

## Genetic Structure at the Major Histocompatibility Complex in the Endangered Barrens Topminnow (*Fundulus julisia*)

Carla Hurt<sup>1,\*</sup>, Natalie Ellis<sup>1</sup>, Alexis Harman<sup>1</sup>, and Courtney Savage<sup>1</sup>

**Abstract** - Genetic diversity at the Major Histocompatibility Complex (MHC) is particularly important for species viability as it allows populations to respond to emerging pathogens and infectious disease. Patterns of variation at this gene complex serve as a useful complement to information obtained from neutral loci for planning management and conservation strategies that seek to ensure the adaptive potential of at-risk species. In this study, we investigated patterns of genetic variation at exon 2 of the MHC class II gene in the critically endangered *Fundulus julisia* (Barrens Topminnow). This species has undergone dramatic declines over the last 30 years, leading to its recent proposal for federal protection under the Endangered Species Act. Patterns of nucleotide substitution and phylogenetic analyses revealing trans-species polymorphisms suggest that this locus has been of adaptive importance in the history of this species. Despite recent population declines and documented population bottlenecks, measures of genetic diversity were high in comparison to patterns observed at putatively neutral microsatellite and mitochondrial markers. Results from this study are discussed in the context of recovery plans for the Barrens Topminnow and lend support to the previous designation of evolutionary significant units and management units based on neutral markers.

### Introduction

The preservation of adaptive genetic variation is a fundamental consideration for establishing and prioritizing management units at or below the species level. Selectively neutral genetic markers such as microsatellites or single nucleotide polymorphisms (SNPs) are valuable for determining the demographic histories and distinctiveness of populations. However, neutral markers may not necessarily reflect evolutionarily relevant and adaptive processes (Cote et al. 2005). For example, studies in primates have demonstrated a correlation between a specific MHC supertype and parasite burden, despite a lack of association between overall neutral genetic diversity and overall parasite load (Schwensow et al. 2007). Patterns of nucleotide diversity at the major histocompatibility complex (MHC) suggest that this gene family is particularly important for adaptive fitness in vertebrates (Sommer 2005). MHC class II genes code for proteins that are expressed on the surface of antigen-presenting cells, where they present exogenously derived antigens to CD4+ T helper cells, triggering an immune response. MHC variants may influence

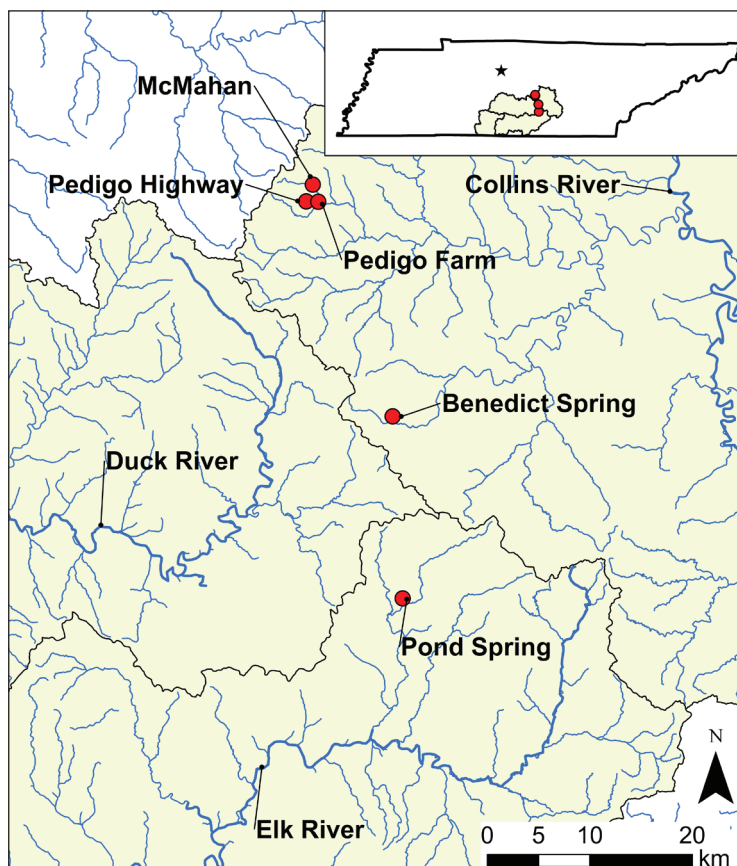
---

<sup>1</sup>Department of Biology, Tennessee Technological University, Cookeville, TN 38505. \*Corresponding author - [churt@tntech.edu](mailto:churt@tntech.edu).

susceptibility to infectious diseases (Harf and Sommer 2005, Wegner 2003), auto-immune responses (Ridgway et al. 1999), mate recognition (Wedekind et al. 1995), and maternal–fetal interactions (Schmidt et al. 1993, see Ujvari and Belov 2011 for further review). As evidence of their functional importance, genes within MHC are the most polymorphic loci known in vertebrates, with as many as 349 identified alleles at a single locus (Robinson et al. 2000). Reduction of functional genetic diversity is a threat to many endangered species as it can increase susceptibility to infectious disease and parasites (Edwards and Potts 1996). MHC variability reflects adaptive processes within and between populations and can be used to inform management strategies that prioritize the maintenance of adaptive genetic variation of endangered species.

*Fundulus julisia* Williams and Etnier (Barrens Topminnow [BTM]) has undergone a dramatic decline since the 1980s. Historically, populations of BTM were known from 3 separate drainages in the Barrens plateau region of middle Tennessee: Duck River, Elk River, and Caney Fork. Prior to 1993, a total of 20 populations throughout all 3 drainages were known to occur (Etnier 1983); these numbers dropped to only a few hundred adults at 7 localities by 1995 (Rakes 1996). Currently, natural populations are known to persist in only 3 areas within the Collins River of the Caney Fork Drainage (Fig. 1; Kuhajda et al. 2014). BTM generally

Figure 1. Location of 5 native sites sampled for genetic analysis: McMahan Creek ( $n = 3$ ), Pedigo Highway ( $n = 15$ ), Pedigo Farm ( $n = 16$ ), Benedict Spring ( $n = 14$ ), and Pond Spring ( $n = 15$ ). Star on map inset indicates Nashville, TN.



inhabit small, spring-fed ponds or streams and are vulnerable to any activity that disturbs groundwater quality or quantity. While loss of habitat has negatively impacted BTM populations, their most significant threat has been the introduction of *Gambusia affinis* (Baird and Gerard) (Western Mosquitofish) throughout their range. Mosquitofish jeopardize the persistence of BTM primarily through predation of juveniles and aggression towards and harassment of adults (Laha and Mattingly 2007). Western Mosquitofish invasion of middle Tennessee has occurred over the past 60 years, and the species was confirmed in the Barrens Fork River system in the 1980s (Etnier and Starnes 2001). All locations where BTM populations had been extirpated and that possessed ideal habitat for BTM had become inhabited by mosquitofish (Rakes 1989, 1996). In response to its dramatic decline, in January 2018 the US Fish and Wildlife proposed to list the BTM as an endangered species under the Endangered Species Act (USFWS 2018).

Fortunately, conservation efforts for the species began in 2001; these efforts have included propagation of captive populations and the establishment of new populations throughout the historical species range (Hurt et al. 2017). To ensure long-term effectiveness of these continuing recovery efforts, maintenance of remaining adaptive genetic variation should be prioritized. Information on the patterns of genetic variation within and between populations is necessary for determination of evolutionarily significant units (ESUs) and management units (MUs) within species (Waples 1995) that help to prioritize conservation needs.

Surveys of genetic variation at putatively neutral loci (mitochondrial control region and microsatellites) suggest that population declines in BTM have led to a reduction of overall genetic variation (Adams et al. 2006, Feldheim et al. 2014, Hurt et al. 2017). Population-level sequence analysis at a portion of the mitochondrial control region showed that all surveyed natural and introduced populations within the Caney Fork and Duck River drainages share a single haplotype. Individuals from the Elk River drainage were polymorphic for 2 unique haplotypes that differed by a single base substitution and a 1 base-pair (bp) indel from the Caney Fork haplotype. Surveys at 14 polymorphic microsatellite loci also showed low levels of heterozygosity within populations relative to observations from other congeneric species. There was significant structuring of microsatellite allele frequencies across loci indicating that Elk River BTM should be managed as a distinct ESU (Hurt et al. 2017). However, nothing is known for BTM about levels of genetic diversity and partitioning of genetic variation in putatively adaptive regions of the genome. Patterns of variation and divergence at adaptive loci frequently differ from patterns observed in neutral markers (Pfrender et al. 2000, Ujvari and Belov 2011). Balancing selection may counteract the effects of genetic drift in small populations and retain functionally important diversity at regions of the genome under selection (Aguilar et al. 2004).

Here we examine patterns of genetic variation at a class II MHC B gene from the 3 remaining natural populations (4 sites) and 1 recently extirpated population of BTM. We looked for evidence of historical balancing selection by comparing levels of heterozygosity at MHC to heterozygosity estimates obtained from microsatellite

loci in the same populations. We also compared partitioning of genetic variation between populations to test for evidence of differential selection at MHC compared to neutral loci in geographically isolated populations of BTM. Results from this study are interpreted in the context of management and conservation planning.

### Materials and Methods

A total of 63 individuals from the 5 remaining natural sites of BTM were sampled for the present study (Fig. 1). These include 4 extant sites in the Collins River of the Caney Fork drainage: Benedict Spring, Pedigo Farm, Pedigo Highway, and McMahan Creek. In addition, we included samples from the recently extirpated population at Pond Spring in the Elk River drainage. Fin clips were obtained during routine monitoring of populations by collaborators at the Tennessee Aquarium in Chattanooga, TN. Tissues from Pond Spring fish were collected by USFWS in 2011 prior to the apparent extirpation of this population. All DNA samples used in this work were used in a previous study, where they were genotyped at 14 microsatellite loci (Hurt et al. 2017). Samples from the Duck River were not available for study as no Duck River BTMs are known to exist in the wild or in captivity. Tissues were stored in 95% ethanol at -20 °C prior to extraction of DNA. We isolated DNA from fin clips using a modification of the protocol by Wang and Storm (2006) and diluted the samples to a concentration of 50–100 ng/μl.

PCR amplification of the MHC class II B gene utilized a newly designed forward primer, Fuju-408 (5'- CTGATCAGCTGTTTATTTCAGAGTC -3'), and degenerate reverse primer MRS (5'- RCTYACCTGATTTAKMCAGA -3'), designed for the congeneric *Fundulus heteroclitus* (L.) (Mummichog) (Cohen 2002). These primers amplify the 3' end of intron 1 and nearly all of exon 2 in the β1 unit, with the exception of 14 bases at the 3' end of exon 2 that are included in the primer sequence. We selected this region for study as it is thought to code for functionally important peptide binding region (PBR; Sommer 2005). For the PCR reaction, we followed manufacturer's recommended conditions using Q5 High-Fidelity DNA polymerase in a 50-μl reaction. PCR cycling conditions were 98 °C for 30 s; 35 cycles at 98 °C for 10 s, 54 °C for 30 s, and 72 °C for 30 s; and a final extension at 72 °C for 2 minutes.

We ligated PCR products into NEB PCR cloning vector (New England Biolabs, Ipswich, MA). We screened clones in PCR reactions with cloning analysis forward and reverse primers (New England Biolabs). We selected between 8 and 16 clones of the appropriate sizes for sequencing and cleaned them using shrimp alkaline phosphatase-exonuclease reaction prior to cycle sequencing on an ABI 3730 automated sequencer (MCLAB). We imported and visualized sequence chromatograms using SEQUENCHER version 5.2 (Gene Codes Corp., Ann Arbor, MI). Sequences were exported to the program Bioedit (Hall 1999) and aligned using the ClustalW multiple alignment algorithm (Larkin et al. 2007), and we adjusted the alignment by eye.

We aligned all sequenced clones to each other and to sequences generated directly from genomic PCR amplification to obtain individual genotypes. We checked individual alignments for PCR recombination events and for PCR errors that can arise

during sequencing of cloned PCR products. We identified recombinant products as sequences in which the first segment was identical to one allele and the second segment was identical to a different allele, and PCR errors as sequences that differed by 1–2 nucleotides from redundant sequences and contained peaks that were not corroborated by sequences generated from PCR from genomic template DNA.

### Statistical analysis

*Variation at MHC within populations.* Population allele frequencies, observed and expected heterozygosities ( $H_o$  and  $H_e$ ), and  $F_{IS}$  values were estimated with the program FSTAT (Goudet 1995). We performed tests for Hardy-Weinberg equilibrium with the program Genepop 4.7 (Rousset 2008), and characterized genetic diversity using allelic richness with the program FSTAT and nucleotide diversity using the program Arlequin (Excoffier et al. 2005).

*Tests for selection.* We tested selection at the MHC using 3 different methods. First, we calculated the ratio of non-synonymous to synonymous substitutions ( $d_N/d_S$ ). If nonsynonymous mutations are maintained in populations by balancing selection and/or positive selection, then  $\omega$  is predicted to be greater than 1. We used the Nei-Gojobori method with Jukes Cantor correction to calculate the overall average  $\omega$  for all sites, the peptide binding region (PBR), and non-PBR separately as implemented in MEGA 5 (Tamura et al. 2011). We inferred the PBR codons from the human model of the MHC class 2 DRB1 locus (Fig. 2; Brown et al. 1993). We determined the significance of  $d_N > d_S$  using a Z-test for selection where the null hypothesis is  $d_N - d_S = 0$ .

Next, we used a maximum likelihood method (Yang et al. 2005) to identify individual codons under selection using the program CODEML (Yang 1997). In this approach, a likelihood-ratio test (LRT) is used to compare 2 pairs of models of sequence evolution. In the first pair of models, the M1 (nearly neutral) model experiences weak purifying selection, so that  $\omega$  can vary between  $<1$  and 0, and the M2 (selection) model is an extension of M1 where a proportion of sites experience positive selection ( $d_N/d_S > 1$ ). In the second pair of models, the M7 (beta) model allows  $\omega$  to vary in a beta distribution between 0 and 1, while the M8 (beta  $d_N/d_S$ ) model extends the M7 model and accounts for positive selection by assigning a



Figure 2. Amino acid sequences inferred from DNA sequences of BTM class II B exon 2. Inferred PBR sites are indicated at the top of the alignment with asterisks. Numbering corresponds with the mature protein (Ono et al. 1993). Highlighted/shaded columns indicate sites with codons that were significant for positive selection using the Bayesian Empirical Bayes (BEB) method for both model M2 and model M8.

subset of sites with  $\omega > 1$ . A LRT is used to compare the M1 model to the M2 model and to compare the M7 model to the M8 model. Positive selection is inferred if the difference in log-likelihood values between the 2 models is greater than the  $\chi^2$  critical value with  $df = 1$ . We then used the Bayes empirical Bayes (BEB) approach to identify the subset of codons that had experienced positive selection (Yang et al. 2005). Using MEGA, we constructed a neighbor-joining tree under the Kimura-2-parameter model of nucleotide substitution to use as an input phylogeny for CODEML analyses.

Finally, balancing selection can alter the distribution of allele frequencies in populations. Specifically, under balancing selection, alleles are maintained at intermediate frequency resulting in a positive value for Tajima's D, which we calculated using the software DNAsp v6.10.01 (Rozas et al. 2003). We obtained statistical significance of the D statistic using a 2-tailed test and assuming that D follows a beta distribution (Tajima 1989).

*Phylogenetic analysis.* We estimated phylogenetic relationships among MHC alleles within BTM by employing the neighbor-net algorithm based on Kimura 2-parameter genetic distance using the software SplitsTree 4.14.5 (Huson and Bryant 2006). Traditional phylogenetic methods may not appropriately represent the history of MHC alleles due to incompatible phylogenetic signals. This approach may more accurately reflect the relationships between alleles at the MHC as it reveals conflicting signals that result from gene conversion and recombination.

In a second analysis, we blasted alleles found in BTM against all *Fundulus* alleles available in GenBank to detect instances of shared functional variation across the genus. We used maximum likelihood and Bayesian phylogenetic reconstructions to explore how BTM alleles are related to congeneric heterospecific MHC alleles. The best-fit model of nucleotide substitution was determined as the Kimura 2-parameter Gamma model using Bayesian Information Criterion (BIC) as performed by MEGA. We performed the maximum likelihood reconstruction in MEGA with 5000 bootstrap replicates, and the Bayesian reconstruction in MrBayes 3.2 (Ronquist and Huelsenbeck 2003). We performed 2 independent runs of 4 Metropolis coupled Markov chain Monte Carlo for  $1 \times 10^7$  generations and sampled every 1000 generations. The burn-in fraction was set at 0.25, resulting in 15,000 trees.

### **Population differentiation and comparisons with microsatellites**

To investigate differences in partitioning of putatively adaptive and neutral genetic variation, we used a hierarchical analysis of molecular variance (AMOVA; Weir and Cockerham 1984) performed separately for microsatellites and MHC sequence data. We used  $\phi$ -statistics to estimate the proportion of genetic variation found among populations ( $\phi_{ST}$ ), among populations within drainages ( $\phi_{SC}$ ), and among drainages ( $\phi_{CT}$ ). We evaluated significance of the fixation indices through random allelic permutation (1000 permutations).

We evaluated the correlation between measures of population differentiation at microsatellites and MHC by performing a Mantel test on estimates of pairwise  $F_{ST}$  values between all populations using the R package Ecodist (Goslee 2007).

For microsatellite genotype data, Weir and Cockerham's (1984) estimator of  $F_{ST}$  values between all pairs of populations was calculated in FSTAT 2.9.3.2 (Goudet 1995). We calculated pairwise MHC  $F_{ST}$  values in Arlequin using only haplotype frequency data.

## Results

### MHC sequences and polymorphism

A total of 63 individuals were cloned and sequenced for a 398-bp amplified gene fragment. Comparisons with other *Fundulus* species confirmed that the amplified sequences are from the exon 2 of a class II MHC gene. No stop codons or indels were detected within the exon region, and a maximum of 2 alleles were detected within an individual. Thus, the alleles found in this study are inferred to represent a single functional locus. This finding is consistent with MHC class II surveys in the congeneric Mummichog, which showed amplification of a single locus using the same reverse primer and a species-specific forward primer.

A total of 12 unique alleles were identified across the 5 sampled populations (Table 1; Genbank accessions MK391390–MK391401). The number of alleles per population varied from 2 in McMahan Spring to 6 in Pond Spring. Pond Spring possessed 4 private alleles not sequenced in other populations. Benedict Spring and McMahan Spring both possessed a single private allele. All other alleles were shared across multiple populations. Tests for Hardy–Weinberg equilibrium detected a significant deficiency of heterozygotes in 2 populations (Pond Spring and Pedigo Highway). Allelic richness was highest for Pond Spring ( $A_R = 3.66$ ) and lowest for McMahan Spring ( $A_R = 2.00$ ), averaging 2.83 across the 5 populations examined (Table 2). Nineteen of the 84 amino-acid positions examined were variable (22.6%) (Fig. 2). The proportion of variable amino acids at putative binding sites was higher (29.2%; 7 out of 24 sites) than the proportion of variable amino acids at non-peptide binding sites (20.0%; 12 out of 60 sites). Nucleotide diversity was highest for Pond

Table 1. The frequency of the 12 MHC alleles found in 63 individuals from the 5 natural *Fundulus julisia* (Barrens Topminnow) sites. Sample sizes are shown in parentheses next to site names.

Allele	Sites					Mean
	Benedict (14)	McMahan (3)	Pedigo F. (16)	Pedigo H. (15)	Pond (15)	
Fuju001	0.679	-	-	-	-	0.136
Fuju002	0.107	-	0.219	0.100	-	0.085
Fuju003	0.036	0.333	0.281	0.667	0.167	0.297
Fuju004	0.143	-	0.094	0.067	-	0.061
Fuju005	0.036	-	-	-	-	0.007
Fuju006	-	0.667	-	-	-	0.133
Fuju007	-	-	0.375	0.167	-	0.108
Fuju008	-	-	0.031	-	0.133	0.033
Fuju009	-	-	-	-	0.233	0.047
Fuju010	-	-	-	-	0.367	0.073
Fuju011	-	-	-	-	0.067	0.013
Fuju012	-	-	-	-	0.033	0.007

Spring ( $\pi = 0.034$ ) and lowest for Pedigo Highway ( $\pi = 0.011$ ), averaging 0.023 across populations.

### Tests for selection

The average nonsynonymous distance ( $d_N$ ) was greater than the average synonymous distance ( $d_S$ ) for both PBR codons and non-PBR codons; however, this difference was only significant for non-PBR segments (Table 3). The  $d_N/d_S$  ratios for PBR and non-PBR codons were 1.265 ( $P = 0.305$ ) and 4.083 ( $P = 0.014$ ), respectively. The LRT for models M1 ( $\ln L = -643.14$ ) and M2 ( $\ln L = -619.59$ ) and for models M7 ( $\ln L = -644.89$ ) and M8 ( $\ln L = -623.71$ ) were significantly positive ( $P < 0.0001$ ) for both tests indicating the evidence of positive selection at a subset of sites within the sequence alignment. In the BEB analysis, the same subset of 13 codons showed evidence of positive selection with both the M2 and M8 model. Six of these codons (3, 7, 32, 55, 80, 83) were within putative PBR in the exon, while the remaining 7 codon (6, 23, 30, 33, 56, 81, 84) were not part of the putative PBR. However, all 7 of the positively selected codons not part of the putative PBR were directly adjacent to putative PBR codons. Estimates of Tajima's D were positive for all populations; however, statistically significant estimates were only obtained for 2 populations, Benedict Spring and Pedigo Farm.

### Phylogenetic reconstruction

The neighbor-net network analysis resulted in a boxed network implying that conflicting signals were present in the dataset and suggesting that recombination

Table 2. Summary of genetic variability at MHC: sample size ( $n$ ), number of alleles (NA), allelic richness ( $A_R$ ), nucleotide diversity ( $\pi$ ), observed heterozygosity ( $H_o$ ), expected heterozygosity ( $H_e$ ), inbreeding coefficient ( $F_{IS}$ ), Tajima's D (D). Significance denoted by \* $P < 0.05$ , \*\* $P < 0.01$ .

Site	$n$	NA	$A_R$	$\pi$	$H_o$	$H_e$	$F_{IS}$	D
Benedict Spring	14	5	2.60	0.027	0.36	0.52	0.37	1.958*
McMahan Creek	3	2	2.00	0.022	0.00	0.53	1.00	1.337
Pedigo Farm	16	5	3.31	0.020	0.75	0.75	-0.01	2.907**
Pedigo Highway	15	4	2.57	0.011	0.13	0.53	0.76*	0.119
Pond Spring	15	6	3.66	0.034	0.47	0.79	0.42*	1.189
Mean		4.40	2.83	0.023	0.34	0.62		

Table 3. Results from the  $d_N/d_S$  ratio test for selection in five natural sites of *Fundulus julisia* (Barrens Topminnow), presented separately for all codons, PBR codons and non-PBR codons.

Statistic	All codons	PBR codons	Non-PBR codons
Mean number of synonymous differences between sequences	1.58	1.50	0.43
Mean number of non-synonymous differences between sequences	10.91	4.28	6.62
Substitutions per synonymous site ( $d_S$ )	0.029	0.068	0.012
Substitutions per non-synonymous site ( $d_N$ )	0.059	0.086	0.049
$d_N/d_S$	2.034	1.265	4.083
Z	0.029	0.511	2.237
P-value	0.013	0.305	0.014



between alleles and/or gene conversion may have occurred (Fig. 3A). The allele *Fuju009* was connected by a single long edge to all other BTM alleles and was exclusively identified in Pond Spring.

The Mummichog was the only other species within the genus *Fundulus* for which homologous MHC class II alleles were available. Although there were no shared alleles between the 2 taxa, traditional phylogenetic reconstructions of all 27 recovered Mummichog MHC alleles and the 12 BTM alleles revealed multiple instances of polyphyly from both maximum likelihood (data not shown) and



Figure 3. (A) SPLITSTREE neighbor network of 12 MHC alleles identified from BTM. (B) Bayesian reconstruction of 12 MHC alleles from BTM and *Fundulus* alleles from Genbank. Triangles denote sequences obtained from BTM. Posterior probabilities shown for nodes with 95% support or greater. Taxonomic identifiers for sequences from Genbank correspond to Genbank accession numbers.

Bayesian reconstructions (Fig. 3B). Specifically, the BTM allele *Fuju009* was more closely related to 3 alleles from Mummichog (Genbank accession numbers AF529617, XM\_012860333, AF529588) than to other BTM alleles. BTM alleles *Fuju001* and *Fuju006* formed a clade with 2 Mummichog alleles (Genbank accession numbers XM\_012860362 and XM\_021314911). Finally the Mummichog alleles XM\_012860177 and XM012860241 were part of a well-supported clade in both ML and Bayesian topologies that included all 12 BTM alleles. Alleles found within BTM did not cluster by population; 5 out of the 12 BTM alleles were shared across multiple BTM populations.

*Population differentiation and comparison with microsatellites.* Results from AMOVA of microsatellite genotype data indicated that genetic variation was partitioned both by drainages (21%) and by populations within drainages (23%) (Table 4). All variance components were significantly greater than zero, and the overall  $\phi_{ST}$  at microsatellites was 0.436. The AMOVA results for MHC sequence data indicated less structuring by populations and drainages than results from microsatellite data (Table 4). The percent of variation attributed to differentiation among drainages was slightly negative (-0.18%), which can occur by chance in the absence of genetic structure. Differentiation between populations within drainages accounted for 32% of the total variation. The fixation indices  $\phi_{SC}$  and  $\phi_{ST}$  were both significantly greater than zero; however,  $\phi_{CT}$  was not significant. The overall  $\phi_{ST}$  at MHC was 0.313, slightly less than what was found at microsatellite markers.

Patterns of population differentiation in putatively adaptive MHC markers mirrored what was found at microsatellites (Fig. 4). Pairwise  $F_{ST}$  values for the 2 classes of marker were significantly correlated ( $r_M = 0.849$ ,  $P = 0.020$ ). Average pairwise  $F_{ST}$  values across all population comparisons were higher for microsatellite genotypes than for MHC sequences (0.421 and 0.299, respectively), indicating that populations were more divergent, on average, at microsatellite loci than at this MHC gene.

Table 4. Summary of the AMOVA results that partition genetic variation of the MHC class II alleles among 5 populations of *Fundulus julisia* (Barrens Topminnow). The significance values for fixation indices were obtained by permutation of alleles 1023 times.

Source of variation	d.f.	Sum of squares	Variance component	% variation	Fixation indices	P-value
<b>Microsatellites</b>						
Among drainages	1	82.99	0.583 $V_a$	20.62	$\phi_{CT} = 0.206$	0.202
Among populations within drainages	3	61.14	0.612 $V_b$	23.01	$\phi_{SC} = 0.290$	< 0.001
Within populations	191	286.20	1.498 $V_c$	56.37	$\phi_{ST} = 0.436$	< 0.001
Total	195	430.33	2.658			
<b>MHC</b>						
Among drainages	1	58.68	-0.010 $V_a$	-0.18	$\phi_{CT} = 0.002$	0.401
Among populations within drainages	3	137.40	1.858 $V_b$	31.47	$\phi_{SC} = 0.314$	< 0.001
Within populations	121	490.72	4.055 $V_c$	68.72	$\phi_{ST} = 0.313$	< 0.001
Total	125	686.79	5.903			

## Discussion

Implementing evolutionary principles into effective conservation of at-risk species requires an understanding of patterns of adaptive variation within and among populations. Here we present results from an examination of diversity, selection, and population structure of a putative adaptive locus, exon 2 of the MHC class II B gene, in the imperiled BTM. Because of the dramatic decline of this species, conservation planning requires a multifaceted approach that includes artificial propagation, translocations, and habitat improvement. Identification of source populations for captive stocks and translocations, and prioritization of populations for habitat monitoring and improvement should consider the ability of species and populations to respond to selection. Our survey provides a first glimpse at the organization of adaptive genetic variation in the BTM that can be used as a guide for informing conservation practices that will maintain adaptive potential of BTM.

### Genetic diversity at the MHC

Levels of allelic variation among the 5 surveyed sites of BTM were moderate compared to observations from the congeneric Mummichog. We identified 12 unique alleles from 64 individuals of BTM. In a similar survey of MHC class II B gene in 5 populations (42 individuals) of Mummichog, 41 unique alleles were identified, so that only a few alleles were found more than once in different individuals

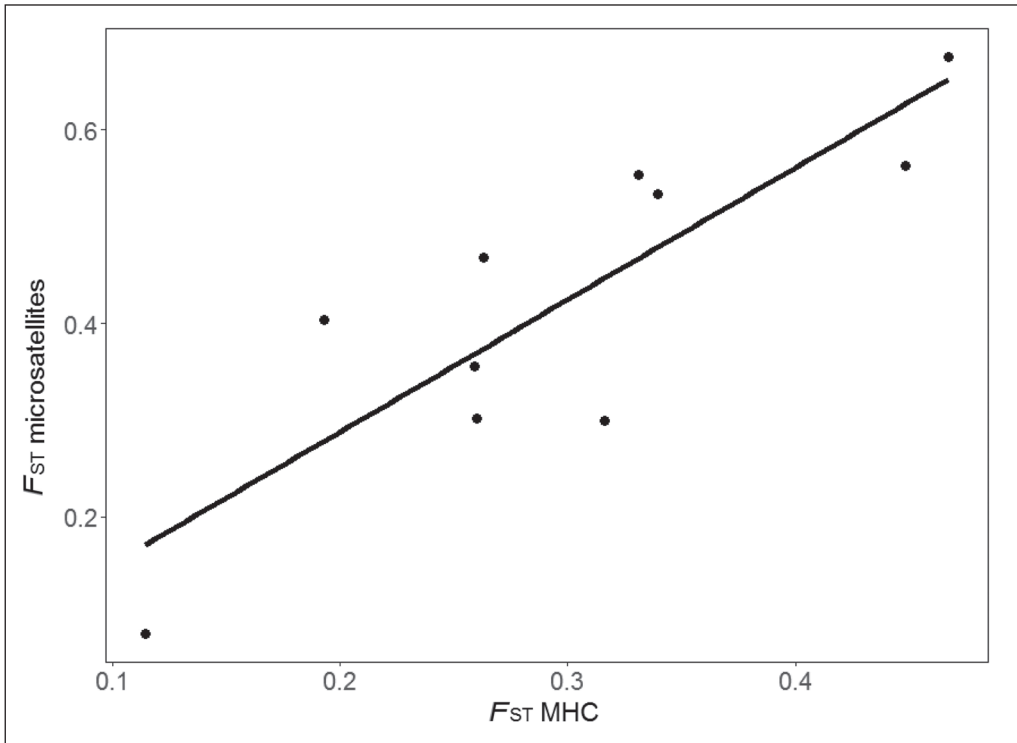


Figure 4. Correlation between the levels of genetic differentiation (pairwise  $F_{ST}$ ) between populations of BTM for microsatellite genotypes and MHC sequence data.

(Cohen 2002). However, Mummichog is likely to harbor greater levels of genetic variation overall as it is a widespread species known for its hardiness and tolerance to environmental change, and is likely to support larger populations. MHC surveys in other imperiled fishes with restricted habitats and similar histories of population decline to the BTM report lower levels of allelic diversity, comparable to what we observed here. In a survey of 5 relict lineages (103 individuals) of the threatened *Oncorhynchus gilae gilae* (Miller) (Gila Trout) only 5 unique alleles were identified across all populations. In the federally endangered *Poeciliopsis occidentalis* (S.F. Baird and Girard) (Gila Topminnow), Hedrick and Parker (1998) describe 9 unique alleles identified from 40 individuals across 4 populations. Allelic diversity at the MHC in the BTM is likely limited by its inherently small population sizes as well as a recent history of demographic decline.

Despite moderate allelic diversity relative to MHC surveys in more widespread species, levels of expected heterozygosity at the MHC were significantly higher than heterozygosity estimates from putatively neutral microsatellite loci at the same populations. The average expected heterozygosity for microsatellites and for MHC across the 5 study sites was 0.333 and 0.624, respectively (1-way ANOVA:  $P = 0.012$ ). Mitochondrial variation was also remarkably low; 4 out of 5 of the natural populations examined here were fixed for a single mitochondrial control region haplotype (Hurt et al. 2017). Collectively these results are consistent with the selective maintenance of genetic variation at the MHC above background neutral loci.

### Selection at the MHC

Tests for balancing selection at the MHC can be categorized as tests that detect selection over the history of species and tests that allow detection of selection in contemporary populations and in the current generation. We found multiple lines of evidence that positive and/or balancing selection has influenced patterns of MHC diversity over species and population timescales; however, evidence for selection in the present generation is lacking. The most commonly applied test for positive selection on evolutionary timescales is the test for  $d_N/d_S > 1$ . It was significantly greater than 1 when considering the entire codon sequence and in the LRT comparing models with and without selection. However, it is not possible to pinpoint the timing of this selection as mutational events that result in elevated  $d_N/d_S$  ratios can take hundreds of thousands of generations to accumulate, and these patterns can take even longer to be erased (Garrigan and Hedrick 2003). Secondly, the retention of ancestral polymorphisms that result in polyphyletic topologies in gene trees spanning multiple related species is also interpreted as evidence of balancing selection (Klein et al 1998). Again, the timing of processes that result in trans-species polymorphisms span species boundaries and do not necessarily imply recent selection.

Selection at MHC over the history of populations may shape the extent of population subdivision and result in patterns of population differentiation that differ from patterns found at neutral loci. Under balancing selection, differentiation between populations is predicted to be reduced compared to neutral loci because

(1) fewer alleles are lost to drift as alleles are kept at equal frequencies and (2) incoming migrant alleles may be selected for and successfully maintained in a population (Schierup et al. 2000). In the BTM, comparisons between measures of population structure at MHC show reduced differentiation at the population level relative to patterns from microsatellite loci. Results from the AMOVA analysis show that at the MHC, sequence variation is not partitioned by drainages, and 69% of the variance is found within populations compared to 56% for microsatellites. Pairwise  $F_{ST}$  values were highly correlated for MHC and microsatellites; however, BTM populations were more divergent at microsatellites than at MHC. Finally, within the current population, an excess of heterozygotes relative to expectations under Hardy–Weinberg proportions is expected under balancing selection. We did not find an excess of heterozygosity in any of the populations examined; instead, 2 out of 5 populations showed a heterozygote deficiency. Collectively, these results suggest that selection at the MHC has historically influenced patterns of nucleotide substitutions at some time in the history of this species and may have reduced that overall level divergence among populations. However, correlations of pairwise  $F_{ST}$  values imply that random genetic drift has also shaped patterns of differentiation between populations at the MHC and may be overriding selection in the current population.

One unexpected result from the analysis of patterns of nucleotide substitution was the elevated  $d_N/d_S$  ratio in non-PBR codons within the MHC exon. Typically  $d_N/d_S$  ratios are highest at sites that interact with foreign peptides (Aguilar and Garza 2006, Bernatchez and Landry 2003, Cohen 2002). Surprisingly we found  $d_N/d_S$  to be only slightly, but non-significantly, elevated at PBR sites, and higher  $d_N/d_S$  were recovered at non-PBR regions. In addition, of the 13 sites identified as being under selection in the BEB analysis, 7 were not part of the PBR but were next to PBR codons. Similar results were obtained in a study of MHC variation in Gila Trout where  $d_N/d_S$  was elevated in codons adjacent to putative PBR sites that were identified by alignment with human HLA-DR1 genes (Peters and Turner 2008). These results cast doubt on the identity of the PBR codons in BTM and suggest that models of MHC protein structure in humans may not adequately describe functional peptide-binding sites in other species.

### **Population structure and units of conservation**

The designation of ESUs and MUs below the species level should incorporate information about both the adaptive differences between populations as well as patterns at neutral markers (Hedrick et al. 2001). Partitioning of alleles among populations at the MHC support the previous recommendation based on microsatellite and mitochondrial markers that Pond Spring be managed as a separate ESU. Pond Spring possessed the most private alleles of any of the natural sites examined here. Four of the 6 alleles identified from Pond Spring were unique to this site, including the allele *Fuju009*, which was highly diverged in the neighbor-net reconstruction and formed a clade with Mummichog alleles in the Bayesian analysis (Fig. 3). Pond Spring was also the most genetically diverse population at MHC as measured by

allelic richness ( $A_R = 3.66$ ), expected heterozygosity ( $H_e = 0.79$ ), and nucleotide diversity ( $\pi = 0.034$ ). Unfortunately, surveys over the last 3 years indicate that this population has been extirpated, most likely due to occupation of this site by Western Mosquitofish (Kuhajda 2014). However, captive populations initiated with brood fish from this site are currently being maintained, and a number of stocked populations have been initiated from this captive population. Our results emphasize the importance of the continued maintenance of this lineage.

Our results also support the earlier recommendations based on microsatellites and mitochondrial loci of 2 separate MU's within the Caney Fork River drainage. The first MU would include Benedict Spring site, which is formed by a spring emanating from a cave in the headwaters of the West Fork of Hickory Creek and is the type locale for the species (Goldsworthy and Bettoli 2005). Here we identified 2 unique alleles not found in any other BTM population. Benedict Spring had the highest average pairwise  $F_{ST}$  estimate (0.40) at the MHC, due to the high frequency of the unique allele *Fuju001*. This site has undergone a series of bottlenecks due to recent droughts; this spring has dried up 3 times since 2006. During each drought, between 120 and 1000 fish were removed and held in captivity until they could be returned (Kuhajda 2014). Likely as a result of these bottlenecks, Benedict Spring possessed the lowest levels of genetic variation at microsatellites of any natural population. At the MHC, however, this site demonstrated substantial diversity ( $\pi = 0.027$ ,  $H_e = 0.52$ ) consistent with balancing selection maintaining functional allelic diversity despite demographic declines and loss of neutral variation. Benedict Spring has also served as a source for brood fish for a separate captive population that has been used to stock numerous sites within the historical distribution of BTM. The second MU within the Caney Fork is composed of the Pedigo Farm and Pedigo Highway sites, which are adjacently located in Lewis Creek within the Barren Fork watershed. These 2 sites possessed nearly overlapping sets of alleles and both shared a private allele (*Fuju007*) not found at other sites. Due to the limitation of available tissues samples, we are unable to make robust conclusions about the genetic structure of McMahan Spring. It is worthwhile to note that 2 out of 3 individuals possessed a unique allele, *Fuju006*, suggesting that this allele is likely to be at high frequency in this population, although further sampling of this site is needed to assess the genetic composition of this population. In general, results from our survey of MHC lend support for the designation of conservation units based on microsatellite and mitochondrial markers.

## Conclusions

Information on patterns of adaptive genetic variation is increasingly playing a role in conservation and management planning for imperiled species (Gebremedhin et al. 2009). Genes from the MHC offer unparalleled opportunity for understanding adaptive differences in natural populations (Bernatchez and Landery 2003, Ujvari and Belov 2011). Results from this survey of a single MHC locus in the endangered BTM suggest that this gene has been of adaptive significance in the history of the species, although drift likely also influenced current patterns of differentiation.

Recent demographic declines have reduced neutral variation as evidenced at both microsatellite and mitochondrial markers, yet substantial allelic diversity was found at the MHC. Nevertheless, information provided by any single locus survey may not reflect the genome as a whole. Management decisions should take into account as many genetic markers as possible. As genomic tools continue to be applied to non-model systems, we expect to see a shift away from analysis of single candidate genes towards the identification and surveys of genome-wide neutral and adaptive loci that enable monitoring of microevolutionary processes in at-risk species.

### Acknowledgments

We thank Bernard Kuhajda, the Tennessee Aquarium, and Conservation Fisheries Inc. for obtaining tissue samples of BTMs. We thank Robert Paine for construction of maps used in this publication. We also thank Philip Hedrick for providing comments on an earlier version of this manuscript. Funding for this project was provided by the US Fish and Wildlife service and by the URECA program at Tennessee Technological University.

### Literature Cited

- Adams, S.M., J.B. Lindmeier, and D.D. Duvernell. 2006. Microsatellite analysis of the phylogeography, Pleistocene history, and secondary contact hypotheses for the Killifish, *Fundulus heteroclitus*. *Molecular Ecology* 15:1109–1123.
- Aguilar, A., and J.C. Garza. 2006. A comparison of variability and population structure for major histocompatibility complex and microsatellite loci in California coastal Steelhead (*Oncorhynchus mykiss Walbaum*). *Molecular Ecology* 15:923–937.
- 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.
- Brown, J.H., T.S. Jardetzky, J.C. Gorga, L.J. Stern, R.G. Urban, J.L. Strominger, and D.C. Wiley. 1993. Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* 364:33–39.
- 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.
- Cote, S.D., A. Stien, R.J. Irvine, J.F. Dallas, F. Marshall, O. Halvorsen, R. Langvatn, and S.D. Albon. 2005. Resistance to abomasal nematodes and individual genetic variability in reindeer. *Molecular Ecology* 14:4159–4168.
- Edwards, S.V., and W.K. Potts. 1996. Polymorphism of genes in the major histocompatibility complex (MHC): Implications for conservation genetics of vertebrates. Pp. 214–237. *In* T.B. Smith and R.K. Wayne (Eds.). *Molecular Genetic Approaches in Conservation*. Oxford University Press, Oxford, UK.
- Etnier, D.A. 1983. Status report for the Barrens Topminnow (*Fundulus julisia* Williams and Etnier). Tennessee Wildlife Resources Agency, Nashville, TN. PCAN No. 0064. 21 pp.
- Etnier, D.A., and W.C. Starnes. 2001. *The Fishes of Tennessee*. University of Tennessee Press, Knoxville, TN. 696 pp.
- Excoffier, L., G. Laval, and S. Schneider. (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1:47.

- Feldheim, K.A., B.R. Kreiser, B. Schmidt, D.D. Duvernell, and J.F. Schaefer. 2014. Isolation and characterization of microsatellite loci for the Blackstripe Topminnow, *Fundulus notatus*, and their variability in two closely related species. *Journal of Fish Biology* 85:1726–1732.
- Garrigan, D., and P.W. Hedrick 2003. Perspective: Detecting adaptive molecular polymorphism—Lessons from the MHC. *Evolution* 57:1707–1722.
- Gebremedhin, B., G.F. Ficetola, S. Naderi, H.R. Rezaei, C. Maudet, D. Rioux, G. Luikart, Ø. Flagstad, W. Thuiller, and P. Taberlet. 2009. Frontiers in identifying conservation units: From neutral markers to adaptive genetic variation. *Animal Conservation* 12:107–109.
- Goldsworthy, C., and P.W. Bettoli. 2005. The fate of stocked Barrens Topminnows, *Fundulus julisia* (Fundulidae), and status of wild populations. Report submitted to Tennessee Wildlife Resources Agency, Nashville, TN.
- Goslee, S., and D. Urban. 2007. The ecodist Package: Dissimilarity-based functions for ecological analysis. *Journal of Statistical Software* 22:1–19.
- Goudet, J. 1995. FSTAT (Version 1.2): A computer program to calculate *F*-statistics. *Journal of Heredity* 86:485–486.
- Hall, T.A. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41:95–98
- Harf, R; Sommer, S. 2005. Association between major histocompatibility complex class II DRB alleles and parasite load in the Hairy-footed Gerbil, *Gerbillurus paebe*, in the southern Kalahari. *Molecular Ecology* 14:85–91.
- 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. 2001. Using microsatellite and MHC variation to identify species, ESUs, and MUs in the endangered Sonoran Topminnow. *Molecular Ecology* 10:1399–1412.
- Hurt, C., B. Kuhajda, A. Harman, N. Ellis, and M. Nalan. 2017. Genetic diversity and population structure in the Barrens Topminnow (*Fundulus julisia*): Implications for conservation and management of a critically endangered species. *Conservation Genetics* 18:1–12.
- Huson, D.H., and D. Bryant. 2006. Application of Phylogenetic Networks in Evolutionary Studies. *Molecular Biology and Evolution* 23:254–267.
- Klein J., A. Sato, S. Nagl, and C. O’huigin 1998. Molecular trans-species polymorphism. *Annual Review of Ecology and Systematics* 29:1–21.
- Kuhajda, B., D. Neely, and K. Alford. 2014. Population status and monitoring of the imperiled Barrens Topminnow, *Fundulus julisia*: 2013 monitoring report. Report to US Fish and Wildlife Service, Cookeville TN. 122 pp.
- Laha, M., and H.T. Mattingly. 2006. Identifying environmental conditions to promote species coexistence: An example with the native Barrens Topminnow and invasive Western Mosquitofish. *Biological Invasions* 8:719–725.
- Larkin, M.A., G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, and J.D. Thompson. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948.
- Ono, H., Klein, D., V. Vincek, F. Figueroa, C. O’huigin, H. Tichy, and J. Klein. 1992. Major histocompatibility complex class II genes of zebrafish. *Proceedings of the National Academy of Sciences of the USA* 89:11886–11890.
- Peters, M.B., and T.F. Turner. 2008. Genetic variation of the major histocompatibility complex (MHC class II  $\beta$  gene) in the threatened Gila Trout, *Oncorhynchus gilae gilae*. *Conservation Genetics* 9:257–270.



- Pfrender M.E., K. Spitze, J. Hicks, K. Morgan, L. Latta, M. Lynch 2000. Lack of concordance between genetic diversity estimates at the molecular and quantitative-trait levels. *Conservation Genetics* 1:263–269.
- Rakes P.L. 1989. Life history and ecology of the Barrens Topminnow, *Fundulus julisia* Williams and Etnier (Pisces, Fundulidae). Masters Thesis. University of Tennessee, Knoxville, TN.
- Rakes, P.L. 1996. Status survey and review with recommendations for management of the Barrens Topminnow for Arnold Engineering Development Center, Coffee and Franklin counties, including surrounding areas. Final Report to The Nature Conservancy, Nashville, TN.
- Ridgway, W.M., M. Fassò, and G.C. Fathman. 1999. A new look at MHC and autoimmune disease. *Science* 284:749–751.
- Robinson J., A. Malik, P. Parham, J.G. Bodmer, and S.G. Marsh. 2000. IMGT/HLA Database—A sequence database for the human major histocompatibility complex. *Tissue Antigens* 55:280–287.
- Ronquist, F., and J.P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.
- Rousset, F. 2008. Genepop'007: A complete re-implementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources* 8:103–106.
- Rozas, J., J.C. Sánchez-DelBarrio, X. Messeguer, and R. Rozas. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496–2497.
- 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.
- Schmidt, C.M., and H.T. Orr. 1993. Maternal/fetal interactions: The role of the MHC class I molecule HLA-G. *Critical Reviews in Immunology* 13:207–224.
- 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. *99:265–277*.
- Sommer, S. 2005. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Frontiers in Zoology* 2:16.
- Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.
- 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.
- Ujvari, B., and K. Belov 2011. Major histocompatibility complex (MHC) markers in conservation biology. *International Journal of Molecular Science* 12:5168–5186.
- United States Fish and Wildlife Service (USFWS). 2018. Endangered and threatened wildlife and plants; endangered species status for Barrens Topminnow. *Federal Register* 83:490–498
- Wang, Z., and D.R. Storm. 2006. Extraction of DNA from mouse tails. *BioTechniques* 41:410–412.
- Waples, R.S. 1995. Evolutionarily significant units and the conservation of biological diversity under the Endangered Species Act. *American Fisheries Society Symposium* 17:8–27.
- Wedekind, C., T. Seebeck, F. Bettens, and A.J. Paepke. 1995. MHC-dependent mate preferences in humans. *Proceedings of the Royal Society of London B* 260:245–249.

- Wegner, K.M.; M. Kalbe, J. Kurtz, T.B.H. Reusch, and M. Milinski. 2003 Parasite selection for immunogenetic optimality. *Science* 301:1343.
- Weir, B.S., and C. Cockerham. 1984. Estimating  $F$ -statistics for the analysis of population structure *Evolution* 38:1358–1370.
- Yang, Z. 1997. PAML: A program package for phylogenetic analysis by maximum likelihood. *Bioinformatics* 13:555–556.
- Yang, Z., Wong, W.S., and R. Nielsen. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution* 22:1107–1118.