is consistent with maintenance of MHC polymorphism through sexual selection or ABC evolution. Altogether, these findings highlight that a novel avenue in MHC research should be undertaken that investigates the role of alternative agents of selection. Both ABC evolution and sexual selection are prime candidates in future research on the MHC in wild populations.

Genetic differentiation between populations was higher for MHC than for neutral microsatellite loci. This finding may indicate maintenance of MHC polymorphism through selection that varies in time and space. In addition, this finding is at odds with a previous finding in the Mid and Low populations from 2001 (van Oosterhout et al. 2006b). In the next chapter, I analyse MHC variation in the Mid and Low populations sampled in 2001 and 2007 in order to distinguish the mode of selection acting on these guppy populations of the Aripo River, Trinidad.

Spatio-temporal analysis of MHC and microsatellite variation in Aripo River guppy populations

Chapter 4

Abstract

Major Histocompatibility Complex (MHC) variation in two guppy (*Poecilia reticulata*) populations of the Aripo River (Trinidad) is believed to be maintained by balancing selection. However, genetic differentiation between populations is lower than neutral microsatellite variation in a sample from 2001 and higher in a 2007 sample. These data are indicative of maintenance of MHC alleles through alternative modes of balancing selection. The present study aims to distinguish between the different models of balancing selection, in particular overdominance and negative frequency dependent selection versus selection that varies in time and/or space (hereafter called fluctuating selection). These models are evaluated by comparing the observed MHC and microsatellite data with the population genetic predictions of the alternative models. The relatively high temporal variation in MHC compared to microsatellites suggests that the MHC has a higher effective migration rate which effectively replaces MHC alleles over time (allelic turnover). This observation is consistent with fluctuating selection, i.e. changes in parasite selection pressures over time. However, it is also concordant with the models that favour establishment of rare alleles into the population (i.e. overdominance and negative frequency dependent selection) in a metapopulation with migration. The relatively low level of spatial genetic differentiation at the MHC observed in 2001 is consistent with overdominance and negative frequency dependent selection, but not with fluctuating selection. By contrast, large spatial genetic differentiation of MHC compared to that of neutral loci (in 2007) suggests a role for fluctuating selection. Indeed, the 2007 data are inconsistent with overdominance and negative frequency dependent selection. Altogether, these data suggest that the alternative models of balancing selection are not mutually exclusive, and that they may operate at different times in the same populations. This study furthermore highlights that the impact of migration can differ substantially between neutral and selected genes, and that low levels of migration can result in a temporally highly dynamic MHC polymorphism.

Introduction

Balancing selection can explain the maintenance of genetic variation at a locus above that expected under a neutral model of evolution (see Kimura et al. 1963; Ford 2002; Garrigan and Hedrick 2003; Charlesworth 2009). Theoretical studies have demonstrated that both negative frequency dependent selection (NFDS) and overdominance tend to reduce spatial genetic differentiation (Schierup 1998; Schierup et al. 2000; Muirhead 2001; see also review by Spurgin and Richardson 2010). Selection that fluctuates over a spatio-temporal scale (i.e. fluctuating selection) has also been shown to maintain a balanced polymorphism in theoretical studies (e.g. Hedrick 2002). This mechanism assumes MHC alleles confer resistance to different parasites which vary spatially and/or temporally (Hedrick 2002). The crucial difference between fluctuating selection and both other types of balancing selection is that the former is not a coevolutionary model, in that parasites fluctuate independently from the MHC.

NFDS and overdominance are expected to reduce genetic differentiation over time by maintaining alleles in the population (see Takahata and Nei 1990). Rare alleles are thought to offer a higher fitness in both the NFDS, as well as in the overdominance model because such rare alleles are found more often in a heterozygote condition (Apanius et al. 1997; see also Meyer and Thomson 2001). Therefore, both NFDS and overdominance exert positive selection on rare alleles, thereby reducing the rate of loss of such rare alleles through genetic drift.

On the other hand, fluctuating selection predicts high temporal genetic differentiation because the parasite fauna changes over time. Allelic variation at the MHC can be maintained by spatial or temporal fluctuations in pathogen fauna because distinct alleles are preserved in separate populations at different points in time (Hedrick et al. 1976; Hedrick 2002). In this fluctuation model of balancing selection, a balanced polymorphism is maintained over longer timescales by favouring alternative alleles in different places and/or times (fluctuating selection). These changes in directional selection result in large temporal genetic differentiation in MHC relative to neutral loci.

The distinction between fluctuating selection on the one hand, versus NFDS and overdominance on the other, might become less distinct in a metapopulation where novel MHC alleles are introduced by immigrants. Theoretically, the introduction of novel MHC alleles may cause turnover of such alleles not only under fluctuating selection, but also under NFDS and overdominant selection (see Chapter 5). Given that novel immigrant alleles are favoured because they are rare, such alleles will rapidly reach the (migration-selection-drift) equilibrium frequency. From this point onwards, immigrant alleles and resident alleles have an equal probability to be lost by random genetic drift. Consequently, immigrant alleles that become established in the population ultimately will replace resident alleles. If the gene pool of the source population is large, this replacement through immigrants effectively results in the fast turnover of alleles. That is, established resident alleles will be replaced with novel immigrant alleles from elsewhere in the metapopulation. This process of allelic turnover is accelerated by all models of balancing selection (compared to neutral evolution), rather than slowed down. This means that increased temporal genetic differentiation can be explained under all models of balancing selection in a metapopulation with a large gene pool of MHC alleles.

Barson et al. (2009) assessed population size and migration rates between eight populations of the guppy metapopulation in the Caroni drainage. The authors found that while migration was biased in the downstream direction, upstream migration was significantly larger than zero, and

that admixture ranged from 5-21% in the lowlands (Barson et al. 2009). Furthermore, Fraser & Neff (2010) found a large number of novel MHC alleles in the Caroni drainage, and by combining all datasets, I estimated that there are approximately 85 MHC class II*B* alleles in the entire metapopulation. This was done using the Michaelis-Menten equation of enzyme kinetics, originally devised to describe the rate of an irreversible enzymatic reaction (see Appendix 6). Given the metapopulation structure of this system and the large number of MHC alleles, novel MHC alleles are likely to be rapidly introduced by NFDS, overdominance and/or fluctuating selection. Hence, all models of balancing selection are expected to increase the temporal genetic divergence of MHC relative to that of neutral microsatellite loci. The significance of comparing temporal genetic differentiation of the MHC to that of microsatellites is to show that balancing selection acts on the MHC. Furthermore, it establishes the fact that the impact of migration differs substantially between neutral and selected genes in a metapopulation.

In contrast to temporal genetic differentiation, the expectations of spatial genetic divergence are more distinct for the alternative models of balancing selection (see Table 1 Spurgin and Richardson 2010). Under models that favour rare alleles (NFDS and overdominance), the effective migration rate is increased and population differentiation is thus reduced compared to microsatellite loci (Schierup 1998; Schierup et al. 2000; Muirhead 2001). On the other hand, fluctuating selection is predicted to increase genetic differentiation of the MHC on a spatial scale (Hedrick 2002, see also reviews by Meyer Thomson 2001; Spurgin and Richardson 2010).

The limited amount of empirical data on spatio-temporal MHC variation gives us some insights into the processes that govern the evolution of these genes (see e.g. Wegner 2004; Westerdahl et al. 2004; Charbonnel and Pemberton 2005; Hansen et al. 2007; Oliver et al. 2009a). Higher than expected fluctuations of MHC allele frequencies have been shown in great reed warblers (Westerdahl et al. 2004). The authors suggest these fluctuations cannot be explained by demographic stochasticity alone, but that they are due to changes in direction and intensity of selection caused by temporal fluctuations in pathogen fauna. Further evidence for fluctuating selection pressure is shown in a 13 year study of a free living island population of Soay sheep (Charbonnel and Pemberton 2005). This study measured separate populations (i.e. hefts) of sheep and demonstrated both higher and lower levels of genetic differentiation of MHC compared to neutral loci. A spatial difference in selection pressure between sheep hefts was suggested to explain the instances where the MHC was more differentiated than neutral genes. The authors suggest that selection for the same alleles in different hefts reduces MHC genetic differentiation compared to neutral loci. The sparse spatio-temporal empirical MHC data thus appears to support fluctuating selection acting on both spatial and temporal scales (but see Wegner 2004; Oliver et al. 2009a).

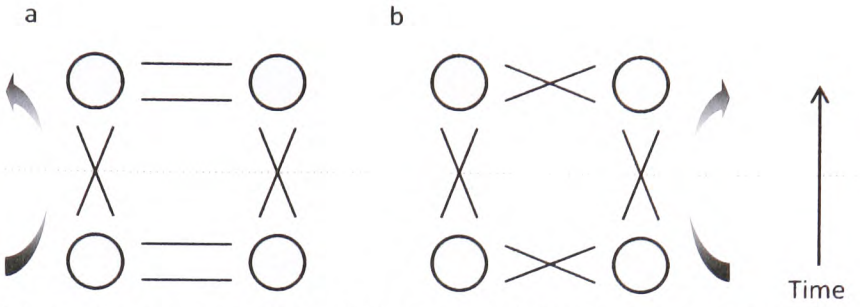


Figure 1. Schematic representation of the expected level of genetic differentiation (G'_{ST}) spatially and temporally, at MHC loci relative to that of neutral genetic differentiation. Circles represent two populations separated by space and time, lines represent the transfer of MHC alleles between populations (parallel = low MHC G'_{ST}; crossed = high MHC G'_{ST}). Over time, migrants may also enter these populations from outside the system (metapopulation = shaded arrows). NFDS and overdominance rescue rare immigrant alleles which may be from the other spatial population, reducing spatial MHC genetic differentiation or the metapopulation increasing temporal genetic differentiation at the MHC compared to neutral loci (a). Fluctuating selection favours different MHC alleles in different populations at different times thus increase MHC genetic divergence relative to neutral loci under all comparisons (b).

In this study, I aim to explore the model of balancing selection acting on MHC in the guppy (*Poecilia reticulata*) in 2001 and 2007 by investigating the spatial and temporal variation of MHC alleles. I examine the effects of migration and balancing selection on two populations of guppies and compare genetic variation at both adaptive (MHC) and neutral microsatellite loci between the Mid Naranjo (Mid) and Lower Aripo (Low) River populations of the Caroni drainage. I study the level of genetic differentiation at both a spatial (G'_{ST}) and temporal ($G'_{ST\,(temp)}$) scale. I investigate the hypothesis that the temporal fluctuations in MHC class II*B* allele frequency are more pronounced than fluctuations of neutral microsatellite alleles because of the interplay between balancing selection, migration and random genetic drift (Fig. 1). In particular, I examine whether the temporal changes in MHC allele frequencies are more pronounced than those of microsatellite alleles. This would be consistent with a high effective migration rate fuelled by balancing selection from a metapopulation with a large gene pool of MHC alleles. I also test whether the level of genetic differentiation between spatial populations is significantly different for the MHC than for the microsatellite loci, and I use this comparison to evaluate which model of balancing selection best explains the data.

Materials and methods

Guppy collection and DNA extraction

Two samples of wild caught guppies, collected in October 2001 and June 2007, were analysed for MHC and neutral variation at six microsatellite loci in order to measure the temporal effects of random genetic drift and selection on the MHC. Guppies were captured using a seine net, given a lethal dose of MS222 and stored separately in molecular grade ethanol. In order to minimise allele frequency biases associated with sampling families or cryptic social structure, each site was sampled at multiple locations in order to maximise the chance of catching guppies from separate shoals. Guppies are from two populations in the Aripo River, which is part of the Caroni Drainage in the Northern Mountain Range of Trinidad (grid reference: Mid Naranjo [Mid], PS693100E and 1181800N; Lower Aripo [Low], PS6914000E and 1177700N).

Molecular analyses of the 2001 sample were reported by van Oosterhout et al. (2006b). van Oosterhout et al. (2006b) sampled the Mid Naranjo (Mid) and Lower Aripo (Low) populations in 2001 in order to measure both MHC diversity (Mid, n = 21; Low, n = 35) and neutral microsatellite genetic variation (Mid, n = 44; Low, n = 56). Here, I repeat MHC and microsatellite analysis on the 2007 sample from the Mid Naranjo (MHC, n = 20; microsatellites, n = 45) and Lower Aripo (MHC, n = 19; microsatellites, n = 45) populations. Genomic DNA was extracted from the caudal fin using the HotSHOT protocol of Truett et al. (2000).

Microsatellite analysis

Extracted DNA was amplified at 13 microsatellite loci (see Chapter 2). Here, just six of those loci are used to measure neutral divergence, they include two interrupted repeats *Pr39* and *Pr92* (Becher et al. 2002); three dinucleotide repeats *Pret-69, pret-77* (Watanabe et al. 2003), *Hull 9-1*; and a tetranucleotide repeat *Hull 70-2* (van Oosterhout et al. 2006b). The 2007 dataset is reduced here from 13 to 6 loci. This is for comparison to the 2001 dataset which consisted of only six microsatellite loci (see van Oosterhout et al. 2006b).

MHC analysis

Extracted DNA was PCR amplified, cloned and sequenced (see Chapter 3). Errors can be introduced to sequences by heteroduplex mismatch repair, chimera sequences or sporadic substitutions caused by *Taq* polymerase misincorporations (Keohavong and Thilly 1989; Kobayashi et al. 1999; Kanagawa 2003; Cummings et al. 2010). In the present study, I accounted for errors associated with cloning and sequencing by confirming MHC alleles only when they were observed in two or more independent PCRs (Lukas et al. 2004; Lukas and Vigilant 2005). This conservative method is only feasible when large numbers of individuals and clones per individual have been sequenced. An alternative method, which involves clustering MHC sequences based on similarity (Acinas et al. 2004; Blais et al. 2007), was used by van Oosterhout et al. (2006b). In this clustering method, sequences that differ by three base pairs or less were grouped together and one representative sequence was used. For consistency, I used this clustering method to confirm the 2001 sequences that differed by three base pairs or less to MHC sequences analysed in the 2007 sample (see Fig. 2, see also Fig 3 and 4 for MHC allele counts after clustering). A network was built (Network 4.5.6.1 - http://www.fluxus-technology.com) using median joining to confirm the numbers of mutations between 2001 and 2007 MHC class II*B* alleles (Fig. 2).

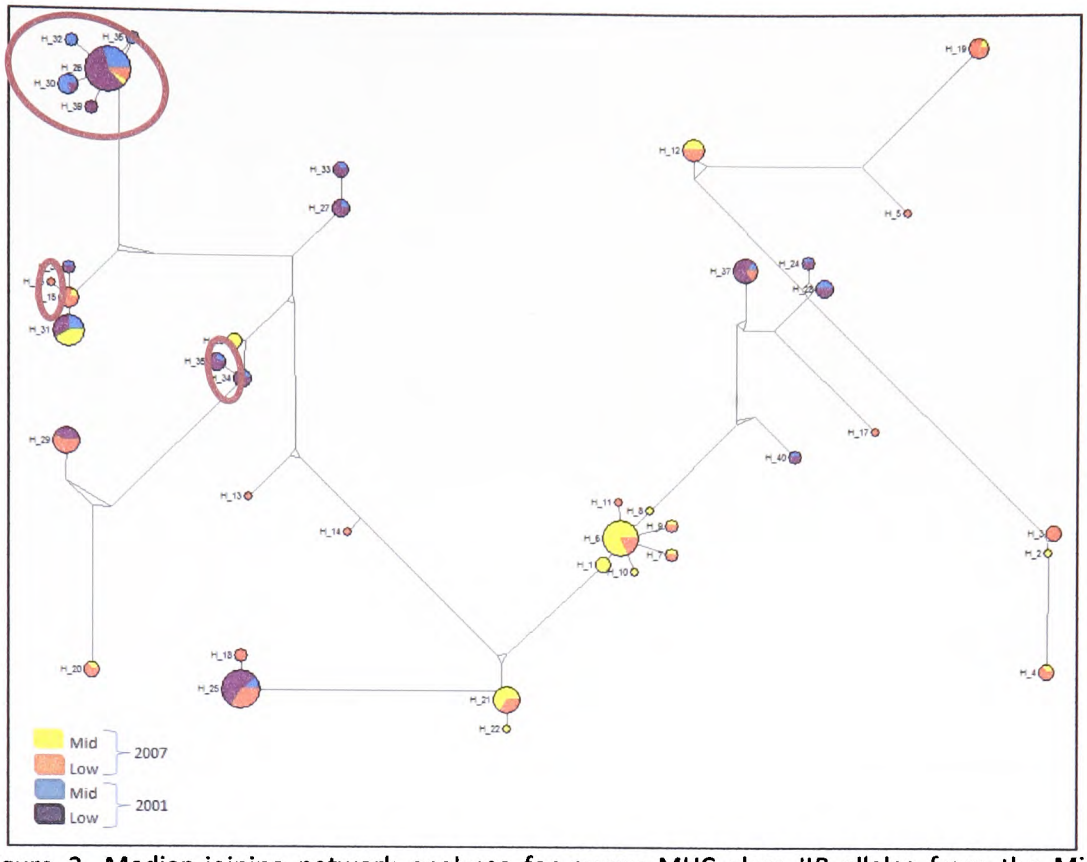


Figure 2. Median-joining network analyses for guppy MHC class II*B* alleles from the Mid Naranjo (Mid) and Aripo (Low) Rivers, Trinidad. Guppies were sampled in October 2001 and June 2007 and had between one and six MHC class II*B* alleles. Note that numbers of individuals (2001 Mid = 21, Low = 35; 2007 Mid = 20, Low= 19) and numbers of clones sequenced per individual (2001 ≈ 6; 2007 ≈ 16) are different. Brown ovals represent clusters of alleles, less than three base pairs apart, that were condensed between datasets ([26, 30, 32, 36, 39] [15, 38] [23 34]).

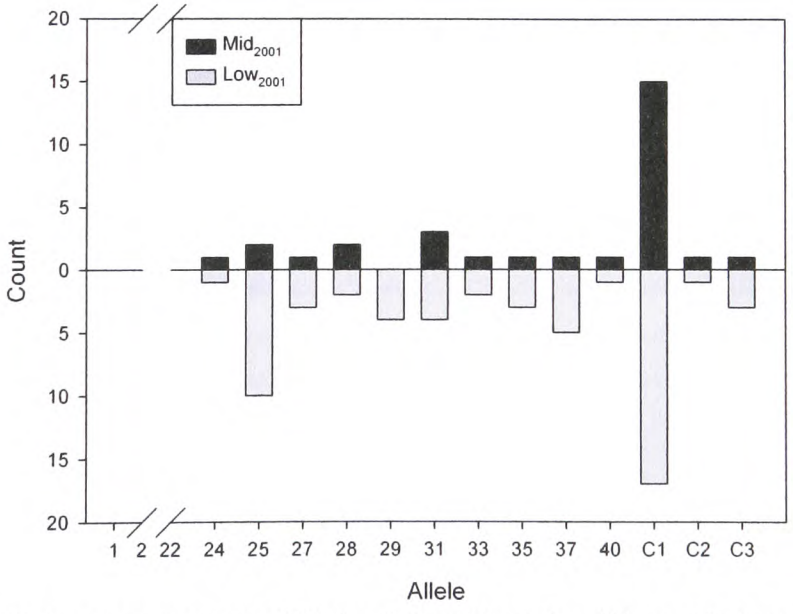


Figure 3. MHC allele counts of the Mid (n = 21) and Low (n = 35) populations in 2001. Alleles between the 2001 and 2007 data sets similar by three nucleotides or less were clustered into one allele confirmed in 2007 (C1-C3). The effect of this clustering method was to reduce the number of alleles in 2001. Compare with Figure 1 in van Oosterhout et al. (2006b).

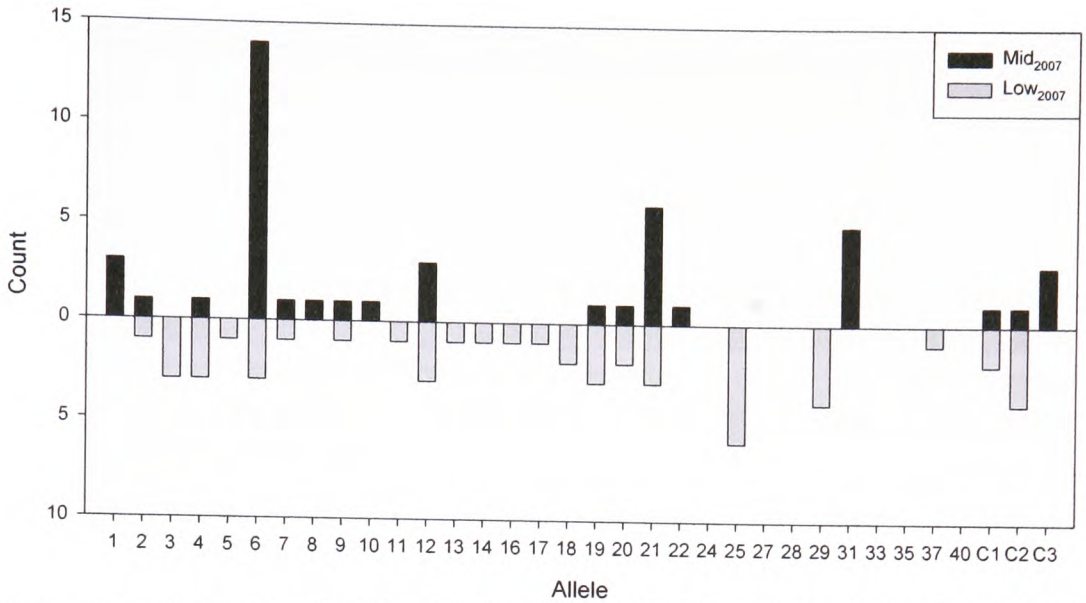


Figure 4. MHC allele counts of the Mid (n = 20) and Low (n = 19) populations in 2007. Alleles between the 2001 and 2007 data sets similar by three nucleotides or less were clustered into one allele confirmed in 2007 (C1-C3). (For Mid and Low comparisons across years see Appendix 7)

Population genetic analyses

Microsatellite genotype frequencies from the present study were checked for the presence of null alleles, amplification bias and scoring errors using MICRO-CHECKER 2.2.3 (van Oosterhout et al. 2004). Genotypic disequilibrium was calculated using FSTAT version 2.9.3 (Goudet 1995, 2001) with a Bonferroni adjusted P-value for multiple testing at the 5% nominal level (α = 0.000641).

Genetic differentiation (G'_{ST}), corrected for highly polymorphic loci, (G'_{ST}) (Hedrick 2005b) was used to measure differentiation at microsatellite and MHC loci among the Up, Mid and Low populations. The mean simulated G'_{ST} values (with 5 – 95% CI) were calculated by bootstrapping over alleles with replacement and using 1000 runs. Bootstrapping was done over alleles as opposed to individuals because it minimises the effect of separate sampling regimes employed by van Oosterhout et al. (2006b) and the present study. Two MHC alleles were sampled with replacement for each individual, which resulted in a close match between the observed and simulated frequencies of alleles per individual (see Appendix 8). For comparison to temporal MHC, microsatellite genetic variation was also re-analysed here by bootstrapping over alleles. This (over alleles) sampling scheme was not found to change estimates of microsatellite G'_{ST} (see Appendix 9).

Genetic differentiation is calculated for both microsatellite and MHC loci. Comparisons of genetic differentiation are made within population over time (temporal) and between contemporary populations (spatial) using randomisation tests. These levels of comparison allow me to examine the hypothesis that the interplay between migration, drift and balancing selection is driving adaptive evolution and changes in MHC allele frequency in these guppy populations. The pattern of genetic differentiation observed between MHC and microsatellites

may differ. Reduced MHC to microsatellite genetic differentiation is indicative of balancing selection mediated by rare allele advantage (NFDS or overdominance). On the other hand, increased MHC to microsatellite genetic differentiation is indicative of balancing selection mediated by selection that varies over space and/or time (Spurgin and Richardson 2010). The measurement of spatial genetic differentiation allows the distinction of these modes of balancing selection.

Results

Micro-Checker 2.2.3 (van Oosterhout et al. 2004) analysis indicated null alleles in *Hull 70-2* and *Hull 9-1*. However, in both cases, the combined probability for all classes was not significant and these frequencies were used uncorrected in future analyses. There was no genotypic disequilibrium at microsatellite loci ($P > 0.0006$, adjusted P-value corresponding to 5% nominal level after Bonferroni correction).

Temporal MHC allele frequencies show a marked difference, with only 7 out of 33 (21.2%) MHC alleles being observed in both the 2001 and 2007 samples (Fig. 5). By contrast, for the six microsatellite loci that were analysed in both 2001 and 2007, there were 39 out of 69 (56.5%) alleles shared between the temporal samples of the Aripo populations (see Appendix 10). To further quantify and interpret these results two comparisons of genetic differentiation were made: (1) within population over time (temporal) (Fig.6), (2) between populations within time (spatial) (Fig.7). These spatio-temporal comparisons are made for both the MHC, as well as for the microsatellite loci.

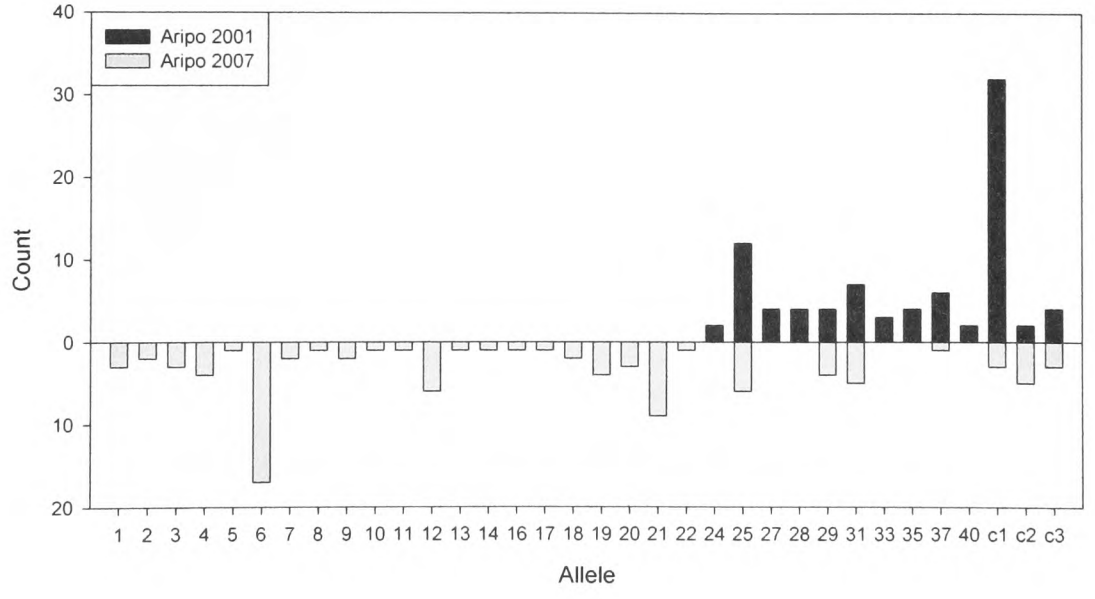


Figure 5. MHC allele counts of the Mid ($n = 21$) and Low ($n = 35$) populations in 2001 and the Mid ($n = 20$) and Low ($n = 19$) populations in 2007. Just 7 out of 33 (21.2%) MHC alleles are observed in both the 2001 and 2007 samples.

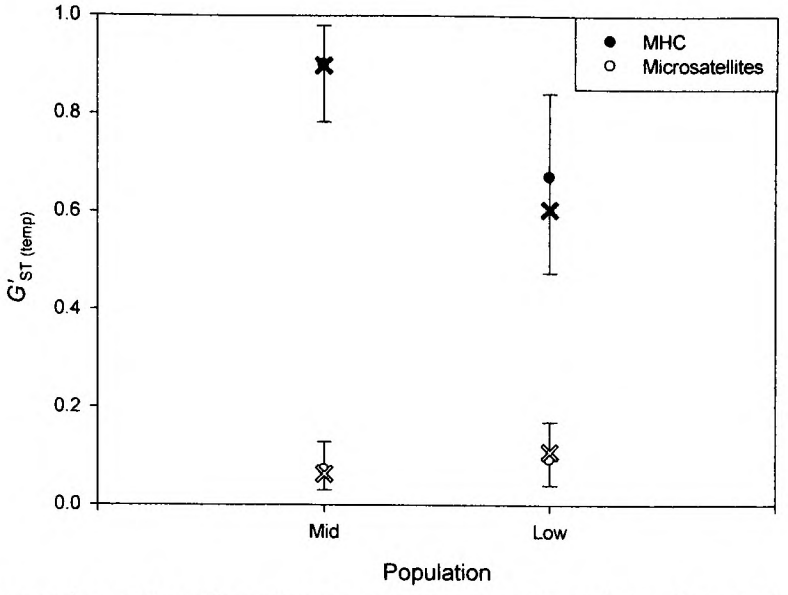


Figure 6. Temporal genetic differentiation expressed as G'_{ST} values of neutral variation at six microsatellite loci (open symbols) and at MHC loci (solid symbols). Observed and simulated (mean, 5 - 95% CI) measures are represented by crosses and points, respectively. Measures of genetic differentiation are within populations (Mid or Low) between 2001 and 2007 sampling periods. Temporal genetic differentiation is significantly higher for the MHC than for the microsatellites.

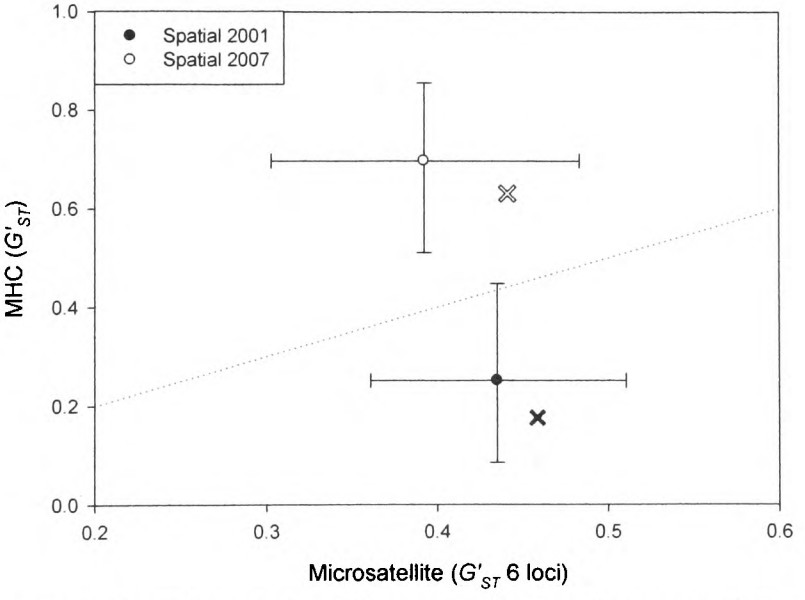


Figure 7. Spatial genetic differentiation between the Mid and Low populations of the Aripo River expressed as G'_{ST} values of neutral variation at six microsatellite loci and at MHC loci. The dotted line (MHC G'_{ST} = Microsatellite G'_{ST}) shows where the MHC and neutral microsatellite divergence are identical. Observed and simulated (mean, 5 - 95% CI) measures are represented by crosses and points, respectively. Spatial MHC genetic differentiation is lower than neutral in 2001 (solid symbols below dotted line) as predicted for loci under balancing selection (Schierup 1998; Schierup et al. 2000; Muirhead 2001). However, this pattern is diametrically different in 2007 (open symbols), where the opposite pattern is observed. The pattern in 2007 is consistent with fluctuating selection, e.g. due to different local adaptations at the MHC in the Mid and Low populations.

Genetic divergence (G'_{ST})

Temporal genetic differentiation: MHC *vs.* microsatellites

Between 2001 and 2007, temporal genetic divergence of neutral microsatellite markers is significantly lower than the temporal differentiation of the MHC (Randomisation test: $P < 0.001$), (Fig.6 and Table 1). This confirms that temporal changes in MHC allele frequencies are more pronounced than those of microsatellite alleles as predicted by all models of balancing selection. Thus, this observation is consistent with selection that varies over time (fluctuating selection), as well as with NFDS or overdominance. Importantly this finding indicates the introduction of MHC alleles from populations other than the Mid or Low (such as the Caroni metapopulation).

Table 1. Genetic differentiation (G'_{ST}, mean (5 – 95% CI)) at MHC and microsatellite loci among populations of guppies of the Upper and Mid Naranjo and Aripo Rivers, Trinidad.

		Temporal			Spatial			
		***			***			
		MHC	Microsatellites		MHC		Microsatellites
Mid	*	0.900 (0.784-0.980)	0.071 (0.030-0.129)	**2001**	0.251 (0.085-0.448)	**	0.435 (0.361-0.511)
Low	*	0.672 (0.477-0.842)	0.094 (0.040-0.171)	**2007**	0.697 (0.511-0.856)	**	0.392 (0.303-0.483)
					**		

* Values are significant at $P < 0.050$
** Values are significant at $P < 0.010$
*** Values are significant at $P < 0.001$

Spatial and temporal genetic differentiation: Microsatellites

Genetic divergence at microsatellite loci is significantly higher at the spatial (overall mean $G'_{ST} = 0.450$) (Fig.7) than at the temporal scale (overall mean $G'_{ST} = 0.086$) (Fig.6 and Table 1), (Randomisation test: $P < 0.001$). This high level of spatial genetic differentiation for the microsatellites is consistent with low levels of gene flow between the Mid and the Low populations. Furthermore, the low temporal G'_{ST} values shows that the microsatellite allele frequencies distributions are stable over time. In summary, these data indicate that microsatellite alleles show a relatively little allelic turnover, and that there is little gene flow between the Mid and Low populations judging by the large level of spatial differentiation.

Spatial and temporal genetic differentiation: MHC

Temporal genetic differentiation at the MHC is significantly higher in the Mid population than in the Low population (Randomisation test: $P = 0.041$) (Fig.6 and Table 1). This finding suggests there is a higher turnover rate of MHC alleles in the Mid than in the Low. This relatively high turnover of MHC in the Mid may be indicative of larger changes in parasite fauna. It is also consistent with larger stochasticity and allele frequency fluctuations in a smaller population, as observed in previous studies (van Oosterhout et al. 2006; Barson et al. 2009). Spatial genetic differentiation at the MHC is significantly lower in 2001 compared to 2007 (Randomisation test: $P = 0.006$) (Fig.7). This suggests that the efficacy of balancing selection in homogenising gene pools of populations varies significantly across years, which may be indicative of a change in the mode of balancing selection.

Spatial genetic differentiation in 2001 and 2007: MHC *vs.* microsatellites

Figure 7 furthermore shows that the spatial differentiation at the MHC is lower than neutral microsatellite differentiation in 2001 (solid symbols below dotted line). The observed MHC G'_{ST} (0.176) was outside the lower CI of the microsatellite G'_{ST} ($G'_{ST\ 95\%\ CI}$ = 0.361) which suggests the level of MHC differentiation was significantly lower than microsatellite differentiation. This observation is consistent with overdominance and NFDS models of balancing selection (Schierup 1998; Schierup et al. 2000; Muirhead 2001). Interestingly, however, the pattern observed in 2007 (open symbols), shows that the MHC is spatially more diverged than the microsatellites (Randomisation test: P = 0.005). The pattern in 2007 is consistent with fluctuating selection, e.g. due to different local adaptations at the MHC in the Mid and Low populations. The 2001 MHC data are thus most consistent with overdominance and/or NFDS, whereas the 2007 MHC data suggest large spatial differences in selection, concordant with the fluctuating selection model of MHC evolution and local adaptation.

Spatial and temporal genetic differentiation: MHC *vs.* microsatellites

The most striking difference between MHC and microsatellites is observed when comparing the relative magnitudes of the level of spatial versus temporal differentiation (Fig. 8). The spatial differentiation (G'_{ST}) is relatively similar for the microsatellites and for the MHC. However, temporally, the MHC diverges almost an order of magnitude faster than the neutral microsatellite markers (see Fig. 8). This level of temporal divergence of MHC to microsatellite loci suggests that higher effective migration of MHC alleles is a driving factor in temporal variation.

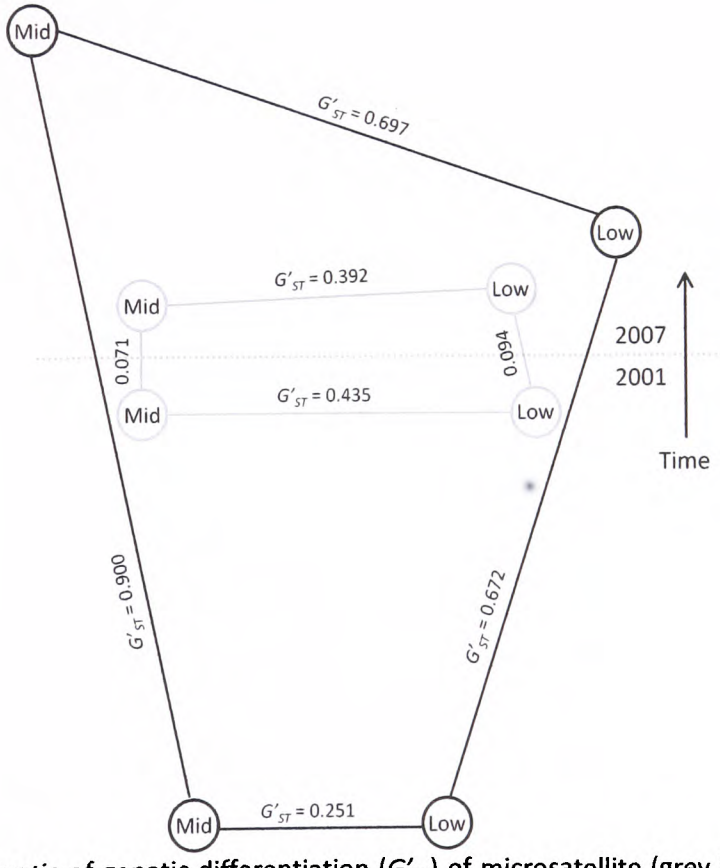


Figure 8. Schematic of genetic differentiation (G'_{ST}) of microsatellite (grey bars) and MHC loci (black bars) between the Mid Naranjo (Mid) and Lower Aripo (Low) guppy populations sampled in 2001 and 2007. Pairwise distances between population samples are relative to their G'_{ST} values.

Discussion

In the present study, I analysed genetic differentiation (G'_{ST}) at six neutral microsatellite loci and at selectively non-neutral MHC loci. I tested whether gene flow is more efficient in homogenising genetic variation of genes under balancing selection than it is at neutral loci. This paradigm was examined at both the spatial and the temporal scale by comparing the temporal fluctuations in MHC and microsatellite allele frequencies of wild guppy populations. The data suggests that the interplay between random genetic drift, migration and alternate modes of balancing selection is driving adaptive evolution and MHC allele frequency changes in these guppy populations. Changes in MHC allele frequencies over time are more pronounced than those of microsatellite alleles. High temporal MHC (relative to microsatellite) variation indicates the introduction of novel alleles from populations outside of this study. However high temporal MHC genetic differentiation is consistent with fluctuating selection as well as NFDS and overdominance (see Fig. 1). Spatial genetic differentiation in 2001 is indicative of NFDS or overdominant processes where a higher effective migration rate of MHC alleles reduced population differentiation compared to neutral loci (Schierup 1998; Schierup et al. 2000; Muirhead 2001). Conversely, the opposite pattern of genetic divergence between MHC and microsatellite loci is observed in 2007 where the MHC is more diverged than that of microsatellites. This is indicative of fluctuating selection over a spatial scale in a metapopulation with gene flow (i.e. fluctuating selection). Altogether, temporal replacement (turnover) of MHC alleles indicates introduction of novel MHC alleles though immigration aided by a higher effective migration rate. The type of balancing selection however may have changed from overdominance or NFDS in 2001, to spatial variation in the direction of selection (fluctuating selection) in 2007.

Three nonexclusive models of balancing selection

Overdominance, NFDS and fluctuating selection have all been proposed as a means of maintenance of alleles under balancing selection (reviewed by Meyer and Thomson 2001; Spurgin and Richardson 2010). Overdominance and NFDS maintain genetic variation through distinctly different processes (see Doherty and Zinkernagel 1975b; Slade and McCallum 1992; Hughes and Yeager 1998; Sommer 2005), but the predicted effects on gene diversity and genetic differentiation are empirically and theoretically almost indistinguishable (Slade and McCallum 1992; Apanius et al. 1997; but see Borghans et al. 2004; De Boer et al. 2004). In a closed system (only the Mid and Low populations) both models lead to low levels of genetic differentiation over time as rare alleles are "rescued" because of their increased fitness, hence, reducing the allelic turnover rate (Kojima 1971; Takahata and Nei 1990). In open populations (*e.g.* stepping stone metapopulations), overdominance and NFDS can however increase genetic divergence of a population over time (see verbal model below). Similarly, rapid allelic turnover rate is expected when selection varies over time and space (fluctuating selection) (Hedrick et al. 1976; Hedrick 2002) because a change in the direction and intensity of selection favours alternative MHC alleles (Spurgin and Richardson 2010). This type of fluctuating selection may elevate the spatial and temporal genetic divergence of the MHC above that of neutral microsatellite genes.

Temporal differences in genetic differentiation (G'_{ST})

In the present study, temporal genetic differentiation is much higher for the MHC than for the microsatellites. The high level of MHC differentiation is explained by the large amount of allelic turnover (78.8% introduced MHC alleles). van Oosterhout et al. (2006b) confirmed MHC by

grouping alleles similar by three base pairs or less. This effectively reduces the estimated number of alleles in the 2001 dataset. Concordantly, the high temporal variation estimated here is a conservative measure of the actual amount of allele turnover. The considerable temporal genetic variation of MHC alleles is consistent with another study on these guppy populations that was based on a 2006 and a 2007 sample (see Fraser et al. 2010b).

Increased temporal MHC allele frequency variation (compared to neutral variation) can be explained by temporally fluctuating selection (Hedrick et al. 1976; Hedrick 2002). However, the introduction of so many alleles requires that the population is open and exchanges genes with different source populations in a metapopulation. Given a metapopulation structure, temporal MHC divergence is also consistent with the introduction of alleles by immigration in combination with either NFDS or overdominance (see Introduction and Chapter 5). Large fluctuations in allele frequencies over time have been documented in two temporal studies (Westerdahl et al. 2004; Charbonnel and Pemberton 2005). These data are thus consistent with overdominance and NFDS (in a metapopulation with migration; see verbal model below), as well as with selection that varies in time and space.

Spatial differences in genetic differentiation (G'_{ST})

Spatial genetic differentiation in 2001 was lower for MHC than for microsatellites, which is indicative of balancing selection though overdominance or NFDS. The opposite is true for 2007 and indicates selection fluctuating over a spatial scale. Overall, MHC genetic differentiation between populations (spatial) was lower than temporal but this difference was not statistically significant. Conversely, for neutral variation spatial genetic differentiation was significantly higher than temporal. High spatial to temporal microsatellite variation indicates there is little migration between populations.

Fraser et al. (2010a) demonstrated lower than expected MHC divergence between ten guppy populations from five separate rivers in three drainages. The authors concluded that (parasite mediated) balancing selection had acted to homogenise MHC genetic differentiation between populations. Fraser and Neff (2010) identified gyrodactylids as a potential agent of selection, suggesting these parasites select for similar MHC alleles in separate populations. This is consistent with overdominance and NFDS, and not with the predictions of fluctuating selection. Interestingly, Fraser et al. (Fraser et al. 2010b) also noted a change in the direction of selection for their 2007 sample of guppies.

Two accounts of maintenance of MHC variation in the Aripo River

Fluctuating selection model

Heterogeneity in selection pressure implies that intensity and direction of selection by parasite fauna has changed in the years between 2001 and 2007. Theoretical studies have shown that fluctuating selection (i.e. selection fluctuating over a spatio-temporal scale) can maintain a balanced polymorphism (Hedrick 2002). Gyrodactylids represent a large proportion of guppy parasite fauna (Cable In press) and many studies have used them as a proxy for parasite selection in the guppy (van Oosterhout et al. 2006b; van Oosterhout et al. 2007a; van Oosterhout 2009a; Fraser and Neff 2010; Fraser et al. 2010b). Gyrodactylids are ubiquitous guppy ectoparasites, which complete their entire life cycle on a single host (Cable and Harris 2002; Cable In press). Gyrodactylid infections reduce swimming ability (van Oosterhout et al. 2003) and increase susceptibility to secondary bacterial infection (Xu et al. 2007). We found

that these factors severely reduce an individual's potential to avoid being flushed downstream (van Oosterhout et al. 2007a). Furthermore, a correlation was found between a specific type of MHC allele and an individual's gyrodactylid load (Fraser and Neff 2010). Gyrodactylid and other parasites (i.e. digenean and fungal) burdens have been shown to vary significantly within population over time (see Chapter 3). Unfortunately, specific changes in parasite abundance in 2001 are not available and so it is not possible to make a direct comparison between parasite fauna of 2001 and 2007. Nevertheless, it is possible that changes in parasite composition over the sampling period have influenced the direction of selection in MHC variation with time.

NFDS and overdominant selection: a verbal model

Increased efficiency of migration can result in a fast turnover rate of MHC alleles, particularly when the effective population size fluctuates over time. Large populations tend to maintain more MHC alleles than populations with smaller effective size (assuming mutation-selection-drift balance). When a population experiences a bottleneck (e.g. during a wet season rains, see van Oosterhout et al. 2007a; see also Barson et al. 2009), its effective population size is reduced and MHC alleles will be lost. High effective migration rate of MHC alleles facilitates the invasion of new MHC alleles during subsequent population expansion (in the dry season). The immigrant alleles rapidly increase in frequency driven by both the population expansion and the effects of balancing selection (in particular, overdominance and NFDS). In the next bottleneck, resident and immigrant alleles have similar (equilibrium) frequencies in the population and are thus equally likely to be lost from the gene pool. Annual population bottlenecks (during wet season spate events) may remove MHC alleles, which combined with the high effective migration rate of alleles under balancing selection could perpetuate the turnover of MHC alleles in small upstream (Mid) habitats.

In Lowland populations, the turnover of alleles could be explained by the same process where introduction of alleles is from upstream habitats or other rivers in the drainage (Barson et al. 2009), or even from other drainages (Suk and Neff 2009; Willing et al. 2010). The rapid turnover rate of MHC could be a process that acts over the entire Caroni drainage, or even across drainages. This could explain why similar MHC alleles have been found in separate guppy populations within and between drainages (see Fraser and Neff 2010). Babik et al. (2008) highlighted the importance of population demographic processes (drift and migration) in MHC variation of the Alpine newt. The present study provides further evidence for balancing selection acting to rescue MHC depauperate populations through higher effective migration rate of MHC alleles compared to neutral genetic variation.

Concluding remarks

In the present study, I find that evidence for the maintenance of MHC variation that is consistent with different modes of selection in different years. Clearly further analysis of the data is required in order to determine the mode of selection acting on the MHC. However, these hypotheses require temporal variation in MHC alleles but not neutral microsatellite alleles. Such a signal is indicative of a higher effective migration rate of the MHC. In Chapter 5, I will examine whether changes in parasite selection pressure are required to explain the fast turnover of MHC alleles. The reason to test this is that although temporal samples differ in parasite loads and prevalence, the same species (or genera) of parasites are found in the same population in different years. Furthermore, one guppy population (Up) that has little or no parasites, appears to posses very similar MHC alleles as the parasitized Mid and Low

populations. Finally, the vast change in MHC alleles needs to be fuelled by immigration, because the mutation rate cannot provide so many novel genetic variants in only few generations. Altogether, these data cast doubt on the driving force of parasite-mediated selection and Red Queen Dynamics, and this begs the question of whether the combination of immigration and balancing selection could explain the rapid temporal changes in MHC alleles.

A theoretical study on the population genetics of MHC in a metapopulation

Faster than the Red Queen

Chapter 5

Abstract

The co-evolutionary arms race between parasite virulence genes and host immune genes is known as Red Queen dynamics. Documenting Red Queen dynamics empirically is particularly challenging, and instead, most studies demonstrate a pattern of variation in genotypes and allele frequency fluctuations that is consistent with host-parasite co-evolution. Here computer simulations and empirical data of the Major Histocompatibility Complex (MHC) of guppies (*Poecilia reticulata*) are used to study how the turnover rate of alleles (temporal G'_{ST}) is affected by balancing selection, the level of genetic polymorphism in the metapopulation, the migration rate and the effective population size. Although large fluctuations in allele frequencies are consistent with Red Queen dynamics, simulations show that other demographic and population genetic processes can also account for such data. The combined impacts of evolutionary forces and stochastic events can dramatically alter the expectation for genes under balancing selection compared to that of neutral genes. In particular, the commonly held assumption that balancing selection homogenises gene frequencies and reduces the level of genetic differentiation (G'_{ST}) is not always correct, and it depends on the interaction between evolutionary forces and population demography. Furthermore, although balancing selection plays an important role in the population genetics of the MHC, parasites are not the only agents of selection. The present study demonstrates that balancing selection can explain a rapid turnover in MHC alleles, and that this observation should not be taken as evidence of Red Queen dynamics through host-parasite co-evolution.

Introduction

Currently, a major challenge in evolutionary biology is to understand why the level of genetic variation differs among genes that occur within the same genome (e.g. Hohenlohe et al. 2010). The effects of random genetic drift and migration are typically ruled out based on the logic that all genes occur in the same genomic environment. Hence, these stochastic processes are assumed to affect all genes to more or less the same extent and therefore, represent the null model by which other forces are measured (reviewed by Ford 2002; Nei 2005). Polymorphism generated in a population through mutation or recombination is eroded by genetic drift (reviewed by Charlesworth 2009). Natural selection can act to increase or reduce polymorphism and therefore, change the expectation under a neutral model of evolution (reviewed by Nielsen 2005; Hurst 2009). Balancing selection is of particular interest because it acts to maintain polymorphism despite the effects of genetic drift (see e.g. Aguilar et al. 2004; van Oosterhout et al. 2006b; but see Bollmer et al. 2007).

There are four main models of balancing selection, overdominance (or heterozygote advantage), negative frequency dependent selection (NFDS) (or rare allele advantage), selection varying in time and space, and associative balancing complex (ABC) evolution (see Kojima 1971; Doherty and Zinkernagel 1975a; Takahata and Nei 1990; Penn and Potts 1999; van Oosterhout 2009a, b). Although these models differ in their assumptions about the agents of selection, they have in common that they balance the frequencies of alleles in the population. Hence, these models explain how a genetic polymorphism can be maintained above that of so-called neutral genes, i.e. genes that are in a mutation-drift equilibrium (Crow and Kimura 1970).

The effects of population demography on polymorphism may vary depending on whether a gene is neutral or subject to balancing selection. For example, compared to neutral alleles, alleles under NFDS or overdominance are predicted to show a greater effective migration rate because selection favours the initially rare immigrant alleles (Schierup 1998; Schierup et al. 2000; Muirhead 2001). Hence, at a spatial scale, genes under balancing selection are predicted to show less population subdivision (Schierup et al. 2000). Although theoretically this effect of balancing selection is well-established, relatively few empirical studies corroborate these expectations (e.g. Sommer 2003; van Oosterhout et al. 2006b; Mona et al. 2008; Fraser and Neff 2010) and there are many exceptions (e.g. Landry and Bernatchez 2001; Cohen 2002; Ekblom et al. 2007; Alcaide et al. 2009; Fraser et al. 2010b).

Balancing selection maintains polymorphism introduced by migration and mutation, thus increasing the frequency of novel genetic variants. It also increases the number of alleles that can be maintained at a locus, although this number remains finite (Crow and Kimura 1970). Hence, once a population is in a mutation-selection-drift equilibrium, any novel allele that becomes established in the population will ultimately replace a resident allele. With high migration rates, balancing selection may thus promote the turnover rate of alleles and increase the genetic differentiation at a temporal scale (i.e. the temporal F_{ST}) relative to that of neutral genes. This prediction appears to be consistent with the few temporal MHC studies done so far (Westerdahl et al. 2004; Bryja et al. 2007; Hess et al. 2007; Fraser et al. 2010b), although Oliver et al. (2009b) found that temporal genetic change of the MHC in a water vole metapopulation was most consistent with neutral factors such as migration and drift.

The rapid turnover rate of alleles in genes under balancing selection might be interpreted as evidence for adaptive evolution. The dynamic co-evolutionary arms race between parasite virulence genes and host immune genes is known as Red Queen dynamics (Bell 1982). To empirically demonstrate Red Queen dynamics is particularly challenging because it requires a time-series analysis of both host and parasite genotypes. Hence, only few studies have shown the co-evolutionary dynamics in nature (e.g. Decaestecker et al. 2007; Paterson et al. 2010), and studies showing co-evolution between parasites and vertebrate hosts are rare (Woolhouse et al. 2002). Instead, most studies simply demonstrate a pattern of variation in genotypes that is consistent with co-evolution (Woolhouse et al. 2002), and in those cases, the rapid turnover rate of alleles is often taken as evidence of Red Queen dynamics. Although large fluctuations in allele frequencies are consistent with Red Queen dynamics, other demographic and population genetic processes might be able to explain such data as well.

Genes of the Major Histocompatibility Complex (MHC) play a central role in the vertebrate immune system. The molecular structure and the function of some MHC genes are conserved

across vertebrates (Kaufman et al. 1999; Hess and Edwards 2002; Danchin et al. 2004; Kelley et al. 2005; Wegner 2008), which has made them ideal candidates for studying selection in non-model organisms in the wild (see reviews by Bernatchez and Landry 2003; Piertney and Oliver 2006). The aim of this study is to explore the factors that can contribute to the remarkable evolutionary dynamics of the MHC besides temporal changes in parasite-mediated selection. For this purpose, I analyse theoretically the interaction between random genetic drift, gene flow and balancing selection, namely (symmetric) overdominant selection. I investigate how these evolutionary forces affect the spatio-temporal genetic differentiation of genes in this highly polymorphic multigene family. I compare the theoretical expectations with the population genetic data of microsatellites (Barson et al. 2009, see also Chapter 2), SNPs (Willing et al. 2010) and MHC (van Oosterhout et al. 2006b; Fraser and Neff 2010; Fraser et al. 2010b, see also Chapter 3) from a guppy (*Poecilia reticulata*) metapopulation in the Caroni Drainage in Trinidad.

Materials and Methods

A computer simulation model was written to study how the allelic turnover rate as estimated by the temporal genetic differentiation (G'_{ST}) is affected by balancing selection, gene flow, population size and the polymorphism at the locus. The population sizes and metapopulation structure simulated was designed to mimic those of many freshwater fish species (e.g. Hanfling and Weetman 2006), in particular that of guppies (*Poecilia reticulata*) in the Caroni Drainage in Trinidad (Barson et al. 2009; Willing et al. 2010). In this source–sink metapopulation (Caroni drainage), the tributary furthest downstream (the Caura catchment) represent a "super-sink" that receives immigrants from several rivers upstream in the drainage. Downstream migration occurs during the wet season rains when the rivers are in spate (van Oosterhout et al. 2007a). After seasonal floods, the census population size of guppies in the Caura catchment is estimated to be tens to hundreds of millions (van Oosterhout pers. comm.). This extremely high density of guppies in the drains of towns in the lowlands has given the guppy its local common name "million fish" (Houde 1997). In the dry season, the fish migrate upstream and these seasonal migration events create a highly dynamic metapopulation structure (Barson et al. 2009).

This metapopulation structure was modelled simulating an upstream and a downstream river population that were connected by gene flow. These populations represent the Mid (upstream) and the Low (downstream) guppy populations. An additional, infinitely large "super-sink" population (the Caroni drainage) contributed to the gene pool of the downstream river. The level of admixture estimated empirically by Barson et al. (2009) ranged from 5-21% (mean 12%) for the downstream populations of guppies. The model simulated three generations per year over a six year period. Within each year, population demography was affected by a flushing event, caused by annual rains (see van Oosterhout et al. 2007a). A variety of alternate models were simulated, however, the general model simulated a flushing event followed by upstream migration and then population growth in each year. Microsatellite variation was affected by population demographic processes and the MHC was also affected by symmetric overdominance. Parameter variations within this model included the selection coefficient ($S = 0 - 0.5$), source sink MHC diversity ($k = 5 - 80$), effective population size ($N_e = 30 - 1000$), migration rate ($Nm = 0.2 - 10.0$) and the time of year that migration took place (continual migration or seasonal migration).

Guppies are thought to possess (up to) three MHC class II*B* gene loci (Chapter 3) which are highly polymorphic (van Oosterhout et al. 2006a; van Oosterhout et al. 2006b; Fraser and Neff 2010; Fraser et al. 2010b). By combining three temporal datasets of the MHC of guppies in the Caroni Drainage (van Oosterhout et al. 2006a; van Oosterhout et al. 2006b; Fraser and Neff 2010; Fraser et al. 2010b) (Chapter 3) and calculating a discovery curve, it is estimated that there are approximately 85 MHC alleles in the guppy populations of the Caroni Drainage (see Appendix 6). The locus-specificity of these alleles remains unknown, and given the evidence of inter-locus recombination (van Oosterhout et al. 2006a), it is assumed each allele can occur at any of these three loci.

Balancing selection was modelled in the form of symmetric overdominance, with the fitness of the homozygotes $w = 1 - s$, and the heterozygote with a fitness of unity ($w = 1$). Given the relatively high migration rate ($Nm \geq 1$, see Chapter 2) and the large number of MHC alleles present in the entire metapopulation ($k = 85$) and the relatively small number of generations that are simulated ($\mu N_e t \ll 1$), the mutation rate is set to zero ($\mu = 0$). There is no genotypic linkage disequilibrium (i.e. free recombination, $c = 0.5$) and fitness is multiplicative across loci. Temporal genetic differentiation was expressed using a conventional F_{ST} statistic (G'_{ST}) which corrects for high levels of polymorphism present at the MHC and the simulated loci (Hedrick 2005b). The temporal G'_{ST} was calculated for the upstream population sampled in two sampling periods separated by six years, which in guppy life history is equivalent to $t = 18$ generations. This represents the two sampling periods (2001 and 2007) of which I have most extensive microsatellite, SNP and MHC data (van Oosterhout et al. 2006b; Barson et al. 2009; Fraser and Neff 2010; Fraser et al. 2010b; Willing et al. 2010, see also Chapters 2 and 3). Simulations were started with genotypes sampled at random from the source population (containing k MHC alleles). Data was recorded after a burn-in period of t = 1000 generations.

Results

Selection coefficient

First, the effect of balancing selection (symmetric overdominance) on the level of temporal genetic differentiation (G'_{ST}) for simulated upstream populations with constant size ($N_e = 100$). In these simulations, the gene flow between the upstream and downstream populations equals $Nm = 1$, and there are 85 alleles in the metapopulation. Figure 1 shows that G'_{ST} increases the higher the coefficient of balancing selection. Apparently, with stronger selection favouring the initially rare immigrant alleles, the allelic turnover rate and therefore the temporal G'_{ST} increase.

Amount of MHC polymorphism

Temporal genetic differentiation (G'_{ST}) furthermore increases with an increased level of polymorphism in the metapopulation (Fig. 2). If there are more distinct MHC alleles, immigrants are more likely to introduce novel, selectively favourable MHC alleles. The novel immigrant alleles can potentially replace the resident alleles, thereby increasing the temporal G'_{ST}.

Migration rate

Temporal genetic differentiation (G'_{ST}) increases with increasing upstream migration rate (Fig. 3). More upstream migrants introduce more novel MHC alleles. This increases the turnover rate of MHC alleles with time and therefore the temporal G'_{ST} increases.

Population size

The gene pools of large populations are relatively stable and only little affected by stochastic allele frequency fluctuations. However, small populations rapidly loose alleles, which are subsequently replaced by alleles brought in by immigrants. Consequently, the temporal G'_{ST} increases with decreasing upstream population size (Fig. 4).

Finally, I analysed the effects of population size fluctuations and fluctuations in migration rates on temporal genetic differentiation. In these simulations, the harmonic mean population size of the constant and fluctuating population are similar. Figure 5 shows that when the migration rate is constant, population size fluctuations do not affect the temporal G'_{ST}. However, population size fluctuations do increase the level of temporal genetic differentiation when migration occurs after a population-crash (Fig. 5). The MHC alleles that are introduced by immigrants after the wet season floods are likely to become established when the population is rapidly expanding after the population-crash in the wet season. Hence, the allelic turnover rate is highest when migration occurs seasonally, just after the population crash.

Comparison between theoretical expectation and empirical data

Figure 6 shows the genetic differentiation values of the upstream simulated (Mid) population that increases from 30, 300 to 3000 from the wet to dry season (harmonic mean population size $N_H = 81$), and a Lowland population (Low) that increases from 120, 1200 to 12000 ($N_H = 324$). In the simulations presented in Figure 6, the upstream and downstream migration rates were $Nm = 3.3$, and balancing selection operated on the MHC with selection coefficient $s = 0.5$. Selection coefficients $s < 0.2$ and migration rates $Nm < 2$ were inconsistent with the observed empirical G'_{ST} values (data not shown).

The observed high level of temporal G'_{ST} of the MHC, and the low level of temporal G'_{ST} of the microsatellites can be explained by the simulations. Note, however, that the observed spatial G'_{ST} of the MHC falls outside the 95% CI of the simulated values, indicating that other (unidentified) sources of variance may contribute to the spatial genetic differentiation in these guppy populations.

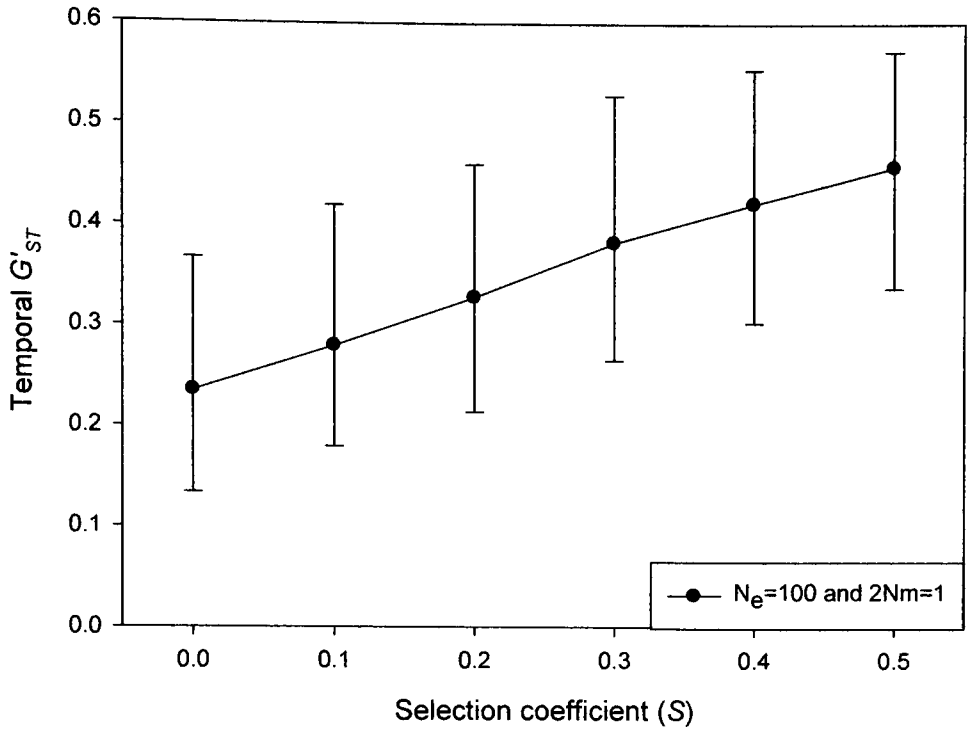


Figure 1. Temporal MHC G'_{ST} for simulated populations with constant size (N_e = 100) and with gene flow (Nm = 1) from a source populations with 85 distinct MHC alleles at a locus subject to balancing selection with $0 \leq s \leq 0.5$.

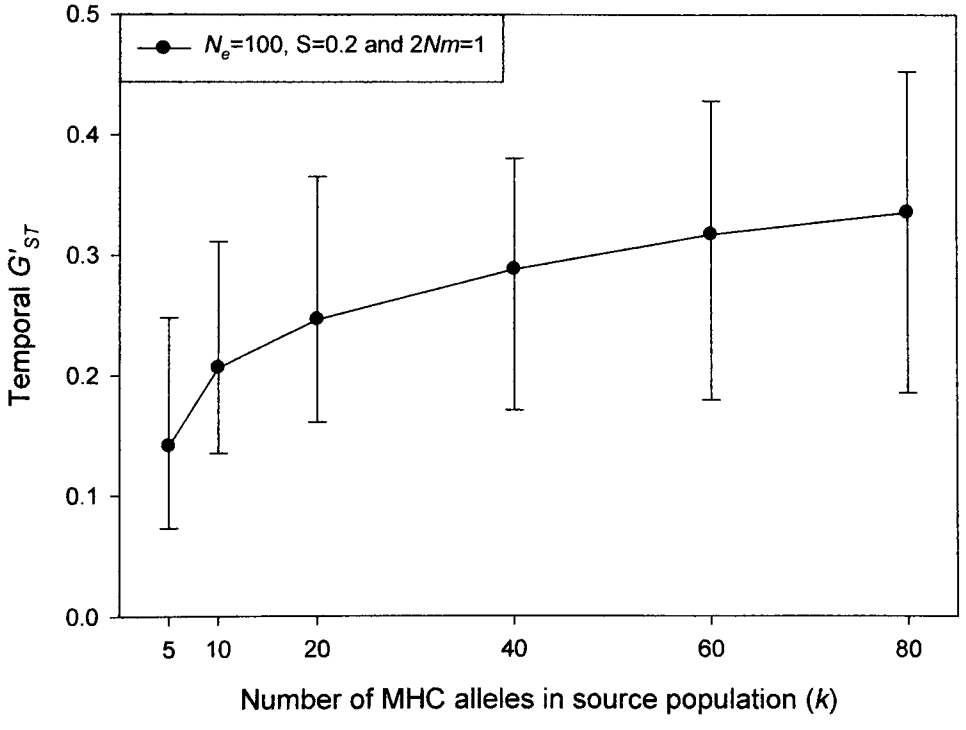


Figure 2. Temporal MHC G'_{ST} for simulated upstream populations with size N_e = 100 and with gene flow (Nm = 1) from a source populations with k = 5, 10, 20, 40, 60 and 80 distinct MHC alleles at a locus that is subject to balancing selection with s = 0.2.

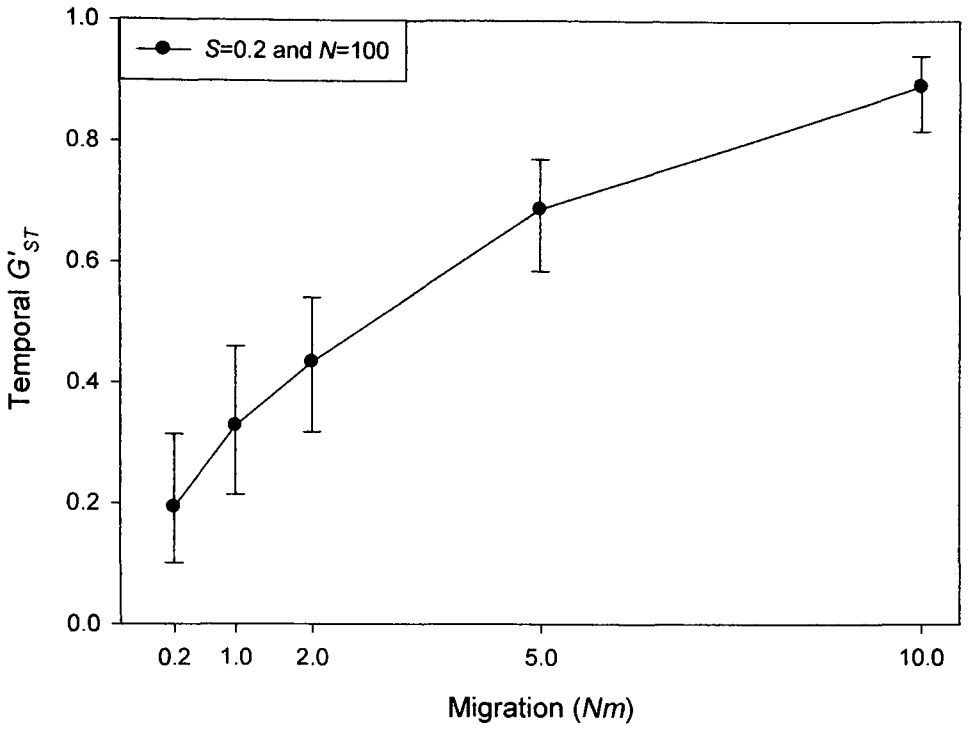


Figure 3. Temporal MHC G'_{ST} for simulated populations with size N_e = 100 and from a source populations with 85 distinct MHC alleles at a locus with balancing selection coefficient S = 0.2, and with an upstream migration rate of 0.2, 1.0, 2.0, 5.0 and 10.0 individuals per generation.

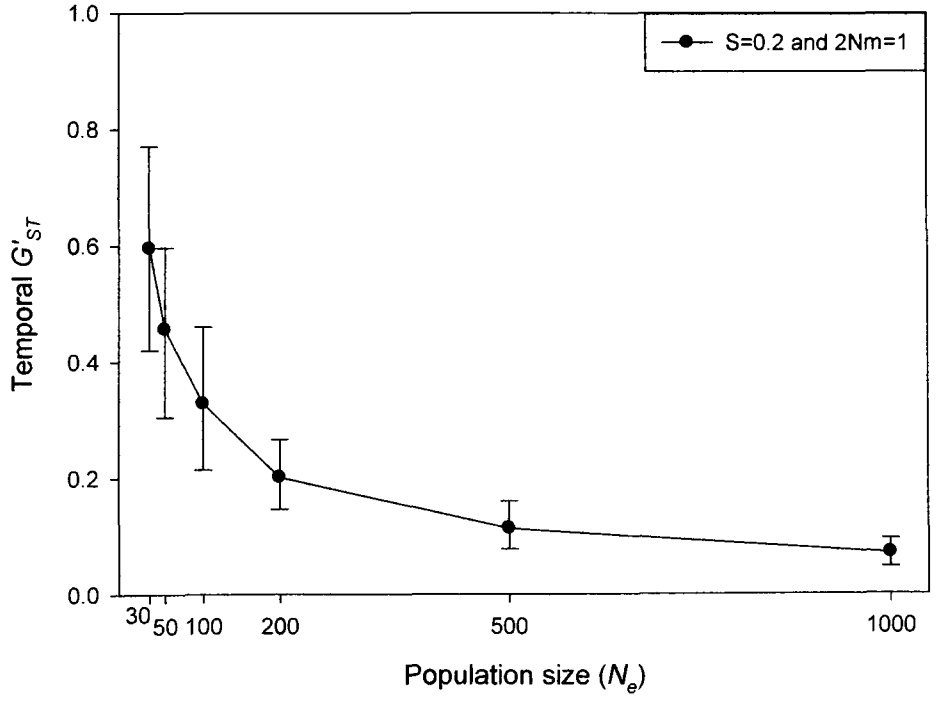


Figure 4. Temporal MHC G'_{ST} for simulated populations with gene flow (Nm = 1) from a source populations with 85 distinct MHC alleles at a locus with balancing selection coefficient S = 0.2, and with population size 30, 50, 100, 200, 500 and 1000.

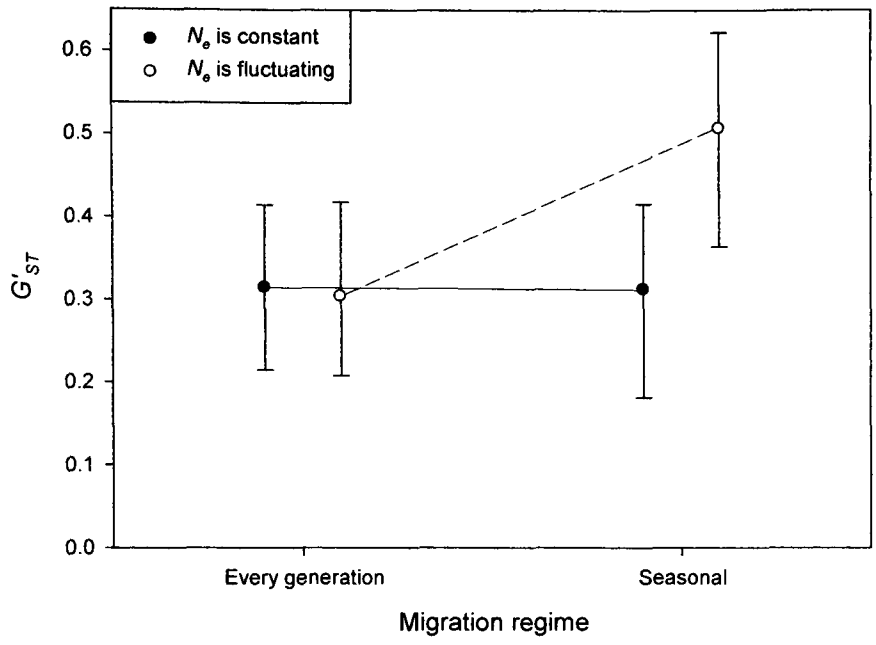


Figure 5. Temporal MHC G'_{ST} for simulated populations with a constant size N_e = 100 (solid circles), and fluctuating population sizes (N = 34, 2500, and 5400 in a continuous 3-generation cycle) that is equivalent to a harmonic mean size N_H = 100 (open circles). The gene flow rate is either constant every generation (Nm = 1) or seasonal (Nm = 3, 0 and 0). Gene flow is from a source populations with k = 85 distinct MHC alleles at a locus that is subject to balancing selection with S = 0.2, and the G'_{ST} was calculated by sampling the population at an 18 generation time interval.

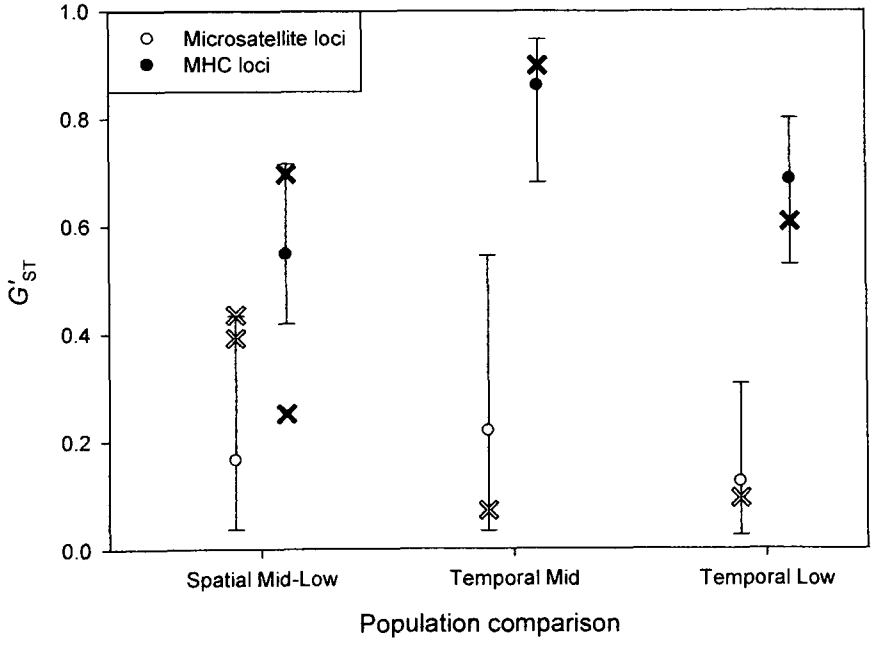


Figure 6. Simulated (mean G'_{ST} (5-95% CI)) and the mean observed genetic differentiation (G'_{ST}) (circles and crosses, respectively) between the Mid Naranjo [Mid] and Lower Aripo [Low] populations. Shown are the spatial genetic differentiation between Mid and Low (Spatial Mid-Low), as well as the temporal genetic differentiation within populations over time (2001-2007) (Temporal Mid and Temporal Low). Genetic differentiation is measured at microsatellite (open symbols) and MHC (solid symbols).

Discussion

The present study aimed to determine whether the interactions between demographic processes and evolutionary forces could explain the rapid turnover rate of MHC alleles in a metapopulation of guppies in Trinidad. Computer simulations were used to study temporal genetic differentiation (G'_{ST}), which quantifies the turnover rate of alleles. Simulations show that the temporal G'_{ST} increased with the coefficient of balancing selection (symmetric overdominance), the level of genetic polymorphism in the metapopulation, the migration rate, and with decreasing effective population size. When migration occurred seasonally in a population with fluctuating size, the temporal G'_{ST} increased, indicating that novel immigrant alleles replaced the resident alleles that were lost during the bottleneck.

Here I used symmetric overdominance and annual variation in population size to show that MHC alleles may be replaced over time. The replacement of MHC alleles relies on a property of alleles maintained through a rare-allele-advantage-type mechanism which is that they have a higher effective migration rate. Here symmetric overdominance was used which is arguably consistent with selection by parasites (Takahata and Nei 1990; but see De Boer et al. 2004) but is also consistent with modes of non-parasite selection such as ABC evolution (see van Oosterhout 2009a). Importantly though, this model does not require a change in parasite fauna, as suggested by the fluctuating selection model, to change MHC alleles rapidly over relatively short periods (i.e. 6 years or 18 generations of guppies). These results demonstrate that large temporal genetic differentiation of MHC genes is not necessarily evidence of Red Queen dynamics because it can be explained by the population genetic characteristics of this multigene family and the demography of metapopulations. This study furthermore shows that the combined impact of evolutionary forces and stochastic events can be different for genes under balancing selection than for neutral genes. It demonstrates, in particular, that the high level of temporal genetic divergence (G'_{ST}) of MHC compared to that of microsatellites (see Chapter 4) is consistent with the observed data and that this phenomenon does not require a parasite-mediated runaway process as in Red Queen dynamics.

Simulations were able to account for the level of spatial and temporal divergence at both neutral microsatellite and MHC loci (see Chapter 4). However, in order to attain the observed level of turnover in the Mid population, migration rate was higher than previous estimates (simulated values: Nm = 3.3 versus empirical values Nm = 0.13 and Nm = 0.288, see van Oosterhout et al. 2006b; and Barson et al. 2009, respectively). Effective population size estimates were also somewhat lower than previous estimates (simulated values: Ne = 81 versus empirical values Ne = 104 and Ne = 1090, see van Oosterhout et al. 2006b; and Barson et al. 2009, respectively). Demographic estimates with large effective population size or smaller migration rates were inconsistent with the empirical data and could either not explain the large genetic differentiation of the MHC or they failed to explain the relatively low genetic differentiation of the microsatellites. Estimates of effective population size and migration rate vary between van Oosterhout et al. (2006b), Barson et al. (2009) and the present study (see Chapter 2). This is because of (1) differences in the assumptions of the software used (IM (Hey and Nielsen 2004) and MIGRATE (Beerli and Felsenstein 1999, 2001)), (2) differences in the number of subpopulations analysed and (3) real differences in the demography of the system. The demographic scenario used in the simulation is conceivable given the biology of the system, and it provides an explanation for the high turnover rate of MHC alleles and the low rate of allelic turnover at the neutral microsatellite loci. This simulation also provides an

explanation for the high turnover rate of MHC alleles observed by Fraser et al. (2010b). Nevertheless, the large difference between the 2001 and 2007 samples in the observed spatial differentiation of the MHC is not fully accounted for in this simulation model. The simulated values (5-95% CI G'_{ST} = 0.419 – 0.716) do not encompass the observed values from 2001 (Mid-Up 2001 G'_{ST} = 0.251; Mid-Up 2007 G'_{ST} = 0.697). This suggests that other (not yet identified) sources of variance contribute that can either reduce or inflate the spatial genetic differentiation between populations for the MHC.

Results from the present study should not be interpreted as evidence that Red Queen dynamics does not occur, nor that the co-evolutionary arms race between parasite virulence genes and vertebrate immune genes is not important in the evolution of the MHC. The empirical evidence supporting host-parasite co-evolution is strong (e.g. Dionne et al. 2007), albeit often indirect. Relatively few studies have directly correlated the infection rate of hosts to their MHC genotype or to particular MHC alleles (Wegner et al. 2003b; Westerdahl et al. 2004; Bonneaud et al. 2006; Tollenaere et al. 2008; Oliver et al. 2009b; see also review by Spurgin and Richardson 2010). This shortage of studies simply illustrates the difficulty demonstrating the causal link between the immunological function of the MHC and the rich parasite biodiversity. Furthermore, host-parasite co-evolution may also differ fundamentally between systems. For example, in the minimum essential chicken MHC, causal links between parasites and MHC alleles were established, a feat probably facilitated by the relative simple genomic architecture of the MHC in chickens (Kaufman et al. 1999). Studies on sticklebacks show that an optimal rather than a maximum number of MHC alleles is associated to the lowest parasite load or highest lifetime reproductive success (Aeschlimann et al. 2003; Wegner et al. 2003a; Kalbe et al. 2009). Those results highlight that organisms may differ in the way selection operates on their immune genes. Nevertheless, where the MHC is polymorphic, balancing selection is likely to play a crucial role in maintaining this diversity. However, balancing selection should not be equated to parasite selection (or sexual selection), as other selective agents may act on the immune genes (van Oosterhout 2009a).

Recent theoretical work on the MHC (van Oosterhout 2009a) and the *S*-locus of plants (Uyenoyama 2005; Llaurens et al. 2009) demonstrated that these multigene families evolving under balancing selection are prone to the accumulation of recessive deleterious mutations. There are two characteristics that make these multigene families prone to this Muller's ratchet-like process (Muller 1932): (1) their high levels of heterozygosity and (2) areas with linkage disequilibria (van Oosterhout 2009a). Purifying selection acts against the most common MHC alleles, because compared to rare ones, high-frequency alleles are more likely to occur in homozygous state and express their recessive mutations. Selection thus balances the allele frequencies, which maintains the polymorphism. This model is called Associative Balancing Complex (ABC) evolution, and it proposes that balancing selection operates continuously, irrespective of parasites. In MHC research, parasite selection is often inferred in cases where the null model of neutrality is rejected (Langefors et al. 1998; Bryja et al. 2007; de Eyto et al. 2007). The present study demonstrates that large fluctuations in MHC allele frequencies are not necessarily indicative of changes in parasite selection pressures, and that such temporal variation is inherent to the population genetics in a metapopulation of a multigene family in which genes are under balancing selection.

General Discussion

Chapter 6

In a study of the population demography of guppies (*Poecilia reticulata*), I find that upstream populations of guppies have small population sizes. Gene flow is biased in the downstream direction but upstream gene flow is significantly larger than zero (Chapter 2). These small upstream populations show significant reduction in microsatellite allelic richness. However, MHC allelic richness does not decline with that of microsatellites. Instead, the MHC allelic richness remains the same across all three guppy populations (Chapter 3). This is particularly interesting because the smallest population, the Upper Naranjo (Up) was found to have a significantly reduced parasite fauna whilst maintaining MHC polymorphism despite the effects of random genetic drift (Chapter 3).

Spatio-temporal genetic differentiation of MHC and microsatellite loci was measured between the Mid Naranjo (Mid) and Lower Aripo (Low) guppy populations in 2001 and 2007. Genetic differentiation was consistent with negative frequency dependent selection or overdominance in 2001 and of fluctuating selection in 2007 (Chapter 4). This change in the mode of balancing selection prompted the use of a simulation model to account for both the 2001 and the 2007 results using overdominant selection (Chapter 5). Importantly, high temporal variation in MHC, but not microsatellite alleles, implicated gene flow within a metapopulation system (Chapter 4). With the combination of overdominant selection and a source-sink metapopulation, simulations were able to account for spatial genetic divergence previously proposed to be produced by fluctuating selection (Chapter 5).

Altogether, these results highlight firstly, that other sources of balancing selection should be considered in addition to parasite selection and secondly, that population demography can have a different impact on the population genetics of the MHC compared to that of neutral loci.

Spatial variation in microsatellites, MHC and parasite selection

Microsatellite variation in guppy populations within the Aripo River declines significantly from Low to Mid to Up. This progressive decline in genetic variation from the lowlands to the uplands is a long established observation in the guppy literature (Carvalho et al. 1991; Shaw et al. 1991; Shaw et al. 1994; Carvalho et al. 1996; Barson et al. 2009) and it is typical for the demography of a riverine environment where movement is largely unidirectional with the water flow (see Hanfling and Weetman 2006). The reduction in genetic diversity in the uplands may be a consequence of past colonisation events and the relatively small effective population size in this habitat. Moreover, the elevated level of genetic variation in the lowlands is a likely result of long term gene flow in the downstream direction from different tributaries, as well the exchange between tributaries of the Caroni River (Barson et al. 2009). Barson et al. (2009) found that upstream sites within the Caroni drainage were in mutation-drift equilibrium indicating that they were not subject to rapid population turnover. However, the authors found that the Mid population did show evidence of population decline. This is also consistent with evidence from a mark-recapture study conducted in the Mid population in 2006, where

the census population size of guppies declined markedly after a spate flushing event (van Oosterhout et al. 2007a, see Appendix 12). The findings by Barson et al. (2009) and van Oosterhout et al. (2007a) support the assumptions made in the temporal simulation study of the MHC that the population dynamics of guppies in the upland may contribute significantly to the microevolution and turnover of the MHC (see Chapter 5).

The level of MHC diversity within the Up, Mid and Low populations is however not different between sites, and consequently, cannot be explained by neutrality and demography alone (Chapter 3). Balancing selection through MHC recognition of parasite fauna is most frequently suggested as the mechanism by which MHC variation is maintained (see Bernatchez and Landry 2003; Piertney and Oliver 2006, and references therein). However, in the Up population we observe maintenance of MHC despite not finding any parasites in 2003 or 2006 (Chapter 3). In 2007, no digenean infections, no fungal infections and only six guppies that carried a gyrodactylid infection were found (Chapter 3). This infection incidence is significantly lower than any other gyrodactylid infection found in the Mid or Low population, in any year. Furthermore, Up guppies are exceptionally susceptible to (experimental) *Gyrodactylus* infections (Cable pers. comm.). This suggests that the small number of worms found on six guppies in 2007 may not have been virulent *G. turnbulli* or *G. bullatarudis* parasites, but rather a cross-infection from a new gyrodactylid genus from *Rivulus hartii* (Cable pers. comm.). This new taxon is currently being described by Cable et al. (in prep.). *Gyrodactylus* and digenean parasites represent the majority of the species diversity of parasites known to infect wild guppies (Cable In press). The combination of the high MHC diversity in the Up population and its significantly reduced parasite fauna suggests that MHC polymorphism can be maintained with little or no parasite selection. Alternatively, this study has overlooked a number of important microparasites that exercise strong selection on the MHC and which are absent in the Up but present in the Mid and Low populations.

Bacteria are an important constituent of the parasite fauna that were not taken into account in the present study. Bacterial diversity has been shown to co-vary with MHC diversity in Atlantic salmon (Dionne et al. 2007) and resistance and susceptibility associations have been made between bacteria and specific MHC alleles (Langefors et al. 2001; Lohm et al. 2002; Grimholt et al. 2003). However, analysis of the variation of the microbial fauna falls outside the confines of the present study. Indeed, this will be a huge challenge also for future studies, a viewpoint emphasised by the following quote by Eckburg et al. (2005): "*No complex microbial community in nature has been sampled to completion*" (Eckburg et al. 2005). The challenge lies in the fact that the microbial fauna may vary between sites, but also within a site, on different substrates and at different water depths. In addition to spatial variation, there are different levels of temporal variation, including annual, seasonal and diurnal variation in microbe biodiversity (see e.g. Shiah and Ducklow 1994; Cebron et al. 2004; Hullar et al. 2006; Winter et al. 2007). Moreover, bacterial diversity is not indicative of pathogenisity and phylogenetic units may not represent ecologically (biologically) significant units (see e.g. Acinas et al. 2004). It is superfluous to survey bacterial species without a priori knowledge of the taxonomic units that are pathogenic. In addition, bacterial pathogenisity may be opportunistic, commensal at some times but pathogenic at others (Tenaillon et al. 2010).

If we assume that (the un-sampled) bacteria and viruses are primarily responsible for the maintenance of polymorphism in the Up population, then we must also assume that gyrodactylid, digenean and fungal infections in the Mid and Low do not exercise strong

selection on the MHC. However, recent evidence suggests that this is not the case. For example, reduced gyrodactylid loads have been linked with a particular (a-type) MHC allele both in the wild and under laboratory conditions (Fraser and Neff 2009, 2010). Altogether, the different lines of evidence suggest that parasite selection alone cannot be responsible for the high level of MHC diversity in the Up population. Sexual selection and Associative Balancing Complex (ABC) evolution are therefore more likely to explain the observed level of MHC diversity in wild guppy populations.

Temporal variation in MHC polymorphism

Theory shows that balancing selection through overdominance selection and NFDS will reduce MHC genetic divergence (G'_{ST}) at a spatial scale (Takahata and Nei 1990; Schierup et al. 2000; Muirhead 2001). However, this thesis shows that balancing selection tends to increase temporal genetic variation, particularly in an open metapopulation structure because novel immigrant alleles are likely to replace resident alleles over time (see Table 1 in Spurgin and Richardson 2010, see also Chapter 5). In the Mid and Low populations, the level of MHC turnover between 2001 and 2007 is large with only 7 out of 33 (21.2%) MHC alleles being observed in both the 2001 and 2007 samples (Chapter 4). This pattern was not observed for the neutral microsatellite genes, which had a significantly lower rate of turnover with 56.5% of alleles represented in both datasets.

The computer simulation outlined in Chapter 5 showed that the combination of large fluctuations in census population size, balancing selection through overdominance and immigration from a large gene pool of MHC alleles could explain the observed level of MHC turnover. The same demographic scenario could also explain the population genetics of (neutral) microsatellite genes. The crucial difference between MHC and microsatellite genes is that balancing selection promotes the invasion of rare immigrant alleles. This effect is further augmented by rapid population expansion after a population crash. Without fluctuations in population size, immigrant alleles would enter a population that is already in selection-drift equilibrium. The probability that a novel (rare) immigrant allele is lost from the population by chance is considerably higher in a steady-state population than in a population that increases in size. Indeed, when a rare immigrant allele enters a population after a crash or bottleneck, overdominant (or NFDS) selection will help it to increase its frequency during subsequent population expansion.

A comparison of genetic differentiation between the Mid and Low guppy populations in 2001 and 2007, revealed differences in the relative divergence between microsatellite and MHC loci. In 2001, genetic differentiation was lower for MHC than microsatellite loci, a pattern indicative of balancing selection through overdominance (or NFDS) (Schierup 1998; Schierup et al. 2000; Muirhead 2001). In 2007, the pattern had reversed and MHC loci were more diverged than microsatellite loci. This pattern is indicative of fluctuating selection acting over a spatial scale (Hedrick et al. 1976; Hedrick 2002). Variation in the mode of selection has been demonstrated in previous temporal studies (e.g. Charbonnel and Pemberton 2005; Bryja et al. 2007; Fraser et al. 2010b). It is possible that a spatial differentiation in parasite selection has occurred in 2007 which is consistent with the parasite data.

Limitations and future directions

In the present study, the MHC class II*B* of guppies was quantified using an intensive cloning and sequencing procedure. Now such a study could be done with more individuals from more populations using second generation technology. Until recently, the time and cost these sorts of experiments were prohibitive. However, using a PCR tagging procedure combined with illumina sequencing of multiple individuals is both time and cost effective. Using this sort of method, individuals could be PCR'd for exon 2 of the MHC class II*B*, individually tagged and then pooled in an illumina run. This new technology also allows wild studies, similar to the one conducted here, to considerably increase the number of samples and sample sizes without the relative increase in investment of time and money.

The broad interpretations of the findings of the present study may be somewhat limited in their application to the wider MHC community. The present study demonstrates the maintenance of MHC variation in a parasite depauperate site and that temporal MHC variation may be driven without fluctuating selection. However, this study focused on a limited number of populations. The reason for the reduced number of populations is twofold. Firstly, the numbers of populations used in the present study was limited because the investment (cloning and sequencing) in individual guppies was high. The present study was the first to identify guppies with more than four MHC class II*B* alleles. This cloning effort reduced the number of samples but increased the reliability of estimates of population genetic differentiation at the MHC. The importance of estimating as accurately as possible the actual level of MHC variation within each population was paramount because one of the objectives was to observe reduced levels of polymorphism in a parasite depauperate site. The second reason that the number of populations used here was limited was that the present study focussed on the UN, the parasite depauperate site. This site and those downstream from it were sampled because they represent a rare opportunity to test the effects of parasite selection within a river system. Should additional upstream parasite depauperate sites be identified they would make excellent replicates for the work done in the present study.

Parasitism is implicated as a source of natural selection in the maintenance of polymorphism in the MHC (Piertney and Oliver 2006). MHC alleles have been identified for resistance (and susceptibility) to specific parasites in a number of systems (Langefors et al. 2001; Lohm et al. 2002; Penn et al. 2002; Trachtenberg et al. 2003; Harf and Sommer 2005; Fraser and Neff 2009). A number of these studies use host survival within a mesocosm, with parasites, in order to identify resistant and susceptible alleles (Langefors et al. 2001; Lohm et al. 2002; Fraser and Neff 2009), or the increased fitness of heterozygous individuals (Penn et al. 2002). While these studies are useful in indentifying parasites as a source of natural selection on the MHC, they do not aid our understanding of how host parasite dynamic coevolution maintains MHC polymorphism in the wild. This is because firstly, there is no history of host and parasite interaction in the mesocosm and secondly, other factors (sexual selection and ABC evolution) act in addition to parasite selection in the wild. Studies that observe correlations between MHC allele frequencies and parasites in the wild (e.g. Fraser and Neff 2010), must explore these relationships further in order to demonstrate causality (e.g. Fraser and Neff 2009). Spurious correlations between MHC allele and parasite frequencies may occur where MHC alleles are taken as biologically significant units where in fact they may be grouped into supertypes (see e.g. Trachtenberg et al. 2003; Schwensow et al. 2007). In the present study, parasites are grouped into biologically significant units based on species but this may also be

an oversimplification. The recognition capabilities of a single MHC allele may target: a particular group of parasite individuals within a species, all the individuals within a species, or groups of individuals within multiple species. Consequently, the evolution and maintenance of polymorphism at the MHC is likely due to selection from different groups that are difficult to quantify.

Mesocosm experiments over multiple generations allow controlled environments both with and without parasites. Guppies are an ideal model to set up multiple generation mesocosm experiments because of their short generation time (circa 3 months). Maintenance of MHC polymorphism through means other than parasite selection can be tested in a parasite free mesocosm. This, combined with a parallel experiment where guppies are part of one of three breeding regimes, inbred, outbred and a random breeding regime (see e.g. van Oosterhout et al. 2007b), allow the determination of the relative effects of sexual selection and ABC evolution. The predictions of this experiment are, that over time the mesocosm guppies would lose neutral genetic variation due to random genetic drift. However, any maintenance of MHC variation would be attributable to selection by ABC processes and/or sexual selection. There are two ways to delineate the combined effects of ABC and sexual selection. The first would be to sample multiple generations of offspring in the mesocosm. Neutral markers would be used to identify parentals, and the effect of sexual selection would be confirmed where females were shown to select males based on their MHC genotypes (for increased heterozygosity, optimality, or rare alleles). The effect of ABC evolution would be identified where MHC homozygotes are reduced compared to Mendelian expectations. A prediction of ABC evolution is that MHC homozygous individuals are less fit. Therefore, a finding that progenies had a heterozygote excess would indicate a role for ABC evolution. The second way to delineate ABC evolution and sexual selection would be to use the guppies in the breeding regime. In this experiment, guppies are denied the ability to select a mate and instead any maintenance of MHC variation above that of a neutral expectation is due to ABC evolution. The effects of a bottleneck on neutral genetic variation from the inbreeding regime would be predicted to be greater than the effect on genetic variation at the MHC. Outbreeding and random breeding regimes would be compared to the results from the mesocosm in order to better determine the mating strategy of the mesocosm females (random or outbreeding).

Simulations that incorporate demographic processes such as migration and fluctuations in census population size could be constructed. These simulations would also include the observed effects of sexual selection and ABC evolution on the maintenance of MHC variation. A comparison between the maintenance of MHC variation in the simulation and that observed in the wild would allow one to account for the extra source of selection contributed by parasites. Indeed, the confirmation that MHC polymorphism can be maintained without parasite selection would already be an advance in our understanding of the perpetual forces (sexual selection and ABC evolution) of balancing selection.

Conclusion

The research conducted in this thesis challenges two major paradigms in MHC research.

1. Based on my analysis of a population with unusually-low parasite diversity, is unlikely that parasite selection can solely explain the evolution of the MHC.

2. Red Queen Dynamics are often inferred from MHC-parasite data without a causal link. Here I show using both a computer simulation and empirical data that such Red Queen Dynamics of the MHC can also be explained by the interaction of the evolutionary forces, characteristics of highly polymorphic multigene families, and demographic factors typical of source-sink metapopulations.

Both conclusions warrant future studies on other selective forces acting on the MHC, in particular sexual selection and ABC evolution.

References

Acinas, S. G., V. Klepac-Ceraj, D. E. Hunt, C. Pharino, I. Ceraj, D. L. Distel, and M. F. Polz. 2004. Fine-scale phylogenetic architecture of a complex bacterial community. Nature 430:551-554.

Aeschlimann, P. B., M. A. Haberli, T. B. H. Reusch, T. Boehm, and M. Milinski. 2003. Female sticklebacks *Gasterosteus aculeatus* use self-reference to optimize MHC allele number during mate selection. Behavioral Ecology and Sociobiology 54:119-126.

Aguilar, A., G. Roemer, S. Debenham, M. Binns, D. Garcelon, and R. K. Wayne. 2004. High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. Proceedings of the National Academy of Sciences of the United States of America 101:3490-3494.

Alcaide, M., S. V. Edwards, L. Cadahia, and J. J. Negro. 2009. MHC class I genes of birds of prey: isolation, polymorphism and diversifying selection. Conservation Genetics 10:1349-1355.

Alexander, H. J., J. S. Taylor, S. S. T. Wu, and F. Breden. 2006. Parallel evolution and vicariance in the guppy (*Poecilia reticulata*) over multiple spatial and temporal scales. Evolution 60:2352-2369.

Anderson, E. C., and E. A. Thompson. 2002. A model-based method for identifying species hybrids using multilocus genetic data. Genetics 160:1217-1229.

Anisimova, M., R. Nielsen, and Z. Yang. 2003. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. Genetics 164:1229-1236.

Apanius, V., D. Penn, P. R. Slev, L. R. Ruff, and W. K. Potts. 1997. The nature of selection on the major histocompatibility complex. Critical Reviews in Immunology 17:179-224.

Archard, G. A., I. C. Cuthill, J. C. Partridge, and C. van Oosterhout. 2008. Female guppies (*Poecilia reticulata*) show no preference for conspecific chemosensory cues in the field or an artificial flow chamber. Behaviour 145:1329-1346.

Arkush, K. D., A. R. Giese, H. L. Mendonca, A. M. McBride, G. D. Marty, and P. W. Hedrick. 2002. Resistance to three pathogens in the endangered winter-run chinook salmon (*Oncorhynchus tshawytscha*): effects of inbreeding and major histocompatibility complex genotypes. Canadian Journal of Fisheries and Aquatic Sciences 59:966-975.

Artamonova, II, and M. S. Gelfand. 2007. Comparative genomics and evolution of alternative splicing: The pessimists' science. Chemical Reviews 107:3407-3430.

Babik, W., M. Pabijan, and J. Radwan. 2008. Contrasting patterns of variation in MHC loci in the Alpine newt. Molecular Ecology 17:2339-2355.

Bakke, T. A., J. Cable, and P. D. Harris. 2007. The biology of gyrodactylid monogeneans: The "Russian-doll killers". Advances in Parasitology 64:161-367.

Balakirev, E. S., and F. J. Ayala. 2003. Pseudogenes: Are they "Junk" or functional DNA? Annual Review of Genetics 37:123-151.

Barber, I., D. Hoare, and J. Krause. 2000. Effects of parasites on fish behaviour: a review and evolutionary perspective. Reviews in Fish Biology and Fisheries 10:131-165.

Barrick, J. E., D. S. Yu, S. H. Yoon, H. Jeong, T. K. Oh, D. Schneider, R. E. Lenski, and J. F. Kim. 2009. Genome evolution and adaptation in a long-term experiment with Escherichia coli. Nature 461:1243-1249.

Barson, N. J., J. Cable, and C. van Oosterhout. 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. Journal of Evolutionary Biology 22:485-497.

Bassar, R. D., M. C. Marshall, A. Lopez-Sepulcre, E. Zandona, S. K. Auer, J. Travis, C. M. Pringle, A. S. Flecker, S. A. Thomas, D. F. Fraser, and D. N. Reznick. 2010. Local adaptation in Trinidadian guppies alters ecosystem processes. PNAS 107:3616-3621.

Becher, S. A., S. T. Russell, and A. E. Magurran. 2002. Isolation and characterization of polymorphic microsatellites in the Trinidadian guppy (*Poecilia reticulata*). Molecular Ecology Notes 2:456-458.

Beerli, P. 2002. MIGRATE: documentation and program, part of LAMARC[Internet]. Available from: http://evolution.genetics. washington.edu/lamarc.html. Accessed 12 Feb 2010.

Beerli, P. 2004. Effect of unsampled populations on the estimation of population sizes and migration rates between sampled populations. Molecular Ecology 13:827-836.

Beerli, P., and J. Felsenstein. 1999. Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. Genetics 152:763-773.

Beerli, P., and J. Felsenstein. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. Proceedings of the National Academy of Sciences of the United States of America 98:4563-4568.

Bell, G. 1982. The Masterpiece of Nature: the Evolution and Genetics of Sexuality. University of California Press, Berkeley, Ca.

Bernatchez, L., and C. Landry. 2003. MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? Journal of Evolutionary Biology 16:363-377.

Birkhead, T. R., and T. Pizzari. 2002. Postcopulatory sexual selection. Nature Reviews Genetics 3:262-273.

Bjorkman, P. J., M. A. Saper, B. Samraoui, W. S. Bennett, J. L. Strominger, and D. C. Wiley. 1987a. The foreign antigen-binding site and T-cell recognition regions of class-I histocompatibility antigens. Nature 329:512-518.

Bjorkman, P. J., M. A. Saper, B. Samraoui, W. S. Bennett, J. L. Strominger, and D. C. Wiley. 1987b. Structure of the human class-I histocompatibility antigen, *HLA-A2*. Nature 329:506-512.

Blais, J., C. Rico, C. van Oosterhout, J. Cable, G. F. Turner, and L. Bernatchez. 2007. MHC adaptive divergence between closely related and sympatric african cichlids. Plos ONE 2:e734.

Bly, J. E., S. M. A. Quiniou, and L. W. Clem. 1997. Environmental effects on fish immune mechanisms. Fish Vaccinology 90:33-43.

Bollmer, J. L., F. H. Vargas, and P. G. Parker. 2007. Low MHC variation in the endangered Galapagos penguin (*Spheniscus mendiculus*). Immunogenetics 59:593-602.

Bonneaud, C., J. Perez-Tris, P. Federici, O. Chastel, and G. Sorci. 2006. Major histocompatibility alleles associated with local resistance to malaria in a passerine. Evolution 60:383-389.

Borghans, J. A. M., J. B. Beltman, and R. J. De Boer. 2004. MHC polymorphism under host-pathogen coevolution. Immunogenetics 55:732-739.

Boss, J. M., and P. E. Jensen. 2003. Transcriptional regulation of the MHC class II antigen presentation pathway. Current Opinion in Immunology 15:105-111.

Brown, J. H., T. S. Jardetzky, J. C. Gorga, L. J. Stern, R. G. Urban, J. L. Strominger, and D. C. Wiley. 1993. 3-Dimensional structure of the human class-II histocompatibility antigen *HLA-DR1*. Nature 364:33-39.

Bryja, J., N. Charbonnel, K. Berthier, M. Galan, and J. F. Cosson. 2007. Density-related changes in selection pattern for major histocompatibility complex genes in fluctuating populations of voles. Molecular Ecology 16:5084-5097.

Buckling, A., R. C. Maclean, M. A. Brockhurst, and N. Colegrave. 2009. The Beagle in a bottle. Nature 457:824-829.

Cable, J. In press. Poeciliid Parasites *in* J. Evans, A. Pilastro, and I. Schlupp, eds. Ecology and Evolution of Livebearing Fishes (Poeciliidae). Chicago University Press.

Cable, J., and P. D. Harris. 2002. Gyrodactylid developmental biology: historical review, current status and future trends. International Journal for Parasitology 32:255-280.

Cable, J., and C. van Oosterhout. 2007a. The impact of parasites on the life history evolution of guppies (Poecilia reticulata): The effects of host size on parasite virulence. International Journal for Parasitology 37:1449-1458.

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

Carius, H. J., T. J. Little, and D. Ebert. 2001. Genetic variation in a host-parasite association: Potential for coevolution and frequency-dependent selection. Evolution 55:1136-1145.

Carvalho, G. R., P. W. Shaw, L. Hauser, B. H. Seghers, and A. E. Magurran. 1996. Artificial introductions, evolutionary change and population differentiation in Trinidadian guppies (*Poecilia reticulata*:Poeciliidae). Biological Journal of the Linnean Society 57:219-234.

Carvalho, G. R., P. W. Shaw, A. E. Magurran, and B. H. Seghers. 1991. Marked genetic divergence revealed by allozymes among populations of the guppy *Poecilia reticulata* (Poeciliidae), in Trinidad. Biological Journal of the Linnean Society 42:389-405.

Cebron, A., M. Coci, J. Garnier, and H. J. Laanbroek. 2004. Denaturing gradient gel electrophoretic analysis of ammonia-oxidizing bacterial community structure in the lower Seine River: Impact of Paris wastewater effluents. Applied and Environmental Microbiology 70:6726-6737.

Charbonnel, N., and J. Pemberton. 2005. A long-term genetic survey of an ungulate population reveals balancing selection acting on MHC through spatial and temporal fluctuations in selection. Heredity 95:377-388.

Charlesworth, B. 2009. Effective population size and patterns of molecular evolution and variation. Nature Reviews Genetics 10:195-205.

Clarke, B., and D. R. S. Kirby. 1966. Maintenance of histocompatibility complex polymorphisms. Nature 211:999-1000.

Cohen, S. 2002. Strong positive selection and habitat-specific amino acid substitution patterns in Mhc from an estuarine fish under intense pollution stress. Molecular Biology and Evolution 19:1870-1880.

Colwell RK, Coddington JA (1994) Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society B-Biological Sciences* **345**, 101-118.

Comeron, J. M., A. Williford, and R. M. Kliman. 2008. The Hill-Robertson effect: evolutionary consequences of weak selection and linkage in finite populations. Heredity 100:19-31.

Crispo, E., P. Bentzen, D. N. Reznick, M. T. Kinnison, and A. P. Hendry. 2006. The relative influence of natural selection and geography on gene flow in guppies. Molecular Ecology 15:49-62.

Crow, J. F., and M. Kimura. 1970. An introduction to population genetics theory. Harper & Row Publishers, New York.

Crow, R. T., and N. R. Liley. 1979. Sexual pheromone in the guppy, *Poecilia-reticulata* (PETERS). Canadian Journal of Zoology-Revue Canadienne De Zoologie 57:184-188.

Cummings, S. M., M. McMullan, D. A. Joyce, and C. van Oosterhout. 2010. Solutions for PCR, cloning and sequencing errors in population genetic analysis. Conservation Genetics 11:1095-1097.

Danchin, E., V. Vitiello, A. Vienne, O. Richard, P. Gouret, M. F. McDermott, and P. Pontarotti. 2004. The major histocompatibility complex origin. Immunological Reviews 198:216-232.

Darwin, C. 1859. On The Origin of Species. John Murray, London.

Dausset, J. 1958. Iso-leuko-antibodies. Acta Haematol 20:156-166.

De Boer, R. J., J. A. M. Borghans, M. van Boven, C. Kesmir, and F. J. Weissing. 2004. Heterozygote advantage fails to explain the high degree of polymorphism of the MHC. Immunogenetics 55:725-731.

de Eyto, E., P. McGinnity, S. Consuegra, J. Coughlan, J. Tufto, K. Farrell, H. J. Megens, W. Jordan, T. Cross, and R. J. M. Stet. 2007. Natural selection acts on Atlantic salmon

major histocompatibility (MH) variability in the wild. Proceedings of the Royal Society B-Biological Sciences 274:861-869.

Decaestecker, E., S. Gaba, J. A. M. Raeymaekers, R. Stoks, L. Van Kerckhoven, D. Ebert, and L. De Meester. 2007. Host-parasite 'Red Queen' dynamics archived in pond sediment. Nature 450:870-U816.

Dionne, M., K. M. Miller, J. J. Dodson, F. Caron, and L. Bernatchez. 2007. Clinal variation in MHC diversity with temperature: evidence for the role of host-pathogen interaction on local adaptation in Atlantic salmon. Evolution 61:2154-2164.

Dixon, B. 2008. Rene Stet's impact on the study of teleost major histocompatibility genes: evolution from loci to populations. Immunogenetics 60:77-82.

Doherty, P. C., and R. M. Zinkernagel. 1975a. A biological role for the major histocompatibility complex antigens. Lancet 305:1406-1409.

Doherty, P. C., and R. M. Zinkernagel. 1975b. Enhanced immunological surveillance in mice heterozygous at the *H-2* gene complex. Nature 256:50-52.

Dorak, M. T., T. Lawson, H. K. G. Machulla, K. I. Mills, and A. K. Burnett. 2002. Increased heterozygosity for MHC class II lineages in newborn males. Genes and Immunity 3:263-269.

Dybdahl, M. F., and C. M. Lively. 1998. Host-parasite coevolution: Evidence for rare advantage and time-lagged selection in a natural population. Evolution 52:1057-1066.

Earl, D. A. 2011. Structure harvester v0.6.1. Available
at http://taylor0.biology.ucla.edu/struct_harvest/.

Eckburg, P. B., E. M. Bik, C. N. Bernstein, E. Purdom, L. Dethlefsen, M. Sargent, S. R. Gill, K. E. Nelson, and D. A. Relman. 2005. Diversity of the human intestinal microbial flora. Science 308:1635-1638.

Edwards, S. V., and P. W. Hedrick. 1998. Evolution and ecology of MHC molecules: from genomics to sexual selection. Trends in Ecology & Evolution 13:305-311.

Ekblom, R., S. A. Saether, P. Jacobsson, P. Fiske, T. Sahlman, M. Grahn, J. A. Kalas, and J. Hoglund. 2007. Spatial pattern of MHC class II variation in the great snipe (Gallinago media). Molecular Ecology 16:1439-1451.

Ellegren, H. 2000. Microsatellite mutations in the germline: implications for evolutionary inference. Trends in Genetics 16:551-558.

Ellis, S. A., R. E. Bontrop, D. F. Antczak, K. Ballingall, C. J. Davies, J. Kaufman, L. J. Kennedy, J. Robinson, D. M. Smith, M. J. Stear, R. J. M. Stet, M. J. Waller, L. Walter, and S. G. E. Marsh. 2006. ISAG/IUIS-VIC Comparative MHC Nomenclature Committee report, 2005. Immunogenetics 57:953-958.

Endler, J. A. 1980. Natural-selection on color patterns in *Poecilia reticulata*. Evolution 34:76-91.

Endler, J. A. 1995. Multiple-trait coevolution and environmental gradients in guppies. Trends in Ecology & Evolution 10:22-29.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14:2611-2620.

Fajen, A., and F. Breden. 1992. Mitochondrial-DNA Sequence variation among natural-populations of the Trinidad guppy, *Poecilia-reticulata*. Evolution 46:1457-1465.

Figueroa, F., W. E. Mayer, A. Sato, Z. Zaleska-Rutczynska, B. Hess, H. Tichy, and J. Klein. 2001. MHC class I genes of swordtail fishes, Xiphophorus: variation in the number of loci and existence of ancient gene families. Immunogenetics 53:695-708.

Figueroa, F., W. E. Mayer, H. Sultmann, C. O'hUigin, H. Tichy, Y. Satta, N. Takezaki, N. Takahata, and J. Klein. 2000. MHC class IIB gene evolution in East African cichlid fishes. Immunogenetics 51:556-575.

Flajnik, M. F., Y. Ohta, C. Namikawa-Yamada, and M. Nonaka. 1999. Insight into the primordial MHC from studies in ectothermic vertebrates. Immunological Reviews 167:59-67.

Ford, M. J. 2002. Applications of selective neutrality tests to molecular ecology. Molecular Ecology 11:1245-1262.

Fraser, B. A., and B. D. Neff. 2009. MHC class IIB additive and non-additive effects on fitness measures in the guppy Poecilia reticulata. Journal of Fish Biology 75:2299-2312.

Fraser, B. A., and B. D. Neff. 2010. Parasite mediated homogenizing selection at the MHC in guppies. Genetica 138:273-278.

Fraser, B. A., I. W. Ramnarine, and B. D. Neff. 2010a. Selection at the MHC class IIB locus across guppy (Poecilia reticulata) populations. Heredity 104:155-167.

Fraser, B. A., I. W. Ramnarine, and B. D. Neff. 2010b. Temporal variation at the MHC class IIB in wild populations of the guppy (*Poecilia reticulata*). Evolution 64:2086-2096.

Froeschke, G., and S. Sommer. 2005. MHC class II *DRB* variability and parasite load in the striped mouse (*Rhabdomys pumilio*) in the southern Kalahari. Molecular Biology and Evolution 22:1254-1259.

Galindo, C. L., L. J. McIver, J. F. McCormick, M. A. Skinner, Y. Xie, R. A. Gelhausen, K. Ng, N. M. Kumar, and H. R. Garner. 2009. Global Microsatellite Content Distinguishes Humans, Primates, Animals, and Plants. Molecular Biology and Evolution 26:2809-2819.

Garrigan, D., and P. W. Hedrick. 2003. Perspective: Detecting adaptive molecular polymorphism: Lessons from the MHC. Evolution 57:1707-1722.

Gomez, D., P. Conejeros, S. H. Marshall, and S. Consuegra. 2010. MHC evolution in three salmonid species: a comparison between class II alpha and beta genes. Immunogenetics 62:531-542.

Goudet, J. 1995. FSTAT (Version 1.2): A computer program to calculate F-statistics. Journal of Heredity 86:485-486.

Goudet, J. 2001. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from http://www.unil.ch/izea/softwares/fstat.html. Updated from Goudet (1995).

Gregersen, J. W., K. R. Kranc, X. Y. Ke, P. Svendsen, L. S. Madsen, A. R. Thomsen, L. R. Cardon, J. I. Bell, and L. Fugger. 2006. Functional epistasis on a common MHC haplotype associated with multiple sclerosis. Nature 443:574-577.

Griffiths, S. W., and A. E. Magurran. 1999. Schooling decisions in guppies (*Poecilia reticulata*) are based on familiarity rather than kin recognition by phenotype matching. Behavioral Ecology and Sociobiology 45:437-443.

Grimholt, U., S. Larsen, R. Nordmo, P. Midtlyng, S. Kjoeglum, A. Storset, S. Saebo, and R. J. M. Stet. 2003. MHC polymorphism and disease resistance in Atlantic salmon (*Salmo salar*); facing pathogens with single expressed major histocompatibility class I and class II loci. Immunogenetics 55:210-219.

Gu, X., and M. Nei. 1999. Locus specificity of polymorphic alleles and evolution by a birth-and-death process in mammalian MHC genes. Molecular Biology and Evolution 16:147-156.

Haddrill, P. R., D. L. Halligan, D. Tomaras, and B. Charlesworth. 2007. Reduced efficacy of selection in regions of the Drosophila genome that lack crossing over. Genome Biology 8:R18.

Hale, K. A., and J. V. Briskie. 2007. Decreased immunocompetence in a severely bottlenecked population of an endemic New Zealand bird. Animal Conservation 10:2-10.

Hanfling, B., and D. Weetman. 2006. Concordant genetic estimators of migration reveal anthropogenically enhanced source-sink population structure in the River Sculpin, Cottus gobio. Genetics 173:1487-1501.

Hansen, M. M., O. Skaala, L. F. Jensen, D. Bekkevold, and K. L. D. Mensberg. 2007. Gene flow, effective population size and selection at major histocompatibility complex genes: brown trout in the Hardanger Fjord, Norway. Molecular Ecology 16:1413-1425.

Harf, R., and S. Sommer. 2005. Association between major histocompatibility complex class II *DRB* alleles and parasite load in the hairy-footed gerbil, *Gerbillurus paeba*, in the southern Kalahari. Molecular Ecology 14:85-91.

Hedrick, P. W. 2002. Pathogen resistance and genetic variation at MHC loci. Evolution 56:1902-1908.

Hedrick, P. W. 2005a. Genetics of Populations. Jones & Bartlett Publishers, Sudbury, MA, USA.

Hedrick, P. W. 2005b. A standardized genetic differentiation measure. Evolution 59:1633-1638.
Hedrick, P. W. 2011. Genetics of Populations. Jones & Bartlett Publishers, Sadbury, MA, USA.
Hedrick, P. W., M. E. Ginevan, and E. P. Ewing. 1976. Genetic polymorphism in heterogeneous environments. Annual Review of Ecology and Systematics 7:1-32.
Hedrick, P. W., T. J. Kim, and K. M. Parker. 2001a. Parasite resistance and genetic variation in the endangered Gila topminnow. Animal Conservation 4:103-109.
Hedrick, P. W., and K. M. Parker. 1998. MHC variation in the endangered Gila topminnow. Evolution 52:194-199.
Hedrick, P. W., K. M. Parker, and R. N. Lee. 2001b. Using microsatellite and MHC variation to identify species, ESUs, and MUs in the endangered Sonoran topminnow. Molecular Ecology 10:1399-1412.
Hess, C. M., and S. V. Edwards. 2002. The evolution of the major histocompatibility complex in birds. Bioscience 52:423-431.
Hess, C. M., Z. S. Wang, and S. V. Edwards. 2007. Evolutionary genetics of Carpodacus mexicanus, a recently colonized host of a bacterial pathogen, Mycoplasma gallisepticum. Genetica 129:217-225.
Hey, J., and R. Nielsen. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D-persimilis*. Genetics 167:747-760.
Hohenlohe, P. A., S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson, and W. A. Cresko. 2010. Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. Plos Genetics 6:23.
Houde, A. E. 1997. Sex, Color and Mate Choice in Guppies. Princeton University Press, Princeton, NJ.
Houde, A. E., and A. J. Torio. 1992. Effect of Parasitic Infection on Male Color Pattern and Female Choice in Guppies. Behavioral Ecology 3:346-351.
Hubisz, M. J., D. Falush, M. Stephens, and J. K. Pritchard. 2009. Inferring weak population structure with the assistance of sample group information. Molecular Ecology Resources 9:1322-1332.
Hughes, A. L. 2002. Natural selection and the diversification of vertebrate immune effectors. Immunological Reviews 190:161-168.
Hughes, A. L., and M. Yeager. 1998. Natural selection at major histocompatibility complex loci of vertebrates. Annual Review of Genetics 32:415-435.
Hullar, M. A. J., L. A. Kaplan, and D. A. Stahl. 2006. Recurring seasonal dynamics of microbial communities in stream habitats. Applied and Environmental Microbiology 72:713-722.
Hurst, L. D. 2009. Genetics and the understanding of selection. Nature Reviews Genetics 10:83-93.
Jameson, S. C., K. A. Hogquist, and M. J. Bevan. 1994. Specificity and flexibility in thymic selection. Nature 369:750-752.
Jensen, L. F., M. M. Hansen, K. L. Mensberg, and V. Loeschcke. 2008. Spatially and temporally fluctuating selection at non-MHC immune genes: evidence from TAP polymorphism in populations of brown trout (*Salmo trutta*, L.). Heredity 100:79-91.
Jordan, M. S., A. Boesteanu, A. J. Reed, A. L. Petrone, A. E. Holenbeck, M. A. Lerman, A. Naji, and A. J. Caton. 2001. Thymic selection of CD4(+)CD25(+) regulatory T cells induced by an agonist self-peptide. Nature Immunology 2:301-306.
Jorde, L. B., A. R. Rogers, M. Bamshad, W. S. Watkins, P. Krakowiak, S. Sung, J. Kere, and H. C. Harpending. 1997. Microsatellite diversity and the demographic history of modern humans. Proceedings of the National Academy of Sciences of the United States of America 94:3100-3103.
Kalbe, M., C. Eizaguirre, I. Dankert, T. B. H. Reusch, R. D. Sommerfeld, K. M. Wegner, and M. Milinski. 2009. Lifetime reproductive success is maximized with optimal major histocompatibility complex diversity. Proceedings of the Royal Society B-Biological Sciences 276:925-934.

Kalinowski, S. T. 2009. How well do evolutionary trees describe genetic relationships among populations? Heredity 102:506-513.

Kalinowski, S. T. 2011. The computer program STRUCTURE does not reliably identify the main genetic clusters within species: simulations and implications for human population structure. Heredity 106:625-632.

Kanagawa, T. 2003. Bias and artifacts in multitemplate polymerase chain reactions (PCR). Journal of Bioscience and Bioengineering 96:317-323.

Kaufman, J., S. Milne, T. W. F. Gobel, B. A. Walker, J. P. Jacob, C. Auffray, R. Zoorob, and S. Beck. 1999. The chicken *B* locus is a minimal essential major histocompatibility complex. Nature 401:923-925.

Kavaliers, M., E. Choleris, A. Agmo, and D. W. Pfaff. 2004. Olfactory-mediated parasite recognition and avoidance: linking genes to behavior. Hormones and Behavior 46:272-283.

Kelley, J., L. Walter, and J. Trowsdale. 2005. Comparative genomics of major histocompatibility complexes. Immunogenetics 56:683-695.

Keohavong, P., and W. G. Thilly. 1989. Fidelity of DNA-polumerases in DNA amplification. Proceedings of the National Academy of Sciences of the United States of America 86:9253-9257.

Kimura, M., T. Maruyama, and J. F. Crow. 1963. The mutation load in small populations. Genetics 48:1303-1312.

Klein, J., R. E. Bontrop, R. L. Dawkins, H. A. Erlich, U. B. Gyllensten, E. R. Heise, P. P. Jones, P. Parham, E. K. Wakeland, and D. I. Watkins. 1990. Nomenclature for the major histocompatibility complexes of different species - a proposal. Immunogenetics 31:217-219.

Klein, J., H. Ono, D. Klein, and C. O'hUigin. 1993. The accordion model of Mhc evolution. Progress in Immunology 8:137-143.

Klein, J., and A. Sato. 1998. Birth of the major histocompatibility complex. Scandinavian Journal of Immunology 47:199-209.

Klein, J., A. Sato, and W. E. Mayer. 2000. Jaws and AIS. Pp. 3-26 *in* M. Kasahara, ed. Major Histocompatibility Complex: Evolution, Structure and Function. Springer, Tokyo.

Klein, J., A. Sato, S. Nagl, and C. O'hUigin. 1998. Molecular trans-species polymorphism. Annual Review of Ecology and Systematics 29:1-21.

Kobayashi, N., K. Tamura, and T. Aotsuka. 1999. PCR error and molecular population genetics. Biochemical Genetics 37:317-321.

Kojima, K. I. 1971. Is there a constant fitness value for a given genotype? No! Evolution 25:281-285.

Lafferty, K. D., and A. K. Morris. 1996. Altered behavior of parasitized killifish increases susceptibility to predation by bird final hosts. Ecology 77:1390-1397.

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

Landry, C., D. Garant, P. Duchesne, and L. Bernatchez. 2001. 'Good genes as heterozygosity': the major histocompatibility complex and mate choice in Atlantic salmon (*Salmo salar*). Proceedings of the Royal Society of London Series B-Biological Sciences 268:1279-1285.

Langefors, A., J. Lohm, M. Grahn, O. Andersen, and T. von Schantz. 2001. Association between major histocompatibility complex class IIB alleles and resistance to Aeromonas salmonicida in Atlantic salmon. Proceedings of the Royal Society of London Series B-Biological Sciences 268:479-485.

Langefors, A., J. Lohm, T. von Schantz, and M. Grahn. 2000. Screening of Mhc variation in Atlantic salmon (*Salmo salar*): a comparison of restriction fragment length polymorphism (RFLP), denaturing gradient gel electrophoresis (DGGE) and sequencing. Molecular Ecology 9:215-219.

Langefors, A., T. Von Schantz, and B. Widegren. 1998. Allelic variation of Mhc class II in Atlantic salmon; a population genetic analysis. Heredity 80:568-575.

Librado, P., and J. Rozas. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451-1452.

Lindholm, A. K., F. Breden, H. J. Alexander, W. K. Chan, S. G. Thakurta, and R. Brooks. 2005. Invasion success and genetic diversity of introduced populations of guppies *Poecilia reticulata* in Australia. Molecular Ecology 14:3671-3682.

Lively, C. M., and M. F. Dybdahl. 2000. Parasite adaptation to locally common host genotypes. Nature 405:679-681.

Llaurens, V., L. Gonthier, and S. Billiard. 2009. The Sheltered Genetic Load Linked to the S Locus in Plants: New Insights From Theoretical and Empirical Approaches in Sporophytic Self-Incompatibility. Genetics 183:1105-1118.

Lohm, J., M. Grahn, A. Langefors, O. Andersen, A. Storset, and T. von Schantz. 2002. Experimental evidence for major histocompatibility complex-allele-specific resistance to a bacterial infection. Proceedings of the Royal Society of London Series B-Biological Sciences 269:2029-2033.

Long, K. D., and G. Rosenqvist. 1998. Changes in male guppy courting distance in response to a fluctuating light environment. Behavioral Ecology and Sociobiology 44:77-83.

Lopez, S. 1998. Acquired resistance affects male sexual display and female choice in guppies. Proceedings of the Royal Society of London Series B-Biological Sciences 265:717-723.

Lopez, S. 1999. Parasitized female guppies do not prefer showy males. Animal Behaviour 57:1129-1134.

Lukas, D., B. J. Bradley, A. M. Nsubuga, D. Doran-Sheehy, M. M. Robbins, and L. Vigilant. 2004. Major histocompatibility complex and microsatellite variation in two populations of wild gorillas. Molecular Ecology 13:3389-3402.

Lukas, D., and L. Vigilant. 2005. Reply: Facts, faeces and setting standards for the study of MHC genes using noninvasive samples. Molecular Ecology 14:1601-1602.

Magurran, A. E. 2005. Evolutionary ecology: the Trinidadian guppy. Oxford University Press, Oxford.

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

Martin, C. H., and S. Johnsen. 2007. A field test of the Hamilton-Zuk hypothesis in the Trinidadian guppy (Poecilia reticulata). Behavioral Ecology and Sociobiology 61:1897-1909.

McConnell, T. J., U. B. Godwin, and B. J. Cuthbertson. 1998a. Expressed major histocompatibility complex class II loci in fishes. Immunological Reviews 166:294-300.

McConnell, T. J., U. B. Godwin, S. F. Norton, R. S. Nairn, S. Kazianis, and D. C. Morizot. 1998b. Identification and mapping of two divergent, unlinked major histocompatibility complex class II B genes in Xiphophorus fishes. Genetics 149:1921-1934.

McLennan, D. A. 2003. The importance of olfactory signals in the gasterosteid mating system: sticklebacks go multimodal. Biological Journal of the Linnean Society 80:555-572.

McMullan, M., and C. van Oosterhout. in press. Evolutionary genetics and molecular ecology of the Major Histocompatibility Complex (MHC) genes of poeciliids *in* J. Evans, A. Pilastro, and I. Schlupp, eds. Ecology and Evolution of Livebearing Fishes (Poeciliidae). Chicago University Press.

Meyer, D., and G. Thomson. 2001. How selection shapes variation of the human major histocompatibility complex: a review. Annals of Human Genetics 65:1-26.

Meyer, J. H., and N. R. Liley. 1982. The control of production of a sexual pheromone in the female guppy, *Poecilia-reticulata*. Canadian Journal of Zoology-Revue Canadienne De Zoologie 60:1505-1510.

Milinski, M. 2003. The function of mate choice in sticklebacks: optimizing Mhc genetics. Journal of Fish Biology 63:1-16.

Milinski, M. 2006. The major histocompatibility complex, sexual selection, and mate choice. Annual Review of Ecology Evolution and Systematics 37:159-186.

Milinski, M., S. Griffiths, K. M. Wegner, T. B. H. Reusch, A. Haas-Assenbaum, and T. Boehm. 2005. Mate choice decisions of stickleback females predictably modified by MHC peptide ligands. Proceedings of the National Academy of Sciences of the United States of America 102:4414-4418.

Miller, H. C., and D. M. Lambert. 2004. Genetic drift outweighs balancing selection in shaping post-bottleneck major histocompatibility complex variation in New Zealand robins (Petroicidae). Molecular Ecology 13:3709-3721.

Miller, K. M., K. H. Kaukinen, T. D. Beacham, and R. E. Withler. 2001. Geographic heterogeneity in natural selection on an MHC locus in sockeye salmon. Genetica 111:237-257.

Modrek, B., A. Resch, C. Grasso, and C. Lee. 2001. Genome-wide detection of alternative splicing in expressed sequences of human genes. Nucleic Acids Research 29:2850-2859.

Mona, S., B. Crestanello, S. Bankhead-Dronnet, E. Pecchioli, S. Ingrosso, S. D'Amelio, L. Rossi, P. G. Meneguz, and G. Bertorelle. 2008. Disentangling the effects of recombination, selection, and demography on the genetic variation at a major histocompatibility complex class II gene in the alpine chamois. Molecular Ecology 17:4053-4067.

Muirhead, C. A. 2001. Consequences of population structure on genes under balancing selection. Evolution 55:1532-1541.

Muller, H. J. 1932. Some Genetic Aspects of Sex. American Naturalist 66:118-138.

Neff, B. D., J. W. Garner, and D. D. Health. 2008. The MHC and non-random mating in a captive population of Chinook salmon. Heredity 101:175-185.

Nei, M. 1987. Molecular Evolutionary Genetics. Columbia University Press, New York.

Nei, M. 2005. Selectionism and neutralism in molecular evolution. Molecular Biology and Evolution 22:2318-2342.

Nei, M., and T. Gojobori. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Molecular Biology and Evolution 3:418-426.

Nei, M., X. Gu, and T. Sitnikova. 1997. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. Proceedings of the National Academy of Sciences of the United States of America 94:7799-7806.

Nei, M., and A. P. Rooney. 2005. Concerted and birth-and-death evolution of multigene families. Annual Review of Genetics 39:121-152.

Nielsen, R. 2005. Molecular signatures of natural selection. Annual Review of Genetics 39:197-218.

Nowak, M. A., K. Tarczyhornoch, and J. M. Austyn. 1992. The optimal number of major histocompatibility complex-molecules in an individual. Proceedings of the National Academy of Sciences of the United States of America 89:10896-10899.

Ohta, Y., K. Okamura, E. C. McKinney, S. Bartl, K. Hashimoto, and M. F. Flajnik. 2000. Primitive synteny of vertebrate major histocompatibility complex class I and class II genes. Proceedings of the National Academy of Sciences of the United States of America 97:4712-4717.

Oliver, M. K., X. Lambin, T. Cornulier, and S. B. Piertney. 2009a. Spatio-temporal variation in the strength and mode of selection acting on major histocompatibility complex diversity in water vole (Arvicola terrestris) metapopulations. Molecular Ecology 18:80-92.

Oliver, M. K., S. Telfer, and S. B. Piertney. 2009b. Major histocompatibility complex (MHC) heterozygote superiority to natural multi-parasite infections in the water vole (Arvicola terrestris). Proceedings of the Royal Society B-Biological Sciences 276:1119-1128.

Paterson, S., T. Vogwill, A. Buckling, R. Benmayor, A. J. Spiers, N. R. Thomson, M. Quail, F. Smith, D. Walker, B. Libberton, A. Fenton, N. Hall, and M. A. Brockhurst. 2010. Antagonistic coevolution accelerates molecular evolution. Nature 464:275-278.

Penn, D. J., K. Damjanovich, and W. K. Potts. 2002. MHC heterozygosity confers a selective advantage against multiple-strain infections. Proceedings of the National Academy of Sciences of the United States of America 99:11260-11264.

Penn, D. J., and W. K. Potts. 1999. The evolution of mating preferences and major histocompatibility complex genes. American Naturalist 153:145-164.

Piertney, S. B., and M. K. Oliver. 2006. The evolutionary ecology of the major histocompatibility complex. Heredity 96:7-21.

Posada, D., and K. A. Crandall. 2002. The Effect of Recombination on the Accuracy of Phylogeny Estimation. Journal of Molecular Evolution 54:396-402.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945-959.

Reusch, T. B. H., M. A. Haberli, P. B. Aeschlimann, and M. Milinski. 2001. Female sticklebacks count alleles in a strategy of sexual selection explaining MHC polymorphism. Nature 414:300-302.

Reusch, T. B. H., and A. Langefors. 2005. Inter- and intralocus recombination drive MHC class IIB gene diversification in a teleost, the three-spined stickleback *Gasterosteus aculeatus*. Journal of Molecular Evolution 61:531-541.

Reusch, T. B. H., H. Schaschl, and K. M. Wegner. 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.

Reznick, D. N., and H. Bryga. 1987. Life-history evolution in guppies (*Poecilia-reticulata*) .1. Phenotypic and genetic changes in an introduction experiment. Evolution 41:1370-1385.

Reznick, D. N., and J. A. Endler. 1982. The impact of predation on life-history evolution in Trinidadian Guppies (*Poecilia reticulata*). Evolution 36:160-177.

Reznick, D. N., F. H. Shaw, F. H. Rodd, and R. G. Shaw. 1997. Evaluation of the rate of evolution in natural populations of guppies (*Poecilia reticulata*). Science 275:1934-1937.

Robinson, J., M. J. Waller, P. Parham, N. de Groot, R. Bontrop, L. J. Kennedy, P. Stoehr, and S. G. E. Marsh. 2003. IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. Nucleic Acids Research 31:311-314.

Roney, K. E., B. J. Cuthbertson, U. B. Godwin, S. Kazianis, L. Della Coletta, G. G. Rosenthal, M. J. Ryan, M. Schmidt, and T. J. McConnell. 2004. Alternative splicing of major histocompatibility complex class II *DXB* transcripts in Xiphophorus fishes. Immunogenetics 56:462-466.

Saitou, N., and M. Nei. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4:406-425.

Salomonsen, J., D. Marston, D. Avila, N. Bumstead, B. Johansson, H. Juul-Madsen, G. D. Olesen, P. Riegert, K. Skjodt, O. Vainio, M. V. Wiles, and J. Kaufman. 2003. The properties of the single chicken MHC classical class II alpha chain (*B-LA*) gene indicate an ancient origin for the *DR/E*-like isotype of class II molecules. Immunogenetics 55:605-614.

Sambrook, Y. G., R. Russell, Y. Umrania, Y. K. J. Edwards, R. D. Campbell, G. Elgar, and M. S. Clark. 2002. Fugu orthologues of human major histocompatibility complex genes: a genome survey. Immunogenetics 54:367-380.

Sato, A., F. Figueroa, B. W. Murray, E. Malaga-Trillo, Z. Zaleska-Rutczynska, H. Sultmann, S. Toyosawa, C. Wedekind, N. Steck, and J. Klein. 2000. Nonlinkage of major histocompatibility complex class I and class II loci in bony fishes. Immunogenetics 51:108-116.

Sato, A., F. Figueroa, C. OhUigin, D. N. Reznick, and J. Klein. 1995. Identification of major histocompatibility complex genes in the guppy, *Poecilia reticulata*. Immunogenetics 43:38-49.

Schaschl, H., M. Tobler, M. Plath, D. J. Penn, and I. Schlupp. 2008. Polymorphic MHC loci in an asexual fish, the Amazon molly (*Poecilia formosa*; Poeciliidae). Molecular Ecology 17:5220-5230.

Schierup, M. H. 1998. The number of self-incompatibility alleles in a finite, subdivided population. Genetics 149:1153-1162.

Schierup, M. H., X. Vekemans, and D. Charlesworth. 2000. The effect of subdivision on variation at multi-allelic loci under balancing selection. Genetical Research 76:51-62.

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

Scott, M. E. 1982. Reproductive potential of *Gyrodactylus bullatarudis* (Monogenea) on guppies (*Poecilia reticulata*). Parasitology 85:217-236.

Selkoe, K. A., and R. J. Toonen. 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecology Letters 9:615-629.

Shaw, P. W., G. R. Carvalho, and A. E. Magurran. 1991. Population differentiation in Trinidadian guppies (*Poecilia reticulata*): patterns and problems. Journal of Fish Biology 39:203-209.

Shaw, P. W., G. R. Carvalho, A. E. Magurran, and B. H. Seghers. 1994. Factors affection the distribution of genetic-variability in the guppy, *Poecilia-reticulata*. Journal of Fish Biology 45:875-888.

Shen, X. Y., G. P. Yang, and M. J. Liao. 2006. Development of 51 genomic microsatellite DNA markers of guppy (*Poecilia reticulata*) and their application in closely related species. Molecular Ecology Notes 7:302-306.

Shiah, F. K., and H. W. Ducklow. 1994. Temperature regulation of heterotrophic bacterioplankton abundance, production, and specific growth-rate in Chesapeake bay. Limnology and Oceanography 39:1243-1258.

Shiina, T., H. Inoko, and J. K. Kulski. 2004. An update of the HLA genomic region, locus information and disease associations: 2004. Tissue Antigens 64:631-649.

Shikano, T., D. Ando, and N. Taniguchi. 2005. Relationships of vertebral deformity with genetic variation and heterosis in the guppy *Poecilia reticulata*. Aquaculture 246:133-138.

Shikano, T., and N. Taniguchi. 2002. Relationships between genetic variation measured by microsatellite DNA markers and a fitness-related trait in the guppy (Poecilia reticulata). Aquaculture 209:77-90.

Shohet, A. J., and P. J. Watt. 2004. Female association preferences based on olfactory cues in the guppy, *Poecilia reticulata*. Behavioral Ecology and Sociobiology 55:363-369.

Skarstein, F., I. Folstad, S. Liljedal, and M. Grahn. 2005. MHC and fertilization success in the Arctic charr (*Salvelinus alpinus*). Behavioral Ecology and Sociobiology 57:374-380.

Slade, R. W. 1992. Limited MHC polymorphism in the southern elephant seal - implications for MHC evolution and Marine mammal population biology. Proceedings of the Royal Society of London Series B-Biological Sciences 249:163-171.

Slade, R. W., and H. I. McCallum. 1992. Overdominant vs frequency-dependent selection at MHC loci. Genetics 132:861-862.

Slatkin, M. 2005. Seeing ghosts: the effect of unsampled populations on migration rates estimated for sampled populations. Molecular Ecology 14:67-73.

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. Molecular Ecology 12:2845-2851.

Sommer, S. 2005. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Front Zool 2:16-34.

Spielman, D., B. W. Brook, D. A. Briscoe, and R. Frankham. 2004. Does inbreeding and loss of genetic diversity decrease disease resistance? Conservation Genetics 5:439-448.

Spurgin, L. G., and D. S. Richardson. 2010. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. Proceedings of the Royal Society B-Biological Sciences 277:979-988.

Stenzel, A., T. Lu, W. A. Koch, J. Hampe, S. M. Guenther, F. M. De La Vega, M. Krawczak, and S. Schreiber. 2004. Patterns of linkage disequilibrium in the MHC region on human chromosome 6p. Human Genetics 114:377-385.

Stern, L. J., J. H. Brown, T. S. Jardetzky, J. C. Gorga, R. G. Urban, J. L. Strominger, and D. C. Wiley. 1994. Crystal-structure of the human class II MHC protein *HLA-DR1* complexed with an influenza virus peptide. Nature 368:215-221.

Stet, R. J. M., B. de Vries, K. Mudde, T. Hermsen, J. van Heerwaarden, B. P. Shum, and U. Grimholt. 2002. Unique haplotypes of co-segregating major histocompatibility class II A and class II B alleles in Atlantic salmon (*Salmo salar*) give rise to diverse class II genotypes. Immunogenetics 54:320-331.

Stet, R. J. M., C. P. Kruiswijk, and B. Dixon. 2003. Major histocompatibility lineages and immune gene function in teleost fishes: The road not taken. Critical Reviews in Immunology 23:441-471.

Stetefeld, J., and M. A. Ruegg. 2005. Structural and functional diversity generated by alternative mRNA splicing. Trends in Biochemical Sciences 30:515-521.

Suk, H. Y., and B. D. Neff. 2009. Microsatellite genetic differentiation among populations of the Trinidadian guppy. Heredity 102:425-434.

Summers, K., K. E. Roney, J. da Silva, G. Capraro, B. J. Cuthbertson, S. Kazianis, G. G. Rosenthal, M. J. Ryan, and T. J. McConnell. 2008. Divergent patterns of selection on the *DAB* and *DXB* MHC class II loci in Xiphophorus fishes. Genetica Advanced online:DOI 10.1007/s10709-10008-19284-10704.

Tajima, F. 1989. Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585-595.

Takahata, N., and M. Nei. 1990. Allelic genealogy under overdominant and frequency-depentent selection and polymorphism of major histocompatinility complex loci. Genetics 124:967-978.

Tamura, K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24:1596-1599.

Tenaillon, O., D. Skurnik, B. Picard, and E. Denamur. 2010. The population genetics of commensal Escherichia coli. Nature Reviews Microbiology 8:207-217.

The-MHC-sequencing-consortium. 1999. Complete sequence and gene map of a human major histocompatibility complex. Nature 401:921-923.

Ting, J. P. Y., and J. Trowsdale. 2002. Genetic control of MHC class II expression. Cell 109:S21-S33.

Tobler, M., and I. Schlupp. 2005. Parasites in sexual and asexual mollies (*Poecilia*, Poeciliidae, Teleostei): a case for the Red Queen? Biology Letters 1:166-168.

Tobler, M., T. Wahli, and I. Schlupp. 2005. Comparison of parasite communities in native and introduced populations of sexual and asexual mollies of the genus Poecilia. Journal of Fish Biology 67:1072-1082.

Tollenaere, C., J. Bryja, M. Galan, P. Cadet, J. Deter, Y. Chaval, K. Berthier, A. R. Salvador, L. Voutilainen, J. Laakkonen, H. Henttonen, J. F. Cosson, and N. Charbonnel. 2008. Multiple parasites mediate balancing selection at two MHC class II genes in the fossorial water vole: insights from multivariate analyses and population genetics. Journal of Evolutionary Biology 21:1307-1320.

Trachtenberg, E., B. Korber, C. Sollars, T. B. Kepler, P. T. Hraber, E. Hayes, R. Funkhouser, M. Fugate, J. Theiler, Y. S. Hsu, K. Kunstman, S. Wu, J. Phair, H. Erlich, and S. Wolinsky. 2003. Advantage of rare HLA supertype in HIV disease progression. Nature Medicine 9:928-935.

Truett, G. E., P. Heeger, R. L. Mynatt, A. A. Truett, J. A. Walker, and M. L. Warman. 2000. Preparation of PCR-quality mouse genomic DNA with hot sodium hydroxide and tris (HotSHOT). Biotechniques 29:52-+.

Uyenoyama, M. K. 2005. Evolution under tight linkage to mating type. New Phytologist 165:63-70.

van Oosterhout, C. 2009a. A new theory of major histocompatibility complex evolution: beyond selection on the immune genes. Proceedings of the Royal Society B-Biological Sciences 276:657-665.

van Oosterhout, C. 2009b. Transposons in the MHC: the Yin and Yang of the vertebrate immune system. Heredity.

van Oosterhout, C., P. D. Harris, and J. Cable. 2003. Marked variation in parasite resistance between two wild populations of the Trinidadian guppy, *Poecilia reticulata* (Pisces : Poeciliidae). Biological Journal of the Linnean Society 79:645-651.

van Oosterhout, C., W. F. Hutchinson, D. P. M. Wills, and P. Shipley. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes 4:535-538.

van Oosterhout, C., D. A. Joyce, and S. M. Cummings. 2006a. Evolution of MHC class IIB in the genome of wild and ornamental guppies, *Poecilia reticulata*. Heredity 97:111-118.

van Oosterhout, C., D. A. Joyce, S. M. Cummings, J. Blais, N. J. Barson, I. W. Ramnarine, R. S. Mohammed, N. Persad, and J. Cable. 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., R. S. Mohammed, H. Hansen, G. A. Archard, M. McMullan, D. J. Weese, and J. Cable. 2007a. Selection by parasites in spate conditions in wild Trinidadian guppies (*Poecilia reticulata*). International Journal for Parasitology 37:805-812.

van Oosterhout, C., A. M. Smith, B. Hanfling, I. W. Ramnarine, R. S. Mohammed, and J. Cable. 2007b. The guppy as a conservation model: Implications of parasitism and inbreeding for reintroduction success. Conservation Biology 21:1573-1583.

von Salomé, J., U. Gyllenstwn, and T. F. Bergstrom. 2007. Full-length sequence analysis of the *HLA-DRB1* locus suggests a recent origin of alleles. Immunogenetics 59:261-271.

Watanabe, T., M. Yoshida, M. Nakajima, and N. Taniguchi. 2003. Isolation and characterization of 43 microsatellite DNA markers for guppy (*Poecilia reticulata*). Molecular Ecology Notes 3:487-490.

Wegner, K. M. 2004. Major histocompatibility genes, polymorphism and balancing selection. Christian-Albrechts-Univ, Kiel.

Wegner, K. M. 2008. Historical and contemporary selection of teleost MHC genes: did we leave the past behind? Journal of Fish Biology 73:2110-2132.

Wegner, K. M., M. Kalbe, J. Kurtz, T. B. H. Reusch, and M. Milinski. 2003a. Parasite selection for immunogenetic optimality. Science 301:1343-1343.

Wegner, K. M., M. Kalbe, H. Schaschl, and T. B. H. Reusch. 2004. Parasites and individual major histocompatibility complex diversity - an optimal choice? Microbes and Infection 6:1110-1116.

Wegner, K. M., T. B. H. Reusch, and M. Kalbe. 2003b. Multiple parasites are driving major histocompatibility complex polymorphism in the wild. Journal of Evolutionary Biology 16:224-232.

Westerdahl, H., B. Hansson, S. Bensch, and D. Hasselquist. 2004. Between-year variation of MHC allele frequencies in great reed warblers: selection or drift? Journal of Evolutionary Biology 17:485-492.

Willing, E. M., P. Bentzen, C. van Oosterhout, M. Hoffmann, J. Cable, F. Breden, D. Weigel, and C. Dreyer. 2010. Genome-wide single nucleotide polymorphisms reveal population history and adaptive divergence in wild guppies. Molecular Ecology 19:968-984.

Wilson, D. J., and G. McVean. 2006. Estimating diversifying selection and functional constraint in the presence of recombination. Genetics 172:1411-1425.

Winter, C., T. Hein, G. Kavka, R. L. Mach, and A. H. Farnleitner. 2007. Longitudinal changes in the bacterial community composition of the Danube River: a whole-river approach. Applied and Environmental Microbiology 73:421-431.

Woolhouse, M. E. J., J. P. Webster, E. Domingo, B. Charlesworth, and B. R. Levin. 2002. Biological and biomedical implications of the co-evolution of pathogens and their hosts. Nature Genetics 32:569-577.

Xu, D. H., C. A. Shoemaker, and P. H. Klesius. 2007. Evaluation of the link between gyrodactylosis and streptococcosis of Nile tilapia, *Oreochromis niloticus* (L.). Journal of Fish Diseases 30:233-238.

Zhang, J. Z. 2003. Evolution by gene duplication: an update. Trends in Ecology & Evolution 18:292-298.

Appendix

Appendix 1. *Genetic differentiation – macro check*

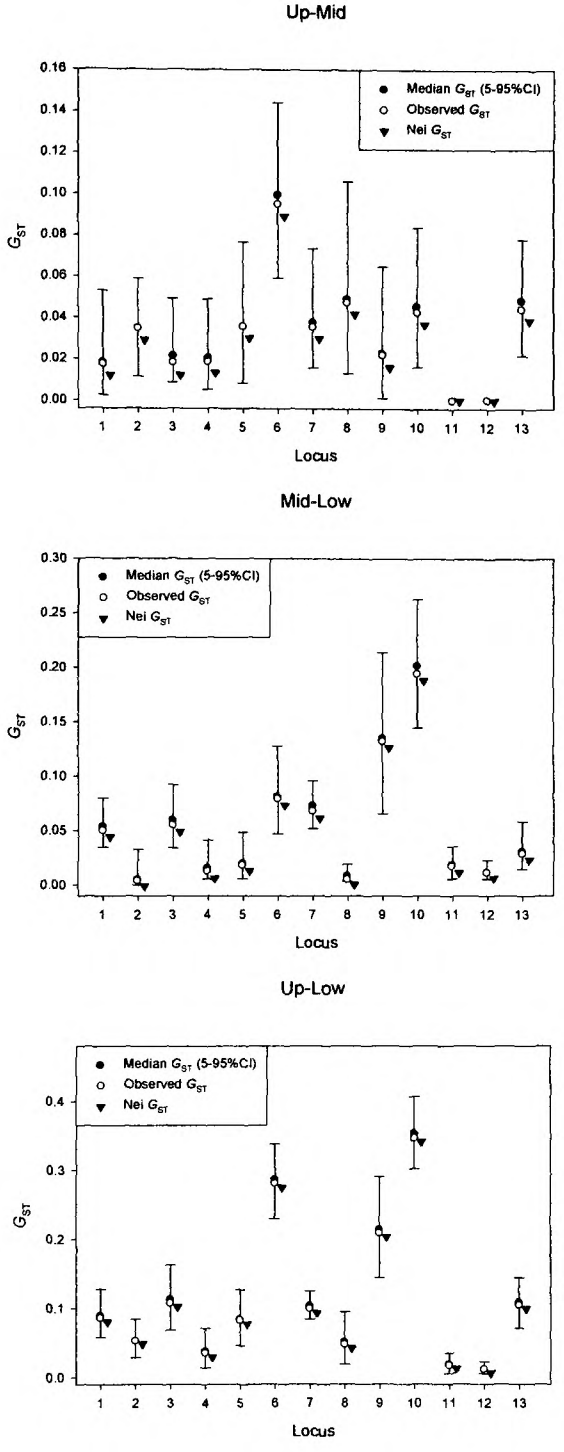


Figure 1. Genetic differentiation calculated between Up and Mid populations for each of 13 neutral microsatellite loci. Observed and median G_{ST} (5-95% CI) are calculated by bootrapping (with replacement) over individuals using 1000 replicates. In addition the validity of the bootstrap method is confirmed with the addition of G_{ST} using the method of Nei (1987) within FSTAT version 2.9.3 (Goudet 1995, 2001).

Appendix 2. *Microsatellite allele counts per locus (2007)*

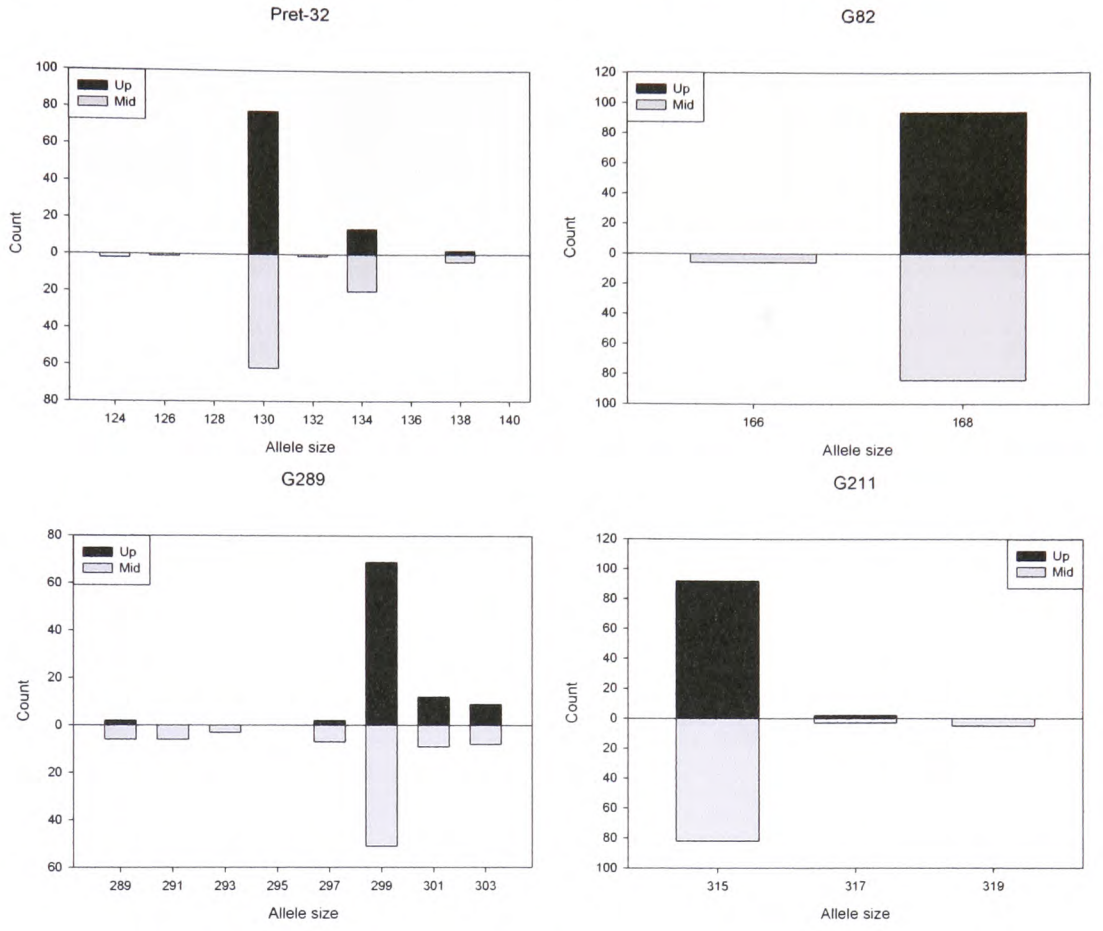
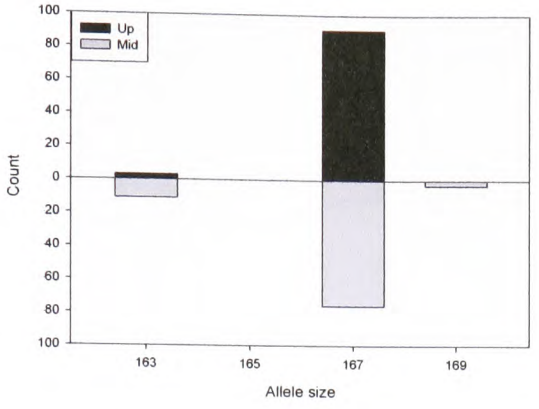


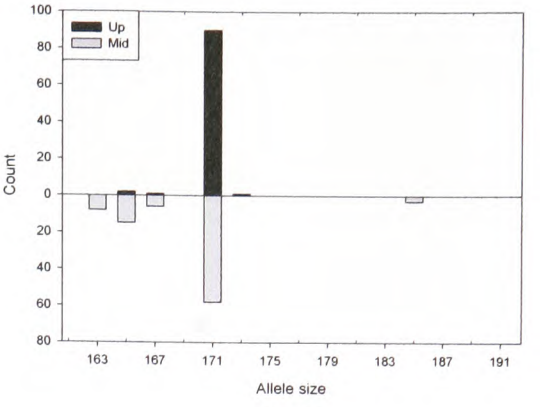
Figure 2. Counts of alleles for 11 out of 13 microsatellite loci (Pret32, G82, G289, G211, Pret-77, Pr39, Hull70-2, Pret-69, Pr92, Hull9-1 and Pret-46)of guppies in the Upper Naranjo (Up: n = 47) and Mid Naranjo (n = 45), regions of the Aripo River. Two loci are not included because they are fixed for the same allele. Genetic differentiation between these populations is low (G'_{ST} = 0.066) but significantly greater than zero. Figure continued over.

Pret-77

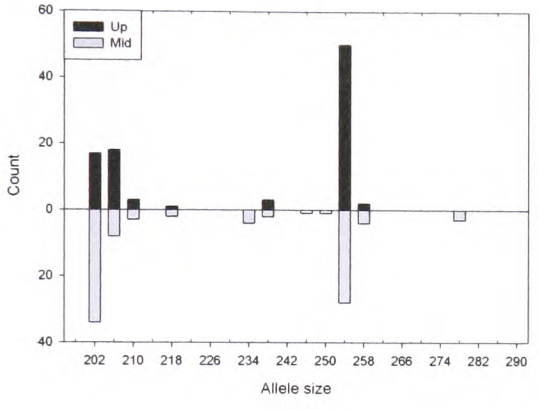


Hull 70-2

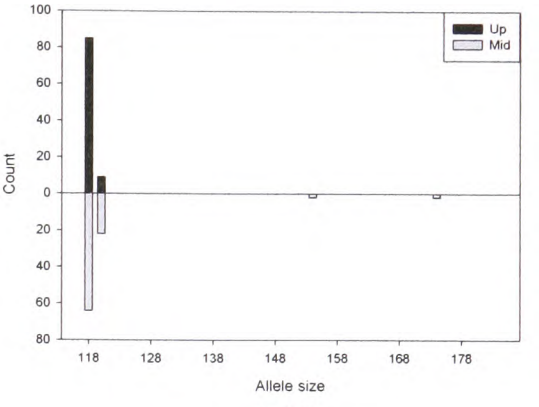
Pr39



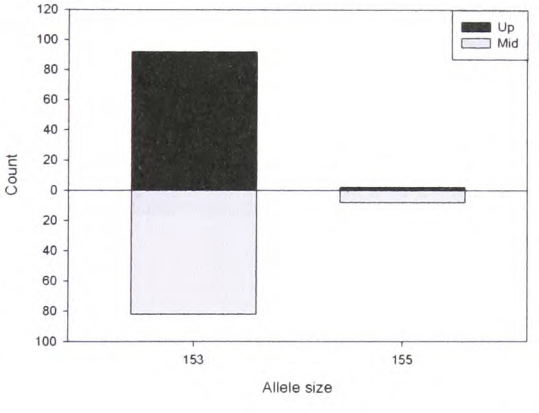
Pret-69

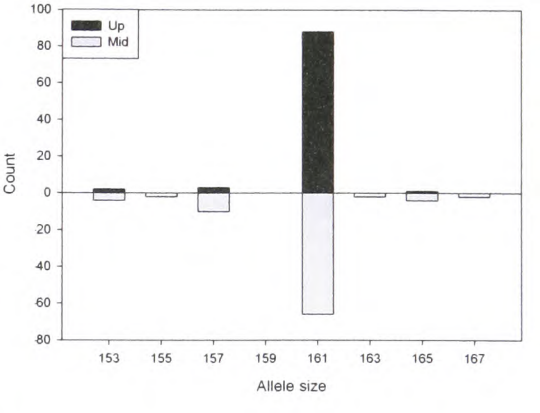


Pr92

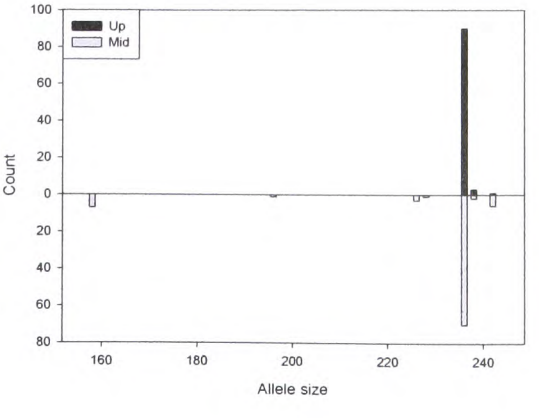


Hull 9-1





Pret-46



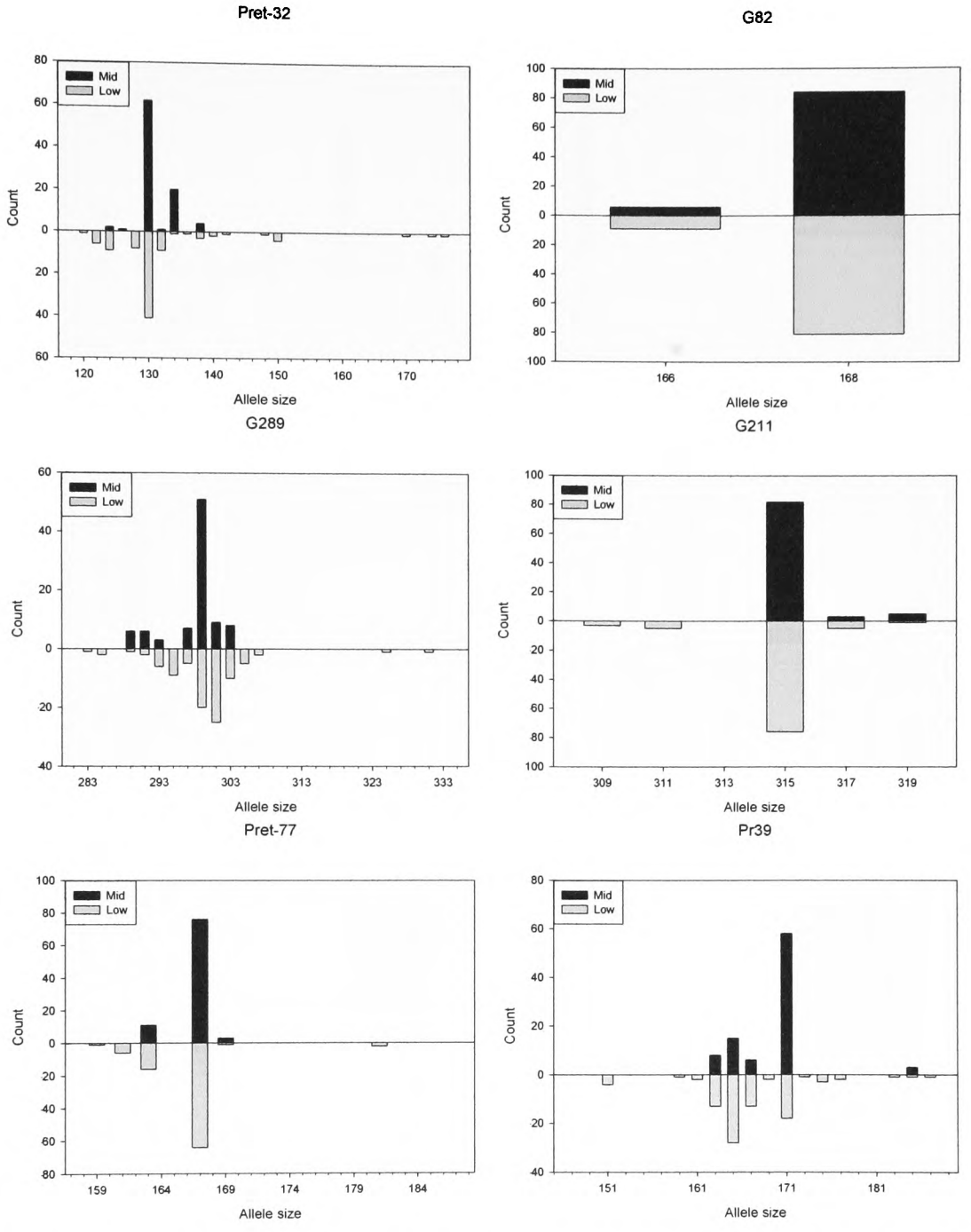
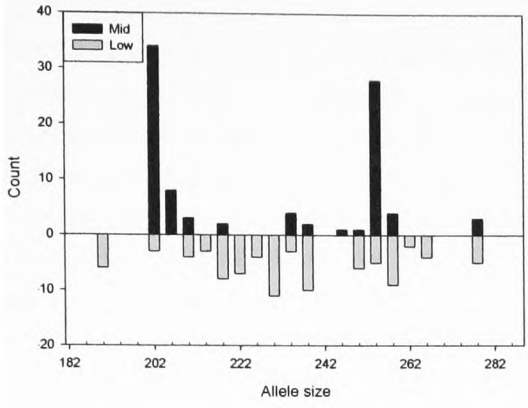


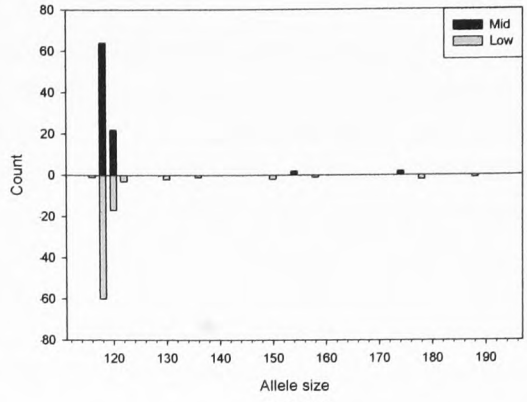
Figure 3. Counts of alleles of 13 microsatellite loci (Pret32, G82, G289, G211, Pret-77, Pr39, Hull70-2, Pret-69, Pr92, Hull9-1, G72, G350 and Pret-46) from guppies in the Mid Naranjo (Mid: n = 45) and Lower Aripo (Low: n = 45) regions of the Aripo River (G'_{ST} = 0.232). Figure continued over.

Hull 70-2

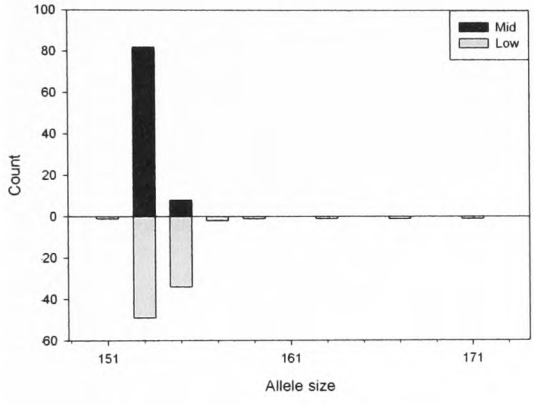


Pr92

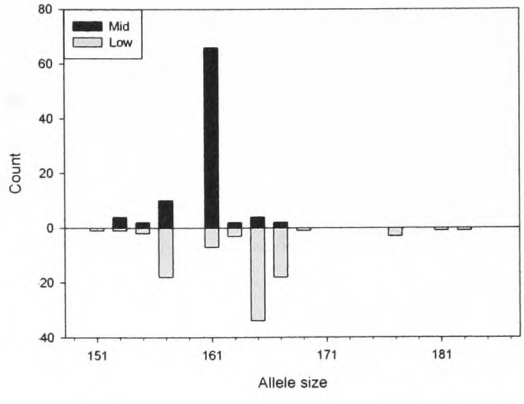
Pret-69

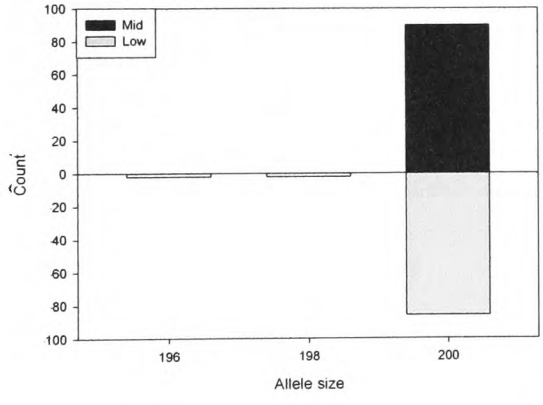


Hull 9-1

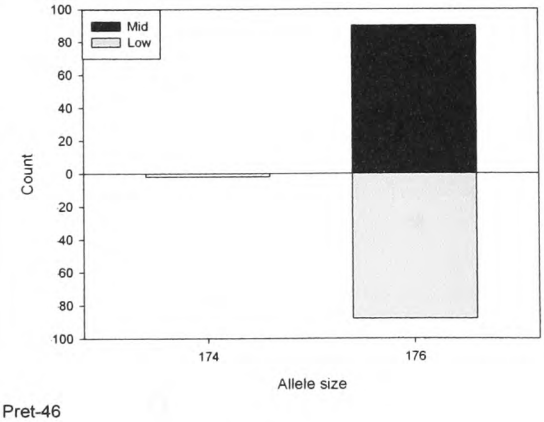


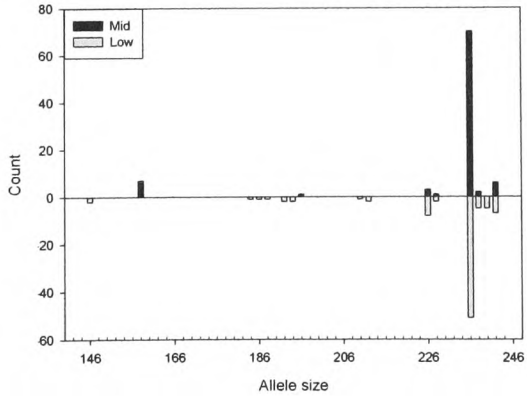
G350





Pret-46





Appendix 3. *Results controlled for cloning effort - 12 clones per individual*

Allelic richness (*A*) and genetic differentiation (G'_{ST}) estimates were based on all clones available per individual. However, 12 was the minimum number of clones sequenced per individual. The cloning effort in the Up and Mid populations was approximately equal (number of clones per population: Up = 342, Mid = 336) but the cloning effort in the Low population was reduced (number of clones per population: Low = 288). Here I use the same bootstrapping protocol (see Chapter 3) except cloning effort is capped at a maximum of 12 clones per individual. This did not significantly affect MHC allelic richness (*A*) or genetic differentiation (G'_{ST}) (compare Fig. 7 and Fig. 8 in Chapter 3) of populations (Randomisation test: *A*, Up P = 0.297, Mid P = 0.159, Low P = 0.238; G'_{ST}, Up-Mid P = 0.446, Mid-Low P = 0.525, Up-Low P = 0.481).

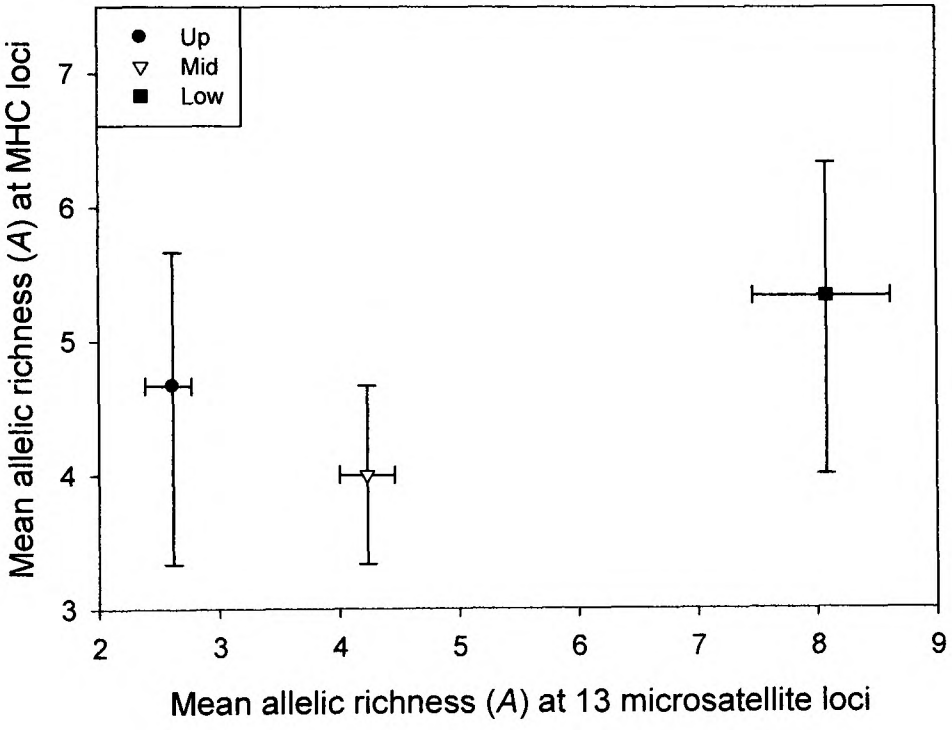


Figure 4. Allelic richness (*A*, 5-95% CI) of the MHC and microsatellite DNA variation in the Upper Naranjo, Mid Naranjo and Lower Aripo guppy populations. The mean *A* for the MHC was calculated based on an estimated 3 loci. The 95% confidence intervals are calculated by bootstrapping over individuals. An attempt was made to control for cloning effort and therefore 12 clones per individual was used. There was no significant difference between methods utilising all clones per individual and limiting to 12 clones per individual (see text and Fig. 7 in MHC chapter)

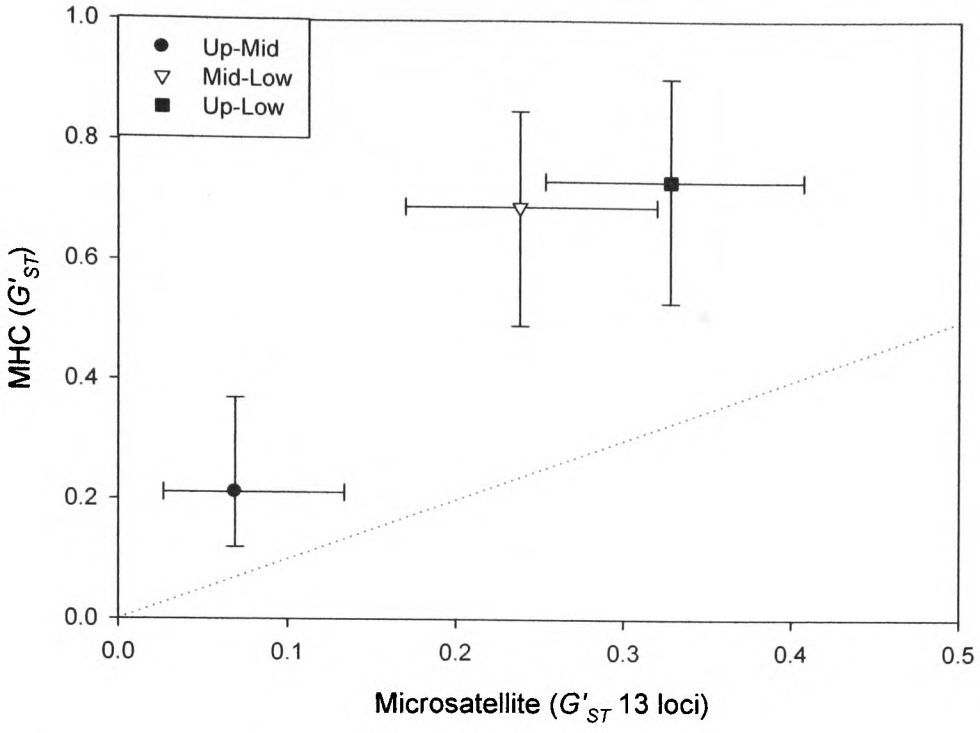


Figure 5. Pairwise comparisons among the Up, Mid and Low populations were made for genetic differentiation (G'_{ST}, with 5-95% CI) of MHC and 13 microsatellite loci by bootstrapping over individuals. The dotted line passes from 0 to 1 and represents the expectation where divergence of neutral variation and MHC is equally partitioned between these loci. Here an attempt was made to control for MHC cloning effort and therefore 12 clones per individual was used. There was no significant difference between methods utilising all clones per individual and those limiting to 12 clones per individual (see text and Fig. 8 in MHC chapter)

Appendix 4. *Genetic differentiation, GST against G'ST*

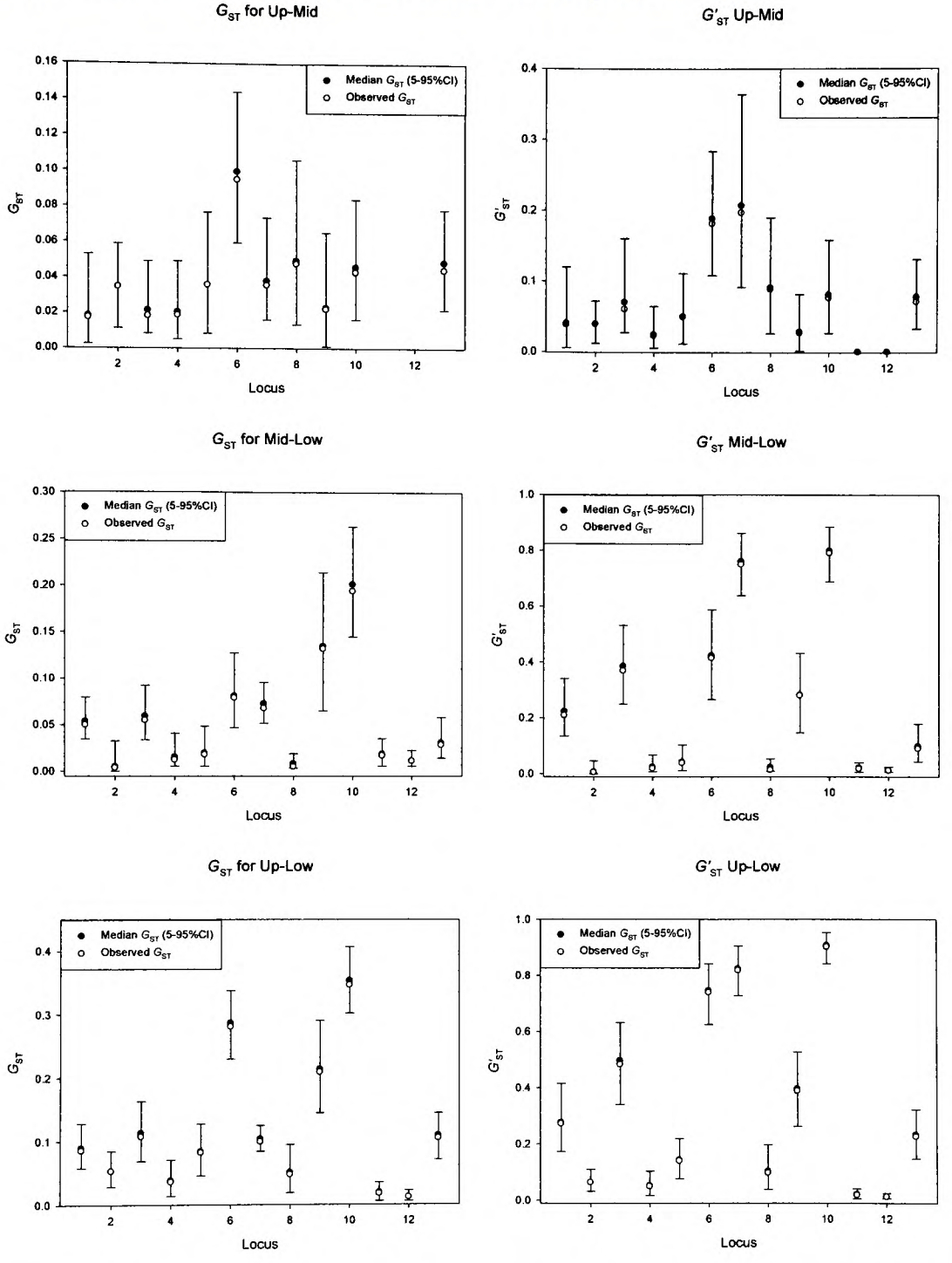


Figure 6. G_{ST} against G'_{ST} 13 loci. Genetic differentiation (G_{ST} and G'_{ST} with 5-95% CI) of neutral variation calculated by bootstrapping over individuals. Note that because G_{ST} values can only be less than the average within population homozygosity (Hedrick 2005b), the most polymorphic loci (see Table 1. Chapter 2) don't have representative G_{ST} values. These loci have G'_{ST} values representative of the polymorphism.

Appendix 5. *PCR bias and gene duplication*

PCR amplification bias is expected to be more common where multiple templates (alleles/loci) are amplified using degenerate primers. A positive correlation between the proportion of clones of an allele (within an individual) and the number of individuals found to have that allele can be taken as evidence of PCR bias. The rationale for this is that alleles that are preferentially amplified will make up a higher proportion of clones within an individual in which they are present and therefore, are more likely to be observed in individuals compared to other alleles. However, this finding can also indicate a locus fixed for a polymorphism that has been duplicated. Here I demonstrate the positive correlation between the proportion of individuals with a given allele and the mean proportion of clones that allele is represented within. Guppies of the Mid and Low populations of the Aripo River were cloned and sequenced for their MHC diversity (see Chapter 3 Materials & Methods). One guppy was excluded from the analysis because it contained a single allele confirmed using an MHC allele outside the current dataset.

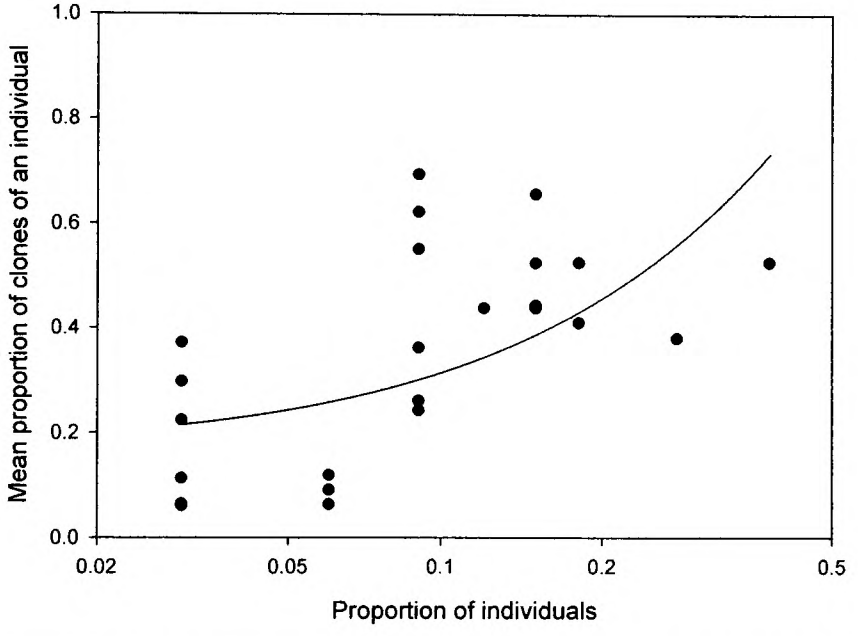


Figure 7. Regression of the proportion of individuals with a copy of a given allele against the mean proportion of clones (within an individual) containing that allele. There is a strong positive relationship between observing an allele in many individuals and the proportion of clones containing that allele (Regression: R2=0.342; F1,26=13.53; p=0.001). Relationship is indicative of favourable amplification of particular alleles or the fixation of alleles at particular loci.

The positive relationship between the proportion of individuals with an allele and the number of clones containing that allele is indicative of both PCR amplification bias and fixation of alleles at duplicated loci (Fig. 7). Interestingly, the observation that individuals possess an intermediate number of MHC alleles (Fig. 8) is also consistent with both variation in locus number with fixation of alleles at multiple loci and, preferential amplification of alleles during PCR. PCR bias should be considered before MHC optimality in such cases where allele number indicates locus variation between individual. This is especially true when one set of primers is used at multiple loci.

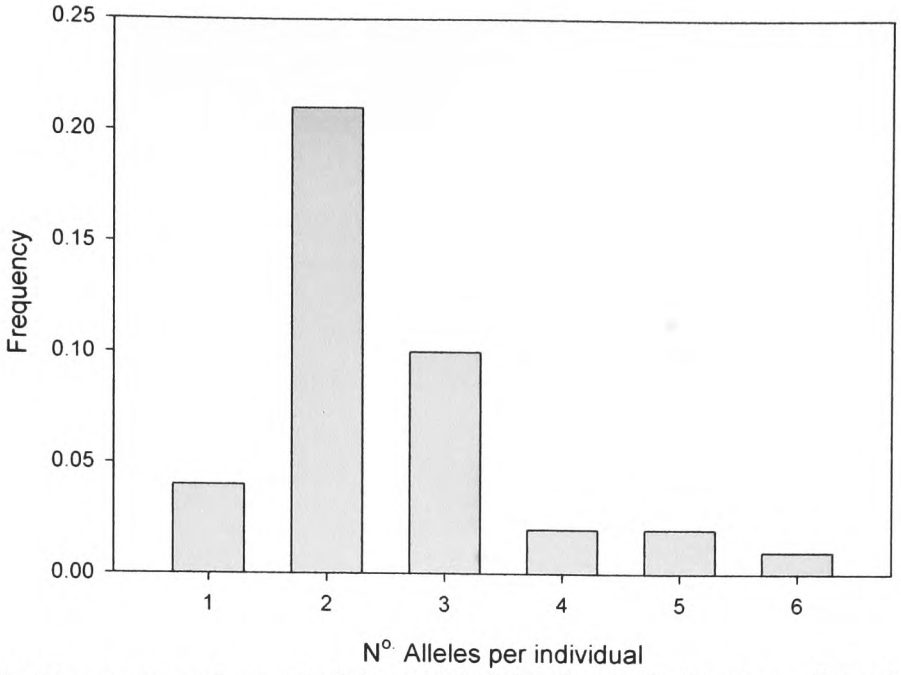


Figure 8. Observed numbers of alleles per individual reveals that two alleles is the most common number of alleles per individual. Only individuals from the Up Mid and Low populations for which 18 clones were sampled are included. Variation in the number of alleles per individual is consistent with both a variation in the number of loci per individual and reduced observation of alleles that are unfavourably amplified during PCR.

Appendix 6. *Estimation of the number of Caroni MHC alleles using the Michaelis-Menten rate equation.*

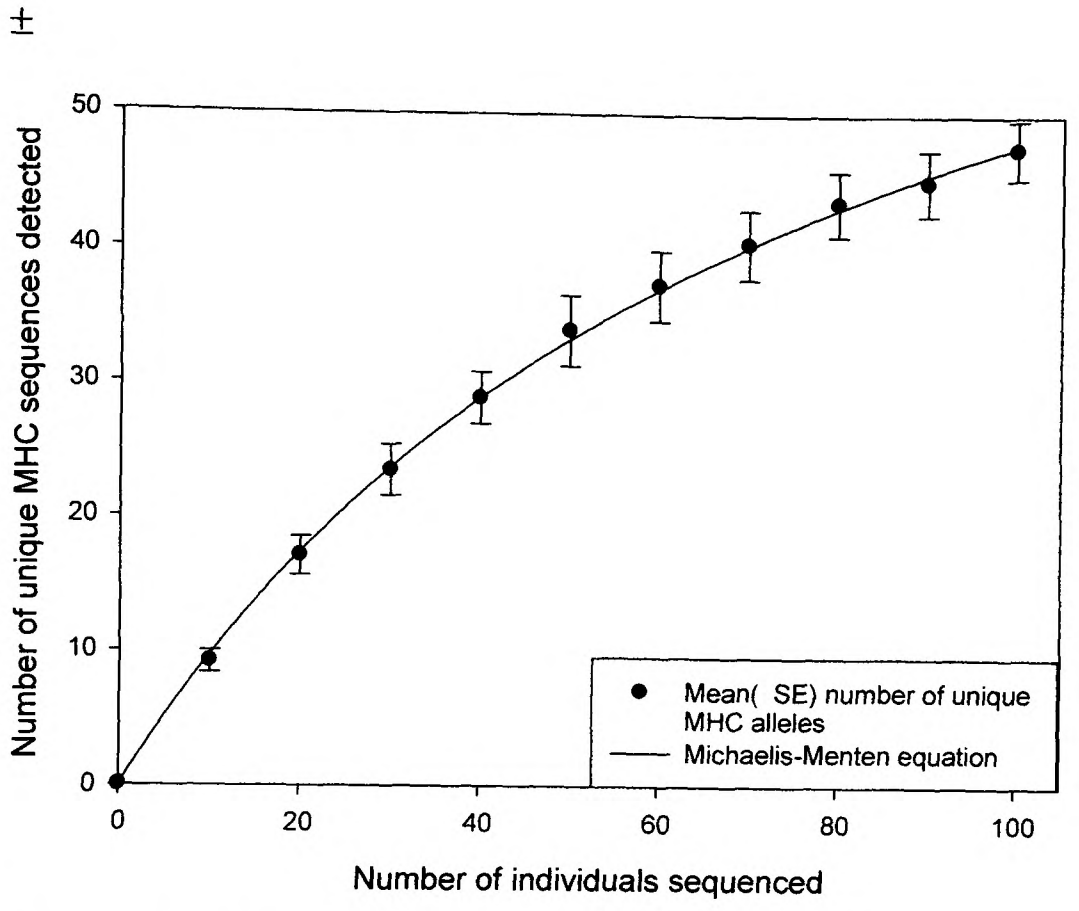
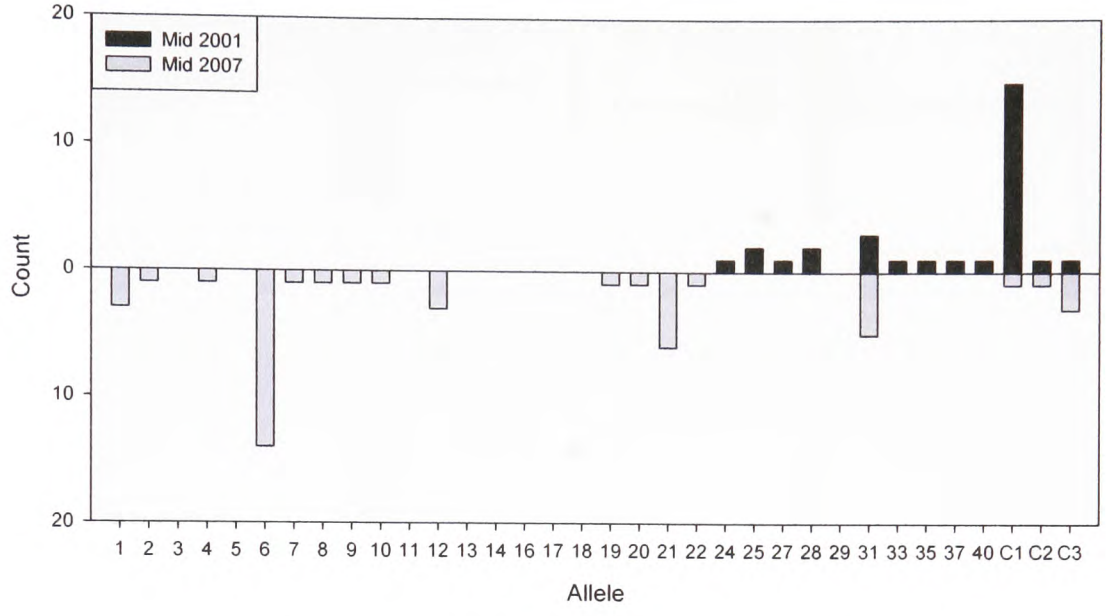


Figure 9. Exploration curve using a Michaelis-Menten equation to estimate the maximum number of MHC alleles present in the Caroni Drainage (Colwell and Coddington 1994). The two-parameter hyperbola, S(Ind.) = (S_{max} Ind.)/(B + Ind.) was fitted using the observed MHC allele frequencies in the Caroni Drainage based on data from van Oosterhout et al. (2006), Fraser & Neff (2009), Fraser et al. (2010) and Chapter 3. The equation with the best fit and R^2=0.9999 is S(ind.) = 84.9Ind./(77.7 + Ind.). This suggests there is an asymptotic maximum of 84.9 unique MHC sequences present in the Caroni.

Allele counts of 2001 and 2007 populations of the Mid
(n 2001=21; n 2007=20)



Allele counts of 2001 and 2007 populations of the Low
(n 2001=35; n 2007=19)

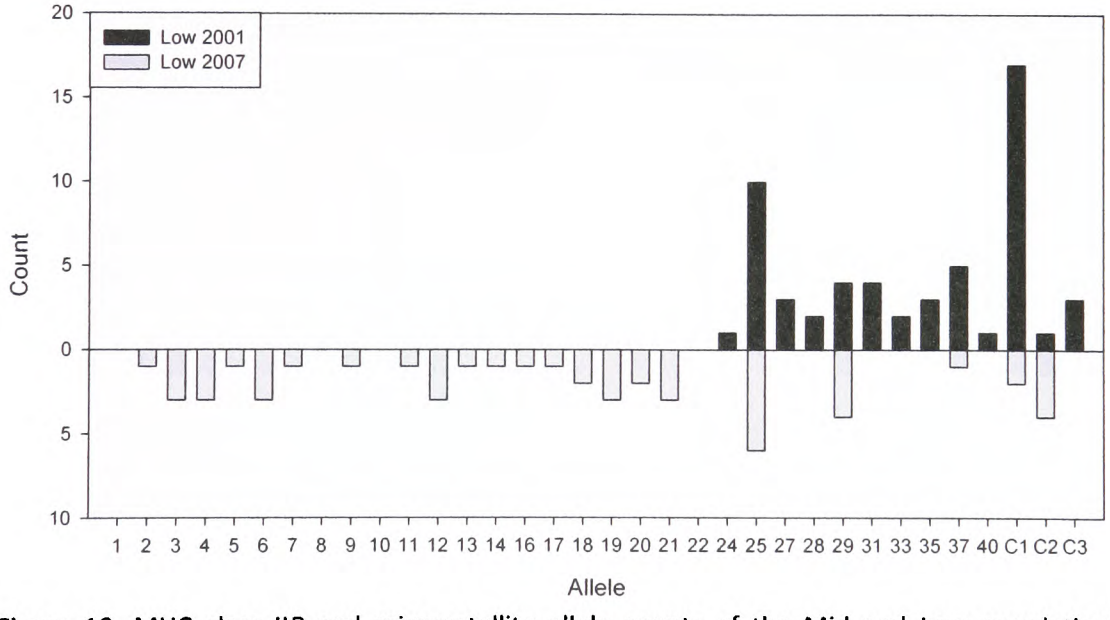
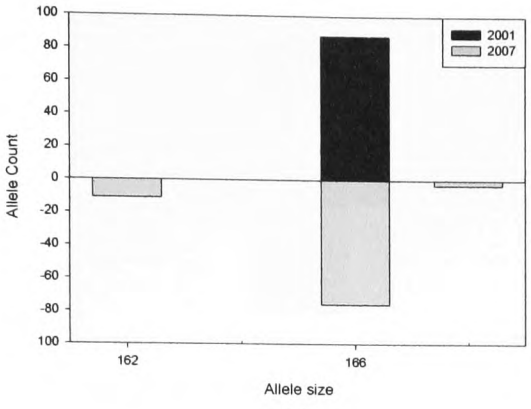
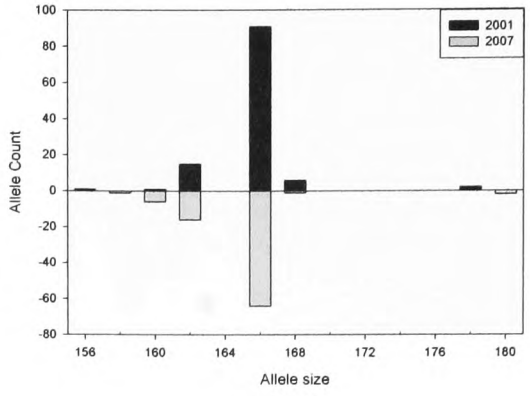


Figure 10. MHC class II*B* and microsatellite allele counts of the Mid and Low populations compared across years (2001 – 2007). MHC alleles correspond to those given by Network (see Fig. 1 Chapter 3). Note that counts of each allele are obtained from different numbers of individuals. Figure continued over.

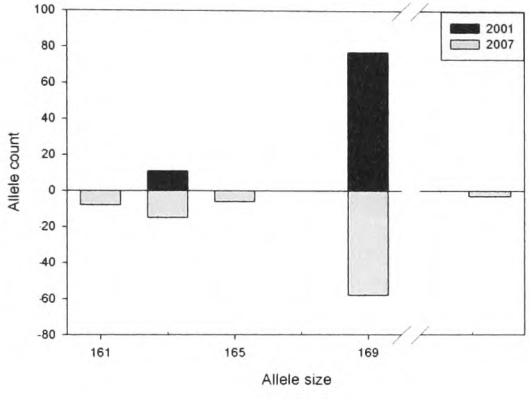
Pret-77
Mid



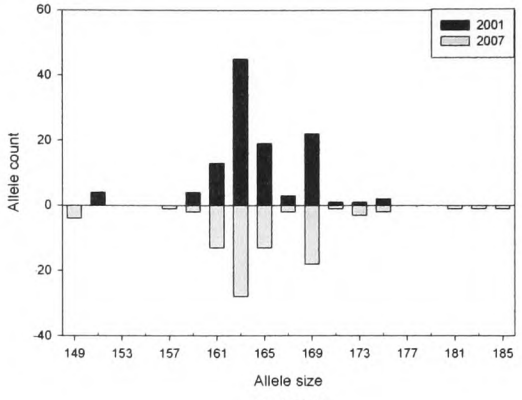
Pret-77
Low



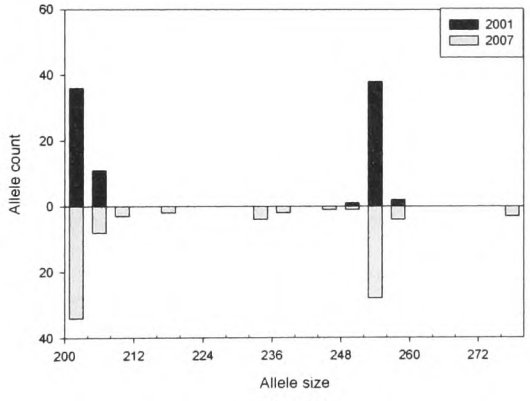
Pr39
Mid



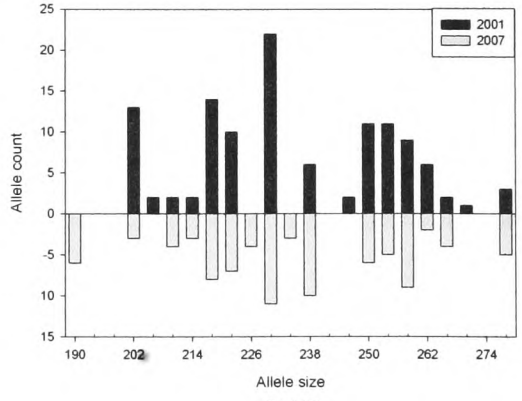
Pr39
Low



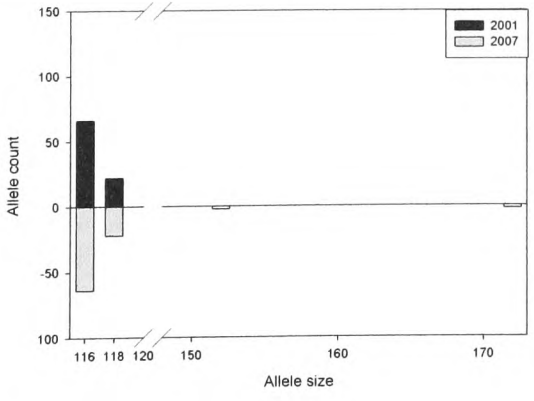
Hull 70-2
Mid



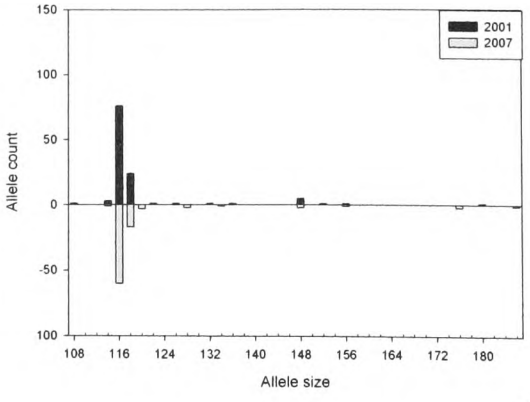
Hull 70-2
Low



Pret-69
Mid



Pret-69
Low



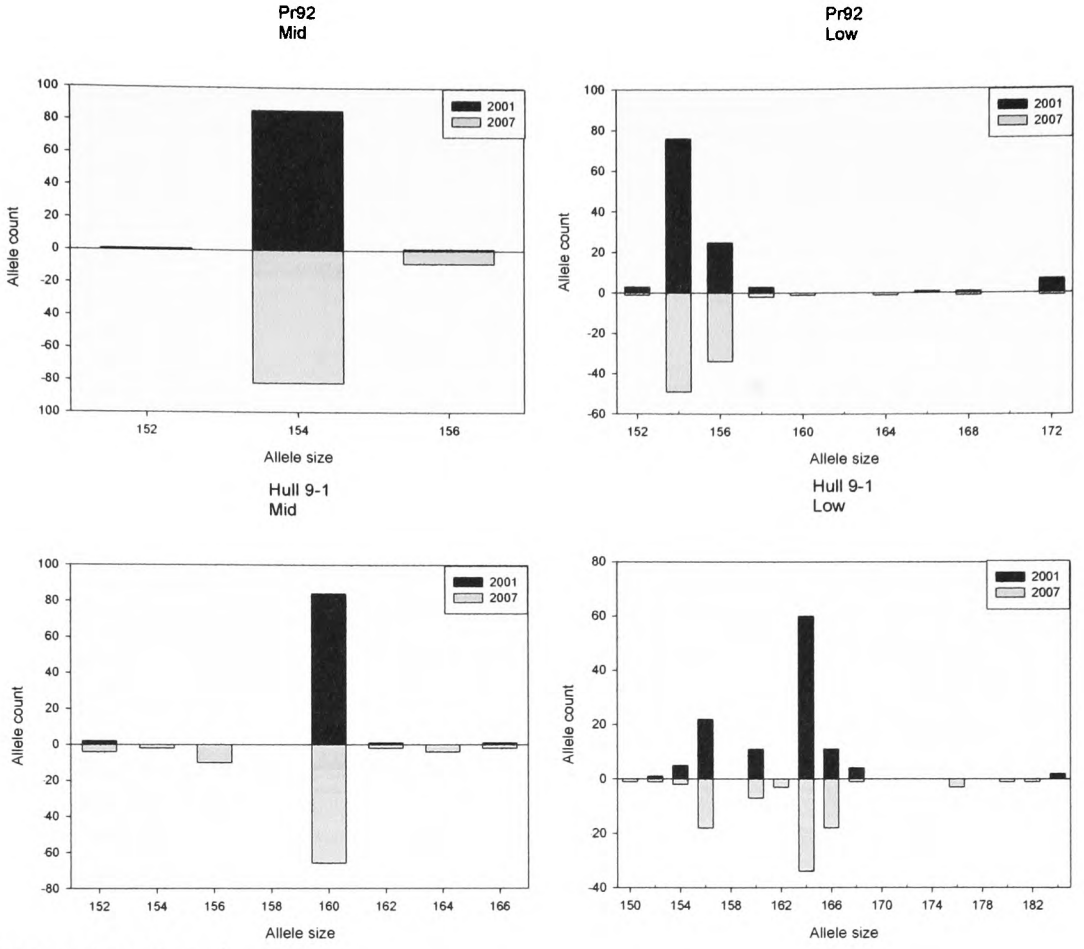


Figure 10 continued

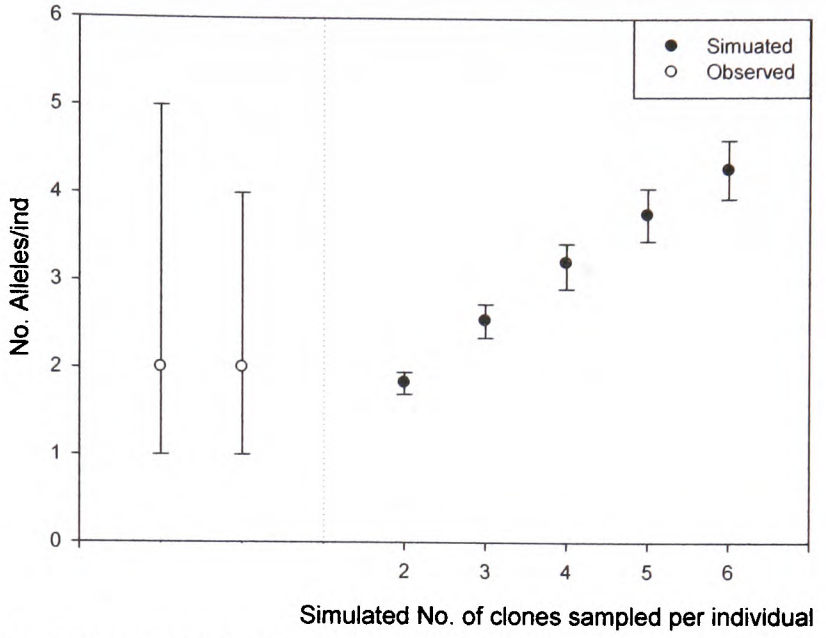


Figure 11. A re-sampling technique used to ascertain the number of clones per individual to sample to best match the observed number of alleles per individual. Bootstrapping over alleles reveals that selecting just two alleles per individual is the most similar way of achieving the median from the observed data.

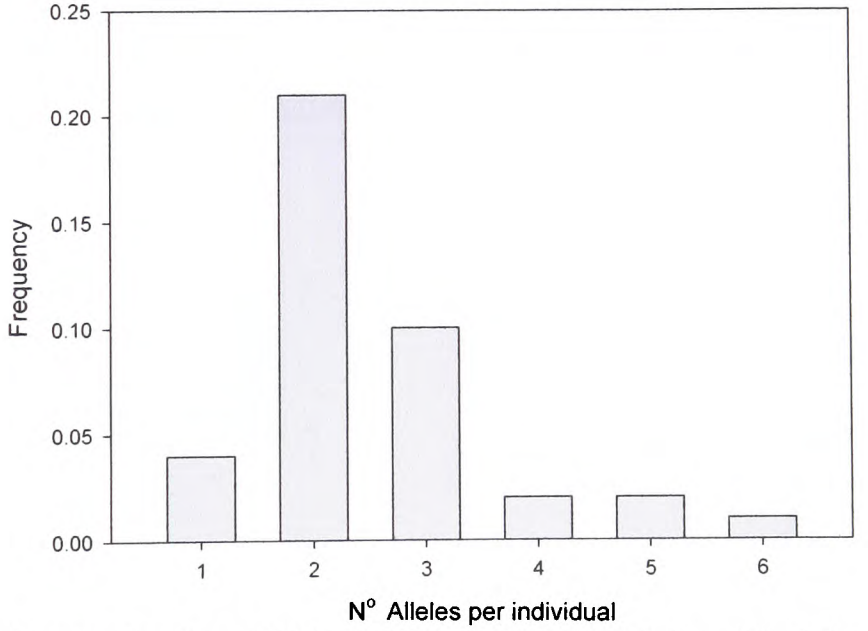


Figure 12. Observed numbers of alleles per individual reveals that two alleles is the most common number of per individual. Only individuals from the Up Mid and Low populations for which 18 clones were sampled are included. Variation in the number of alleles per individual is consistent with both a variation in the number of loci per individual and reduced observation of alleles that are unfavourably amplified during PCR.

Appendix 9. *Reconciling G'ST estimates based on 6 and 13 microsatellite loci*

For the temporal analysis, I used the same loci that have been used in the previous study, the original 6 loci. This allows for a like-with-like comparison in a nested analysis. In response to the examiner's comments, I now added a new analysis that employed a resampling technique of all 13 loci from 2007 or from a subset of the original 6 loci, used in the previous study. I tried three variations of resampling techniques. The first was sampling over individuals (Ind). This technique involved sampling all loci (6 or 13) from each individual. The second technique sampled alleles from within each locus (Allele) and each locus was sampled separately. The third technique involved sampling alleles from within loci (Alleles technique) but sorting each locus separately before getting 5-95 % CI for G'_{ST} (Sort Allele technique). This has the effect of increasing the CI because each locus is sorted separately and then the mean G'_{ST} is calculated across loci.

Results

The 'Ind' and 'Allele' resampling techniques have little effect on G'_{ST} and as expected, the 'Sort Allele' technique increases the CI of the 'Allele' technique. However, the use of the original 6 loci used in a previous study significantly increases the G'_{ST} (Fig. 13). These 6 loci have, on average, 1.4 more alleles per locus than the remaining seven loci (Table 1). This could explain why they produce a higher G'_{ST} (Fig. 12).

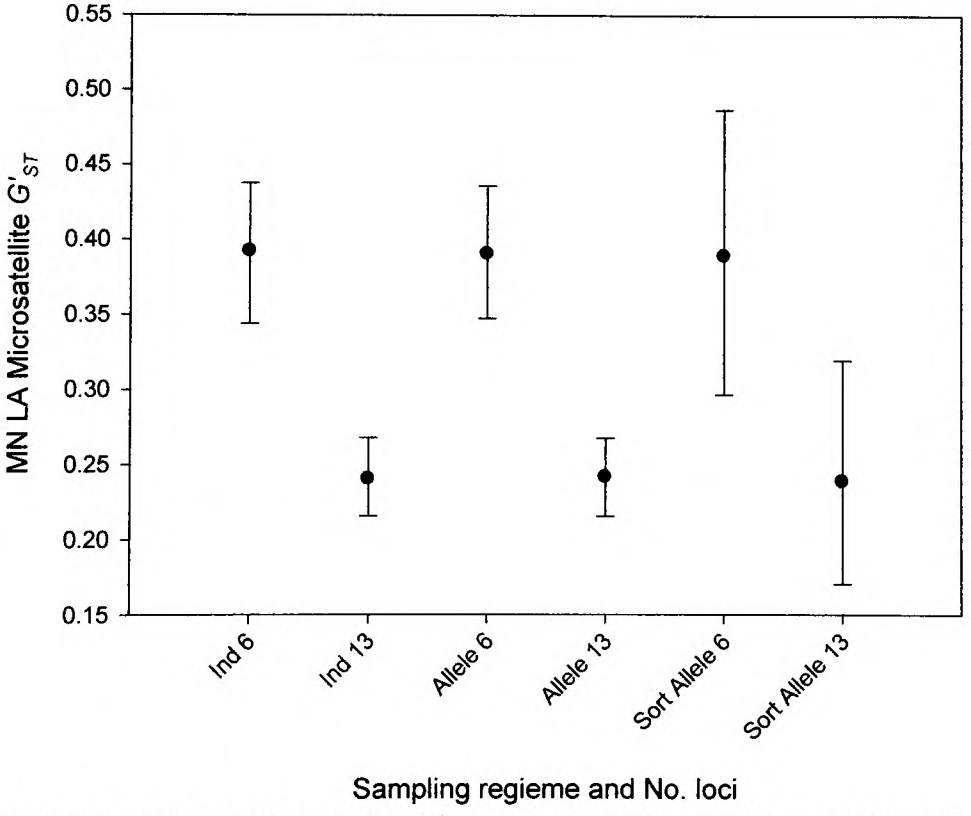


Figure 13. Microsatellite G'_{ST} calculated between the Mid and Low populations of guppies (sampled 2007) using three sampling regimes, Ind, Allele, and Sort Allele. Each regime has been repeated for both the original 6 and all 13 loci. G'_{ST} calculated using the original 6 loci is significantly greater than that calculated using all 13.

Table 1. Allelic richness of 13 microsatellite loci from guppies of the Up, Mid and Low populations, Trinidad (sampled 2007). The additional 7 loci are highlighted in grey and the original 6 loci sampled in a previous study are highlighted in green. The original 6 loci are on average 1.4 times more polymorphic than the other 7 loci, which could explain the increased G'_{ST} between populations when assessing these loci alone.

	Allelic richness at the **additional 7** loci								Allelic richness at the **original 6** loci							
	Pret-32	G82	G289	G211	G72	G350	Pret-46	Mean	Pret-77	Pr39	Hull 70-2	Pret-69	Pr92	Hull 9-1	Mean	Overall Mean
Up	3	1	5	2	1	1	3	2.286	2	4	7	2	2	4	3.5	2.846
Mid	6	2	7	3	1	1	7	3.857	3	5	11	4	2	7	5.333	4.538
Low	16	2	14	5	3	2	14	8	6	14	16	10	8	12	11	9.385
								14.143							19.833	

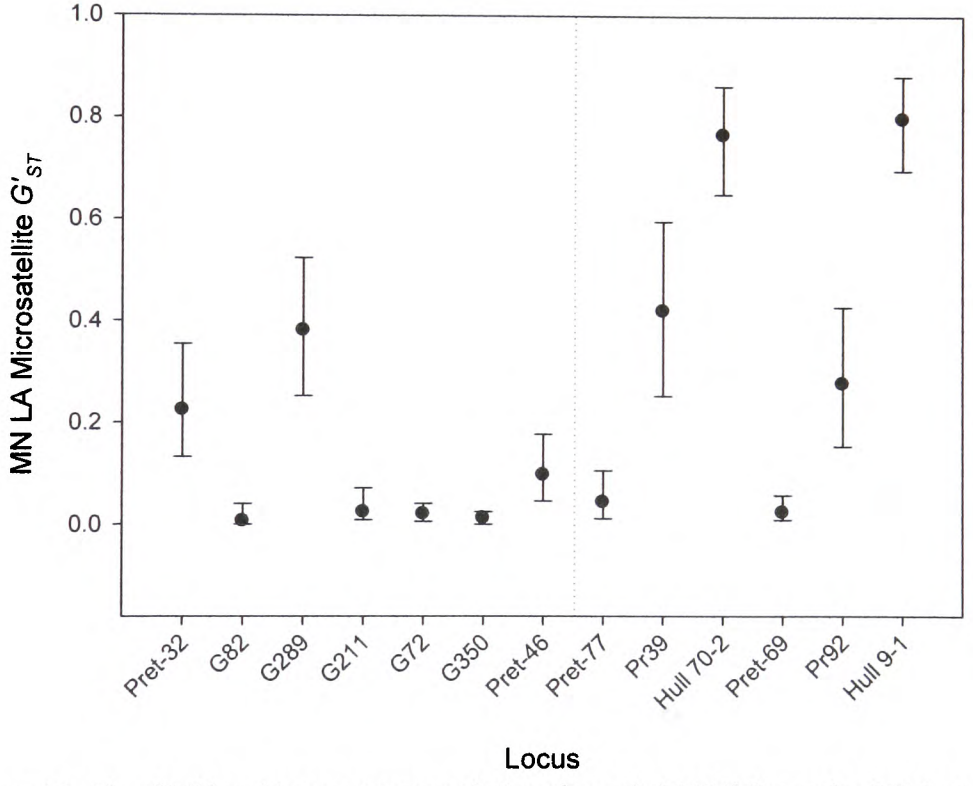


Figure 12. G'_{ST} of Mid and Low guppy populations (sampled 2007) for each of 13 loci. a dotted line separates 7 of 13 loci used in later studies (Additional 7, left) and 6 of 13 loci used in previous studies (Original 6, right).

Conclusion

The higher level of polymorphism of the original 6 loci elevated the G'_{ST} estimate. The assignment bias may be explained by the fact that the original 6 loci were selected to analyse differences between the Mid and Low, whereas the later 7 loci were selected to study polymorphism over a much wider geographic scale. In order to avoid bias, I have included the analysis with the original 6 loci only, as this avoids the complication caused by the comparison of different subsets of loci.

Appendix 10

Table 2. Microsatellite allele frequencies at each of six loci from the Mid Naranjo and Lower Aripo Guppy populations from 2001 and 2007.

	Pret77 Frequency					Pret69 Frequency			
Size	Mid 2001	Low 2001	Med 2007	Low 2007	**Size**	Mid 2001	Low 2001	Med 2007	Low 2007
162	0.00	0.02	0.00	0.02	**114**	0.00	0.02	0.00	0.00
166	1.00	0.84	0.93	0.91	**116**	0.55	0.50	0.49	0.42
168	0.00	0.10	0.07	0.02	**118**	0.45	0.31	0.42	0.33
178	0.00	0.03	0.00	0.00	**120**	0.00	0.00	0.00	0.04
180	0.00	0.00	0.00	0.04	**122**	0.00	0.02	0.00	0.00
	Pr39 Frequency				**126**	0.00	0.02	0.00	0.00
159	0.00	0.00	0.00	0.02	**128**	0.00	0.00	0.00	0.04
161	0.00	0.05	0.02	0.07	**134**	0.00	0.00	0.00	0.02
163	0.00	0.30	0.07	0.27	**136**	0.00	0.02	0.00	0.00
165	0.00	0.19	0.02	0.16	**148**	0.00	0.07	0.00	0.04
167	0.00	0.04	0.00	0.04	**152**	0.00	0.02	0.04	0.00
169	1.00	0.35	0.82	0.29	**156**	0.00	0.02	0.00	0.02
171	0.00	0.02	0.00	0.00	**172**	0.00	0.00	0.04	0.00
173	0.00	0.02	0.00	0.07	**176**	0.00	0.00	0.00	0.04
175	0.00	0.04	0.00	0.02	**180**	0.00	0.02	0.00	0.00
181	0.00	0.00	0.00	0.02	**186**	0.00	0.00	0.00	0.02
183	0.00	0.00	0.07	0.02		**Pr92 Frequency**			
185	0.00	0.00	0.00	0.02	**152**	0.00	0.02	0.00	0.00
	Hull70-2 Frequency				**154**	0.98	0.48	0.84	0.29
190	0.00	0.00	0.00	0.02	**156**	0.02	0.29	0.16	0.58
202	0.18	0.02	0.24	0.00	**158**	0.00	0.05	0.00	0.04
206	0.09	0.02	0.04	0.00	**160**	0.00	0.00	0.00	0.02
210	0.00	0.00	0.02	0.02	**164**	0.00	0.00	0.00	0.02
218	0.00	0.07	0.02	0.02	**166**	0.00	0.02	0.00	0.00
222	0.00	0.07	0.00	0.07	**168**	0.00	0.02	0.00	0.02
230	0.00	0.17	0.00	0.09	**172**	0.00	0.12	0.00	0.02
234	0.00	0.00	0.04	0.02		**Hull9-1 Frequency**			
238	0.00	0.05	0.00	0.16	**156**	0.00	0.10	0.04	0.09
246	0.00	0.03	0.00	0.00	**160**	0.95	0.05	0.78	0.04
250	0.00	0.14	0.02	0.07	**162**	0.02	0.00	0.04	0.02
254	0.68	0.16	0.47	0.09	**164**	0.00	0.60	0.09	0.40
258	0.05	0.10	0.07	0.20	**166**	0.02	0.17	0.04	0.31
262	0.00	0.07	0.00	0.04	**168**	0.00	0.03	0.00	0.02
266	0.00	0.03	0.00	0.09	**176**	0.00	0.00	0.00	0.07
270	0.00	0.02	0.00	0.00	**180**	0.00	0.00	0.00	0.02
278	0.00	0.05	0.07	0.11	**182**	0.00	0.00	0.00	0.02
					184	0.00	0.03	0.00	0.00

Appendix 11

Figure 13 (see over). Nucleotide MHC class II*B* sequences of *Poecilia reticulata*. Identity with the *#Pore 101* sequence is indicated by *dots*, unavailability of information by a *dash* (first base in codon 1). The *number above* the sequence indicates codon positions.

	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42
#Pore_101	-TC	AAA	GAC	ATT	CAG	TTC	ATC	GAC	TCC	TAT	TAT	TAC	AAC	AAG	CTG	GAG	TTC	CTG	AGG	TTT	GAC
#Pore_102	-.G	.C.	...	..C	...	...	...	AG.	...	..C	AT.	...	...	...	A.A	...	.AT	...	...	...	...
#Pore_103	-.G	.C.	...	..C	...	...	...	AG.	...	..C	AT.	...	...	...	A.A	...	.AT	...	...	...	...
#Pore_104	-.G	.C.	...	..C	...	...	...	AG.	...	..C	AT.	...	...	...	A.A	...	.GG	GCC	...	...	...
#Pore_105	-..	...	...	...	...	.AT	G.T	...	...	...	.G.	...	G..	...	...	..A	.A.	A.C	...	...	...
#Pore_106	-.G	..G	...	..C	G..	...	...	AGG	...	A.C	...	...	...	...	...	...	C.G	TAC	...	..C	AG.
#Pore_107	-..	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
#Pore_108	-..	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
#Pore_109	-..	..G	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
#Pore_110	-..	.G.	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
#Pore_111	-..	...	..T	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
#Pore_112	-..	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	.C.	...	...	...	...
#Pore_113	-..	...	...	...	...	...	...	...	.T.	...	...	...	...	...	...	...	...	...	...	...	...
#Pore_114	-..	...	...	...	...	C..	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
#Pore_115	-..	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
#Pore_116	-..	...	...	..C	...	.AT	G.T	...	...	...	.G.	...	...	...	...	...	...	A.C	...	...	...
#Pore_117	-.G	..G	...	..C	...	.A.	...	T..	...	ATC	...	...	...	...	...	..A	...	GCC	...	...	...
#Pore_118	-..	...	...	...	...	.A.	...	TT.	...	ATG	...	...	...	...	...	..A	.A.	GCC	...	...	...
#Pore_119	-.G	..G	...	..C	...	.A.	...	...	...	...	...	...	...	...	...	...	...	A.C	...	...	...
#Pore_120	-.G	..G	...	..C	G..	...	...	AGG	...	A.C	...	...	...	...	...	...	C.G	TAC	...	..C	AG.
#Pore_121	-.G	..G	...	..C	G..	...	...	AGG	...	A.C	...	...	...	.G.	...	...	C.G	TAC	...	..C	AG.
#Pore_122	-..	...	...	...	...	.A.	...	AGG	...	...	.G.	...	...	..A	...	...	...	A.C	...	...	...
#Pore_124	-.G	..G	...	...	...	.A.	G..	T..	...	ATG	...	...	...	...	...	..A	.A.	A..	...	...	...
#Pore_125	-.G	..G	...	...	.G.	.A.	G..	T..	...	ATG	...	...	...	...	...	..A	.A.	A..	...	...	...
#Pore_126	-.G	..G	...	..C	...	...	...	AGA	...	A.C	...	...	...	...	...	...	A..	GCC	...	...	...
#Pore_127	-..	...	...	..C	...	.A.	G..	TT.	...	A.C	...	...	...	...	...	...	A..	GCC	...	...	...
#Pore_129	-.G	..G	...	..C	...	.AT	G.T	...	...	...	.G.	...	G..	...	...	..A	.A.	GCC	...	...	...

#Pore_131	-..	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
#Pore_132	-.G	...	...	...	...	.A.	...	...	...	AGC	...	..T	...	...	...	...	C..	T.C	...	...	...
#Pore_133	-..	...	.G.	...	...	.A.	G..	T..	...	ATG	...	...	...	...	...	..A	.A.	A.C	...	...	...
#Pore_134	-..	...	.G.	...	...	.A.	G..	T..	...	ATG	...	.G.	...	...	...	..A	.A.	A.C	...	...	...
#Pore_136	-.G	..G	...	..C	...	...	...	AGA	...	A.C	...	...	...	...	...	...	A..	GCC	...	...	...
#Pore_138	-.G	..G	...	...	...	...	...	AGG	...	ATC	...	...	...	...	...	...	A..	A.C	...	..C	AC.

	43	**44**	**45**	**46**	**47**	**48**	**49**	**50**	**51**	**52**	**53**	**54**	**55**	**56**	**57**	**58**	**59**	**60**	**61**	**62**	**63**
#Pore_101	AGC	AAC	CTG	GGT	AAA	TAT	GTT	GGA	TAC	ACG	GAG	CTG	GGA	GTG	AAG	AAC	GCT	GAG	CGG	GCC	AAC
#Pore_102	...	...	...	..G	..G	.T.	...	...	...	...	...	T.T	...	...	...	.CT	...	..A	..A	TGG	...
#Pore_103	...	...	...	..G	..G	.T.	...	...	...	...	...	T.T	...	...	...	..T	...	..A	...	TGG	...
#Pore_104	...	...	...	..G	..G	.T.	...	...	...	...	...	T.T	...	...	...	.CT	...	..A	..A	TGG	...
#Pore_105	...	...	G..	..G	...	...	...	...	...	..A	...	TAC	...	...	...	...	..A	...	...	TT.	...
#Pore_106	...	.GT	T..	..C	.GG	.T.	...	...	...	..A	...	TAC	...	...	...	C.G	..C	...	TAC	TT.	...
#Pore_107	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
#Pore_108	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
#Pore_109	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
#Pore_110	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
#Pore_111	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
#Pore_112	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
#Pore_113	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
#Pore_114	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
#Pore_115	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
#Pore_116	...	...	G..	..G	..G	.T.	...	...	...	...	...	...	...	...	...	...	..A	...	ATC	TGG	...
#Pore_117	...	...	G..	..G	..G	.T.	...	...	...	...	...	T.T	...	...	...	C.G	..A	A.C	TAC	TGG	...
#Pore_118	...	...	...	..G	...	...	...	...	...	..A	...	TAT	...	...	...	C.G	..A	A.C	TAC	TGG	...
#Pore_119	...	...	G..	..A	...	...	...	...	...	...	...	...	...	...	...	...	...	...	ATC	TGG	...
#Pore_120	...	.GT	T..	..C	.GG	.T.	...	...	...	..A	...	TAC	...	...	...	C.G	..C	...	TAC	TT.	...

#Pore_121	...	.GT	T..	..C	.GG	.T.	...	...	...	..A	...	TAC	...	...	...	C.G	..C	...	TAC	TT.	...
#Pore_122	...	...	G..	..G	...	...	...	...	...	..A	...	...	...	...	...	...	..A	...	ATC	TGG	...
#Pore_124	...	...	G..	..G	..G	.T.	...	...	...	...	AC.	TAT	...	...	...	...	...	...	.AC	TGG	...
#Pore_125	...	...	G..	..G	..G	.T.	...	...	...	...	AC.	TAT	...	...	...	...	...	...	.AC	TGG	...
#Pore_126	...	...	..C	...	...	...	...	...	...	..A	...	T.T	...	...	...	C.G	..A	A.C	TAC	TGG	...
#Pore_127	...	...	G..	...	.G.	...	...	...	...	..A	...	TAC	...	...	...	C.G	..A	A.C	TAC	TGG	...
#Pore_129	...	...	G..	..G	..G	.T.	...	...	...	..A	...	TAC	...	...	...	...	...	..A	...	CT.	...
#Pore_131	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
#Pore_132	...	...	...	..C	.GG	.T.	...	...	...	..A	...	TAC	...	...	...	C.G	..C	...	TAC	AGG	...
#Pore_133	...	...	...	...	...	...	...	...	...	...	AC.	T.T	...	...	...	...	...	...	...	...	...
#Pore_134	...	...	...	...	...	...	...	...	...	...	AC.	T.T	...	...	...	...	...	...	...	...	...
#Pore_136	...	...	..C	...	...	...	...	...	...	..A	...	T.T	...	...	...	C.G	..A	A.C	TAC	TGG	...
#Pore_138	...	.G.	...	..C	.G.	.T.	...	...	...	..A	...	TAT	...	...	.T.	...	...	...	..C	TGG	...

	64	**65**	**66**	**67**	**68**	**69**	**70**	**71**	**72**	**73**	**74**	**75**	**76**	**77**	**78**	**79**	**80**	**81**	**82**	**83**	**84**
#Pore_101	AGC	AAC	CCA	TCA	GAG	ATC	GCT	GGG	CTG	AAA	GCT	CAG	AGG	GAG	ACC	TAC	TGT	CTG	AAC	AAC	GTC
#Pore_102	..A	G.T	..T	...	T.T	..A	..A	TC.	A..	.G.	...	...	.A.	...	...	...	..C	...	...	C..	..T
#Pore_103	..A	G.T	..T	...	T.T	..A	..A	TC.	A..	.G.	...	...	.A.	...	...	...	..C	...	...	C..	..T
#Pore_104	..A	G.T	..T	...	T.T	..A	..A	TC.	A..	.G.	...	...	.A.	...	...	...	..C	.AA	T..	...	A.T
#Pore_105	.AA	G..	A.G	...	T.T	...	...	.C.	...	..C	...	...	.A.	...	G..	...	...	...	...	...	A..
#Pore_106	...	G..	..G	...	ATT	..T	.GA	.CA	A..	...	...	...	...	...	...	...	...	...	...	...	...
#Pore_107	...	...	..G	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
#Pore_108	...	...	..G	...	...	...	...	...	...	.G.	...	...	...	...	...	...	...	...	...	...	...
#Pore_109	...	...	..G	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
#Pore_110	...	...	..G	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
#Pore_111	...	...	..G	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
#Pore_112	...	...	..G	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
#Pore_113	...	...	..G	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

#Pore_114	...	...	..G	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
#Pore_115	...	...	..G	...	...	...	...	...	...	...	...	...	...	...	...	...	C..	...	...	...
#Pore_116	.A.	...	..G	...	T.T	...	...	.C.	A..	...	...	...	...	...	...	...	...	...	C..	...
#Pore_117	...	G..	A.G	...	TTT	...	...	...	AG.	.G.	...	...	...	...	G..	...	...	...	C..	...
#Pore_118	...	...	..G	...	...	...	T..	...	AG.	.G.	...	...	...	...	G..	...	...	...	C..	...
#Pore_119	.A.	...	..G	...	C..	...	..A	TCA	GA.	.G.	...	...	...	...	G..	...	...	...	.C.	...
#Pore_120	...	G..	..G	...	ATT	..T	.GA	.CA	A..	...	...	...	...	...	G..	...	...	...	...	...
#Pore_121	...	G..	..G	...	ATT	..T	.GA	.CA	A..	...	...	...	...	...	G..	...	...	...	...	...
#Pore_122	.AA	G..	..G	...	C..	...	...	.C.	A..	.CG	...	...	...	...	GGT	...	...	...	...	...
#Pore_124	.AA	G..	A.G	...	C..	...	...	.C.	AG.	...	...	...	..T	...	...	GT.	...	...	C..	...
#Pore_125	.AA	G..	A.G	...	C..	...	...	.C.	AG.	...	...	...	..T	...	...	GT.	...	...	C..	...
#Pore_126	...	...	..G	...	T.T	...	...	...	...	...	...	...	...	...	G..	...	...	...	...	...
#Pore_127	.A.	...	..G	...	ATT	..T	.GA	TCA	AG.	...	...	...	...	...	G..	...	...	...	.C.	.C.
#Pore_129	.AA	G..	A.G	...	TTT	...	...	...	...	.G.	...	...	...	...	...	...	..C	...	...	C.T
#Pore_131	...	...	..G	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
#Pore_132	.A.	...	..T	...	ATT	..T	.GA	TCA	AG.	..G	...	...	...	...	G..	...	..C	...	.C.	...
#Pore_133	...	...	...	...	...	...	...	...	AG.	...	...	...	...	...	G..	...	...	...	...	...
#Pore_134	...	...	...	...	...	...	...	...	AG.	...	...	...	...	...	G..	...	...	...	...	...
#Pore_136	...	G..	A.G	...	T.T	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
#Pore_138	.AA	G..	A.G	...	ATT	..T	.GA	.CA	A..	GC.	...	...	...	...	GGT	...	..C	A..	...	...

	85	86	87	88	89	90	91	92	93	94	95	96
#Pore_101	GGT	GTC	GAC	TAC	CAA	ACC	GCT	CTG	GAT	AAA	TCA	G
#Pore_102	A..	A.A	...	...	..G	.AT	.T.	...	AC.	...	...	.
#Pore_103	A..	A.A	...	...	..G	.AT	.T.	...	AC.	...	...	.
#Pore_104	...	...	..G	...	..G	.AT	.T.	...	AC.	...	...	.
#Pore_105	...	...	T..	...	...	...	...	...	...	...	...	.
#Pore_106	...	A..	...	...	..G	GTG	...	...	...	...	...	.
#Pore_107	...	...	...	...	...	...	...	...	...	...	...	.
#Pore_108	...	...	...	...	...	...	...	...	...	...	...	.
#Pore_109	...	...	...	...	...	...	...	...	...	...	...	.
#Pore_110	...	...	...	...	...	...	...	...	...	...	...	.
#Pore_111	...	...	...	...	...	...	...	...	...	...	...	.
#Pore_112	...	...	...	...	...	...	...	...	...	...	...	.
#Pore_113	...	...	...	...	...	...	...	...	...	...	...	.
#Pore_114	...	...	...	...	...	...	...	...	...	...	...	.
#Pore_115	...	...	...	...	...	...	...	...	...	...	...	.
#Pore_116	...	C..	TGG	...	.C.	.AT	ATG	...	AC.	...	...	.
#Pore_117	.A.	A..	TGG	...	GGG	.AT	C.A	...	.C.	...	...	.
#Pore_118	.A.	A..	TGG	...	GGG	.AT	C.A	..A	AC.	...	...	.
#Pore_119	...	A..	...	...	...	...	...	...	...	...	...	.
#Pore_120	...	A..	...	...	..G	GTG	...	...	...	...	...	.
#Pore_121	...	A..	...	...	..G	GTG	...	...	...	...	...	.
#Pore_122	...	AA.	...	...	.G.	.A.	AT.	...	...	...	...	.
#Pore_124	...	A..	...	...	..G	G.G	...	...	...	...	...	.
#Pore_125	...	A..	...	...	..G	G.G	...	...	...	...	...	.
#Pore_126	...	A..	...	...	...	...	...	...	...	...	...	.
#Pore_127	...	A..	...	...	..G	G.G	...	...	...	...	...	.
#Pore_129	AC.	.CT	TT.	...	.C.	.A.	...	...	...	...	...	.

#Pore_131	...	...	...	...	...	...	.T.	...	...	...	...	.
#Pore_132	...	A..	...	...	..G	GTG	...	...	...	...	...	.
#Pore_133	...	A..	TGG	...	GGG	.AT	C.A	...	.C.	...	...	.
#Pore_134	...	A..	TGG	...	GGG	.AT	C.A	...	.C.	...	...	.
#Pore_136	...	A..	...	...	...	...	...	...	...	...	...	.
#Pore_138	...	A..	...	...	G..	...	...	...	...	...	...	.

Appendix 12

van Oosterhout, C., R. S. Mohammed, H. Hansen, G. A. Archard, M. McMullan, D. J. Weese, and J. Cable. 2007. Selection by parasites in spate conditions in wild Trinidadian guppies (*Poecilia reticulata*). International Journal for Parasitology 37:805-812.

Available online at www.sciencedirect.com
ScienceDirect
International Journal for Parasitology 37 (2007) 805–812
www.elsevier.com/locate/ijpara

Selection by parasites in spate conditions in wild Trinidadian guppies (*Poecilia reticulata*)

C. van Oosterhout [a,*], R.S. Mohammed [b], H. Hansen [c], G.A. Archard [c,1], M. McMullan [a], D.J. Weese [d], J. Cable [c]

[a] *Molecular Ecology and Fisheries Genetics Laboratory, University of Hull, Hull HU6 7RX, UK*
[b] *Department of Life Sciences, University of the West Indies, St. Augustine, Trinidad, West Indies*
[c] *School of Biosciences, Cardiff University, Cardiff CF10 3TL, UK*
[d] *Department of Biological Sciences, University of Maine, Orono ME 04469, USA*

Received 27 October 2006; received in revised form 20 December 2006; accepted 21 December 2006

Abstract

Gyrodactylids are ubiquitous fish parasites and yet, with the notable exception of *Gyrodactylus salaris*, few studies have reported the effect of these parasites on host survival in natural populations. Here, we assess the impact of the parasite load of gyrodactylids (*G. turnbulli* and *G. bullatarudis*) on the survival and migration of guppies (*Poecilia reticulata*) in their natural habitat of the Aripo River in Trinidad. The recapture rate of males declined by 19% with every additional parasite, a remarkably high figure given that the parasite load in this study ranged from zero to 20 worms. In addition, with an increased number of parasites, males were more prone to be recovered downstream. In contrast, no effect of parasitism was observed in females. The mean parasite load sharply declined after a series of flushing events during heavy seasonal downpours. The parasite load varied significantly between fish depending on their location in the river, and the size of the fish explained variation in parasite load between individuals. The present study indicates that tropical gyrodactylid parasites can play an important role in the ecology of natural fish populations, causing intense bouts of natural selection in guppies during heavy rains in the wet season.

Keywords: Evolution; Ecology; Mark–recapture; Fitness; MHC; Natural selection

1. Introduction

Parasites play a major role in the ecology and evolution of their host species (Haldane, 1949; Hamilton and Zuk, 1982), but relatively few studies have analysed the impact of parasitism on the survival probability of infected individuals in natural populations (e.g. Brown et al., 1995; Brown and Brown, 2004; Coonan et al., 2005; Forrester and Finley, 2006; Hawlena et al., 2006; Telfer et al., 2002). Mark–recapture analyses are commonly employed to estimate mortality of the host species (e.g. Forrester and Finley, 2006), which provides important data to infer fitness costs and estimate parasite-mediated selection. This is particularly important because the infection status of hosts (i.e. their parasite load) may not translate to a true fitness measure as infected hosts could simply survive and cope with their parasites (e.g. Dale et al., 1996). Quantifying the long-term effects of parasites on host's fitness and assessing spatial and temporal variation in parasite prevalence are essential to our understanding of how natural selection operates in wild populations.

The guppy (*Poecilia reticulata*) is an important ecological and evolutionary model organism that has been used extensively to study adaptive evolution in response to

* Corresponding author. Tel.: +44 (0) 1482 465505; fax: +44 (0) 1482 465458.
E-mail address: c.van-oosterhout@hull.ac.uk (C. van Oosterhout).
[1] Present address: School of Biosciences, University of Bristol, Bristol, BS8 1UG, UK.

0020-7519/$30.00
doi:10.1016/j.ijpara.2006.12.016

natural (Reznick et al., 1996a,b, 2001) and sexual selection (van Oosterhout et al., 2003b; Magurran, 2005). Guppies have an almost global distribution but their natural range is confined to parts of South America, Trinidad and Tobago. In Trinidad, guppy populations occur in distinct high- and low-predation environments, typically in the downstream and upstream parts of rivers, respectively. Low predation populations co-occur with a small killifish, *Rivulus hartii*, that preys on small guppies (particularly males and juveniles), whereas high predation populations co-occur with a number of predators, the most important being a large cichlid, *Crenicichla alta* (Seghers, B.H., 1973. An analysis of geographic variation in the anti predator adaptations of the guppy, *Poecilia reticulata*. Ph.D. Dissertation. University of British Columbia, Vancouver, Canada; Liley and Seghers, 1975; Mattingly and Butler, 1994). Studies have documented differences in a diverse range of traits associated with predation regime and habitat, including colour and life-history patterns, and reproductive behaviours (reviewed in Endler, 1995; Magurran et al., 1995; see also Reznick et al., 2001).

In the Northern Range Mountains of Trinidad, guppies live in many streams in a seasonal tropical rainforest, an environment that is characterised by heavy annual rainfall of up to 4 m per year. In the rainy season (May–December), downpours can transform the aquatic habitat in a matter of minutes. Extensive changes in stream hydrology can have a severe impact on poeciliid populations. For example, Chapman and Kramer (1991) found that *P. gilli* populations in Costa Rica lost on average 75% of their individuals by involuntary flushing during floods and most of these fish died. Annual flooding is likely to represent a bout of extreme selection on guppies, and guppy biomass (in grams per square meter) was between 22% and 92% lower after floods in a study by Grether et al. (2001). Such extreme mortality (or involuntary displacement) will have profound effects on the biology of these fish and yet the effects on the ecology and evolution of guppies have scarcely been investigated.

Gyrodactylids are ubiquitous fish ectoparasites (Bakke et al., in press). *Gyrodactylus turnbulli* and *G. bullatarudis* are important guppy parasites, with a prevalence of up to 54% in some Trinidadian populations (Lyles, A.M., 1990. Genetic variation and susceptibility to parasites: *Poecilia reticulata* infected with *Gyrodactylus turnbulli*. Ph.D. Dissertation, Princeton University). These monogeneans can be directly transmitted between hosts and are highly contagious. Adult worms reproduce *in situ* on the host and give birth to fully-grown offspring that already contain developing embryos (reviewed in Cable and Harris, 2002). Gyrodactylids have a rapid reproduction rate (generation time of only 24 h at 25°C; Scott, 1982) leading to exponential population growth (van Oosterhout et al., 2003a; Cable and van Oosterhout, in press). Infections can cause a reduced feeding response (van Oosterhout et al., 2003a) and marked fin-clamping (Cable et al., 2002) which is likely to reduce the swimming performance of heavily infected fish. Mortality rates of 50% have been recorded in some experimental infections in the laboratory (Houde, 1997). Selection against parasite-infected males in laboratory experiments (Kennedy et al., 1987; Houde and Torio, 1992; Houde, 1997; López, 1998) furthermore suggests that gyrodactylids could be a formidable selective force in natural guppy populations, particularly in upland populations that lack large fish predators. Despite the potential importance of gyrodactylid parasites on the biology of their guppy host, the effect of infections in natural populations has been ignored until now.

Selection by parasites is thought to maintain a high level of immunogenetic variation in the genes of the major histocompatibility complex (MHC) in some wild guppies in Trinidad (van Oosterhout et al., 2006b). The MHC diversity of ornamental guppy strains is less than 20% of that found in wild populations (van Oosterhout et al., 2006a). However, particularly in the upstream habitats, wild guppy populations can be subject to severe random genetic drift because of their small effective population size. For example, the Naranjo Tributary population has an effective population size (N_e) of about 100 (van Oosterhout et al., 2006b). This means that to maintain a high level of immunogenetic diversity, natural selection must counteract the eroding effects of genetic drift. Computer simulations indicated that a selection coefficient(s) greater than 0.2 was required to maintain the observed level of MHC diversity in this population (van Oosterhout et al., 2006b). Although this study showed that the parasite load on wild-caught guppies of the Naranjo Tributary was considerable, no data were available to estimate host mortality rate or selection intensity by parasites.

The aim of the present study was to analyse the effects of gyrodactylid parasites on the migration tendency and recapture rate (i.e. an estimator of survival) of guppies during the transition from the dry to the wet season. We hypothesised that during the first heavy rains, selection against parasites would be severe, and we predicted that heavily infected fish would be flushed downstream. In addition, we analysed the spatial and temporal variation in parasite load and tested whether there were associations between the sex, size and parasite load of hosts and their migration tendency.

2. Materials and methods

2.1. Mark–recapture experiment

A mark–recapture experiment was carried out in the Naranjo Tributary of the Upper Aripo River (Grid reference: PS693100 E and 1181800 N) in the Northern Mountain Range of Trinidad. All adult guppies (191 in total, 82 females and 109 males with standard lengths (SL) of 13–32 and 13–21 mm, respectively) were caught using seine nets. Damage and loss of parasites was minimised by scooping guppies out of the net with plastic pots, thereby minimising contact with the net. The fish were caught by exhaustive

sampling of four pools along a 103-m stretch of the Mid-Naranjo tributary on 21 May 2006 (see Supplementary Figs. S1–S3). As it is possible to collect all of the guppies in an individual pool (Reznick et al., 1996a), we chose pools as the unit of investigation. The pools were of approximately similar size and interconnected by riffles. Due to the relatively fast currents, up to 3.5 m s^{-1} in our focal site, riffles represent unsuitable guppy habitat that hinder migration and guppies are observed in these parts of the river only occasionally. The experiment was conducted during the transition from the dry to the wet season and started just before the first heavy downpour (on 26 May 2006).

The guppies were anaesthetized with 0.02% tricaine methanesulfonate (MS222) and the number of gyrodactylid parasites was counted using a stereo-microscope. Sex was recorded and the SLs of all fish were measured before each individual was marked with a unique combination of two visible implant elastomer (VIE) markings (Northwest Marine Technology Inc.). Sub-dermal injections were made using six colours at three possible positions on the left and/or right hand side of the fish: in the tail muscles below and above the spine, and directly under the dorsal fin. Each mark was a thin ca. 1.5 mm line consisting of less than 0.5 mg elastomer which does not appear to affect the behaviour or mobility of the fish. These marks have been successfully used in other studies on guppies (e.g. Croft et al., 2003). For 2 days after marking, the fish were kept in four 80 L aquaria according to the pool they were captured from and received a preventive broad-spectrum antibiotic treatment (Binox™ 5 g per 100 L and Polyaqua™ 12.5 ml per 100 L). During the marking procedure, fish unavoidably lost on average 29.4% of their gyrodactylids. To make a valid temporal comparison of parasite loads, all parasite numbers reported here are those observed after VIE marking on 21 May 2006. No guppies died in captivity and all fish were released at their original site of capture on 23 May 2006.

During three subsequent recaptures (30 May, 4 , 5, and 26 June 2006), fish were collected from the study site, identified and released back into the pool from which they were captured. These data were used to establish migration rates (Note that we refer to movement between pools as migration even though this may have been involuntary and caused by floods). Parasites were counted on all recaptured (marked) fish on 4, 5, and 26 June. During resampling, guppies were collected from three additional pools to monitor whether fish had moved outside the focal area. These included Pools 0 and 5 located 65 m upstream and 50 m downstream of the original sampling location, respectively, and Pool 6 ca. 500 m downstream of the focal site. Pool 6 was only sampled once on 14 June, when 250 guppies were collected but no marked fish were found. The entire sampling area covered a stretch of 218 m (from Pools 0 to 5) or ca. 670 m (when including Pool 6). During sampling and resampling, pools and intervening riffles were visited repeatedly throughout the day until no more guppies were observed. Only five fish were found in the riffles and all were marked; three males close to the river bank between Pools 3 and 4, and one male and one female in vegetation of the river bank a few metres downstream from Pool 1. These fish were allocated to the nearest pool. On 26 June, a large sample of 416 unmarked guppies collected from Pools 0 to 5 was also screened for parasites.

2.2. Statistical analysis

The data were analysed to test whether the recapture rate and migration pattern of individual fish was dependent upon on an individual's parasite load (expressed as the number of gyrodactylid worms), gender and SL of the guppy. The recapture rate is defined as whether or not a guppy has been recaptured. Migration pattern is the movement of a guppy upstream, stationary or downstream. Data were also used to test spatial and temporal variation in parasite abundance, and to test whether variation in the parasite load was predicted by the size and gender of the host.

Differences between males and females in migration tendency (i.e. upstream versus downstream migration and remaining stationary in one pool) were tested using χ-square analysis and Binomial tests. A Binomial test was also used to test for unequal sex ratio.

A general linear model (GLM) was used to analyse spatial variation in parasite abundance, and test whether variation in the parasite load of guppies was predicted by the size and gender of hosts. Parasite numbers on individual guppies were transformed using natural logarithms, $\mathrm{Ln}(N_t + 0.1)$. This resulted in homogeneity of variances, as established using Bartlett's Test. The location of the sample ('Pool') was used as a random factor with four factor levels. 'SL' was used as covariate and was nested within 'Sex' as a fixed factor. Logarithmic transformations did not result, however, in equal variances of temporal samples, violating the assumption of a parametric test. Hence, Kruskal–Wallis tests were performed to analyse whether guppies differed in parasite load over time, and to assess the impact of heavy rains on parasite and guppy populations.

A binary logistic regression analysis (logit) was used with a dichotomous dependent variable (re-captured or not recaptured) to test whether the recapture of guppies was associated with the number of parasites ('Parasite load'), 'SL' and gender of the guppy ('Sex'). The model has three predictors ('Parasite load' and 'SL' both as covariates, and 'Sex' as a fixed factor). The model was fitted using an iterative re-weighted least squares algorithm to obtain maximum-likelihood estimates of all parameters. The log-likelihood was used to test whether the coefficients of the predictors were significantly different from zero. A logit link function was employed to calculate the odds ratio and its 95% confidence interval (CI). The odds ratio estimates the change in the link function with a one unit change in the predictor (e.g. 1 mm increase in SL or one

additional parasite), holding all other factors and covariates constant (Sokal and Rolf, 1995).

A logit analysis was also used to test whether the decline in recapture rate over time differed between the sexes. In this model, the dependent variable was whether a guppy was recaptured or not recaptured (coded as '1' and '0', respectively). The last sample in which a guppy was observed 'Resample' (original capture coded as '0', and first, second and third recaptures coded as '1', '2' and '3') was used as a covariate and crossed with 'Sex' as the factor. The 'Resample' × 'Sex' interaction tests whether the maximum-likelihood estimates of the slopes differed between the sexes, i.e. whether males and females differed in the decline of recapture rate over time.

An ordinal logistic regression was used to test the association between migration-direction (i.e. upstream migration = +1, static or no movement = 0 and downstream migration = −1) and predictor variables. As before, in this model 'SL' and 'Parasite load' were covariates and 'Sex' was used as the factor. The 'Pool' did not explain differences in migration direction and was omitted from the analysis. Due to significant interactions between 'Sex' and 'SL', and 'Sex' and 'Parasite load', the models were also calculated separately for males and females.

All statistical analyses were performed using Minitab 12.1.

3. Results

Table 1 shows that males were not significantly more mobile than females (males: 29 stationary and 36 mobile; females: 29 stationary and 17 mobile; $\chi^2 = 3.666$, df = 1, $P = 0.056$). However, compared with females, males were more likely to migrate upstream rather than migrate downstream or remain in the same pool ($\chi^2 = 3.96$, df = 1; $P = 0.047$). Despite heavy rains and spate conditions at the start of the wet season, there was no overall tendency for fish to be recovered downstream (Binomial test for females: $P = 0.072$, and for males: $P = 0.434$). The sex ratio in the original sample of marked and released guppies (109 males and 82 females) was significantly male-biased (Binomial test: $P = 0.030$).

Significant spatial variation existed in guppy parasite load between pools (see Table 2), with mean infection intensity (±SEM) ranging from 1.18 (±0.46) in Pool 2 to 5.31 (±0.93) in Pool 3. No significant differences were detected between males and females when 'SL' was used as a covariate nested within 'Sex'. However, 'SL' explained significant variation in 'Parasite load' of males, with larger individuals carrying a higher parasite load ($T = 2.31$; $P = 0.022$, see Table 3). For females there was also a positive relationship between 'Parasite load' and 'SL', but this was not statistically significant ($T = 1.90$; $P = 0.059$, see Table 3).

Significant differences existed in parasite load between temporal samples of all captured fish (i.e. marked and unmarked guppies) (Kruskal–Wallis test: $H = 78.46$;

Table 1
Migration and recapture of female and male *Poecilia reticulata*

Migration	Females	Males
Downstream	12	19
Upstream	5	17
Static	29	29
Not re-caught	36	44
Total released	82	109

Table 2
General linear model with the (Ln-transformed) number of parasites as response variable and 'Pool' as random factor, and standard length 'SL' nested within 'Sex' as covariate

Source	df	MS	F	P
Pool	3	19.100	8.85	<0.001
SL (Sex)	2	9.139	4.39	0.014
Sex	1	6.010	2.89	0.091
Error	184	2.080		
Total	190			

Table 3
Regression analyses of parasite load on standard length for the female and male guppies, separately

Term	Coef	SD	T	P
Constant	−2.3282	0.9005	−2.59	0.010
Females	0.05849	0.03084	1.90	0.059
Males	0.21192	0.09187	2.31	0.022

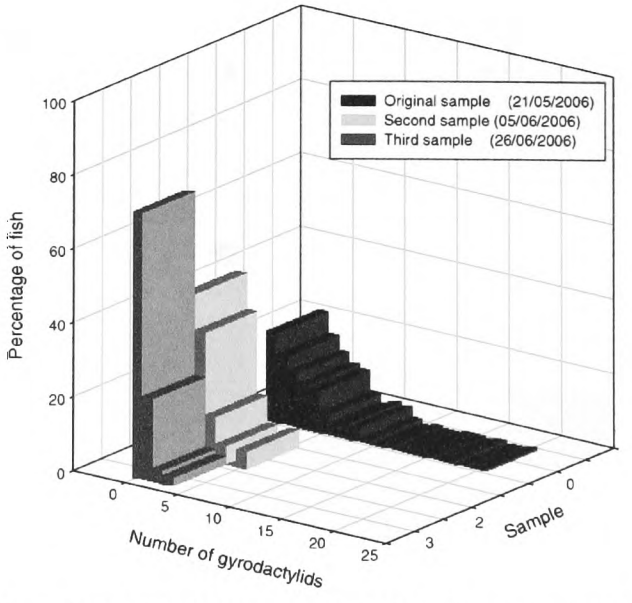


Fig. 1. Temporal variation in parasite load (number of gyrodactylid worms per guppy) in the Naranjo Tributary in the transition from the dry season (original sample collected on 21 May 2006) to the wet season (samples collected on 5 and 26 June 2006).

df = 2; $P < 0.001$). Fig. 1 shows that the number of parasites detected in the later samples was considerably reduced compared with that of early samples. Also, the experimental group of marked guppies appeared to have lost parasites over time. In the original sample of marked guppies collected in the dry season, 145 out of 191 fish (75.9%) were infected with gyrodactylids. This number decreased to 21 out of 84 guppies (25.0%) in the third recapture (wet season) ($\chi^2 = 18.271$; df = 1; $P < 0.001$). Parasite load of marked and unmarked guppies did not differ significantly in the first resampling (Kruskal–Wallis tests: $H = 1.36$; df = 1; $P = 0.245$, n.s.) and third resampling ($H = 0.02$; df = 1; $P = 0.884$, n.s.) events. These results suggest that temporal differences in parasite load are not an artefact caused by the sampling of guppies, the marking procedure or stress associated with handling, but rather that the reduction in parasite numbers was a natural phenomenon, possibly associated with heavy rains in the wet season.

Binary logistic regression (logit) analysis with recapture rate (recaptured or not) as the response variable showed that there was a significant 'Sex' × 'Parasite load' interaction ($Z = 2.86$; $P = 0.004$), indicating that the effect of parasites on recapture rate differs between the sexes. The analysis was therefore performed separately on males and females. In females, neither 'SL' ($Z = -0.12$; $P = 0.904$) nor 'Parasite load' ($Z = 1.66$; $P = 0.096$) explained significant variation in the recapture rate, and similarly, 'SL' of males did not explain variation in recapture rate either ($Z = 0.87$; $P = 0.383$). However, the number of parasites significantly reduced the recapture rate of males (mean (±SEM) $b = -0.214(\pm 0.088)$; $Z = -2.44$; $P = 0.015$) (Fig. 2). The negative coefficient from the logit analysis [mean Odds ratio = 0.81 (95% CI: 0.68–0.96)] indicates that recapture declined with increased numbers of parasites. Infected males were 81% as likely to be recaptured compared with males with one less parasite. In other words, recapture rate declined by 19% with every additional parasite.

A logit analysis showed that the recapture rate of males declined faster over resampling events than that of females ($Z = 2.08$; $P = 0.038$; Odds ratio = 1.47 (95% CI: 1.02–2.10). Of the marked females, 46.3%, 48.8% and 28.0% were recaptured in the first, second and third resampling events, which compare with 39.4%, 40.4% and 16.5%, respectively, of marked males recaptured in the same periods.

An ordinal logistic regression was performed with migration direction (i.e. upstream, stationary or downstream movement) as response variable. This showed there was a significant 'Sex' × 'SL' interaction ($Z = -2.36$; $P = 0.018$), indicating that migration direction differed depending on the gender and size of the guppy. Therefore, this analysis was also performed separately for males and females. Neither 'SL' ($Z = 0.34$; $P = 0.734$) nor the number of parasites ($Z = 0.28$; $P = 0.782$) affected the migration direction or tendency of females. In contrast, with an increased number of parasites, males were more prone to be recovered downstream [$Z = -2.28$; $P = 0.022$; Odds ratio = 0.73 (95% CI: 0.56–0.96)] (Fig. 3). Also, the size of males affected the direction of migration, with larger males going upstream and smaller males going downstream [$Z = 2.67$; $P = 0.008$; Odds ratio = 1.59 (95% CI: 1.13–2.24)] (Fig. 4).

In contrast to the male-biased sex ratio in the original sample of all guppies, the sex ratio of a sample of unmarked guppies collected in Pools 0–5 on 26 June 2006 (third resampling event, wet season) showed a significant

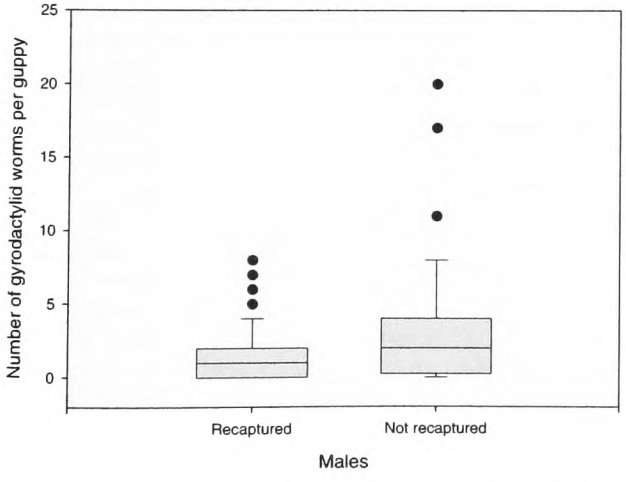


Fig. 2. Number of gyrodactylid worms on male guppies that were either recaptured or not recaptured. The dots represent outliers, the bars the lower and upper limits, and the box represents the first and third quartile value with the median.

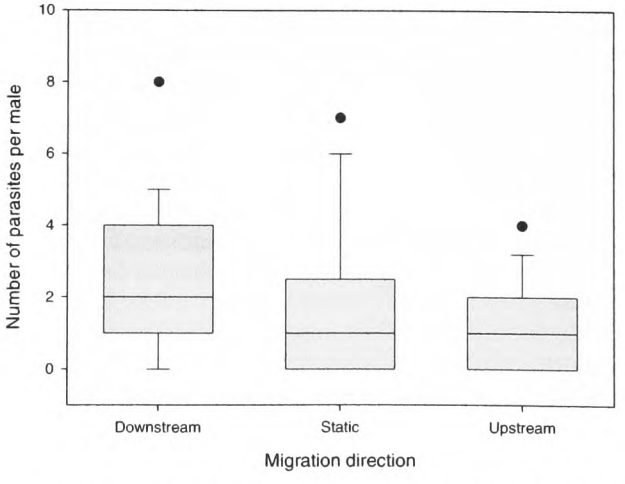


Fig. 3. The number of gyrodactylid parasites of male guppies in relation to the direction of migration (downstream, static in the same pool, or upstream). The dots represent outliers; the bars, the lower and upper limits; and the box represents the first and third quartile value with the median.

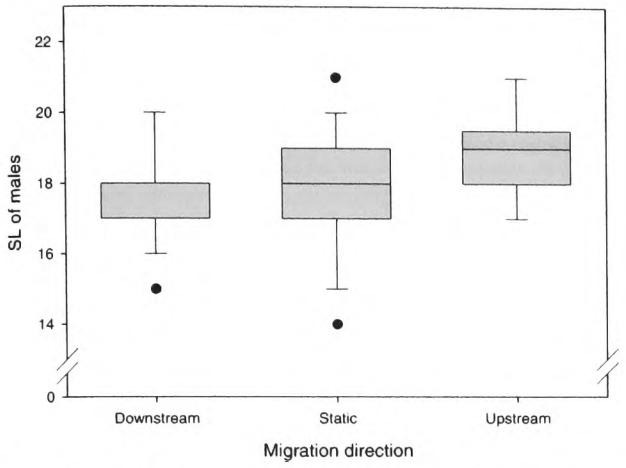


Fig. 4. The standard length (in mm) of male guppies in relation to the direction of migration (downstream, static in the same pool or upstream). The dots represent outliers; the bars, the lower and upper limits; and the box represents the first and third quartile value with the median.

female-biased sex ratio (189 males and 227 females; Binomial probability: $P = 0.0348$).

4. Discussion

The parasite load of gyrodactylids affected the recapture rate of males but not females. The recapture rate of males declined by 19% with every additional parasite, a remarkably high figure given that natural parasite loads ranged from zero to 20 worms in the current study, and up to 100 parasites in a previous study (van Oosterhout et al., 2006b). Compared with the dry season samples, we observed a significant reduction in parasite load in the wet season samples of both marked and unmarked fish. These observations suggest that individual male guppies with high parasite burden were selectively removed from the population during heavy downpours, which reduced the mean parasite load in the population.

The handling of fish during marking resulted in a direct loss of parasites, but we did not observe differences in the parasite burden of marked and unmarked guppies when resampling. It is furthermore conceivable that males were more prone to stress-induced mortality after marking, which could explain the low recapture rate of marked males. However, we found a significant female-biased sex ratio of unmarked guppies in the later wet season samples which suggests that irrespective of marking, strong selection acts on males during the transition from the dry to the wet season. Guppies that were not recaptured possibly survived and simply moved outside the sampled area. However, Croft et al. (2003) recorded movements up to 184 m for males and up to 60 m for females. Given that our research area exceeded the maximum migration rate estimated by Croft et al. (2003), and that we found no marked individuals in a large sample of guppies 500 m downstream from the focal site, we are confident that only a few individuals voluntarily moved outside the sampling area. Furthermore, in another poeciliid, involuntary flushing during floods resulted in the death of most fish (Chapman and Kramer, 1991; see Introduction). In addition, even if the fish that were flushed downstream managed to survive, they were removed from the study population as fast flowing rapids would have hindered upstream migration. Hence, from an evolutionary perspective, these fish effectively died as they would have been unable to reproduce in this population.

A recent analysis of MHC variation in the same guppy population suggested that strong selection has been acting on these immunological genes (van Oosterhout et al., 2006b). The present study corroborates this suggestion and indicates that gyrodactylid parasites can be a formidable selective force in this population. The question remains, however, whether the MHC genotype of guppies is associated with resistance to gyrodactylid infections and/or associated infections, particularly bacteria and fungi (Bakke et al., 2006).

Besides the effects of parasitism, male size also affected migration, with larger individuals tending to move upstream. Larger males are able to move upstream, possibly due to their better ability to resist flushing compared with smaller males. Males were also more likely to migrate upstream than females, but the overall migration tendency did not differ significantly between the sexes. Croft et al. (2003) examined the movement patterns of guppies in the Arima River and in the Northern Mountain Range of Trinidad, and they too found a positive relationship between distance moved and SL. Furthermore, they found male-biased movement, consistent with sexual asymmetry in reproductive investment (Croft et al., 2003).

The current study shows, however, that although males had a significantly higher tendency to migrate upstream than females, no significant male bias migration was detected. Possibly, heavy rainfall resulted in (involuntary) downstream migration or flushing. For example, Chapman and Kramer (1991) found that male and juvenile *P. gilli* were more likely to disperse downstream after floods, and Grether et al. (2001) showed that guppy biomass decreased by between 22% and 92% after floods in the wet season in Trinidad. Such flushing events could have offset any attempts by male guppies to migrate upstream, and may also have inflated the apparent migration rate of females. Indeed, females showed a marginally significant downstream migration, which suggests that dispersal was involuntary and mediated by flushing events, assuming that females have a low tendency to move (Croft et al., 2003). It has been suggested that upstream movement allows riverine populations to maintain positions within a stream (Kaya, 1991). We suggest that natural selection favours upstream migration, and that selection is particularly strong for males as, due to their smaller size, they are more prone to flushing during downpours.

Relatively little is known about how parasitism affects the migration tendency of hosts (see Boulinier et al., 2001 for review). Here, we show that dispersal decisions or involuntary displacement of hosts may depend on several conditions, including the size and gender of the host, as well as its parasite load. The profound impact of flooding events was particularly noteworthy, especially because this may affect many other riverine species. Heavily parasitized guppy hosts display fin clamping (Cable et al., 2002) and their reduced swimming ability probably renders infected guppies less able to hold their position in the river during flushing events. Studying the effects of parasitism in a sheltered and relatively benign laboratory environment is unlikely to reveal the full fitness consequences for infected hosts, and only in a natural setting can the full impact of parasitism be appreciated. Flushing and purging of the most heavily infected individuals (males in particular) could explain the large differences in parasite load in the transition from the dry to the wet season (Fig. 1), and may also be partly responsible for the male-biased sex ratio in the upland habitats in the wet season.

Grether et al. (2001) showed that during wet season floods, both the biomass of algae and guppies were markedly reduced, but there was no net change in algal availability for guppies after floods. They suggested that factors other than, or in addition to, algal production might limit guppy populations and discussed the role of *Rivulus*. Indeed, *Rivulus* appears to be the most serious predator of guppies in these streams (Liley and Seghers, 1975; Reznick et al., 1990), but top–down control of guppies by *Rivulus* seems very unlikely according to Grether et al. (2001), given the negative effect of guppies on *Rivulus* densities (Gilliam et al., 1993). Grether et al. (2001) furthermore suggested that floods keep guppy populations below carrying capacity. This suggestion seems reasonable as in their study guppy biomass was reduced by up to 92%, indicating the profound influence of floods (Grether et al., 2001). The present study suggests that parasites might play an additional regulatory role in the density of guppies in headwater habitats: parasites regulate guppy population size and strong bouts of selection during floods cap the abundance of both parasites and guppies by selectively removing the most heavily infected fish.

Acknowledgements

This study was supported by a European Community Framework Programme 6 Marie Curie Host Fellowship for Transfer of Knowledge (MTKD-CT-2005-030018) and the Natural Environment Research Council, U.K., Ph.D. studentship to M.M. (NER/S/A/2005/13362A) and Advanced Research Fellowship to J.C. (NER/J/S/2002/00706). D.W. was supported by the National Science Foundation (DEB 0235605). All handling and marking of fish was carried out in accordance with UK Home Office regulations under PPL 30/1834.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ijpara.2006.12.016.

References

Bakke, T.A., Cable, J., Harris, P.D., in press. The biology of gyrodactylid monogeneans: the "Russian- Doll Killers". Adv. Parasitol.

Bakke, T.A., Cable, J., Østbø, M., 2006. The ultrastructure of hypersymbionts on *Gyrodactylus salaris* (Monogenea) infecting Atlantic salmon (*Salmo salar*). J. Helminthol. 80, 377–386.

Boulinier, T., McCoy, K.D., Sorci, G., 2001. Dispersal and parasitism. In: Clobert, J., Danchin, E., Dhondt, A.A., Nichols, J.D. (Eds.), Dispersal. Oxford University Press, Oxford, pp. 169–179.

Brown, C.R., Brown, M.B., 2004. Group size and ectoparasitism affect daily survival probability in a colonial bird. Behav. Ecol. Sociobiol. 56, 498–511.

Brown, C.R., Brown, M.B., Rannala, B., 1995. Ectoparasites reduce long-term survival of their avian host. Proc. Roy. Soc. London B 262, 313–319.

Cable, J., Harris, P.D., 2002. Gyrodactylid developmental biology: historical review, current status and future trends. Int. J. Parasitol. 32, 255–280.

Cable, J., van Oosterhout, C., in press. The role of innate and acquired resistance in two natural populations of guppies (*Poecilia reticulata*) infected with the ectoparasite *Gyrodactylus turnbulli*. Biol. J. Linn. Soc.

Cable, J., Scott, E.C.G., Tinsley, R.C., Lazarus, C.M., 2002. Behavior favoring transmission in the viviparous monogenean *Gyrodactylus turnbulli*. J. Parasitol. 88, 183–184.

Chapman, L.J., Kramer, D.L., 1991. The consequences of flooding for the dispersal and fate of poeciliid fish in an intermittent tropical stream. Oecologia 87, 299–306.

Croft, D.P., Albanese, B., Arrowsmith, B.J., Botham, M., Webster, M., Krause, J., 2003. Sex-biased movement in the guppy (*Poecilia reticulata*). Oecologia 137, 62–68.

Coonan, T.J., Schwemm, C.A., Roemer, G.W., Garcelon, D.K., Munson, L., 2005. Decline of an island fox subspecies to near extinction. Southwest Nat. 50, 32–41.

Dale, S., Kruszewicz, A., Slagsvold, T., 1996. Effects of blood parasites on sexual and natural selection in the pied flycatcher. J. Zool. 238, 373–393.

Endler, J.A., 1995. Multiple-trait coevolution and environmental gradients in guppies. Trends Ecol. Evol. 10, 22–29.

Forrester, G.E., Finley, R.J., 2006. Parasitism and a shortage of refuges jointly mediate the strength of density dependence in a reef fish. Ecology 87, 1110–1115.

Gilliam, J.F., Fraser, D.F., Alkins-Koo, M., 1993. Structure of a tropical stream fish community: a role for biotic interactions. Ecology 74, 1856–1870.

Grether, G.F., Millie, D.F., Bryant, M.J., Reznick, D.N., Mayea, W., 2001. Rain forest canopy cover, resource availability, and life history evolution in guppies. Ecology 82, 546–1559.

Haldane, J.B.S., 1949. Disease and evolution. Symposium sui fattori ecologici e genetici della speciazione negli animali. In: Selected Genetic Papers of J. B. S. Haldane. Garland Publ. Inc., New York and London 1990., pp. 325–334.

Hamilton, W.D., Zuk, M., 1982. Heritable true fitness and bright birds: a role for parasites?. Science 218 384–387.

Hawlena, H., Abramsky, Z., Krasnov, B.R., 2006. Ectoparasites and age-dependent survival in a desert rodent. Oecologia 148, 30–39.

Houde, A., 1997. Sex, Color and Mate Choice in Guppies. Princeton University Press, Princeton.

Houde, A.E., Torio, A.J., 1992. Effect of parasitic infection on male color pattern and female choice in guppies. Behav. Ecol. 3, 346–351.

Kaya, C.M., 1991. Rheotactic differentiation between fluvial and lacustrine populations of Arctic grayling (*Thymallus arcticus*), and implications for the only remaining indigenous population of fluvial Montana grayling. Can. J. Fish. Aquat. Sci. 48, 53–59.

Kennedy, C.E.J., Endler, J.A., Poyton, S.L., McMinn, H., 1987. Parasite load in guppies predicts mate choice in guppies. Behav. Ecol. Sociobiol. 21, 291–295.

Liley, N.R., Seghers, B.H., 1975. Factors affecting the morphology and behaviour of guppies in Trinidad. In: Baerends, G., Beer, C., Manning, A. (Eds.), Function and Evolution in Behavior. Clarendon Press, Oxford, UK, pp. 92–118.

López, S., 1998. Acquired resistance affects male sexual display and female choice in guppies. Proc. Roy. Soc. London B 265, 717–723.

Magurran, A.E., 2005. Evolutionary Ecology: The Trinidadian Guppy. Oxford University Press, Oxford.

Magurran, A.E., Seghers, B.H., Shaw, P.W., Carvalho, G.R., 1995. The behavioural diversity and evolution of guppy (*Poecilia reticulata*) populations in Trinidad. Adv. Study Behav. 24, 155–202.

Mattingly, H.T., Butler, M.J., 1994. Laboratory predation on the Trinidadian guppy: Implications for the size-selective predation hypothesis and guppy life history evolution. Oikos 69, 54–64.

Reznick, D.N., Bryga, H., Endler, J.A., 1990. Experimentally induced life-history evolution in a natural population. Nature 346, 357–359.

Reznick, D.N., Butler, I., Rodd, F.H., Ross, P., 1996a. Life history evolution in guppies (*Poecilia reticulata*). 6. Differential mortality as a mechanism for natural selection. Evolution 50, 1651–1660.

Reznick, D.N., Rodd, F.H., Cardenas, M., 1996b. Life history evolution in guppies (*Poecilia reticulata*). IV. Parallelism in life-history phenotypes. Am. Nat. 147, 319–338.

Reznick, D.N., Butler, M.J., Rodd, H., 2001. Life-history evolution in guppies. VII. The comparative ecology of high- and low-predation environments. Am. Nat. 157, 126–140.

Scott, M.E., 1982. Reproductive potential of *Gyrodactylus bullatarudis* (Monogenea) on guppies (*Poecilia reticulata*). Parasitology 85, 217–236.

Sokal, R.R., Rolf, F.J., 1995. Biometry: the principals and practice of statistics in biology research, third ed. W.F. Freeman, New York.

Telfer, S., Bennett, M., Bown, K., Cavanagh, R., Crespin, L., Hazel, S., Jones, T., Begon, M., 2002. The effects of cowpox virus on survival in natural rodent populations: increases and decreases. J. Anim. Ecol. 71, 558–568.

van Oosterhout, C., Harris, P.D., Cable, J., 2003a. Marked variation in parasite resistance between two wild populations of the Trinidadian guppy, *Poecilia reticulata* (Pisces: Poeciliidae). Biol. J. Linn. Soc. 79, 645–651.

van Oosterhout, C., Trigg, R.E., Carvalho, G.R., Magurran, A.E., Hauser, L., Shaw, P.W., 2003b. Inbreeding depression and genetic load of sexually selected traits: how the guppy lost its spots. J. Evol. Biol. 16, 273–281.

van Oosterhout, C., Joyce, D.A., Cummings, S.M., 2006a. Evolution of MHC class IIB in the genome of wild and ornamental guppies, *Poecilia reticulata*. Heredity 97, 111–118.

van Oosterhout, C., Joyce, D.A., Cummings, S.M., Blais, J., Barson, N.J., Ramnarine, I., Mohammed, R., Persad, N., Cable, J., 2006b. Balancing selection, random genetic drift and genetic variation at the Major Histocompatibility Complex (MHC) in two wild populations of guppies (*Poecilia reticulata*). Evolution 60, 2562–2574.

Appendix 13

Cummings, S. M., M. McMullan, D. A. Joyce, and C. van Oosterhout. 2010. Solutions for PCR, cloning and sequencing errors in population genetic analysis. Conservation Genetics 11:1095-1097.

Conserv Genet
DOI 10.1007/s10592-009-9864-6

SHORT COMMUNICATION

Solutions for PCR, cloning and sequencing errors in population genetic analysis

S. M. Cummings · M. McMullan · D. A. Joyce · C. van Oosterhout

Received: 24 November 2008 / Accepted: 19 February 2009

Abstract PCR and sequencing artefacts can seriously bias population genetic analyses, particularly of populations with low genetic variation such as endangered vertebrate populations. Here, we estimate the error rates, discuss their population genetics implications, and propose a simple detection method that helps to reduce the risk of accepting such errors. We study the major histocompatibility complex (MHC) class IIB of guppies, *Poecilia reticulata* and find that PCR base misincorporations inflate the apparent sequence diversity. When analysing neutral genes, such bias can inflate estimates of effective population size. Previously suggested protocols for identifying genuine alleles are unlikely to exclude all sequencing errors, or they ignore genuine sequence diversity. We present a novel and statistically robust method that reduces the likelihood of accepting PCR artefacts as genuine alleles, and which minimises the necessity of repeated genotyping. Our method identifies sequences that are unlikely to be a PCR artefact, and which need to be independently confirmed through additional PCR of the same template DNA. The proposed methods are recommended particularly for population genetic studies that involve multi-template DNA and in studies on genes with low genetic diversity.

Keywords Major histocompatibility complex · MHC · Mutation rate · Null alleles · PCR · Sequencing errors · Population genetic analysis

S. M. Cummings · M. McMullan · D. A. Joyce · C. van Oosterhout (✉)
Evolutionary Biology Group, School of Biological Sciences, University of Hull, Cottingham Rd, Hull HU6 7RX, UK
e-mail: c.van-oosterhout@hull.ac.uk

M. McMullan
e-mail: m.mc-mullan@biosci.hull.ac.uk

Published online: 01 March 2009

Introduction

The generation of sequence artefacts during PCR is an insufficiently recognised problem in conservation genetic studies. PCR error rate is not negligible, with approximately 10% of all sequences containing one or more PCR errors when amplifying a typical 250 bp sequence (Keohavong and Thilly 1989; Kobayashi et al. 1999). PCR and sequence artefacts increase apparent gene diversity and thereby inflate estimates of the effective population size (N_e). Assuming an infinite allele model (IAM), the effective population size for neutral genes can be estimated as $N_e = H_e/[4U(1 - H_e)]$, where U is the mutation rate of the amplified sequence, and H_e the heterozygosity. Given a PCR error rate of 10%, this would lead to an overestimation of N_e of between 20 and 50% for a polymorphic locus with $0.1 \leq H_e \leq 0.5$. The overestimation of the effective population size is most severe for loci with low gene diversity, and hence, PCR errors can significantly affect conservation genetic studies of small populations.

Studies involving multi-template PCR, such as those on the major histocompatibility complex (MHC) multigene family, use various protocols to minimise PCR error. These include minimising the number of PCR cycles (Zylstra et al. 1998) and reamplification of a diluted initial PCR product, a "reconditioning PCR" (Thompson et al. 2002). Separate amplifications with different primer pairs are used to identify potential null alleles (van Oosterhout et al. 2006b) and sequences are confirmed when observed in two or more samples. However, confirmation of sequences is not always obtained by reamplifying the same template DNA, as was suggested by Lukas et al. (2004).

Here we use cloned copies of exon 2 of MHC class II *DAB* alleles amplified from guppy (*Poecilia reticulata*) cDNA to quantify the rate of PCR misincorporations. We

propose a statistically robust method of quantifying the probability of obtaining PCR artefacts and present an example analysis of sequences sharing the same PCR error.

Materials and methods

RNA extraction, amplification and cloning

RNA was extracted and synthesised from the frozen livers of two male guppies (see van Oosterhout et al. 2006b). MHC was amplified using DABdegF and DABR4 in exon 3 (van Oosterhout et al. 2006a) allowing differentiation of expressed from genomic products. Amplification took place in a total volume of 25 μl including 12.5 pmoles of DABdegF, 2.5 pmoles of DABR4, and 0.5 units of CLP *ProofPro* proofreading polymerase. PCR consisted of an initial denaturing step of 94°C for 3 min, followed by 25 or 40 cycles of 94°C: 45 s, 55°C: 1 min, 72°C: 1 min. Products of approximately 300 bp were ligated into pGEM-T (Promega) and transformed into Invitrogen DH5α competent cells according to manufacturer's instructions. Colonies were dipped directly into 25 μl PCR reactions including 10 pmoles of M13 primers and 0.5 units of *ProofPro* polymerase. Sixteen positive colonies were then sequenced in both directions on a Beckman Coulter CEQ8000 using standard M13 primers and fluorescently-labelled terminators. From these sequences we selected three distinct, previously characterised alleles to estimate the PCR error rate in a controlled experiment by counting the number of base misincorporations in these known sequences.

Results and discussion

Sequence artefacts caused by base misincorporations during PCR occurred randomly throughout sequences with no detectable hotspots. The overall rate of base misincorporation was 1.85×10^{-5} misincorporations per base per cycle, which is not significantly different from that reported for other studies (Keohavong and Thilly 1989; Kobayashi et al. 1999). In this study, approximately 11% of all sequences showed PCR error. If these errors are not identified, they will be categorised as novel alleles and bias population genetic analyses.

Acinas et al. (2005) suggested clustering the sequences into 99% sequence similarity groups to reduce the probability including polymerase errors. However, informative sequence data will be ignored by this procedure and the suggested 99% similarity level is chosen arbitrarily. Indeed, Acinas et al. (2005) argue that clustering of sequences based on similarity cut-offs (e.g. the commonly used 97%) may mask micro-diverse clusters, which have been suggested as representing important units of differentiation (Acinas et al. 2004). Similarly, acceptance of only those sequences that were observed in two independent PCRs (Lukas et al. 2004; Lukas and Vigilant 2005) will also result in an underestimation of gene diversity by rejecting rare alleles.

We suggest that researchers implement our new method of calculating the probability of detecting a given number of PCR artefacts in an amplicon, and thus identify sequences with likely base misincorporations. The overall probability of finding a given number of misincorporations in the entire amplicon can be calculated using a cumulative binomial distribution with the BINOMDIST function in MS Excel. The binomial probability mass function of this and all less likely events (i.e. sequences with more errors) equals:

$$P(x \geq k) = 1 - \sum_{i=1}^{k} \binom{N}{k-i} \varepsilon^{k-i} (1-\varepsilon)^{N-(k-i)}$$

in which k is the number of PCR errors, N is the total number of bases in the sequence and ε is the PCR error rate in the entire amplicon. This probability reflects the likelihood of finding a given or large number of misincorporations in the amplicon of N base pairs due to PCR error. If this probability is above the Bonferroni corrected critical value (α'), the observed sequence is likely to be a PCR artefact and can be rejected. However, if the probability is below the critical value, the sequence is unlikely to be a PCR artefact and independent confirmation (e.g. through additional PCR of the same template DNA) is necessary to confirm the sequence is genuine.

For example, suggest we analysed an 800 bp amplicon using a PCR of 30 cycles with 1.85×10^{-5} misincorporations per base per cycle, and found 10 unique sequences that are each 3 bp different from their most similar sequence. The probability that a novel sequences was generated by three independent PCR errors equals $P = 0.0011$. This is smaller than the Bonferroni corrected alpha ($\alpha' = 0.05/10 = 0.005$), and we would therefore conclude these sequences are genuine. Following the suggestion by Lukas et al. (2004), these alleles need to resequenced for confirmation. However, ten sequences that are different by just a 1 or 2 bp will be rejected as PCR errors ($P = 0.074$ and $P = 0.0105$, respectively) because the probability of their occurrence is above the Bonferroni corrected alpha.

Given that a substantial proportion of analysed samples have 1 or 2 bp misincorporations, this procedure can thus dramatically reduce the re-genotyping effort. The method proposed here furthermore reduces the likelihood of rejecting valid sequence information to a commonly accepted statistical type I error of 5%. Finally, although

prior knowledge of the PCR error rate is an advantage, researchers could use the average PCR error rate for the analysed gene in the proposed formula.

Acknowledgments We thank Alan Smith for help with fish husbandry. The work was funded by NERC grants NER/M/S/2002/00101, NER/I/S/2000/00885 and Leverhulme Trust grant F/00 181/F awarded to CVO.

References

Acinas SG, Klepac-Ceraj V, Hunt DE, Pharino C, Ceraj I, Distel DL, Polz MF (2004) Fine-scale phylogenetic architecture of a complex bacterial community. Nature 430:551–554. doi:10.1038/nature02649

Acinas SG, Sarma-Rupavtarm R, Klepac-Ceraj V, Polz MF (2005) PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. Appl Environ Microbiol 71:8966–8969. doi:10.1128/AEM.71.12.8966-8969.2005

Keohavong P, Thilly WG (1989) Fidelity of DNA-polumerases in DNA amplification. Proc Natl Acad Sci USA 86:9253–9257. doi:10.1073/pnas.86.23.9253

Kobayashi N, Tamura K, Aotsuka T (1999) PCR error and molecular population genetics. Biochem Genet 37:317–321. doi:10.1023/A:1018759210666

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. doi:10.1111/j.1365-294X.2005.02459.x

Lukas D, Bradley BJ, Nsubuga AM, Doran-Sheehy D, Robbins MM, Vigilant L (2004) Major histocompatibility complex and microsatellite variation in two populations of wild gorillas. Mol Ecol 13:3389–3402. doi:10.1111/j.1365-294X.2004.02353.x

Thompson JR, Marcelino LA, Polz MF (2002) Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by 'reconditioning PCR'. Nucleic Acids Res 30:2083–2088. doi:10.1093/nar/30.9.2083

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. doi:10.1038/sj.hdy.6800843

van Oosterhout C, Joyce DA, Cummings SM, Blais J, Barson NJ, Ramnarine IW, Mohammed RS, Persad N, Cable J (2006b) Balancing selection, random genetic drift, and genetic variation at the major histocompatibility complex in two wild populations of guppies (*Poecilia reticulata*). Evol Int J Org Evol 60:2562–2574. doi:10.1554/06-286.1

Zylstra P, Rothenfluh HS, Weiller GF, Blanden RV, Steele EJ (1998) PCR amplification of murine immunoglobulin germline V genes: strategies for minimization of recombination artefacts. Immunol Cell Biol 76:395–405. doi:10.1046/j.1440-1711.1998.00772.x

Appendix 14

McMullan, M., and C. van Oosterhout. in press. Evolutionary genetics and molecular ecology of the Major Histocompatibility Complex (MHC) genes of poeciliids *in* J. Evans, A. Pilastro, and I. Schlupp, eds. Ecology and Evolution of Livebearing Fishes (Poeciliidae). Chicago University Press.

CHAPTER 25

Evolutionary genetics and molecular ecology of the Major Histocompatibility Complex (MHC) genes of poeciliids

M. M^{c}MULLAN AND C. VAN OOSTERHOUT

25.1 Introduction

The Major Histocompatibility Complex (MHC) is a large multigene family present in all jawed vertebrates (see Klein et al. 2000). It has a central role in immunocompetence and is implicated in mate choice and sexual selection (Apanius et al. 1997; Edwards & Hedrick 1998; Hughes & Yeager 1998; Penn & Potts 1999; Bernatchez & Landry 2003; Wegner et al. 2004; Sommer 2005). The histocompatibility genes were first studied in relation to transplantation compatibility in humans and mice (Dausset 1958) and then received attention in conservation biology with work on the cheetah (O'Brien et al. 1985). The MHC is implicated with the detection of invading pathogens by recognition of small protein fragments known as peptides, or antigen.

The number of genes required to recognize antigens varies enormously between species. For instance, there are over 200 MHC loci (including pseudogenes) within the human MHC (the HUMAN LEUKOCYTE ANTIGEN or HLA (**Glossary**)) (The-MHC-sequencing-consortium 1999; Shiina et al. 2004) and just 19 within the chicken MHC (B locus) (Kelley et al. 2005). The number of MHC loci can also vary between individuals of the same species (Malaga-Trillo et al. 1998; Figueroa et al. 2001). Variation in the number of genetic loci within the MHC is predicted

by various multigene models, including the ACCORDION MODEL OF MHC EVOLUTION (Klein et al. 1993), the BIRTH-AND-DEATH MODEL and CONCERTED EVOLUTION (see review by Nei & Rooney 2005).

The MHC is characterized by extraordinary levels of polymorphism, as is illustrated for example by the human MHC HLA-B locus, which has over 1000 alleles (Robinson et al. 2003) (IMGT/HLA Database, 2008, http://www.ebi.ac.uk/imgt/hla/stats.html). This level of polymorphism varies within and among populations (Wegner et al. 2003a; van Oosterhout et al. 2006b; Dionne et al. 2007), as well as between different MHC genes (Robinson et al. 2003) (IMGT/HLA Database, 2008). The highest level of polymorphism is observed within a specific region of the MHC known as the PEPTIDE BINDING REGION (PBR) (Hughes & Yeager 1998). This region is responsible for the specificity of the MHC molecule to bind particular peptides from different sources of antigens, including bacteria, viruses and micro- and macroparasites (Bjorkman et al. 1987b; Stern et al. 1994)

Parasitism and host defense against pathogens is generally considered the primary mode of selection maintaining polymorphism at the MHC. The abundance and composition of pathogens within populations is a proposed mechanism maintaining specific MHC alleles, and much of the work on this subject has focused on humans (Shiina et al. 2004) and mice (Penn & Potts 1999). Here, we focus on poeciliids, where the MHC has only recently been unraveled, as well as on other teleosts, such as other Cyprinodontiformes, Gasterosteiformes and Salmoniformes.

In addition to natural selection by parasites, sexual selection may also play a role in maintaining MHC polymorphism (see Rios-Cardenas & Morris, **Chapter 17**, and Evans & Pilastro, **Chapter 18**). Sexual selection for "GOOD GENES" or increased genetic diversity has been shown to maintain MHC polymorphism in many species (Milinski 2006). Furthermore, the variation at the MHC may be an important signal used in mate choice, with individuals showing preference for genetically compatible mating partners which may for example help to avoid inbreeding (Penn & Potts 1999; reviewed by Milinski 2006).

Sexual selection has been studied in extensively in several Poeciliid species (see e.g. Rios-Cardenas & Morris, **Chapter 17**). Mate choice experiments in guppies (*Poecilia reticulata* Peters) (see Houde 1997) have primarily focused on visual cues, and relatively little research has tested the role of olfaction and chemical communication (Shohet & Watt 2004; but see Archard et al. 2008). Olfactory detection of MHC compatible mates has been shown to occur in a number of other animal systems including humans, mice and sticklebacks (Penn & Potts 1998; Aeschlimann et al. 2003; Beauchamp & Yamazaki 2003). Given that poeciliids are suitable model organisms to study the role of olfaction in mate choice, we will briefly discuss here the potential role of the MHC in sexual selection and the maintenance of MHC polymorphism.

25.2 The Major Histocompatibility Complex (MHC)

The MHC multigene family encodes a number of proteins responsible for the preparation, loading and specific binding of antigen fragments (reviewed by Ting & Trowsdale 2002; Boss & Jensen 2003; Danchin et al. 2004). The CLASSICAL MHC GENES are typically highly polymorphic and separated into the MHC CLASS I and MHC CLASS II subfamilies. Both classes of genes form part of the adaptive immune system. Nonclassical MHC genes display low levels of polymorphism and gene expression, and the DNA sequences of these genes commonly show little evidence of positive selection (Summers et al. 2008). The class III region contains a group of genes genetically and evolutionarily unrelated to class I or class II, although some of these genes have immune function, particularly in the innate immune response (Klein & Sato 1998). MHC class I and class II are linked in mammals and birds, but they occur unlinked on different chromosomes in teleost fishes (Sato et al. 2000; Stet et al. 2003). Consequently, some authors refer to teleost MHC as the MH (see Ellis et al. 2006; Dixon 2008), although we prefer to use the more traditional abbreviation, MHC.

Class I and class II molecules differ in the type of antigens they present. Class I molecules are expressed on the surface of all nucleated somatic cells and present peptides originating from within the cell, such as viral peptides. Class II molecules are expressed primarily on antigen presenting cells of the immune system, and they specialize in the presentation of extracellular pathogen peptides (e.g. bacteria and microparasites). These differences in subfamilies are echoed in the specific antigen processing pathways required to load the MHC molecule and to the type of T-cell to which they present. Once degraded, antigen fragments are bound to the MHC glycoprotein and presented to T-cells. T-cells are responsible for the initiation of the adaptive immune response which acts to target the pathogen specifically. T-cells need to respond only to foreign antigens and not to the host's own tissue. In order to recognize "self" from "non-self", autoreactive thymocytes with T-cell receptors that interact with MHC molecules presenting self peptides are eliminated during thymic selection in the developing embryo (Jameson et al. 1994; Jordan et al. 2001).

Much population genetic, immunogenetic and evolutionary research on the MHC has focused on the second exon of classical MHC encoded molecules, as this transcribes a large proportion of the PBR. Positive selection is only evident in this part of the gene, and consequently, this exon shows the highest level of polymorphism of the entire MHC. Other exons of the MHC are relatively conserved, and the sequence variation in these exons is governed mainly by neutral evolutionary forces, such as mutation, genetic drift and RECOMBINATION.

25.3 Evolutionary genetics and molecular ecology of the MHC

25.3.1 Natural selection and the MHC

In the last three decades, molecular ecologists have studied neutral evolution using allozymes, minisatellites, microsatellites and SNPs (Hedrick 2005). Neutral theory asserts that

the majority of new mutations are either deleterious and eliminated from a population by negative selection, or (nearly) neutral and subject to random genetic drift (Ford 2002; Garrigan & Hedrick 2003). The neutral polymorphism is maintained in a mutation-drift equilibrium, and this variation can be used to analyze the population's demography. Neutral marker loci are used to analyze genealogies, establish paternities, track population migration, measure gene flow and estimate effective population size.

More recently, with the advent of economically viable sequencing and the identification of genes under positive selection, research begun to focus on adaptive evolution. Adaptive evolution is driven by positive selection promoting novel (favorable) sequence variants. Positive selection can be inferred using several types of analyses of sequence data, and the d_N/d_S ratio is one commonly used method (Ford 2002; Garrigan & Hedrick 2003). The d_N/d_S value is essentially a ratio of the number of NONSYNONYMOUS MUTATIONS (i.e. amino acid changing mutations) to synonymous ones (also known as silent substitutions). SYNONYMOUS MUTATIONS do not change the protein produced by the gene and are predicted to be selectively neutral. Consequently, synonymous mutations are accumulated and lost in a neutral fashion, and these mutations will occur in drift-mutation equilibrium (Crow & Kimura 1970; Hedrick 2005). In contrast, nonsynonymous mutations can alter protein function and such mutations are thus usually subject to natural selection. Mutations that change the fitness of the organism can be either favorable (positively selected), or deleterious (negatively selected). In conserved genomic regions, nonsynonymous mutations tend to be purged by negative selection as changes to the protein structure are generally deleterious. Synonymous mutations, on the other hand, can persist at a relatively high frequency, and consequently, in conserved gene regions the d_N/d_S ratio will be below unity ($d_N/d_S<1$). However, in gene regions where evolutionary change is favorable, positive selection is predicted to maintain nonsynonymous mutations, resulting in an elevated d_N/d_S ratio ($d_N/d_S\geq1$). In the MHC, the protein region interacting with foreign peptides is transcribed from the codons of the PBR. The codons in this region are thought to accumulate nonsynonymous mutations as these allow the recognition of

novel parasites. By contrast, the areas outside the PBR are conserved, and nonsynonymous mutations are purged from these gene regions. Studies on the MHC have been facilitated by the fact that the PBR codons of the human MHC have been identified through X-ray crystallography (Bjorkman et al. 1987a; b; Brown et al. 1993; Stern et al. 1994), and that the MHC is highly conserved across vertebrates (but see Blais et al. 2007 for differences in PBR codons between humans and cichlid fish). This allows researchers working on the MHC of other vertebrate taxa *a priori* to identify the codons under positive and negative selection.

25.3.2 Balancing selection

Besides positive and negative selection, BALANCING SELECTION plays a pivotal role in the evolution of the MHC. Balancing selection is thought to maintain polymorphism at a locus by one (or combination) of the following three processes: (1) heterozygote advantage (overdominant selection), (2) rare allele advantage (negative frequency dependent selection), or (3) selection varying in time or space (spatial or temporal variable selection). Genes that are evolving under balancing selection are distinct from neutral loci in that they show high levels of heterozygosity and a large numbers of alleles at similar frequencies (hence the term *balancing* selection). Some alleles can be maintained over great lengths of time, possibly exceeding speciation events, a phenomenon called TRANS-SPECIES POLYMORPHISM (reviewed by Klein et al. 1998). Furthermore, balancing selection can also affect the DNA substitution pattern in a similar way as positive selection, resulting in an elevated d_N/d_S ratio (see above).

1. Overdominance selection: Of the three hypotheses suggested to maintain a balanced polymorphism, the model of overdominance selection (heterozygote superiority) has received most attention. The heterozygote advantage hypothesis relates directly to the MHC genotype and the role of specific MHC alleles present at the locus. Heterozygous individuals are thought

to be able to recognize twice as many pathogens as homozygous individuals, and therefore have a superior immune response and higher fitness (Doherty & Zinkernagel 1975). Landry et al. (2001) demonstrated selection for increased proportions of Atlantic salmon offspring heterozygous at the MHC class II *B* locus. Similarly, Arkush et al. (2002) demonstrated increased survival of heterozygous MHC class II *B* Chinook salmon to infectious hematopoietic necrosis virus (IHNV). The level of inbreeding (inbred or out-bred) was not found to correlate with IHNV survival but was associated with increased resistance to another myxozoan pathogen. These findings highlight the importance of heterozygosity at MHC genes and throughout the genome as a means of pathogen defense (Spielman et al. 2004; Hale & Briskie 2007; van Oosterhout et al. 2007b). Penn (2002) performed a meta-analysis and found no clear evidence of MHC heterozygote advantage. Evidence is accumulating that an optimal, rather than a maximal heterozygosity might be providing the highest resistance (see Nowak et al. 1992; Wegner et al. 2003b; Milinski 2006), which begs the question how MHC polymorphism in populations is then maintained.

2. Negative frequency dependent selection: Negative frequency dependent selection is another model of balancing selection that can maintain polymorphisms in populations (Clarke & Kirby 1966; Kojima 1971). Whereas overdominant selection acts on MHC genotypes, the model of frequency dependent selection assumes that each MHC allele recognizes particular parasites. Langefors et al. (2001) infected 4800 *Salmo salar* with a bacterium (*Aeromonas salmonicida*) and identified MHC alleles associated with either high resistance or susceptibility. This shows that besides selection on heterozygous genotypes, parasite selection can also act on single MHC alleles, which may result in a dynamic coevolutionary arms race between the host immune system and parasite virulence genes (Lively & Dybdahl 2000; Carius et al. 2001). Parasites are expected to evolve to avoid detection by common MHC alleles. By contrast, the selection pressure on parasites for avoiding recognition by rare MHC alleles is much lower

(Slade & McCallum 1992). Consequently, rare MHC alleles are more likely to detect pathogens, and hence such alleles are predicted to confer a higher fitness to the host than common alleles (Trachtenberg et al. 2003; Froeschke & Sommer 2005). Over time, parasite selection is expected to increase the frequency of rare alleles as hosts carrying these rare alleles have a higher probability to avoid parasitism. The coevolutionary arms race between the host immune system and parasite virulence results in a dynamic and balanced polymorphism in both host MHC and parasite virulence genes (Sommer 2005). We know of no long term study demonstrating this cyclical pattern of MHC allele frequencies in fish (but for studies on other vertebrates, see Westerdahl et al. 2004). However, given the rapid generation time of poeciliids and the recent advances made by studies on their MHC and parasites, future studies on these fish offer promising insights into host-parasite coevolution and RED QUEEN dynamics (Bell 1982; Penn & Potts 1999).

3. Spatial or temporal variable selection: Spatial and/or temporal heterogeneity in selection pressure also can result in a balanced polymorphism at the MHC (Hedrick 2002). A study of the MHC class I variation over nine years of great reed warblers found variation in allele frequencies over time (Westerdahl et al. 2004). The authors suggest that these fluctuations in MHC allele frequencies are due to temporal fluctuations in pathogen fauna, also consistent with frequency dependent selection (see also Charbonnel & Pemberton 2005). A parasite of the Trinidadian guppy has also been implicated with temporal fluctuation in MHC allele frequency. Miller et al. (2001) sampled MHC polymorphism in 31 populations of sockeye salmon (*Oncorhynchus nerka*) within a river basin. They found evidence of balancing selection acting on the MHC in some but not all populations. The authors were able to separate population demography to rule out population bottlenecks in low MHC diversity populations, showing that the patterns of allele frequencies in some populations were more consistent with directional selection (Landry & Bernatchez 2001).

These three models of balancing selection are not mutually exclusive, but rather, they are likely to act complementarily. De Boer et al. (2004) and Borghans et al. (2004) both modeled negative frequency dependent selection and overdominant selection, finding that overdominance was unlikely to maintain the level of polymorphism observed at the MHC, unless each MHC allele conferred a very similar fitness (e.g. symmetric overdominance). However, it is difficult to discriminate between different types of balancing selection in wild systems. For example, rare alleles that are predicted to be maintained under negative frequency dependent selection are also more likely to be presented in heterozygote condition (Apanius et al. 1997).

These traditional models of balancing selection were developed to understand the evolution and population genetics of a single immune gene. However, the MHC is not a single gene but a multigene family, and these models ignore the potential role of linkage and EPISTATIC gene-gene interactions. The MHC genes are surrounded by linked genetic variation (single nucleotide polymorphisms, SNPs) that is associated with more diseases than any other part of the human genome (de Bakker et al. 2006; Shiina et al. 2006). This suggests that this linked peri-MHC region is under strong selection, and that it could play a potentially important role in the evolution of this multigene family. van Oosterhout (2009) proposed a new theory of MHC evolution that incorporates the impact of selection on the region surrounding the MHC genes. The model is called ASSOCIATIVE BALANCING COMPLEX (ABC) EVOLUTION, and it proposes that selection acts on the deleterious mutations that are associated with the MHC genes (**Box 1**).

25.4 Sexual selection, parasite selection and inbreeding avoidance

According to the Red Queen Hypotheses, the hosts' immune system evolves quickly in response to the rapid evolution of their parasites (Milinski 2006), and balancing selection by parasites can maintain MHC polymorphism as proposed by the various models (see above).

However, mate choice can also result in rapid MHC evolution, or it may help to maintain polymorphism at these genes (Milinski 2006). The MHC is implicated in sexual selection at both pre- and post zygotic levels, for example by means of mate choice for mates with a specific MHC genotype, differential fertilization success, and the selective abortion of MHC-homozygous embryos (see Rios-Cardenas & Morris, **Chapter 17**, and Evans & Pilastro, **Chapter 18**). Studies on a wide variety of animals support the role of MHC in pre-copulatory sexual selection, including research on humans and rodents (reviewed by Penn & Potts 1999; Kavaliers et al. 2004), salmon (Landry et al. 2001; Neff et al. 2008) and sticklebacks (reviewed by Milinski 2003). This work has focused in particular on the role of olfaction, which is thought to enable individuals to determine the compatibility and/or diversity of the MHC of prospective mating partners during courtship rituals. For example, research on sticklebacks showed that females use evolutionarily conserved structural features of MHC molecules to evaluate diversity of the MHC of males (Milinski et al. 2005).

There has been considerable debate on the relative roles of MHC variation and genome-wide heterozygosity and their consequence on fitness (Arkush et al. 2002; Sommer 2005). Landry et al. (2001) were able to separate the effects of MHC heterozygosity and inbreeding avoidance (i.e. the relatedness measured using similarity at five neutral microsatellite loci). They sampled spawning Atlantic salmon and juvenile salmon after hatching, and found that mates selected each other in order to increase genotypic differences in the MHC of juveniles, and not to offer increased heterozygosity in the genome. This supports the role of the MHC in sexual selection for "good genes" (or genetic compatibility), although the authors acknowledge the experiment cannot differentiate between active mate choice, differential fertilization success, and the higher survival of juveniles of heterozygous MHC genotypes (see also de Eyto et al. 2007; Neff et al. 2008). Disentangling the relative importance of inbreeding avoidance, "good gene" selection and parasites resistance in MHC evolution is an important future challenge in MHC research. A particular promising new

avenue of research is offered by the comparison of MHC evolution in a pair of closely-related poeciliids, one species being entirely clonal and the other sexually reproducing (see section 25.5.1 and Schlupp & Riesch, **Chapter 5**).

There is increasing evidence that the outcome of sperm competition can depend on interactions between male and female genotypes (see Rios-Cardenas & Morris, **Chapter 17**, and Evans & Pilastro, **Chapter 18**). Studies on organisms other than poeciliids suggest that MHC may be implicated in non-random fertilization (see Simmons 2005 for review). However, these studies typically score paternity on newborn or adult offspring, making it difficult to disentangle the effect of sperm selection from differential embryo mortality. The only study which directly measured fertilization success in a fish (Arctic charr, *Salvelinus alpinus*) failed to find any effect of MHC genotype on fertilization success in a non-competitive situation (Skarstein et al. 2005).

25.5 Poeciliid MHC

The class I and II genes of the teleosts MHC are unlinked and thought to reside on separate chromosomes, an arrangement that is markedly different from that of tetrapods (Sato et al. 2000; Stet et al. 2003; Dixon 2008). Free recombination between class I and II MHC regions in teleost fish is believed to be a derived characteristic as the class I and II genes in cartilaginous fishes, the oldest class of extant vertebrates, is linked (Ohta et al. 2000). Class III genes are split over several different chromosomes in teleost (Flajnik et al. 1999; Sambrook et al. 2002). The occurrence of separate linkage groups may have important implications for the evolution of teleost MHC, as with free recombination, selection can operate independently on the different gene classes. This suggests that teleost MHC is more versatile, and perhaps less prone to accumulate mutations by GENETIC HITCHHIKING. Relatively little work has been done on MHC class I and III in poeciliids (but see Sato et al. 1995; Sato et al. 2000; Schaschl et al. 2008),

possibly because class I genes are involved in the antigen presentation of more elusive intracellular parasites such as viruses. However, Figueroa et al. (2001) conduct a comprehensive survey of the MHC class I of 22 swordtail fishes. Sequences were found to fall within one of two ancient lineages and the pattern of sequence similarity was consistent with the birth-and-death theory of multigene evolution (Nei & Rooney 2005). In this chapter we will focus on research on MHC class II loci, in particular *DAB*, *DXB* genes and introns (**Box 2**).

25.5.1 Poeciliid MHC *DAB* genes

The MHC multigene family and its large number of genes necessitate a clear classification of alleles and genes. Here, we follow the nomenclature first proposed by Klein et al. (1990). The *DAB* loci in poeciliids (and many other fish species) are possibly the best studied genes. *DAB* specifies the region of the MHC, in which *D* is an abbreviation of Duo, in reference to class II, is *A* represents the family of genes and *B* specifies the β heterodimer (class II *B*). The species from which the MHC sequence was amplified is denoted by abbreviation of genus and species names, for example, *Mhc-Pore* specifies that the MHC sequence was obtained from the organism *Poecilia reticulata* (see Klein et al. 1990).

Sato et al. (1995) first described the MHC of the guppy using ornamental stocks and inbred wild fish to define the class I and class II regions of the MHC. The authors found very low variability at both classes of genes and predicted that guppies possess one MHC class II gene and one or two MHC class I genes. van Oosterhout et al. (2006a) confirmed this finding, showing that ornamental guppy lines possessed between one and three MHC class II *B* alleles. However, they detected up to four alleles in individual wild-caught guppies, which implies there are at least two *DAB* genes. Within populations, the authors detected 15 to 16 MHC alleles, demonstrating these wild guppy populations had a much higher level of MHC polymorphism than ornamental guppy stocks. The disparity in allelic richness between wild

and ornamental fish could not be explained by fixation of alleles by inbreeding in captivity, nor by the presence of non-amplified sequences (i.e. null alleles). Rather, van Oosterhout et al. (2006a) suggested that during many generations in captivity, multigene evolution may have played an important role shaping the organization and polymorphism of the MHC. In the ornamental lines, GENE CONVERSION may have fixed the same allele at duplicated MHC *DAB* genes. (For the role of gene conversion in teleost MHC, see e.g. Reusch et al. 2004; Reusch & Langefors 2005). van Oosterhout et al. (2006a) furthermore hypothesized that a reduced level of parasitism during >100 generations in captivity may have reduced the number of duplicated *DAB* genes in the ornamental strains, a hypothesis consistent with the accordion model of MHC evolution (see Klein et al. 1993).

The MHC has also been well-studied in two other poeciliids, the Gila topminnow (*Poeciliopsis o. occidentalis*) and the Yaqui topminnow (*P. o. sonoriensis*). Work by Hedrick et al. represents one of the first examples for the use of MHC variation in conservation of an endangered species (Hedrick & Parker 1998; Hedrick et al. 2001a; Hedrick et al. 2001b), (see section 25.6.3).

Schaschl *et al.* (2008) have analyzed the MHC class I and class II *B* diversity of two closely related mollies. The Amazon molly (*Poecilia formosa*) is a clonal all-female Poeciliid that lives in sympatry with the sexually reproducing Sailfin molly (*Poecilia latipinna*) (see also Schlupp & Riesch, **Chapter 5**). Both neutral genetic and MHC class I and II diversity of the asexual molly were lower in comparison to its sexual counterpart, although it still maintained a considerable level of MHC diversity. The Red Queen hypothesis suggests that the asexual molly should be unable to provide a "moving target" against rapidly evolving pathogens, as (meiotic) recombination cannot provide novel gene combinations in clonally reproducing organisms. However, previous research showed no difference in parasite prevalence between both sympatric species (Tobler & Schlupp 2005; Tobler et al. 2005). Comparison of MHC diversity between the asexually and sexually reproducing species offers a unique opportunity to

disentangle the relative roles of parasites and sexual selection in the maintenance of MHC polymorphism.

25.5.2 Poeciliid MHC *DXB* gene

In 1998, McConnell et al. (1998b) reported a novel MHC class II *B*-like gene in *Xiphophorus maculatus* and *X. helleri*. This gene was labeled *DXB* as its family designation remains unknown (McConnell et al. 1998b). This locus is unlinked to the *DAB* locus and is expressed in *X. maculatus*, *X. multilineatus* and *X. pygmaeus* (McConnell et al. 1998a; Roney et al. 2004). Genomic copies of this gene have also been identified in *X. helleri* and *P. reticulata*, with features consistent with functional genes (McConnell et al. 1998a; McConnell et al. 1998b). *Xima-DXB* and *Pore-DXB* are thought to be orthologous, sharing 98% sequence similarity in exon 3 (i.e. *Xima-DXB*01* and *Pore-DXB*01*). *DXB* has a sequence similarity of only 61-63% in exon 3 with *DAB* alleles, and both genes are hypothesized to be the result of an ancient gene duplication event. However, unlike *DAB*, the *DXB* genes show very little polymorphism and a weak signal of positive selection (Summers et al. 2008).

The exon-intron structure of the *DXB* is similar to that of amphibian, chicken and human class II*B* structure, which suggests that a *DXB*-like lineage is phylogenetically related to tetrapod MHC class II *B* genes (McConnell et al. 1998a; McConnell et al. 1998b). Interestingly, Roney et al. (2004) identified two alternative splice patterns of the *DXB* gene (as well as one truncated version from a sample of *X. multilineatus* and *X. pygmaeus*). Alternative splicing is the transcription of only certain exons within a gene, and this allows a single gene to encode many distinct proteins (reviewed by Stetefeld & Ruegg 2005; Artamonova & Gelfand 2007). Alternative splicing may be an important feature of the MHC, potentially increasing its versatility, and in one human nonclassical class II chain molecule (*HLA-DM*) four alternative splice patterns have been detected (Modrek et al. 2001). Gene expression and population

genetics of the *DXB* genes in poeciliids has received relatively little attention. However, the interactions between *DXB* gene products and other MHC class II molecules (see Ting & Trowsdale 2002; Danchin et al. 2004), and the relative conservation of this gene across phylogenetically distinct taxa suggest it plays an important role in MHC evolution and its functioning (Jensen et al. 2008). As recently highlighted by Summers et al (2008), research on fish, and in particular poeciliids, may help to unravel the enigmatic role of (nonclassical) MHC genes in other vertebrates.

The genomic organization of Poeciliid MHC is relatively simple, and is more similar to that of salmonids than that of many other important evolutionary and ecological model species. Salmonids are believed to contain just one *DAB* locus (Langefors et al. 1998; Langefors et al. 2000; Landry & Bernatchez 2001; Stet et al. 2002), which makes detecting putative associations between resistance and MHC alleles or genotypes easier than for species with many duplicated MHC genes. For example, the three-spined stickleback is thought to have four *DB* gene loci (*Gaac-DAB*, *-DBB*, *-DCB* & *-DDB*) (Reusch & Langefors 2005), while cichlids may have up to 17 loci (Malaga-Trillo et al. 1998). Consequently, in these species associations between resistance and MHC can only be detected when analyzing multilocus genotypes. Although several studies aimed to find the evolutionary or ecological cause for the high and variable number of *DB* loci, PCR amplifying alleles of a large number of loci is a technically challenging task. Indeed, the relative simplicity of Poeciliid MHC makes this family an excellent model for future immunological, parasitological and evolutionary studies. Furthermore, the apparent difference in genomic organization between ornamental and wild guppies makes this an interesting species to examine the role of different models of multigene evolution.

25.6 Poeciliid MHC, molecular evolution and ecology

25.6.1 Parasite selection and MHC in poeciliids

In natural populations, all genetic variation is effected by mutation and random genetic drift. In addition to these evolutionary forces, MHC genes are also affected by selection (Ford 2002; Garrigan & Hedrick 2003). To study the effects of selection, researchers tend to compare the genetic variation at MHC loci to that of neutral genetic markers (e.g. microsatellites). The signal of natural selection should be detectable only in the MHC, while population demographic forces (such as genetic drift, migration and mutation) will be evident at both neutral and immunogenetic loci. This allows one to disentangle these demographic effects from the effects of selection on the MHC (van Oosterhout et al. 2006b).

van Oosterhout et al. (2006) assessed MHC class II *B* diversity in two wild populations of guppies, whilst also recording differences in parasite fauna. Populations inhabited the upper ($N_e \approx 100$) and lower ($N_e \approx 2400$) regions of the Aripo River in Trinidad. Both sites are characterized by a distinctly different parasite fauna (see Cable, **Chapter 8**), more specifically, *Gyrodactylus* burdens of individual fish were markedly different between sites (van Oosterhout et al. 2006b). The upland population, with higher *Gyrodactylus* loads was shown to maintain high levels of MHC diversity despite the effects of random genetic drift on this small population. Using computer simulations that incorporated the demographic estimates based on microsatellite variation, the authors estimated selection coefficients of $S \geq 0.2$ acting on the MHC in the most parasitized population.

In a subsequent study, van Oosterhout et al. (2007a) estimated the potential for parasite selection directly, using a mark-recapture experiment in the same Aripo guppy population in Trinidad. They marked almost 200 guppies and counted their gyrodactylid parasite load. Although individuals with a relatively high parasite burden were recaptured with a significantly reduced rate, the selection coefficient was lower ($S \leq 0.14$) than that required to explain the MHC variation observed in the previous study (van Oosterhout et al. 2006b). This suggests that other selective forces are operating on the MHC, such as sexual selection

through female mate choice, or selection on linked mutations that hitch-hike with the MHC alleles.

25.6.2 Sexual selection and MHC in poeciliids

An impressive body of research on poeciliids has been dedicated to study the role of female mate choice based on male color pattern and ornamentation (Rios-Cardenas & Morris, **Chapter 17**). However, sexual selection based on other senses such as olfaction has received relatively little attention (Crow & Liley 1979; Meyer & Liley 1982; Griffiths & Magurran 1999; Shohet & Watt 2004; Archard et al. 2008). Nevertheless, studies on other fish species have indicated that multimodal interactions between visual and chemosensory cues potentially play a role in mate choice and sexual selection (e.g., three-spined sticklebacks, *Gasterosteus aculeatus*, McLennan 2003). Archard et al. (2008) showed that high concentrations of chemosensory stimulus were required to elicit a behavioral response in guppies. Such high concentrations can only be attained when the fish are in close proximity. This may arise when guppies shoal, or when males display to females, which occurs within distances of a few centimeters (Endler 1995; Long & Rosenqvist 1998). Shohet & Watt (2004) used an experimental flume to study the attraction of female guppies to conspecific chemosensory cues, and found that females were attracted to cues from other females, and that visually attractive males were chemically unattractive, and vice versa. The potential role of the MHC in this model and other Poeciliid model systems has however not been investigated. Future research into MHC and sexual selection is possibly best studied in Poeciliid species with little variation in male color patterns, since in these species olfaction may play a more prominent role than visual communication.

25.6.3 MHC in conservation

MHC diversity is believed to play an important role in conservation genetics, particularly in captive populations bred for reintroduction programs (O'Brien & Evermann 1988; Hedrick et al. 2001b; Miller et al. 2001). Surprisingly, although the success rates of reintroduction programs is low, with only 11% reaching viable population sizes (Beck et al. 1994), few studies have investigated the efficacy of captive breeding regimes for the release of captive-bred vertebrates. Contemporary breeding strategies for endangered species typically aim to avoid loss of genetic variation, limit the adaptation to captive conditions, and maximize evolutionary potential (van Oosterhout et al. 2007b). Management of the MHC diversity in captive gene pools has been suggested as an alternative breeding protocol (Hughes 1991; but see Miller & Hedrick 1991), although this suggestion has not been tested empirically. Indeed, whether the level of MHC variation affects the viability of populations and future reintroduction success of individuals is not yet clearly established (Gutierrez-Espeleta et al. 2001; Miller & Lambert 2004), although reduced immunocompetence (Hale & Briskie 2007) and increased risk of disease outbreaks (O'Brien & Evermann 1988; Spielman et al. 2004) have been noted in several bottlenecked and/or inbred populations.

van Oosterhout et al. (2007b) tested the effects of breeding regimes on the reintroduction success, using wild caught guppies from two populations in Trinidad. The authors bred fish for four generations in captivity and released the fish in a mesocosm in Trinidad. However, they did not examine whether particular MHC genotypes or alleles affected the survival rate of the reintroduced fish. The authors identified parasites as the main threat to survival in the wild, which suggests that MHC variation could play an important role. Guppies and other poeciliids offer an excellent model to test the effectiveness of breeding programs, and in particular, test the importance of MHC diversity for viability of reintroduced populations.

Hedrick & Parker (1998) examined polymorphism at the MHC in the endangered Gila topminnow (*Poeciliopsis o. occidentalis*). They examined four watersheds in the USA, and

found no MHC variation in the most severely bottlenecked population. Because the studied populations differed significantly in their MHC divergence, the authors suggested further research was warranted to establish whether the watersheds should be considered for independent conservation and management. In a subsequent study, Hedrick et al. (2001b) used both microsatellites and MHC to study the conservation genetics of the endangered Sonoran topminnow. The Sonoran topminnow (*Poeciliopsis occidentalis*) consisted of two putative subspecies, the Gila topminnow (*Poeciliopsis o. occidentalis*) and the Yaqui topminnow (*P. o. sonoriensis*). The authors used both sets of highly polymorphic markers to recommend that both species should be recognized as species (rather than subspecies), and identified as different evolutionarily significant units (ESUs) and management units (MUs). Hedrick et al. (2001a) also showed that gene diversity of the MHC plays a potentially important role in immunocompetence in Gila topminnows, showing that fish with homozygote MHC genotypes were marginally more susceptible to gyrodactylid parasite infections than heterozygous individuals. However, Giese & Hedrick (2003) did not find an association between MHC genotype and resistance to a bacterial infection in the Gila topminnow.

25.7 Concluding remarks

Studies on Poeciliid MHC may unravel some of the most enigmatic questions in evolutionary biology, including the roles of parasite driven selection and sexual selection in adaptive evolution. Poeciliid fishes are particularly well-suited models to examine those questions, given their often (relative) simple MHC organization, well-known parasite faunas and their tractability in behavioral and field experiments. For example, contrasting the mother/embryo interactions between poeciliids with superfetation versus those with ovoviviparity may serve as a model for studying the role of MHC in embryo survival and MHC-mediated prenatal selection in newborns. Research on MHC in poeciliids is furthermore important for conservation biology, not only because these fish are excellent models to

examine captive breeding programs, but also because several Poeciliid species are vulnerable or endangered themselves.

Glossary

ACCORDION MODEL OF MHC EVOLUTION: A model of MHC multigene evolution in which the number of MHC genes is assumed to be expanding (due to gene duplication) or contracting (by gene deletion) depending on the need to protect the host from ever-changing parasite pressure.

BALANCING SELECTION: Balancing selection increases gene diversity and 'balances' allele frequencies so that, compared to neutral evolution, more alleles can be maintained in a population at any single time. Balancing selection can also delay the loss of alleles such that each allele can persist over longer evolutionary time. However, with a high mutation rate, balancing selection can result in a fast turnover of alleles as novel mutations can replace extant alleles. Overdominant selection (also known as heterozygous advantage), negative frequency dependent selection (i.e. rare allele advantage) and selection favoring different alleles in time or space are all types of balancing selection.

BIRTH-AND-DEATH MODEL: This model of multigene evolution proposes that MHC genes are produced by gene duplications. Some of the duplicated genes diverge functionally, whereas others become pseudogenes due to the accumulation of deleterious mutations, or they are deleted from the genome. The birth-and-death model of evolution predicts that sequence variation of member genes will diverge, and that similar sequences are shared among species in the phylogenetic tree.

CONCERTED EVOLUTION: Multigene families are thought to be subject to concerted evolution, in which all member genes of a family evolve as a unit in concert. Genes in a multigene family are thought not to evolve independently from one another, but rather, a mutation that occurs in one of the gene copies is believed to spread through other member genes by repeated occurrence of unequal crossover or gene conversion. Concerted evolution predicts that

sequence variation of member genes is homogenized, and that sequences of alleles at different member genes cluster within species in the phylogenetic tree.

EPISTATIC SELECTION: In population genetics, epistasis is the interaction between two or more genes resulting in fitness values that are different from additively interacting genes. Epistatic selection for favorable gene combinations (i.e. gene combinations with relatively high average fitness) will increase linkage disequilibria between those genes. Functional epistasis between loci can exist, for example, when the loci code for different parts of polypeptide chains that associate to form heterodimers, such as the *DQA1* and *DQB1* loci in the human MHC.

GENE CONVERSION: Gene conversion is a mechanism of homologous recombination that involves the unidirectional transfer of genetic material from a 'donor' sequence to a homologous 'acceptor'. Interlocus gene conversion is the transfer between different loci, for example, paralogous gene sequences in multigene families. Interallelic gene conversion occurs between alleles that reside on homologous chromosomes.

GENETIC HITCHHIKING: A process whereby a selectively neutral or deleterious allele or mutation may spread through the gene pool by virtue of being linked to a beneficial allele.

GOOD GENES HYPOTHESIS: A model of sexual selection in which (generally) females prefer males with great ornaments which indicate greater disease resistance, better foraging abilities, or a more efficient metabolism. The "good genes" of the male indirectly benefit the female as they are inherited to the offspring.

HAPLOTYPE BLOCK (OR HAPLOBLOCK): A large stretch of DNA that is characterized by strong linkage disequilibria that is separated by recombination hotspots. The size of haplotype blocks in the MHC is *ca.* 10^5 bp.

HUMAN LEUKOCYTE ANTIGEN (HLA): The human MHC is commonly referred to as the HLA.

LINKAGE DISEQUILIBRIUM: Linkage disequilibrium (*LD*) is the non-random association of alleles at two or more loci. *LD* can be generated because the loci are physically linked (e.g. they are in close proximity on the same chromosome) or because selection favors or disfavors a particular combination of alleles.

MHC CLASS I GENES: These genes code for molecules that are expressed on the surface of all nucleated somatic cells and which present peptides originating from within the cell (i.e. viral peptides).

MHC CLASS II GENES: MHC genes that code for molecules which are expressed primarily on antigen presenting cells of the immune system. These molecules specialize in the presentation of extracellular pathogen peptides (e.g. bacteria and microparasites).

MHC CLASSICAL GENES: The class I and class II subfamilies of the MHC are divided in classical and nonclassical genes. Classical genes are typically highly polymorphic and expressed. Nonclassical MHC genes display much reduced levels of polymorphism and gene expression. The BIRTH-AND-DEATH MODEL and ABC model of MHC evolution propose that these genes may be in the process of evolving into pseudogenes.

MULLER'S RATCHET: Deleterious mutations can accumulate in small, asexual populations because by chance, not a single individual without any mutations manages to reproduce. Without recombination (and ignoring back-mutations), all individuals in future generations will thus carry at least one mutation. This means that the ratchet has clicked, and it can continue to do so resulting in an accumulation of deleterious mutations (i.e. mutational meltdown). Muller's ratchet is not unique to asexuals, and it can occur also in regions with little or no effective recombination, such as the human Y chromosome and in haplotype blocks.

NONSYNONYMOUS MUTATIONS: Mutations that result in an amino acid replacement which may alter protein function. These mutations are usually thus subject to natural selection.

PEPTIDE BINDING REGION (PBR): A region in the MHC gene that codes for the area of the MHC molecule that binds foreign peptides from bacteria, viruses and micro- and macroparasites. This region is also known as the Antigen Recognition Site (ARS).

PURIFYING SELECTION: Selection against deleterious mutations (i.e. mutations that reduce the fitness of its bearer).

RECOMBINATION: Recombination is one of the four evolutionary processes (together with mutation, selection and random genetic drift). It can occur during crossover between paired chromosomes by which a strand of DNA is broken and rejoined to the homologous strand. Recombination results in the reshuffling of genes, increases the genetic variance in fitness and thereby improves the efficiency of purifying selection. Gene conversion is a type of recombination but leaves the donating chromosome unchanged.

RED QUEEN HYPOTHESIS: The co-evolutionary arms race between hosts immune genes and parasite virulence genes rarely results in fixation of a single genetic variant that confers superior resistance (or infectivity). This process is referred to Red Queen process or Red Queen dynamics in reference to the Red Queen's race in Lewis Carroll's "Through the Looking-Glass" from 1871. The Red Queen said, "*It takes all the running you can do, to keep in the same place.*"

SYNONYMOUS MUTATIONS: These mutations do not change the protein produced by the gene and they are predicted to be selectively neutral. Synonymous mutations are therefore also referred to as silent mutations.

TRANS-SPECIES EVOLUTION: The sharing of MHC allelic lineages between long-diverged species.

Box 1

The model of Associative Balancing Complex (ABC) evolution (van Oosterhout 2009) proposes that recessive deleterious mutations accumulate nearby MHC genes. These mutations maintain polymorphism at the MHC and surrounding region by associative balancing selection. The recessive deleterious mutations can accumulate in the MHC because the effective rate of recombination in some areas of the MHC is very low (see e.g. Stenzel et al. 2004; Gregersen et al. 2006). This reduces the efficiency of PURIFYING SELECTION to remove recessive deleterious mutations (Haddrill et al. 2007). Some of those mutations become fixed in all copies of a particular HAPLOBLOCK in a process analogous to MULLER'S RATCHET (Muller 1932). ABC evolution is inspired by theoretical studies of self-incompatibility loci (*S*-loci) of plants (Uyenoyama 2003), and it offers a plausible explanation for the long-ranging LINKAGE DISEQUILIBRIA and MHC-disease associations in well-researched organisms (humans, dogs and mice) (van Oosterhout 2009). The new theory also elucidates a number of evolutionary properties of the MHC that are not well explained by the traditional theories, such as (1) the shape of the MHC genealogies and trans-species evolution, (2) the unexpected high levels of genetic differentiation of MHC genes under balancing selection, and (3) the high selection coefficients required to maintain MHC polymorphism in small, isolated populations (see van Oosterhout 2009).

Box 2

Introns are found throughout the eukaryote genome and they have been proposed to fulfill a number of functions. Introns may increase the functional diversity of genes through alternative splicing (see above), and they are thought to play a role regulating the recombination rate (Duret 2001; Roy & Gilbert 2006). Microsatellites have been identified in a number of introns of the classical MHC of humans (Balas et al. 2005), chimps (Bak et al. 2006), Atlantic herring (Stet et al. 2008), Atlantic salmon (Stet et al. 2002) and sticklebacks (Reusch et al. 2004), and these repetitive elements are believed to affect protein expression and increase the recombination rate (Arnold et al. 2000; Reusch & Langefors 2005; Majumder et al. 2008).

Recombination has been shown to occur both within and between MHC loci (Hughes et al. 1993; Richman et al. 2003; Reusch & Langefors 2005). The effects of recombination on genetic polymorphism are dependent on the type of selection acting on the gene (Reusch & Langefors 2005). Balancing selection is likely to preserve polymorphisms that are generated by recombination events, while polymorphisms are expected to be lost by random genetic drift at a neutrally evolving locus. Recombination is believed to play an important role introducing polymorphism to the PBR (Charlesworth 2006), providing balancing selection with novel variation above that of the background mutation rate. Evidence of these recombination events leading to the generation of PBR polymorphism may be present in the surrounding introns (Cereb et al. 1997; Bergstrom et al. 1998; Hughes & Yeager 1998; Bergstrom et al. 1999; Elsner et al. 2002; von Salomé et al. 2007)

Besides the functional aspects of intron variation, the repetitive elements often observed within introns are potentially useful markers for population genetic studies. Microsatellites within the MHC region are receiving increased attention as proxy measures of the level of polymorphism at the actual MHC genes themselves (Stet et al. 2002; Santucci et al. 2007; Stet et al. 2008). The screening of microsatellite loci that are tightly linked to the MHC is a particularly time and cost-effective method to study the role of MHC polymorphism in mate

choice and parasite resistance. Future work on Poeciliid MHC would benefit from the development of such linked microsatellite markers.

Reference Cited

Aeschlimann, P. B., Haberli, M. A., Reusch, T. B. H., Boehm, T. and Milinski, M. 2003. Female sticklebacks *Gasterosteus aculeatus* use self-reference to optimize MHC allele number during mate selection. Behavioral Ecology and Sociobiology 54: 119-126

Apanius, V., Penn, D., Slev, P. R., Ruff, L. R. and Potts, W. K. 1997. The nature of selection on the major histocompatibility complex. Critical Reviews in Immunology 17: 179-224

Archard, G. A., Cuthill, I. C., Partridge, J. C. and van Oosterhout, C. 2008. Female guppies (*Poecilia reticulata*) show no preference for conspecific chemosensory cues in the field or an artificial flow chamber. Behaviour 145: 1329-1346

Arkush, K. D., Giese, A. R., Mendonca, H. L., McBride, A. M., Marty, G. D. and Hedrick, P. W. 2002. Resistance to three pathogens in the endangered winter-run chinook salmon (*Oncorhynchus tshawytscha*): effects of inbreeding and major histocompatibility complex genotypes. Canadian Journal of Fisheries and Aquatic Sciences 59: 966-975

Arnold, R., Maueler, W., Bassili, G., Lutz, M., Burke, L., Epplen, T. J. and Renkawitz, R. 2000. The insulator protein CTCF represses transcription on binding to the (gt)(22)(ga)(15) microsatellite in intron 2 of the HLA-DRB1*0401 gene. Gene 253: 209-214

Artamonova, II and Gelfand, M. S. 2007. Comparative genomics and evolution of alternative splicing: The pessimists' science. Chemical Reviews 107: 3407-3430

Bak, E. J., Ishii, Y., Omatsu, T., Kyuwa, S., Tetsuya, T., Hayasaka, I. and Yoshikawa, Y. 2006. Identification and analysis of MHC class II DRB-1 (Patr-DRB1) alleles in chimpanzees. Tissue Antigens 67: 134-142

Balas, A., Aviles, M. J., Alonso-Nieto, M., Zarapuz, L., Blanco, L., Garcia-Sanchez, F. and Vicario, J. L. 2005. HLA-DQA1 introns 2 and 3 sequencing: DQA1 sequencing-based typing and characterization of a highly polymorphic microsatellite at intron 3 of DQA1*0505. Human Immunology 66: 903-911

Beauchamp, G. K. and Yamazaki, K. 2003. Chemical signalling in mice. Biochemical Society Transactions 31: 147-151

Beck, B. B., Rapaport, L. G., Stanley Price, M. R. and Wilson, A. C. 1994. Reintroduction of captive-born animals. In P. J. S. Olney,G. M. Mace and A. T. C. Feistner. (eds) *Creative Conservation: Interactive Management of Wild and Captive Animals.* London: Chapman & Hall pp 265-286

Bell, G. 1982. *The Masterpiece of Nature: the Evolution and Genetics of Sexuality*. Berkeley, Ca: University of California Press

Bergstrom, T. F., Josefsson, A., Erlich, H. A. and Gyllensten, U. 1998. Recent origin of HLA-DRB1 alleles and implications for human evolution. Nature Genetics 18: 237-242

Bergstrom, T. F., Erlandsson, R., Engkvist, H., Josefsson, A., Erlich, H. A. and Gyllensten, U. 1999. Phylogenetic history of hominoid DRB loci and alleles inferred from intron sequences. Immunological Reviews 167: 351-365

Bernatchez, L. and Landry, C. 2003. MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? Journal of Evolutionary Biology 16: 363-377

Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L. and Wiley, D. C. 1987a. Structure of the human class-I histocompatibility antigen, HLA-A2. Nature 329: 506-512

Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L. and Wiley, D. C. 1987b. The foreign antigen-binding site and T-cell recognition regions of class-I histocompatibility antigens. Nature 329: 512-518

Blais, J., Rico, C., van Oosterhout, C., Cable, J., Turner, G. F. and Bernatchez, L. 2007. MHC adaptive divergence between closely related and sympatric african cichlids. Public Library of Science One 2: e734

Borghans, J. A. M., Beltman, J. B. and De Boer, R. J. 2004. MHC polymorphism under host-pathogen coevolution. Immunogenetics 55: 732-739

Boss, J. M. and Jensen, P. E. 2003. Transcriptional regulation of the MHC class II antigen presentation pathway. Current Opinion in Immunology 15: 105-111

Brown, C., Laland, K. and Krause, J. 2006. *Fish Cognition and Behaviour*. Oxford: Blackwell

Brown, J. H., Jardetzky, T. S., Gorga, J. C., Stern, L. J., Urban, R. G., Strominger, J. L. and Wiley, D. C. 1993. 3-Dimensional structure of the human class-II histocompatibility antigen HLA-DR1. Nature 364: 33-39

Carius, H. J., Little, T. J. and Ebert, D. 2001. Genetic variation in a host-parasite association: Potential for coevolution and frequency-dependent selection. Evolution 55: 1136-1145

Cereb, N., Hughes, A. L. and Yang, S. Y. 1997. Locus-specific conservation of the HLA class I introns by intra-locus homogenization. Immunogenetics 47: 30-36

Charbonnel, N. and Pemberton, J. 2005. A long-term genetic survey of an ungulate population reveals balancing selection acting on MHC through spatial and temporal fluctuations in selection. Heredity 95: 377-388

Charlesworth, D. 2006. Balancing selection and its effects on sequences in nearby genome regions. Plos Genetics 2: 379-384

Clarke, B. and Kirby, D. R. S. 1966. Maintenance of histocompatibility complex polymorphisms. Nature 211: 999-1000

Crow, J. F. and Kimura, M. 1970. *An introduction to population genetics theory*. New York: Harper & Row Publishers

Crow, R. T. and Liley, N. R. 1979. A sexual pheromone in the guppy, Poecilia reticulata (Peters). Canadian Journal of Zoology 57: 184-188

Danchin, E., Vitiello, V., Vienne, A., Richard, O., Gouret, P., McDermott, M. F. and Pontarotti, P. 2004. The major histocompatibility complex origin. Immunological Reviews 198: 216-232

Dausset, J. 1958. Iso-leuko-antibodies. Acta Haematol 20: 156-166

de Bakker, P. I. W., McVean, G., Sabeti, P. C., Miretti, M. M., Green, T., Marchini, J., Ke, X. Y., Monsuur, A. J., Whittaker, P., Delgado, M., Morrison, J., Richardson, A., Walsh, E. C., Gao, X. J., Galver, L., Hart, J., Hafler, D. A., Pericak-Vance, M., Todd, J. A., Daly, M. J., Trowsdale, J., Wijmenga, C., Vyse, T. J., Beck, S., Murray, S. S., Carrington, M., Gregory, S., Deloukas, P. and Rioux, J. D. 2006. A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. Nature Genetics 38: 1166-1172

De Boer, R. J., Borghans, J. A. M., van Boven, M., Kesmir, C. and Weissing, F. J. 2004. Heterozygote advantage fails to explain the high degree of polymorphism of the MHC. Immunogenetics 55: 725-731

de Eyto, E., McGinnity, P., Consuegra, S., Coughlan, J., Tufto, J., Farrell, K., Megens, H. J., Jordan, W., Cross, T. and Stet, R. J. M. 2007. Natural selection acts on Atlantic salmon major histocompatibility (MH) variability in the wild. Proceedings of the Royal Society of London Series B-Biological Sciences 274: 861-869

Dionne, M., Miller, K. M., Dodson, J. J., Caron, F. and Bernatchez, L. 2007. Clinal variation in MHC diversity with temperature: Evidence for the role of host-pathogen interaction on local adaptation in Atlantic salmon. Evolution 61: 2154-2164

Dixon, B. 2008. Rene Stet's impact on the study of teleost major histocompatibility genes: evolution from loci to populations. Immunogenetics 60: 77-82

Doherty, P. C. and Zinkernagel, R. M. 1975. Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. Nature 256: 50-52

Duret, L. 2001. Why do genes have introns? Recombination might add a new piece to the puzzle. Trends in Genetics 17: 172-175

Edwards, S. V. and Hedrick, P. W. 1998. Evolution and ecology of MHC molecules: from genomics to sexual selection. Trends in Ecology & Evolution 13: 305-311

Ellis, S. A., Bontrop, R. E., Antczak, D. F., Ballingall, K., Davies, C. J., Kaufman, J., Kennedy, L. J., Robinson, J., Smith, D. M., Stear, M. J., Stet, R. J. M., Waller, M. J., Walter, L. and Marsh, S. G. E.

2006. ISAG/IUIS-VIC Comparative MHC Nomenclature Committee report, 2005. Immunogenetics 57: 953-958

Elsner, H. A., Rozas, J. and Blasczyk, R. 2002. The nature of introns 4-7 largely reflects the lineage specificity of HLA-A alleles. Immunogenetics 54: 447-462

Endler, J. A. 1995. Multiple-trait coevolution and environmental gradients in guppies. Trends in Ecology and Evolution 10: 22-29

Figueroa, F., Mayer, W. E., Sato, A., Zaleska-Rutczynska, Z., Hess, B., Tichy, H. and Klein, J. 2001. Mhc class I genes of swordtail fishes, Xiphophorus: variation in the number of loci and existence of ancient gene families. Immunogenetics 53: 695-708

Flajnik, M. F., Ohta, Y., Namikawa-Yamada, C. and Nonaka, M. 1999. Insight into the primordial MHC from studies in ectothermic vertebrates. Immunological Reviews 167: 59-67

Ford, M. J. 2002. Applications of selective neutrality tests to molecular ecology. Molecular Ecology 11: 1245-1262

Froeschke, G. and Sommer, S. 2005. MHC class II DRB variability and parasite load in the striped mouse (*Rhabdomys pumilio*) in the southern Kalahari. Molecular Biology and Evolution 22: 1254-1259

Garrigan, D. and Hedrick, P. W. 2003. Perspective: Detecting adaptive molecular polymorphism: Lessons from the MHC. Evolution 57: 1707-1722

Giese, A. R. and Hedrick, P. W. 2003. Genetic variation and resistance to a bacterial infection in the endangered Gila topminnow. Animal Conservation 6: 369-377

Gregersen, J. W., Kranc, K. R., Ke, X. Y., Svendsen, P., Madsen, L. S., Thomsen, A. R., Cardon, L. R., Bell, J. I. and Fugger, L. 2006. Functional epistasis on a common MHC haplotype associated with multiple sclerosis. Nature 443: 574-577

Griffiths, S. W. and Magurran, A. E. 1999. Schooling decisions in gunnies (Poecilia reticulata) are based on familiarity rather than kin recognition by phenotype matching. Behavioral Ecology and Sociobiology 45: 437-443

Gutierrez-Espeleta, G. A., Hedrick, P. W., Kalinowski, S. T., Garrigan, D. and Boyce, W. M. 2001. Is the decline of desert bighorn sheep from infectious disease the result of low MHC variation? Heredity 86: 439-450

Haddrill, P. R., Halligan, D. L., Tomaras, D. and Charlesworth, B. 2007. Reduced efficacy of selection in regions of the Drosophila genome that lack crossing over. Genome Biology 8: R18

Hale, K. A. and Briskie, J. V. 2007. Decreased immunocompetence in a severely bottlenecked population of an endemic New Zealand bird. Animal Conservation 10: 2-10

Hedrick, P. W. and Parker, K. M. 1998. MHC variation in the endangered Gila topminnow. Evolution 52: 194-199

Hedrick, P. W., Kim, T. J. and Parker, K. M. 2001a. Parasite resistance and genetic variation in the endangered Gila topminnow. Animal Conservation 4: 103-109

Hedrick, P. W., Parker, K. M. and Lee, R. N. 2001b. Using microsatellite and MHC variation to identify species, ESUs, and MUs in the endangered Sonoran topminnow. Molecular Ecology 10: 1399-1412

Hedrick, P. W. 2002. Pathogen resistance and genetic variation at MHC loci. Evolution 56: 1902-1908

Hedrick, P. W. 2005. *Genetics of Populations*. Sudbury, MA, USA.: Jones & Bartlett Publishers

Houde, A. E. 1997. *Sex, Color, and Mate Choice in Guppies*. Princeton: Princeton University Press

Hughes, A. L. 1991. MHC polymorphism and the design of captive breeding programs. Conservation Biology 5: 249-251

Hughes, A. L., Hughes, M. K. and Watkins, D. I. 1993. Contrasting roles of interallelic recombination at the HLA-A and HLA-B loci. Genetics 133: 669-680

Hughes, A. L. and Yeager, M. 1998. Natural selection at major histocompatibility complex loci of vertebrates. Annual Review of Genetics 32: 415-435

Jameson, S. C., Hogquist, K. A. and Bevan, M. J. 1994. Specificity and flexibility in thymic selection. Nature 369: 750-752

Jensen, L. F., Hansen, M. M., Mensberg, K. L. and Loeschcke, V. 2008. Spatially and temporally fluctuating selection at non-MHC immune genes: evidence from TAP polymorphism in populations of brown trout (*Salmo trutta*, L.). Heredity 100: 79-91

Jordan, M. S., Boesteanu, A., Reed, A. J., Petrone, A. L., Holenbeck, A. E., Lerman, M. A., Naji, A. and Caton, A. J. 2001. Thymic selection of CD4(+)CD25(+) regulatory T cells induced by an agonist self-peptide. Nature Immunology 2: 301-306

Kavaliers, M., Choleris, E., Agmo, A. and Pfaff, D. W. 2004. Olfactory-mediated parasite recognition and avoidance: linking genes to behavior. Hormones and Behavior 46: 272-283

Kelley, J., Walter, L. and Trowsdale, J. 2005. Comparative genomics of major histocompatibility complexes. Immunogenetics 56: 683-695

Klein, J., Bontrop, R. E., Dawkins, R. L., Erlich, H. A., Gyllensten, U. B., Heise, E. R., Jones, P. P., Parham, P., Wakeland, E. K. and Watkins, D. I. 1990. Nomenclature for the major histocompatibility complexes of different species - a proposal. Immunogenetics 31: 217-219

Klein, J., Ono, H., Klein, D. and O'hUigin, C. 1993. The accordion model of Mhc evolution. Progress in Immunology 8: 137-143

Klein, J., Sato, A., Nagl, S. and O'hUigin, C. 1998. Molecular trans-species polymorphism. Annual Review of Ecology and Systematics 29: 1-21

Klein, J. and Sato, A. 1998. Birth of the major histocompatibility complex. Scandinavian Journal of Immunology 47: 199-209

Klein, J., Sato, A. and Mayer, W. E. 2000. Jaws and AIS. In M. Kasahara. (eds) *Major Histocompatibility Complex: Evolution, Structure and Function*. Tokyo: Springer pp 3-26

Kojima, K. I. 1971. Is there a constant fitness value for a given genotype? No! Evolution 25: 281-285

Landry, C., Garant, D., Duchesne, P. and Bernatchez, L. 2001. 'Good genes as heterozygosity': the major histocompatibility complex and mate choice in Atlantic salmon (*Salmo salar*). Proceedings of the Royal Society of London Series B-Biological Sciences 268: 1279-1285

Landry, C. and 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*). Molecular Ecology 10: 2525-2539

Langefors, A., Von Schantz, T. and Widegren, B. 1998. Allelic variation of Mhc class II in Atlantic salmon; a population genetic analysis. Heredity 80: 568-575

Langefors, A., Lohm, J., von Schantz, T. and Grahn, M. 2000. Screening of Mhc variation in Atlantic salmon (*Salmo salar*): a comparison of restriction fragment length polymorphism (RFLP), denaturing gradient gel electrophoresis (DGGE) and sequencing. Molecular Ecology 9: 215-219

Langefors, A., Lohm, J., Grahn, M., Andersen, O. and von Schantz, T. 2001. Association between major histocompatibility complex class IIB alleles and resistance to Aeromonas salmonicida in Atlantic salmon. Proceedings of the Royal Society of London Series B-Biological Sciences 268: 479-485

Lively, C. M. and Dybdahl, M. F. 2000. Parasite adaptation to locally common host genotypes. Nature 405: 679-681

Long, K. D. and Rosenqvist, G. 1998. Changes in male guppy courting distance in response to a fluctuating light environment. Behavioral Ecology & Sociobiology 44: 77-83

Majumder, P., Gomez, J. A., Chadwick, B. P. and Boss, J. M. 2008. The insulator factor CTCF controls MHC class II gene expression and is required for the formation of long-distance chromatin interactions. Journal of Experimental Medicine 205: 785-798

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

McConnell, T. J., Godwin, U. B. and Cuthbertson, B. J. 1998a. Expressed major histocompatibility complex class II loci in fishes. Immunological Reviews 166: 294-300

McConnell, T. J., Godwin, U. B., Norton, S. F., Nairn, R. S., Kazianis, S. and Morizot, D. C. 1998b. Identification and mapping of two divergent, unlinked major histocompatibility complex class II B genes in Xiphophorus fishes. Genetics 149: 1921-1934

McLennan, D. A. 2003. The importance of olfactory signals in the gasterosteid mating system: sticklebacks go multimodal. Biological Journal of the Linnean Society 80: 555-572

Meyer, J. H. and Liley, N. R. 1982. The control of production of a sexual pheromone in the female guppy, *Poecilia reticulata*. Canadian Journal of Zoology-Revue Canadienne De Zoologie 60: 1505-1510

Milinski, M. 2003. The function of mate choice in sticklebacks: optimizing Mhc genetics. Journal of Fish Biology 63: 1-16

Milinski, M., Griffiths, S., Wegner, K. M., Reusch, T. B. H., Haas-Assenbaum, A. and Boehm, T. 2005. Mate choice decisions of stickleback females predictably modified by MHC peptide ligands. Proceedings of the National Academy of Sciences of the United States of America 102: 4414-4418

Milinski, M. 2006. The major histocompatibility complex, sexual selection, and mate choice. Annual Review of Ecology Evolution and Systematics 37: 159-186

Miller, H. C. and Lambert, D. M. 2004. Genetic drift outweighs balancing selection in shaping post-bottleneck major histocompatibility complex variation in New Zealand robins (Petroicidae). Molecular Ecology 13: 3709-3721

Miller, K. M., Kaukinen, K. H., Beacham, T. D. and Withler, R. E. 2001. Geographic heterogeneity in natural selection on an MHC locus in sockeye salmon. Genetica 111: 237-257

Miller, P. S. and Hedrick, P. W. 1991. MHC polymorphism and the design of captive breeding programs - simple solutions are not the answer. Conservation Biology 5: 556-558

Modrek, B., Resch, A., Grasso, C. and Lee, C. 2001. Genome-wide detection of alternative splicing in expressed sequences of human genes. Nucleic Acids Research 29: 2850-2859

Muller, H. J. 1932. Some Genetic Aspects of Sex. American Naturalist 66: 118-138

Neff, B. D., Garner, S. R., Heath, J. W. and Heath, D. 2008. The MHC and non-random mating in a captive population of Chinook salmon. Heredity 101: 175-185

Nei, M. and Rooney, A. P. 2005. Concerted and birth-and-death evolution of multigene families. Annual Review of Genetics 39: 121-152

Nowak, M. A., Tarczyhornoch, K. and Austyn, J. M. 1992. The optimal number of major histocompatibility complex-molecules in an individual. Proceedings of the National Academy of Sciences of the United States of America 89: 10896-10899

O'Brien, S. J., Roelke, M. E., Marker, L., Newman, A., Winkler, C. A., Meltzer, D., Colly, L., Evermann, J. F., Bush, M. and Wildt, D. E. 1985. Genetic-basis for species vulnerability in the cheetah. Science 227: 1428-1434

O'Brien, S. J. and Evermann, J. F. 1988. Interactive influence of infectious disease and genetic diversity in natural populations. Trends in Ecology & Evolution 3: 254-259

Ohta, Y., Okamura, K., McKinney, E. C., Bartl, S., Hashimoto, K. and Flajnik, M. F. 2000. Primitive synteny of vertebrate major histocompatibility complex class I and class II genes. Proceedings of the National Academy of Sciences of the United States of America 97: 4712-4717

Penn, D. and Potts, W. K. 1998. Untrained mice discriminate MHC-Determined odors. Physiology & Behavior 64: 235-243

Penn, D. J. and Potts, W. K. 1999. The evolution of mating preferences and major histocompatibility complex genes. American Naturalist 153: 145-164

Reusch, T. B. H., Schaschl, H. and Wegner, K. M. 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

Reusch, T. B. H. and Langefors, A. 2005. Inter- and intralocus recombination drive MHC class IIB gene diversification in a teleost, the three-spined stickleback *Gasterosteus aculeatus*. Journal of Molecular Evolution 61: 531-541

Richman, A. D., Herrera, L. G., Nash, D. and Schierup, M. H. 2003. Relative roles of mutation and recombination in generating allelic polymorphism at an MHC class II locus in *Peromyseus maniculatus*. Genetical Research 82: 89-99

Robinson, J., Waller, M. J., Parham, P., de Groot, N., Bontrop, R., Kennedy, L. J., Stoehr, P. and Marsh, S. G. E. 2003. IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. Nucleic Acids Research 31: 311-314

Roney, K. E., Cuthbertson, B. J., Godwin, U. B., Kazianis, S., Della Coletta, L., Rosenthal, G. G., Ryan, M. J., Schmidt, M. and McConnell, T. J. 2004. Alternative splicing of major histocompatibility complex class II DXB transcripts in Xiphophorus fishes. Immunogenetics 56: 462-466

Roy, S. W. and Gilbert, W. 2006. The evolution of spliceosomal introns: patterns, puzzles and progress. Nature Reviews Genetics 7: 211-221

Sambrook, Y. G., Russell, R., Umrania, Y., Edwards, Y. K. J., Campbell, R. D., Elgar, G. and Clark, M. S. 2002. Fugu orthologues of human major histocompatibility complex genes: a genome survey. Immunogenetics 54: 367-380

Santucci, F., Ibrahim, K. M., Bruzzone, A. and Hewit, G. M. 2007. Selection on MHC-linked microsatellite loci in sheep populations. Heredity 99: 340-348

Sato, A., Figueroa, F., OhUigin, C., Reznick, D. N. and Klein, J. 1995. Identification of major histocompatibility complex genes in the guppy, *Poecilia reticulata*. Immunogenetics 43: 38-49

Sato, A., Figueroa, F., Murray, B. W., Malaga-Trillo, E., Zaleska-Rutczynska, Z., Sultmann, H., Toyosawa, S., Wedekind, C., Steck, N. and Klein, J. 2000. Nonlinkage of major histocompatibility complex class I and class II loci in bony fishes. Immunogenetics 51: 108-116

Schaschl, H., Tobler, M., Plath, M., Penn, D. J. and Schlupp, I. 2008. Polymorphic MHC loci in an asexual fish, the Amazon molly (*Poecilia formosa*; Poeciliidae). Molecular Ecology 17: 5220-5230

Shiina, T., Inoko, H. and Kulski, J. K. 2004. An update of the HLA genomic region, locus information and disease associations: 2004. Tissue Antigens 64: 631-649

Shiina, T., Ota, M., Shimizu, S., Katsuyama, Y., Hashimoto, N., Takasu, M., Anzai, T., Kulski, J. K., Kikkawa, E., Naruse, T., Kimura, N., Yanagiya, K., Watanabe, A., Hosomichi, K., Kohara, S., Iwamoto, C., Umehara, Y., Meyer, A., Wanner, V., Sano, K., Macquin, C., Ikeo, K., Tokunaga, K., Gojobori, T., Inoko, H. and Bahram, S. 2006. Rapid evolution of major histocompatibility complex class I genes in primates generates new disease alleles in humans via hitchhiking diversity. Genetics 173: 1555-1570

Shohet, A. J. and Watt, P. J. 2004. Female association preferences based on olfactory cues in the guppy, *Poecilia reticulata*. Behavioral Ecology and Sociobiology 55: 363-369

Simmons, L. W. 2005. The evolution of polyandry: sperm competition, sperm selection and offspring viability. Annual Review of Ecology, Evolution and Systematics 36: 125-146

Skarstein, F., Folstad, I., Liljedal, S. and Grahn, M. 2005. MHC and fertilization success in the Arctic charr (*Salvelinus alpinus*). Behavioral Ecology and Sociobiology 57: 374-380

Slade, R. W. and McCallum, H. I. 1992. Overdominant vs frequency-dependent selection at MHC loci. Genetics 132: 861-862

Sommer, S. 2005. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Front Zool 2: 16-34

Spielman, D., Brook, B. W., Briscoe, D. A. and Frankham, R. 2004. Does inbreeding and loss of genetic diversity decrease disease resistance? Conservation Genetics 5: 439-448

Stenzel, A., Lu, T., Koch, W. A., Hampe, J., Guenther, S. M., De La Vega, F. M., Krawczak, M. and Schreiber, S. 2004. Patterns of linkage disequilibrium in the MHC region on human chromosome 6p. Human Genetics 114: 377-385

Stern, L. J., Brown, J. H., Jardetzky, T. S., Gorga, J. C., Urban, R. G., Strominger, J. L. and Wiley, D. C. 1994. Crystal-structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. Nature 368: 215-221

Stet, R. J. M., de Vries, B., Mudde, K., Hermsen, T., van Heerwaarden, J., Shum, B. P. and Grimholt, U. 2002. Unique haplotypes of co-segregating major histocompatibility class II A and class II B alleles in Atlantic salmon (*Salmo salar*) give rise to diverse class II genotypes. Immunogenetics 54: 320-331

Stet, R. J. M., Kruiswijk, C. P. and Dixon, B. 2003. Major histocompatibility lineages and immune gene function in teleost fishes: The road not taken. Critical Reviews in Immunology 23: 441-471

Stet, R. J. M., Mudde, K., Wynne, J. W., Nooijen, A., Dahlgren, T. G., Ruzzante, D. E. and Andre, C. 2008. Characterization of a major histocompatibility class II A gene (Clha-DAA) with an embedded microsatellite marker in Atlantic herring (*Clupea harengus* L.). Journal of Fish Biology 73: 367-381

Stetefeld, J. and Ruegg, M. A. 2005. Structural and functional diversity generated by alternative mRNA splicing. Trends in Biochemical Sciences 30: 515-521

Summers, K., Roney, K. E., da Silva, J., Capraro, G., Cuthbertson, B. J., Kazianis, S., Rosenthal, G. G., Ryan, M. J. and McConnell, T. J. 2008. Divergent patterns of selection on the DAB and DXB MHC class II loci in Xiphophorus fishes. Genetica Advanced online: DOI 10.1007/s10709-008-9284-4

The-MHC-sequencing-consortium 1999. Complete sequence and gene map of a human major histocompatibility complex. Nature 401: 921-923

Ting, J. P. Y. and Trowsdale, J. 2002. Genetic control of MHC class II expression. Cell 109: S21-S33

Tobler, M. and Schlupp, I. 2005. Parasites in sexual and asexual mollies (*Poecilia*, Poeciliidae, Teleostei): a case for the Red Queen? Biology Letters 1: 166-168

Tobler, M., Wahli, T. and Schlupp, I. 2005. Comparison of parasite communities in native and introduced populations of sexual and asexual mollies of the genus Poecilia. Journal of Fish Biology 67: 1072-1082

Trachtenberg, E., Korber, B., Sollars, C., Kepler, T. B., Hraber, P. T., Hayes, E., Funkhouser, R., Fugate, M., Theiler, J., Hsu, Y. S., Kunstman, K., Wu, S., Phair, J., Erlich, H. and Wolinsky, S. 2003. Advantage of rare HLA supertype in HIV disease progression. Nature Medicine 9: 928-935

Uyenoyama, M. K. 2003. Genealogy-dependent variation in viability among self-incompatibility genotypes. Theoretical Population Biology 63: 281-293

van Oosterhout, C., Joyce, D. A. and Cummings, S. M. 2006a. Evolution of MHC class IIB in the genome of wild and ornamental guppies, *Poecilia reticulata*. Heredity 97: 111-118

van Oosterhout, C., Joyce, D. A., Cummings, S. M., Blais, J., Barson, N. J., Ramnarine, I. W., Mohammed, R. S., Persad, N. and Cable, J. 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, R. S., Hansen, H., Archard, G. A., McMullan, M., Weese, D. J. and Cable, J. 2007a. Selection by parasites in spate conditions in wild Trinidadian guppies (Poecilia reticulata). International Journal for Parasitology 37: 805-812

van Oosterhout, C., Smith, A. M., Hanfling, B., Ramnarine, I. W., Mohammed, R. S. and Cable, J. 2007b. The guppy as a conservation model: Implications of parasitism and inbreeding for reintroduction success. Conservation Biology 21: 1573-1583

van Oosterhout, C. 2009. A new theory of MHC evolution: beyond selection on the immune genes. Proceedings of the Royal Society of London Series B-Biological Sciences 276: 657-665

von Salomé, J., Gyllenstwn, U. and Bergstrom, T. F. 2007. Full-length sequence analysis of the HLA-DRB1 locus suggests a recent origin of alleles. Immunogenetics 59: 261-271

Wegner, K. M., Reusch, T. B. H. and Kalbe, M. 2003a. Multiple parasites are driving major histocompatibility complex polymorphism in the wild. Journal of Evolutionary Biology 16: 224-232

Wegner, K. M., Kalbe, M., Kurtz, J., Reusch, T. B. H. and Milinski, M. 2003b. Parasite selection for immunogenetic optimality. Science 301: 1343-1343

Wegner, K. M., Kalbe, M., Schaschl, H. and Reusch, T. B. H. 2004. Parasites and individual major histocompatibility complex diversity - an optimal choice? Microbes and Infection 6: 1110-1116

Westerdahl, H., Hansson, B., Bensch, S. and Hasselquist, D. 2004. Between-year variation of MHC allele frequencies in great reed warblers: selection or drift? Journal of Evolutionary Biology 17: 485-492