MHC Diversity and Mate Choice in the Magellanic Penguin, Spheniscus magellanicus

Gabrielle J. Knafler, J. Alan Clark, P. Dee Boersma, and Juan L. Bouzat

From the Bowling Green State University, Bowling Green, OH (Knafler and Bouzat), University of Washington, Seattle, WA (Clark and Boersma). J. Alan Clark is now at Fordham University, New York, NY 10458.

Address correspondence to Juan L. Bouzat at the address above, or e-mail: jbouzat@bgsu.edu

Abstract

We estimated levels of diversity at the major histocompatibility complex (MHC) class II \(DRß1\) gene in 50 breeding pairs of the Magellanic penguin and compared those to estimates from Humboldt and Galapagos penguins. We tested for positive selection and 2 conditions required for the evolution of MHC-based disassortative mating: 1) greater MHC diversity between breeding pairs compared to random mating, and 2) associations between MHC genotype and fitness. Cloning and sequencing of the \(DRß1\) gene showed that Magellanic penguins had higher levels of genetic variation than Galapagos and Humboldt penguins. Sequence analysis revealed 45 alleles with 3.6% average proportion of nucleotide differences, nucleotide diversity of 0.030, and observed heterozygosity of 0.770. A gene phylogeny showed 9 allelic lineages with interspersed \(DRß1\) sequences from Humboldt and Galapagos penguins, indicating ancestral polymorphisms. \(d_N/d_S\) ratios revealed evidence for positive selection. Analysis of breeding pairs showed no disassortative mating preferences. Significant MHC genotype/fitness associations in females suggest, however, that selection for pathogen resistance plays a more important role than mate choice in maintaining diversity at the MHC in the Magellanic penguin. The differential effect of MHC heterozygosity on fitness between the sexes is likely associated with the relative role of hatching and fledging rates as reliable indicators of overall fitness in males and females.

Key words: disassortative mating, genetic diversity, major histocompatibility complex

Highly polymorphic major histocompatibility complex (MHC) class I and class II molecules perform an important role in the adaptive immune response of most vertebrates (Klein 1986; Trowsdale 1993). MHC class II alleles code for antigen presentation molecules that can recognize and bind a distinct set of extracellular pathogenic peptides (Wakeland et al. 1990). The immune response is then initiated via antigen presentation to helper T cells. Given that individual alleles and the molecules they encode respond only to a limited number of antigenic peptides, having a diverse MHC genotype may be advantageous for disease resistance because of the broad array of pathogens that polymorphic MHC loci can confront (Doherty and Zinkernagel 1975; Hawley and Fleischer 2012).

The exceptional diversity of class II MHC genes may be conserved by balancing selection, in some cases maintaining allelic lineages for millions of years, with ancestral polymorphisms predating speciation events (Klein 1993). Because MHC genes are responsible for initiating the immune response of vertebrates, pathogens have been directly implicated in maintaining variability at MHC loci through 2 forms of balancing selection: overdominance and negative-frequency dependent selection (Snell 1968; Takahata and Nei 1990). Through overdominance, heterozygotes gain an advantage because their antigen peptides can bind a wide range of pathogens (Penn et al. 2002). However, frequency-dependent selection of rare alleles is advantageous when novel pathogens are unrecognized by common alleles present in a population. These 2 processes result in positive selection favoring sequence variation at functional regions of MHC genes (Hughes and Yeager 1998).

In addition to balancing selection, sexual selection by mate choice may also act to conserve MHC variation. In choosing one male over the other, a mate may indirectly benefit by increasing the fitness of their offspring through enhancing their genetic quality (Trivers 1972). Disassortative mating preferences on MHC may function to increase the immunocompetence of offspring. For example, increasing...
offspring genetic diversity at MHC loci may be advantageous in a population experiencing constant pressure from multiple pathogens or bouts of pressure from particular infectious agents (Eizaguirre et al. 2009; Milinski 2006). In contrast, MHC-mediated mating preferences may function to avoid inbreeding because highly polymorphic MHC genes may allow family members to be rejected as potential mates (Potts and Wakeland 1993).

Evidence of positive selection for MHC variation is consistent with both processes of balancing selection and mate choice favoring variation. Positive selection has been recorded in a number of studies that consistently showed higher proportions of nonsynonymous versus synonymous substitutions (e.g., Aguilar et al. 2004; Babik et al. 2005; Hughes et al. 2008; Worley et al. 2010). Two particular conditions, however, should be necessary for the evolution of MHC disassortative mating preferences in response to pathogenic pressures. The first condition is that, under disassortative mating, breeding pairs are predicted to share fewer alleles than those shared by randomly paired individuals. Higher MHC genotypic diversity between breeding pairs would indicate that females are indeed choosing/avoiding males with dissimilar/similar MHC alleles. A growing body of experimental evidence supports this prediction in a number of organisms. For example, early studies by Penn and Potts (1999) found that house mice prefer to mate with individuals of the opposite sex with dissimilar MHC alleles. Investigations on MHC and mate choice in the Atlantic salmon provided similar evidence (Landry et al. 2001), with the additional findings that individual offspring of artificially bred salmon were more similar at their MHC loci and consistently carried higher parasite loads than offspring of naturally spawning salmon (Consuegra and de Leantiz 2008). Furthermore, field studies on MHC-dependent female fidelity in a wild population of songbirds revealed that MHC-similar mates had greater instances of extra-pair copulations (Freeman-Gallant et al. 2003). The second condition for the evolution of MHC disassortative mating preferences requires that MHC variation effectively translates into increased reproductive success (Hamilton and Zuk 1982). If a mate is not directly benefitting by acquiring resources, then they should be acquiring indirect benefits through increasing the genetic quality of the offspring. Affirmation of this prediction is demonstrated by a study that found an association between MHC diversity and juvenile survivorship in the Seychelles warbler (Acrocephalus sechellensis; Brouwer et al. 2010). Furthermore, the role of MHC diversity in disease resistance has been recently demonstrated by Hawley and Fleischer (2012), who provided evidence for pathogen-mediated balancing selection in a wild songbird. Although several studies have shown experimental evidence consistent with the prediction of greater MHC diversity associated with increased fitness, direct evidence from wild populations is still scarce.

We tested whether MHC diversity in a wild population of Magellanic penguins (Spheniscus magellanicus) could potentially be maintained through disassortative mating. That is, individuals may select their mates based on dissimilar alleles present at the MHC class II DRß1 locus. Because penguins live in dense colonies, which may facilitate pathogen transmission, MHC diversity may increase the fitness of breeding pairs. Pathogen transmission has an important effect on the fitness of wild bird populations (Ewald 1994; Hawley and Fleischer 2012) and, in particular, has been shown to reduce the growth rate of Magellanic penguin chicks (Potti 2002). In addition, Magellanic penguins are monogamous, long-lived, have high fidelity (Boersma 2008), and a skewed sex ratio of 1.5 males to females (Boersma, unpublished data). Therefore, female choice may be evident because of the potential long-term fitness consequences on breeding pairs. If parasites or pathogens are a strong selective force in preserving the broad array of MHC alleles through heterozygote advantage or frequency-dependent selection of rare alleles, then MHC disassortative mating should be selectively favored, further advancing MHC polymorphisms (Penn and Potts 1999; Milinski 2006).

Only a few studies have previously characterized MHC genes in penguins. Tsuda et al. (2001) were the first to report genetic polymorphisms and compare phylogenetic relationships of the DRß-like genes among 5 species of penguin in the genera Pygoscelis and Eudyptula. With the exception of the Galapagos penguin (Bollmer et al. 2007), subsequent studies on the characterization of the MHC in penguins (Kikkawa et al. 2005, 2009) revealed high-sequence diversity in all species sampled. Although these studies reported number of alleles and genetic variation present at DRß-like loci, extrapolating such estimates of genetic diversity to the population level should be done with caution because of small sample sizes.

On the basis of the genotyping and sequencing of 100 Magellanic penguins, we report here population-level estimates of MHC genetic diversity at the class II DRß1 gene in the largest continental breeding colony of this species (Boersma 2008). We compared MHC diversity in the Magellanic penguin with that from previous studies on Humboldt and Galapagos penguins (Kikkawa et al. 2005, 2009; Bollmer et al. 2007), constructed a gene phylogeny to assess the potential presence of ancestral polymorphisms, and tested for positive selection at this locus. Furthermore, based on the MHC DRß1 genotyping of 50 breeding pairs, we assessed whether MHC disassortative mating preferences are operating in the Magellanic penguin. Specifically, we tested 2 major conditions required for MHC-facilitated mate choice: 1) evidence for higher MHC diversity between genotypes of individuals in breeding pairs compared to that of randomly selected pairs, and 2) an association between MHC genotype and fitness. Results from this study provide insights into the potential role of balancing selection and mate choice in maintaining MHC variation in wild populations.

Materials and Methods

Population Sampling and Fitness

Fifty penguin breeding pairs were randomly chosen from a major breeding colony of Magellanic penguins located at

Journal of Heredity 2012:103(6)
Punta Tombo, in the Atlantic shores of southern Patagonia, Argentina (44°26’ S, 65°10’ W). Penguins were individually identified with flipper bands and subsequently followed as part of a long-term ecological study initiated by P.D. Boersma in the early 1980s (Boersma et al. 1990). Fitness was defined as the number of eggs hatched and chicks fledged per individual. We used fitness data from 2002, the year when blood was collected, because we could assign the number of eggs hatched and chicks fledged for every chosen banded breeding pair. Sampling was conducted under ethical protocols approved by local authorities (Division of Fauna and Flora Permit #18/02FyFS, Chubut Province, Argentina) as well as USDA-APHIS importation permit #42579 to Robert C. Faucett and Alan Clark (University of Washington Burke Museum of Natural History).

**MHC Genotyping**

Blood samples (approximately 200 µL) were collected by one of us (JAC) through puncture of the brachial vein and stored in Queen's lysis buffer (10 mmol/L Tris, 10 mmol/L NaCl, 10 mmol/L EDTA, 1% s-lauroylsarcosine pH 7.5 (Seutin et al. 1991) for further DNA analysis. DNA extractions were conducted using standard phenol-chloroform extraction protocols (Sambrook et al. 1989). PCR amplification of the MHC class II DRβ1 gene was performed using primers Lpen.hum1F2 (5′-ACTCTCTGGCACAGCCGCCTTG -3’) and Lpen.hum2R (ACACGCTCTCCCCTCCTGTG) originally developed by Kikkawa et al. (2005, 2009). Primers were designed to amplify exon 2 of the class II DRβ1 gene in the Humboldt penguin, *Spheniscus humboldti*, and other closely related penguin species in the genus *Spheniscus* (Kikkawa et al. 2009). Based on the analysis of 5 Magellanic penguins, Kikkawa et al. (2009) described the successful amplification and characterization of the DRβ1 gene in this species. DNA sequencing techniques showed that the primers were apparently locus-specific, with a single MHC class II DRβ1 gene being amplified rather than multiple MHC class II loci with similar sequences. Despite strong evidence for single locus amplification of the class II DRβ1 gene, we tested the primers for potential multilocus amplification.

Amplification reactions were performed in 25 µL volumes containing 1× Taq buffer, 1.5 mmol/L MgCl₂, 0.2 mmol/L of each dNTP, 0.5 µmol/L of each primer, and 0.1 units of Taq. The PCR amplification profile included an initial denaturing step at 95°C for 2 minutes, and 27 cycles of 94°C for 1 minute, 62°C for 1 minute, and 72°C for 2 minutes, with a final extension step at 72°C for 15 minutes, and an ending step at 4°C for 5 minutes. PCR products were then cloned using the TOPO TA Cloning Kit for Sequencing (Invitrogen Inc., Carlsbad, CA). Multiple bacterial colonies containing cloned PCR products were miniprepped for each individual penguin using GeneJET Plasmid Miniprep Kits (Fermentas Life Sciences, Burlington, Ontario, Canada) and sent to The University of Chicago Cancer Research Center DNA Sequencing Facility for sequencing.

Other studies showed that simultaneous amplification of more than one MHC locus and/or pseudogenes may arise because of similarity in MHC class II sequences originated by gene duplication (Aguilar et al. 2006; Edwards 1999). In addition, the amplification of a single MHC locus may lead to the formation of chimeric sequences because of the potential multi-allelic state of nuclear templates (Lenz and Becker 2008). Estimates of genetic diversity and comparison of MHC similarity between penguin pairs could therefore be confounded by multilocus amplification and the presence of DNA chimeras. To prevent multilocus amplification, we chose primers reported to amplify a single MHC class II gene in the Magellanic penguin (Kikkawa 2009). Furthermore, we sequenced up to 12 clones per individual and found, in all cases, no more than 2 distinct translatable sequences, confirming that the primers indeed amplified a single MHC locus.

To account for the potential misidentification of chimeric sequences as true alleles, we applied the conservative criteria for the characterization of MHC alleles used by the International Society for Animal Genetics (ISAG) Human, Cattle, and Dog Nomenclature Committees (Davies et al. 1997; Kennedy et al. 1999; Marsh et al. 2001). The criteria require that in order for a new allele to be reported it must appear in, at least, 2 individuals or in the same individual from 2 independent PCR reactions. Therefore, all identified alleles in this study were first validated by applying a cut-off in which sequences were not considered true alleles unless they were present in another individual. Subsequently, individuals with unique sequences were subjected to a second round of PCR, cloning and sequencing. Potential alleles were then confirmed only if detected in both PCR reactions.

In summary, to conform with ISAG’s criteria for characterizing MHC alleles, the genotyping of individuals resulted from the initial sequencing of 4–6 clones per individual followed by the sequencing of an additional 4–6 clones from a second independent PCR reaction. Homozygote individuals were further subjected to a third PCR, cloning, and sequencing to account for potential misidentification of heterozygotes as a result of PCR bias, leading to a total of 12 sequenced clones per homozygote and 8–12 sequenced clones per heterozygote. This also allowed us to estimate the rate of misclassifying heterozygotes as homozygotes after initially sequencing 4–6 MHC clones from a single PCR reaction.

**Data Analysis**

All DNA sequences were assembled and aligned using BioEdit, version 7.0.5.3 (Hall 1999). On the basis of the multiple sequence alignment, we estimated the total number and proportion of polymorphic sites. DNA sequences were then imported into DnaSP, v.5.0. (Rozas et al. 2003) and Arlequin, v.3.5.1.2 (Schneider et al. 2000) to calculate standard estimates of genetic diversity at the nucleotide and population levels. Nucleotide diversity (π) and haplotype diversity (θ) were calculated using the program DnaSP, whereas observed and expected heterozygosities (Hₒ and Hₑ) as well as the total number of alleles in the sample were
calculated using Arlequin. To assess ancestral relationships among individual MHC alleles, we performed a phylogenetic analysis using Neighbor-joining and maximum parsimony algorithms (Felsenstein 2004), implemented in MEGA, v.4 (Tamura et al. 2007). The phylogenies included sequences of all DRβ1 alleles reported in this study for the Magellanic penguin (Genbank accession numbers JQ715808-JQ715852) and previously reported sequences for Humboldt and Galapagos penguins (Kikkawa et al. 2005, 2009; Bollmer et al. 2007). The phylogenetic tree was rooted using DRβ1 sequences from the domestic chicken (Gallus gallus; GenBank No. M29763) and the common cactus finch (Geospiza scandens; GenBank No. Z74412), and the confidence of groupings estimated through 500 bootstrap replicates.

To assess balancing selection and MHC-based disassortative mating, we evaluated evidence for positive selection, genetic diversity of breeding pairs, and the potential association between MHC genotype and fitness. Evidence for positive selection was assessed using the ratio of nonsynonymous to synonymous substitutions. We used MEGA, v.4, to measure the relative rate of nonsynonymous (dN) and synonymous (dS) substitutions according to Nei and Gojobori (1986), using Jukes and Cantor’s (1969) correction for multiple hits. To test whether positive selection was operating at this locus, we estimated the variances of dN and dS using the bootstrap method implemented in MEGA (500 replications) and compared the relative abundance of nonsynonymous versus synonymous substitutions using a Z-test. dN/dS ratios were calculated for the entire exon 2 of the DRβ1 gene as well as independently for the antigen-binding regions and nonpeptide-binding regions described for humans (Brown et al. 1993).

To assess levels of MHC diversity of mating pairs versus random mating scenarios, we identified the alleles of the 50 selected breeding pairs and compared them to the alleles of these same 50 males and 50 females if they were randomly paired. We tested the null hypothesis of random pairing between males and females by performing a permutation test in which we compared the test statistic, that is, the number of shared alleles between the 50 true breeding couples, to a sampling distribution generated by randomly permuting the females and computing the number of shared alleles between randomly paired couples (1000 permutations). In addition, the proportion of nucleotide differences (p-distance) between alleles from mating pairs was compared to that of randomly paired males and females to test whether mating pairs were likely to have alleles that differed by more mutations.

To evaluate the potential association of MHC genotype and fitness, we performed an ANOVA to assess differences in fitness (i.e., number of eggs hatched and number of chicks fledged) between homozygote and heterozygote individuals. We performed the analysis considering all individuals together as well as independently for males and females. In addition, we compared the fitness of individuals having “common” alleles (with frequencies ≥8%) with genotypes composed of rare alleles (with frequencies ≤4%). We performed a nonparametric ANOVA to account for the lack of normality and variance heterogeneity of the data.

To assess the potential effects of MHC allelic diversity on individual fitness, we estimated the proportion of nucleotide differences (p-distance) between individual alleles and then performed a regression analysis of individual fitness on the sequence divergence between alleles of individual genotypes. We ran this analysis separately for males and females because parental investment may differ by sex. We used JMP statistical software (SAS Institute Inc., Cary, NC) for all statistical analyses for the relationship between MHC genotype and fitness.

Results

Genetic Diversity and Phylogenetic Relationships of the MHC class II DRβ1 Gene

The DNA amplification, cloning, and sequencing of the MHC DRβ1 gene in 100 Magellanic penguins allowed us to genotype individuals and estimate nucleotide and population-level variation (see Supplementary Materials online for genotype information). The initial sequencing of 4–6 clones per individual suggested that amplification primers were locus specific for the MHC class II DRβ1 gene given that translated sequences revealed no stop codons in the 420 bp region of exon 2, and that we did not detect in any case more than 2 distinct sequences/alleles. The initial count of homozygote individuals (47) suggested, however, that PCR bias or preferential cloning of individual alleles could be underestimated the number of heterozygotes if only a limited number of clones is used to genotype individuals. We therefore performed 2 additional independent PCR reactions resulting in a total of 12 clones sequenced for each individual originally categorized as a homozygote, which decreased the number of homozygotes from 47 to 23, leading to a 45% increase in observed heterozygosity.

DNA sequences of the Magellanic penguin MHC class II DRβ1 gene showed relatively high sequence and population-level diversity (Table 1). Using ISAG’s criteria for defining MHC alleles, 45 alleles were detected in our sample of 100 individuals. We found 54 polymorphic sites within the 420 bp region of exon 2. We estimated the average sequence divergence (p-distance) as 0.036 and nucleotide diversity as 0.030. Twenty one of 45 alleles appeared only once in the sample whereas 3 alleles were present in relatively high frequency (0.08, 0.15, and 0.17). We estimated expected heterozygosity (H_E) and haplotype diversity (θ) at 0.930. Genotype frequencies were significantly different from those expected under Hardy-Weinberg equilibrium.

Table 1 Levels of genetic diversity (V = number of variable sites; A = number of alleles; H_E and H_O = observed and expected heterozygositites; and σ = nucleotide diversity) at the MHC class II DRβ1 locus estimated from 100 Magellanic penguins (50 males and 50 females)

<table>
<thead>
<tr>
<th></th>
<th>V</th>
<th>A</th>
<th>H_E</th>
<th>H_O</th>
<th>σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>37</td>
<td>28</td>
<td>0.760</td>
<td>0.946</td>
<td>0.031</td>
</tr>
<tr>
<td>Female</td>
<td>25</td>
<td>25</td>
<td>0.780</td>
<td>0.914</td>
<td>0.029</td>
</tr>
<tr>
<td>Total</td>
<td>54</td>
<td>45</td>
<td>0.770</td>
<td>0.930</td>
<td>0.030</td>
</tr>
</tbody>
</table>
expected under Hardy–Weinberg equilibrium and showed an excess of homozygote genotypes (observed heterozygosity \(H_O = 0.770\)). Males and females had similar diversity estimates (Table 1).

In comparison to other species in the *Spheniscus* genus, the Magellanic penguin had higher nucleotide and haplotype diversities (Table 2). Average proportions of differences among sequences of Magellanic and Humboldt penguins were similar, but both were higher than the estimated \(p\)-distance for the Galapagos penguin (when compared over the same DNA region). Although this study found 45 MHC alleles for Magellanic penguins, previous studies found 8 alleles for Humboldt penguins (\(n = 20\); Kikkawa et al. 2009) and only 3 alleles for Galapagos penguins (\(n = 30\); Bollmer et al. 2007).

The number of alleles in the Humboldt penguin decreased to 7 when sequences were trimmed to the same length of the newly reported sequences of Magellanic penguins (420 bp). An earlier study of 5 Magellanic penguins (Kikkawa et al. 2009) reported 8 DRß1 alleles. We detected 7 of the 8 alleles in our sample. When Magellanic and Humboldt penguin sequences were trimmed to the length of the MHC sequences reported for the Galapagos penguin (157 bp), both species retained a significantly higher number of alleles (Table 2).

Results from the neighbor-joining phylogenetic analysis of 60 MHC DRß1 sequences from Magellanic, Humboldt, and Galapagos penguins are shown in Figure 1. Clustering of alleles with respect to species is not supported. The spread of Humboldt and Galapagos penguin DRß1 sequences throughout the tree, interspersed with Magellanic penguin sequences, suggests the absence of major allelic lineages at the species level. The clustering of Magellanic penguin sequences allowed the identification of nine allelic lineages characterized by monophyletic groups with bootstrap support higher than 41% (lineages A–I; Figure 1). Bootstrap support was, however, very low at most basal nodes. On the basis of the limited genetic divergence among sequences, some clusters could be combined into major lineages (e.g., A–F and G–I; Figure 1). Three of the Humboldt penguin sequences clustered within 2 of the major allelic lineages defined by the Magellanic sequences (i.e., lineages G and I; Figure 1). Confidence for the grouping of Humboldt allele Sphu004 with Magellanic alleles Smag02 and Smag13 is consistent with previous reports (Bollmer et al. 2007; Kikkawa et al. 2009) of allele sharing between Humboldt and Magellanic penguins. Both individuals having the identical allele Sphu004 and Smag02 were homozygotes in captive populations. Accordingly, this allele was relatively frequent (0.06) in our sample of 100 Magellanic penguins from the Punta Tombo colony. The remaining 5 Humboldt penguin sequences appeared to be related to allele lineage H (Sphu0011, Sphu001, Sphu002) or branched prior to the divergence of lineages A–D and E–F (Sphu006 and Sphu005, respectively; Figure 1), although low bootstrap values provided limited confidence for these groupings. The independent grouping of *S. mendiculus* sequences is most likely the result of the limited number of phylogenetically informative sites (3) reported in these short sequences (157 bp). The phylogenetic tree based on maximum parsimony upheld neighbor-joining allelic lineages B, C, D, F, H, and I (with bootstrap values ≥41%; data not shown). Consistent with the neighbor-joining phylogenetic analysis, Humboldt alleles were interspersed with Magellanic penguin sequences, which indicated that the clustering of alleles with respect to species was not supported.

### Table 2  Levels of MHC class II DRß1 sequence diversity in the Magellanic penguin and 2 other penguin species in the genus *Spheniscus*

<table>
<thead>
<tr>
<th>Species</th>
<th>Sequence Length (bp)</th>
<th>(n)</th>
<th>(A)</th>
<th>(V)</th>
<th>N-Diff (±SE)</th>
<th>(P)-Distance (±SE)</th>
<th>(\pi)</th>
<th>(h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Magellanic</td>
<td>420</td>
<td>100</td>
<td>45</td>
<td>54</td>
<td>14.74 ± 2.1</td>
<td>0.036 ± 0.005</td>
<td>0.030</td>
<td>0.930</td>
</tr>
<tr>
<td>Humboldt</td>
<td>157</td>
<td>100</td>
<td>28</td>
<td>31</td>
<td>10.75 ± 1.9</td>
<td>0.068 ± 0.012</td>
<td>0.061</td>
<td>0.910</td>
</tr>
<tr>
<td>Galápagos</td>
<td>157</td>
<td>20</td>
<td>7</td>
<td>29</td>
<td>12.8 ± 2.3</td>
<td>0.031 ± 0.006</td>
<td>0.022</td>
<td>0.671</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6</td>
<td>20</td>
<td>9.0 ± 1.8</td>
<td>0.057 ± 0.012</td>
<td>0.040</td>
<td>0.659</td>
</tr>
</tbody>
</table>

We show the number of individuals genotyped \((n)\), the number of alleles detected \((A)\), number of variable sites \((V)\), average number of nucleotide differences \((N\text{-Diff})\), average proportion of differences \((P\text{-Distance})\), nucleotide diversity \((\pi)\) and haplotype diversity \((h)\). For proper comparisons, diversity estimates for Magellanic and Humboldt penguins are reported for the 420 bp region of the DRß1 exon 2 amplified in this study as well as for the 157 bp region of exon 2 used in Bollmer’s et al. (2007) analysis of the Galapagos penguin. SE, standard error.

\(a\) Data from Kikkawa et al. (2009).

\(b\) Data from Bollmer et al. (2007); \(x\) and \(b\) could not be estimated because allele frequencies were not reported.

Balancing Selection, Disassortative Mating and Fitness Associations

Nonsynonymous to synonymous substitution rates uncovered evidence for positive selection at the MHC class II DRß1 locus. The \(d_N/d_S\) ratio for the DRß1 exon 2 was significantly greater than 1, with an average of 7.00 higher number of non-synonymous than synonymous substitutions (Table 3). The \(d_N/d_S\) ratio was significantly increased when considering the peptide binding region (PBR), which showed a ratio of 11.20 compared to 2.86 detected for the non-PBR regions of exon 2 (Table 3).

The permutation test comparing MHC genetic diversity between the 50 breeding pairs and the randomly generated distribution of male and female pairs was nonsignificant \((p = .125; 1000 permutations)\). Thus, 12.5% of the random
**Figure 1.** Neighbor-joining gene tree of MHC class II DRβ1 exon 2 sequences from 3 species of penguin in the *Spheniscus* genus. Representative class II β sequences from the domestic chicken (*Gallus gallus*) and the common cactus-finch (*Geospiza scandens*) were used as outgroups. Numbers at nodes are bootstrap values based on 500 replicates. *S. magellanicus* sequences reported in this study are designated as Smag. *S. magellanicus* and *S. humboldti* sequences from Kikkawa et al. (2009) are designated Spma and Sphu, respectively. Galapagos penguin alleles (Spme) are from Bollmer et al. (2007). Solid brackets A-I represent Magellanic penguin allelic lineages defined by monophyletic groups with bootstrap support higher than 41%. Limited bootstrap support for basal branching prevents the characterization of potential major allelic lineages 1 and 2, represented by dotted brackets (see text).
permutations showed a less or equal number of shared alleles than the observed value of 7 shared alleles within all breeding pairs. Similarly, we did not find significant differences in the average proportion of nucleotide differences observed between alleles of breeding pairs compared to that of randomly paired samples of males and females.

Statistical analyses of MHC genotype/fitness associations showed different trends for males and females (see Supplementary Materials online for genotype and fitness data). First, the differential fitness between homozygote and heterozygote females was significant ($p = .022; F_{(1,46)} = 5.64$), with heterozygotes hatching more eggs (mean = 0.89; SE = 0.16) than homozygotes (mean = 0.23; SE = 0.17; Figure 2). Female heterozygotes also showed a non-significant trend to fledge more chicks than homozygotes ($p = .261; F_{(1,18)} = 1.35$); in fact, there were no female homozygotes that fledged any chicks. Conversely, male heterozygotes neither hatched more eggs ($p = .101; F_{(1,46)} = 2.81$) nor fledged more chicks ($p = .214; F_{(1,18)} = 1.66$) than homozygote males (Figure 2). Although limited variation in the fitness data, with parent hatching or fledging only 0, 1, or 2 chicks, could potentially account for the lack of significance in males, it did not prevent finding significant differences that allowed us to link one of the fitness measures to MHC diversity in females. Regression analyses between MHC sequence divergences (i.e., $d_k$-distance) among alleles in individual genotypes and fitness revealed that MHC sequence variation for both males and females could not explain fitness differences in eggs hatched or chicks fledged ($p \geq .076$). Finally, we could not detect significant differences in fitness, either in eggs hatched or chicks fledged, between individuals who had at least one of the 3 most frequent alleles compared to those whose genotypes had rare alleles.

### Discussion

In this study, we found no evidence of MHC-based disassortative mating in a large, naturally occurring breeding colony of the Magellanic penguin. Consistent with other studies on MHC variation, the amplification, cloning, and sequencing of the class II $DR\beta_1$ gene from 100 Magellanic penguins revealed high levels of genetic diversity at both the nucleotide and population levels. We detected 45 alleles with an average sequence divergence of 0.036 differences per substitution site and an expected heterozygosity of 0.930. The relatively large effective population size of this species, with an estimated 1.5 million breeding adults (Gandini et al. 1996; Schiavini et al. 2005), and the recent population expansion of the breeding colony at Punta Tombo (Boersma et al. 1990) likely contributed to maintaining high levels of genetic diversity.

Compared to other penguin species in the genus *Spheniscus*, the Magellanic penguin had considerably greater levels of MHC diversity (in number of alleles, nucleotide diversity, and haplotype diversity) than estimates for the Humboldt and Galapagos penguins when analyzed over the same DNA region (Bollmer et al. 2007; Bollmer et al. 2007). Lack of MHC diversity in the Galapagos penguin has been previously attributed to repeated population bottleneck events and/or reduced pathogenic pressure in their restricted range of the Galapagos Islands (Akst et al. 2002; Bollmer et al. 2007). The greater MHC diversity found in the Magellanic penguin is therefore consistent with this idea, because this species has...
been historically more abundant and has a much wider distribution range. Like the Galapagos penguin, the Humboldt penguin also had reduced MHC variation in comparison to the Magellanic penguin (Kikkawa et al. 2009). Relatively low diversity estimates in the Humboldt penguin could be because of the fact that 15 of the 20 penguins genotyped originated from a captive population in the Kasai Sea Life Park in Japan. Therefore, it would be important to assess MHC genetic diversity of the Humboldt penguin from a larger population sample in the wild, because this species has declined considerably over its geographic range and is now deemed threatened (http://www.iucnredlist.org).

In 2006, Baker et al. published a phylogeny of modern penguins using mitochondrial DNA genes. The species tree illustrated Galapagos and Humboldt penguins clustering together and independent of Magellanic penguins (Baker et al. 2006). The neighbor-joining phylogenetic analysis of DRβ1 sequences reported here revealed that the 45 Magellanic penguin DRβ1 sequences clustered into 9 allelic lineages (Figure 1). Previously reported sequences from the Humboldt penguin were distributed throughout the tree, indicating the presence of ancestral polymorphisms that predated the speciation events within the genus Spheniscus. This trend was also validated by the maximum parsimony phylogenetic analysis. Although species clustering of MHC β sequences among birds is known (Hess and Edwards 2002), transpecific clustering of alleles has also been reported in multiple birds, including the Darwin’s finch and several owl species (Edwards et al. 1999; Burri et al. 2008).

The preservation of ancestral polymorphisms in the MHC is commonly ascribed to balancing selection favoring variation for the immune response of vertebrates. The increased proportion of nonsynonymous substitutions at the DRβ1 gene ($d_N/d_S > 1$) suggests that positive selection has driven the evolution of sequence variation at this specific locus, providing support for balancing selection as well the potential for MHC-based mate choice. Higher nonsynonymous to synonymous substitution rates in the peptide binding region (Table 3), that is, the region specifically associated with the functional recognition of antigens, suggests that adaptive selection for immunological responses against pathogenic agents has played an important role in the evolution of MHC variation.

Although balancing selection provides the foundation for MHC-based mate choice, it does not necessarily lead to the evolution of disassortative mating; for example, specific alleles/combination of alleles rather than dissimilarity may be more adaptive under particular situations. In the Magellanic penguin, genotypic frequencies at the DRβ1 gene deviated significantly from Hardy Weinberg expectations with an excess of homozygote genotypes, which would apparently contradict expectations under balancing selection. However, we found a few “common” alleles with frequencies ≥8% suggesting that selection from dominant pathogenic agents could potentially be operating, increasing the occurrence of certain alleles at this locus. For example, spatial and temporal fluctuations in pathogenic pressure may shape which alleles are favored at different times (Loisel et al. 2007).

Evolutionary processes other than balancing selection can, therefore, contribute to changes in the composition of MHC alleles in a population.

We found no direct evidence for disassortative mating preferences based on the genetic analysis of breeding pairs. Levels of allele sharing between males and females within breeding pairs were not significantly different than those expected by chance. The high degree of variability at this locus, with a high number of alleles and high levels of sequence variation, may limit the potential relevance of the MHC in disassortative mating, particularly if levels of MHC diversity make breeding mates unlikely to harbor similar/identical alleles. This was the case in our sample, in which the majority of breeding pairs (43 of 50) did not share any allele. Furthermore, it is not clear what specific recognition system would be associated with MHC detection in this species. Magellanic penguins, although sexually dimorphic in size, are similar in plumage (Williams and Boersma 1995), decreasing the potential association of MHC genotype with phenotypic traits related to mate choice (see, e.g., Von Schantz et al. 1997; Eizaguirre et al. 2009). In addition, penguins in general do not have a highly developed olfactory system (Van Buskirk and Nevitt 2007; Wenzel 2006), which is one of the sensory systems most commonly associated with MHC recognition (Milinski 2006). Therefore, it is likely that other ecological factors, such as nest quality, are more relevant in determining the formation of breeding pairs in the Magellanic penguin, and ultimately, their reproductive success (Stokes and Boersma 1988). Furthermore, mate selection may also be affected by breeding density, which has been shown to affect offspring body condition and immunocompetence in Magellanic penguins, potentially due to increased competition for food and increased stress from parasites and predation (Tella et al. 2001).

Despite the lack of direct disassortative mating evidence associated with the genetic make-up of breeding pairs, we did find genotype-fitness trends that indicate selection is an important force driving MHC variation. We found significant differences in egg hatching between homozygote and heterozygote females ($p = .022$), with no homozygote females, which hatched eggs, fledging any chicks (Figure 2). These results suggest that heterozygote advantage may be operating at this gene. Interestingly, heterozygote males were not significantly more fit than male homozygotes for the number of eggs hatched and fledged chicks, which seem to contradict the potential role of heterosis in maintaining MHC diversity. The differential effects of MHC heterozygosity on fitness between the sexes is most likely associated with the relative role of hatching and fledging rates as reliable indicators of overall fitness in males and females. For example, hatching success may be a better estimator of female fitness given that females have greater parental investment in the production of eggs. Because parental care in Magellanic penguins is shared equally by both parents and chicks are exposed directly to environmental variability (e.g., associated with predation, food availability, pathogen load, etc.), fledging success may be a better indicator of offspring fitness rather than parental fitness. Given the differential investment of the sexes in
producing eggs as well as the potential role of environmental variation in affecting overall fitness, caution should be taken when assessing fitness components of males and females in relation to MHC genotype. Because we did not have access to offspring blood, we were not able to assess which alleles/combination of alleles may be responsible for the success of offspring.

In summary, strong evidence of positive selection on the MHC class II DRβ1 gene of the Magellanic penguin in addition to the phylogenetic spread of alleles retrieved from different species suggest that balancing selection, either through overdominance or frequency dependence, has maintained both the exceptional degree of MHC class II DRβ1 variability and allelic lineages that preceded speciation. Analyses of MHC diversity of breeding pairs indicate MHC-based dissortative mating preferences are not present in the Magellanic penguin. MHC genotype/fitness associations suggest, however, that selection plays an important role in maintaining levels of MHC diversity, particularly if specific MHC alleles and/or alleles combinations are shown to be directly associated with the health status of individual penguins in the wild. Finally, our study indicates that caution should be taken in deciding what fitness components to use when assessing the role of MHC variation on parental success because of the potential differential investment of sexes on offspring fitness.

Funding

This research was supported with funds from the Wadsworth Endowed Chair in Conservation Biology at the University of Washington and Bowling Green State University (BGSU). Field work was funded by the Wildlife Conservation Society and Friends of the Penguins through the Center for Penguins as Ocean Sentinels at the University of Washington.

Acknowledgments

The authors thank Prof. Scott V. Edwards, Miguel Alcaide, and Bianca Lec (Harvard University), Andres Aguilar (UC Merced), and 2 anonymous reviewers for helpful suggestions on prior versions of this manuscript. The authors also thank Dr. James Albert (BGSU) for advice on the permutation test. This research is part of Gabby Knafler’s M.S. thesis at BGSU. We would also like to thank the BGSU SetGo Program for summer support to G.K.

Supplementary Material

Supplementary material can be found at http://www.jhered.oxfordjournals.org/.

References

Aguilar A. 2006. Patterns of variation in MHC class II beta loci of the little greenhual (Andropadus viridus) with comments on MHC evolution in birds. J Hered. 97: 133–142.


Received September 7, 2011; Revised May 31, 2012; Accepted June 8, 2012

Corresponding Editor: Robert C. Fleischer

768