MHC DIVERSITY AND MATE CHOICE IN THE MAGELLANIC PENGUIN, *SPHENISCUS MAGELLANICUS*

Gabrielle J. Knafler

A Thesis

Submitted to the Graduate College of
Bowling Green State University in partial
fulfillment of the requirements for the degree of

Master of Science

December, 2011

Committee:

Dr. Juan L. Bouzat, Advisor

Dr. Gabriela Bidart-Bouzat

Dr. Ron C. Woodruff

Juan L. Bouzat, Advisor

ABSTRACT

Levels of major histocompatibility complex (MHC) diversity of the class II *DRß1* gene in fifty breeding pairs of the Magellanic penguin, *Spheniscus magellanicus*, were estimated and compared to previously published genetic diversity estimates from two other species in the genus *Spheniscus*. Furthermore, positive selection favoring sequence variation and two conditions required for the evolution of MHC-based disassortative mating preferences were tested: (1) evidence for greater MHC diversity among breeding pairs compared to a random mating, and (2) associations between MHC genotype and fitness. Cloning and sequencing of a 420 bp region of the MHC class II *DRß1* locus showed that Magellanic penguins had higher levels of genetic variation than values published for Galapagos (*S. mendiculus*) and Humboldt (*S. humboldti*) penguins. Sequence analysis revealed 45 alleles with 3.6% nucleotide differences and a nucleotide diversity of 0.030. At the population level, expected and observed heterozygosity (H_E=0.930, H_O=0.770) were high. A gene phylogeny of class II *DRß1* sequences showed nine distinct allelic lineages with interspersed sequences from Humboldt and Galapagos penguins, providing support for ancestral polymorphisms. Evidence for positive selection was revealed through d_N/d_S ratios significantly greater than one. Comparisons of breeding pairs to randomly generated pairs showed that disassortative mating preferences were not present. Males and females showed differential effects of MHC heterozygosity on fitness, likely associated with the relative role of hatching and fledging rates as indicators of overall fitness in both sexes. Significant MHC genotype/fitness associations suggest that selection for pathogen resistance plays a more important role than mate choice in maintaining diversity at the MHC in the Magellanic penguin.

ACKNOWLEDGEMENTS

I would like to thank Dr. Juan L. Bouzat for his excellent advising, patience, and motivation throughout my M.S. experience and growth. I would also like to thank my committee members, Dr. Gabriela Bidart- Bouzat and Dr. Ron C. Woodruff, for their valuable input and encouragement. In addition, I extend my gratitude to Dr. James Albert (BGSU) for advice on the permutation test, Matt Hoostal for help in the lab, Dr. Vipa Phuntumart and Dr. Paul Morris for help in navigating the NCBI website, J. Alan Clark for collecting blood samples used in this study, the Boersma Lab for both blood collection and fitness data, and the BGSU SetGo Program for summer support during my last year as an undergraduate student at BGSU.

Special thanks go to my mother, Angelina, for always reminding to take it one day at a time and my sister, Danielle, for her humor and instrumental advice through this process. With their support, I was able to accomplish my goals thus far and realize that a life in the sciences is attainable.

This research was funded by Bowling Green State University (BGSU) and the Wadsworth Endowed Chair in Conservation Biology at the University of Washington. Field work for the collection of samples 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.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

MHC DIVERSITY AND MATE CHOICE IN THE MAGELLANIC PENGUIN, *SPHENISCUS MAGELLANICUS*

INTRODUCTION

Highly polymorphic major histocompatibility complex (MHC) class I and 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).

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 two 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). On the other hand, frequency dependent selection of rare alleles is advantageous when novel pathogens are unrecognized by common alleles present in a population. These two 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 female may indirectly benefit by increasing the fitness of her 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, since 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 non-synonymous versus synonymous substitutions (e.g., 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. First, under disassortative mating, breeding pairs are predicted to share fewer alleles than those shared by randomly-paired individuals. Higher MHC genotypic diversity among 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 offspring of artificially bred salmon were more similar at their MHC loci and consistently carried

higher parasite loads than naturally spawning salmons (Consuegra and de Leaniz 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 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 benefiting 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 recent study that found an association between MHC diversity and juvenile survivorship in the Seychelles warbler (*Acrocephalus sechellensis*) (Brouwer 2010). Although several studies have shown experimental evidence consistent with the prediction of greater MHC diversity and increased fitness among mates, direct evidence from wild populations is still scarce.

I 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 *DRβ1* locus. Pathogen transmission has an important effect on the fitness of wild bird populations (Ewald 1994) and, in particular, has been shown to reduce the growth rate of Magellanic penguin chicks (Potti 2002). Since penguins live in dense colonies, which may facilitate pathogen transmission, MHC diversity may increase the fitness of breeding pairs. Magellanic penguins are monogamous, long-lived, have high fidelity (Boersma 2008), and a skewed sex ratio of 1.5 males to females (Boersma, unpub. data). Therefore, female choice may be evident due to 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 (Milinski 2006; Penn and Potts 1999).

Only a few studies have previously characterized MHC genes in penguins. Tsuda et al. (2001) were the first to report genetic polymorphisms and to compare phylogenetic relationships of the *DRβ*-like genes among five 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 *DRβ*–like loci, extrapolating such estimates of genetic diversity to the population level should be done with caution because of small sample sizes.

Based on the genotyping and sequencing of 100 Magellanic penguins, I report here population-level estimates of MHC genetic diversity at the class II *DRβ1* gene in the largest continental breeding colony of this species (Boersma 2008). I compared MHC diversity in the Magellanic penguin with previous studies on Humboldt and Galapagos penguins (Bollmer et al. 2007; Kikkawa et al. 2005, 2009), 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 *DRβ1* genotyping of 50 breeding pairs, I assessed whether MHC disassorative mating preferences are operating in the Magellanic penguin. Specifically, I tested two 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. To my knowledge, this is the first

study that reports population-level estimates of MHC diversity, examining the potential role of selection in maintaining MHC variation in a wild population of Magellanic penguins.

MATERIALS AND METHODS

Population Sampling and Fitness

Fifty penguin breeding pairs were randomly chosen from a major breeding colony of Magellanic penguins located at Punta Tombo, in the Atlantic shores of southern Patagonia, Argentina (44° 2' 60 S, 65 ° 10' 60 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 1980's (Boersma et al 1990). Fitness was defined as the number of eggs hatched and chicks fledged per individual. I used fitness data from 2002, the year when blood was collected, because I could assign the number of eggs hatched and chicks fledged for every chosen banded breeding pair.

MHC Genotyping

Blood samples (approximately 200ul) were collected by the Boersma Lab through puncture of the brachial vein and stored in Queen's lysis buffer (10 mM Tris, 10 mM NaCl, 10 mM EDTA, 1% n-lauroylsarcosine pH 7.5; Seutin et al. 1991) for further DNA analysis. DNA extractions were performed using standard Phenol-chloroform extraction protocols (Sambrook et al. 1989). PCR amplification of the MHC class II *DRB1* gene was performed using primers Lpen.hum1F2 (5'-ACTCCTGGCACAGCCGCGTG -3') and Lpen.hum2R (ACACGCTCTCCCCTCCTGTG) originally developed by Kikkawa et al. (2005 and 2009). Primers were designed to amplify exon 2 of the class II *DRß1* 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 five 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 (Kikkawa et al. 2009). Despite strong evidence for single locus amplification of the class II *DRβ1* gene, I tested the primers for potential multi-locus amplification (see below).

Amplification reactions were performed in 25μL volumes containing 1X Taq buffer, 1.5mM $MgCl_2$, 0.2mM of each dNTP, 0.5μM of each primer, and 0.1 units of Taq Polymerase. 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, California). 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 due to 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 due to the potential multi-allelic state of nuclear templates (Lenz and Becker 2008). Estimates of genetic diversity and comparison of MHC similarity among penguin pairs could therefore be confounded by multi-locus

amplification and the presence of DNA chimeras. To prevent multi-locus amplification, I chose primers reported to amplify a single MHC class II gene in the Magellanic penguin (Kikkawa 2009). Furthermore, I sequenced up to 12 clones per individual and found, in all cases, no more than two 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, I 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 for a new allele to be reported it must appear in, at least, two individuals or in the same individual from two 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 me 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). Based on the multiple sequence alignment, I 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 (h) were calculated using the program DnaSP, while observed and expected heterozygosities (H_O and H_E) as well as the total number of alleles in the sample were calculated using Arlequin. Estimates of genetic diversity were generated both before and after applying ISAG's conservative criteria for defining MHC alleles (see Appendix). To assess ancestral relationships among individual MHC alleles, I 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 *DRß1* alleles reported in this study for the Magellanic penguin (Genbank accession numbers pending) and previously reported sequences for Humboldt and Galapagos penguins (Bollmer et al. 2007; Kikkawa et al. 2005, 2009). The phylogenetic tree was rooted using *DRß1* sequences from the domestic chicken (*Gallus gallus*; Genbank M29763) and the common cactus finch (*Geospiza scandens*; Genbank Z74412); the confidence of groupings were estimated through 500 bootstrap replicates.

To assess balancing selection and MHC-based disassortative mating, I 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 non-synonymous to synonymous substitutions. I used MEGA, v.4, to measure the relative

rate of non-synonymous (d_N) and synonymous (d_S) 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, I estimated the variances of d_N and d_S using the bootstrap method implemented in MEGA (500 replications) and compared the relative abundance of synonymous and non-synonymous substitutions using a Z-test. d_N/d_S ratios were calculated for the entire exon 2 of the *DRβ1* gene as well as independently for the antigen binding regions and non-peptide binding regions within the exon (Brown et al. 1993).

To assess levels of MHC diversity of mating pairs *versus* random mating scenarios, I identified the alleles of the 50 selected breeding pairs and compared them to the alleles of these same randomly paired 50 males and 50 females. I tested the null hypothesis of random pairing between males and females by performing a permutation test in which I compared the test statistic, i.e., the number of shared alleles among the 50 true breeding couples, to a sampling distribution generated by randomly permuting the females and computing the number of shared alleles among randomly paired couples (1000 permutations).

To evaluate the potential association of MHC genotype and fitness, I performed an ANOVA to assess differences in fitness (i.e., number of eggs hatched and number of chicks fledged) between homozygote and heterozygote individuals. I performed the analysis considering all individuals together as well as independently for males and females. In addition, I compared the fitness of individuals containing "common" alleles (with frequencies $\geq$8%) with genotypes composed of rare alleles (with frequencies $\leq$4%). I performed a non-parametric 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, I 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. I ran this analysis separately for males and females, since parental investment may differ by sex. I 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 DRβ1 Gene Sequences

Results reported here were obtained from MHC class II *DRβ1* alleles that were confirmed using ISAG's conservative criteria for identifying alleles (see Appendix for genetic diversity estimates including unconfirmed alleles). 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. 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 I did not detect in any case more than two distinct sequences/alleles. The initial count of homozygote individuals (47) suggested, however, that PCR bias or preferential cloning of individual alleles could be underestimating the number of heterozygotes if only a limited number of clones is used to genotype individuals. I therefore performed two 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 (Table1). Using ISAG's criteria for defining MHC

alleles, 45 alleles were detected in our sample of 100 individuals. I found 54 polymorphic sites within the 420 bp region of exon 2. I 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 three alleles were present in relatively high frequency (0.08, 0.15, and 0.17). I estimated expected heterozygosity (H_E) and haplotype diversity (*h*) at 0.930. Genotype frequencies were significantly different from those 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 proportion of differences among sequences of Magellanic and Humboldt penguins were similar but both were higher than the estimated for the Galapagos penguin. While this study found 45 MHC alleles for Magellanic penguins, previous studies found eight alleles for Humboldt penguins (n=20) (Kikkawa et al. 2009) and only three alleles for Galapagos penguins (n=30) (Bollmer et al. 2007). The number of alleles in the Humboldt penguin decreased to seven when sequences were trimmed to the same length of the newly reported sequences of Magellanic penguins. An earlier study of five Magellanic penguins (Kikkawa et al. 2009) reported eight *DRβ1* alleles. Seven of the eight alleles were detected in our sample.

Results from the Neighbor-joining phylogenetic analysis of 60 MHC *DRβ1* 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 *DRβ1* 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. Based on 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 two 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 containing 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 five Humboldt penguin sequences appeared to be related to allele lineage H (Sphu0011, Sphu001, Sphu002) or basal to lineages A-D and E-F (Sphu006 and Sphu005, respectively; Figure 1), though 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 methods upheld (with ≥41% confidence) Neighbor-joining allelic lineages B, H, I, C, D, and F (Figure 2). Allelic lineage I is the only supported lineage containing an allele from two different species of penguins. Humboldt allele Sphu004 was expectedly grouped with identical Magellanic penguin alleles Smag02 and Smag13. However, three Humboldt alleles (Sphu003, Sphu007, and Sphu006) were interspersed with Magellanic penguin sequences consistent with the Neighbor-joining

phylogenetic analysis, which showed that clustering of alleles with respect to species was not supported.

Balancing Selection and Disassortative Mating

Non-synonymous to synonymous substitutions uncovered evidence for positive selection at the MHC class II *DRβ1* locus. The d_N/d_S ratio for the *DRβ1* exon 2 was significantly greater than one, 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; see Figure 3 for PBR sites), 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 non-significant (p=0.125; 1000 permutations). The p-value revealed that 12.5% of the random permutations showed a less or equal number of shared alleles than the observed value of 7 shared alleles within all breeding pairs.

Statistical analyses of MHC genotype/fitness associations showed different trends for males and females. First, the differential fitness among homozygote and heterozygote females was significant (p=0.024), with heterozygotes hatching more eggs than homozygotes (Figure 4). Female heterozygotes also showed a non-significant trend to fledge more chicks than homozygotes (p= 0.065); in fact, there were no female homozygotes that fledged any chicks. Conversely, male homozygotes showed non-significant effects of MHC genotype on fitness either for hatching success (p=0.103) or fledging (p=0.189; Figure 4). Regression analyses between MHC sequence divergences among alleles in individual genotypes and fitness revealed that MHC sequence variation for both males and females could not explain fitness differences in

both eggs hatched and chicks fledged ($p \geq 0.076$) (see Figures 5 and 6). Finally, I could not detect significant differences in fitness, either in eggs hatched or chicks fledged, between individuals who contained at least one of the three most frequent alleles compared to those whose genotypes contained rare alleles (i.e., with frequencies ≤ 0.04).

DISCUSSION

In this study, I 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 *DRβ1* gene from 100 Magellanic penguins revealed high levels of genetic diversity at both the nucleotide and population levels. I 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 (Bollmer et al. 2007; Kikkawa et al. 2009). 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 absence of MHC variation, however, may be the result of limited sampling in the number of individuals and sequence length analyzed, as high MHC diversity can exist regardless of extreme

population bottlenecks (Aguilar et al. 2004). Like the Galapagos penguin, the Humboldt penguin also had reduced MHC variation in comparison to the Magellanic penguin (Kikkawa et al. 2009). Low diversity estimates in the Humboldt penguin could be due to the fact that 15 of the 20 penguins genotyped originated from a captive population in the Kasai Sea Life Park in Japan. It would be important to assess MHC genetic diversity of the Humboldt penguin from a larger population sample in the wild, since this species has declined considerably over its geographic range and is now deemed threatened (http://www.iucnredlist.org).

In 2006, Baker and colleagues 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 nine 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 (Figure 2). Although species clustering of MHC *DRβ1* 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 non-synonymous 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 non-

synonymous to synonymous substitutions in the peptide binding region (Table 3), i.e., the region specifically associated with the functional recognition of antigens (Figure 3), suggest 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; e.g., specific alleles/combination of alleles rather than dissimilarity may be more adaptive under particular situations. In the studied population, genotypic frequencies at the *DRβ1* gene deviated significantly from Hardy Weinberg expectations with an excess of homozygote genotypes. I 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).

I 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 out 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., Eizaguerre et al. 2009; Von Schantz et al. 1997). 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). Other ecological factors, such as nest quality are likely more relevant in determining the formation of breeding pairs in Magellanic penguins and, ultimately, their reproductive success (Stokes and Boersma 1988).

Despite the lack of direct disassortative mating evidence associated with the genetic make-up of breeding pairs, I did find some genotype-fitness trends that indicate selection is an important force driving MHC variation. I found significant differences in egg hatching between homozygote and heterozygote females ($p=0.024$), with no homozygote females that hatched eggs, fledging any chicks (Figure 4). These results suggest that heterozygote advantage may be operating at this gene. Interestingly, heterozygote males were not significantly more fit than male homozygotes either for the number of eggs hatched ($p=0.103$) or chicks fledged ($p=0.189$), which seems 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 indicators of overall fitness in males and females. For example, hatching success may be a better estimator of female fitness, since females have greater parental investment in the production of eggs. On the other hand, fledging success may be a better indicator of offspring fitness because chicks are exposed directly to environmental variability, e.g., associated with food availability, predation, pathogen load, etc. Therefore, caution should be taken when assessing fitness components of males and females in relation to MHC genotype, given the differential investment of the sexes in the production of eggs and the rearing of chicks as well as the potential role of environmental variation affecting

overall fitness. Furthermore, the beneficial effects of maternal and paternal alleles may translate into increased immunocompetence of the offspring against pathogenic agents. Since 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 male homozygotes.

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 disassortative 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 combinations of alleles 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.

TABLES AND FIGURES

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

	V	A	H_O	H_E	π
Male	37	28	0.760	0.946	0.031
Female	25	25	0.780	0.914	0.029
Total	54	45	0.770	0.930	0.030

Table 2. Levels of MHC class II *DRß1* sequence diversity in the Magellanic penguin and two other penguin species in the genus *Spheniscus*. Shown is the number of individuals genotyped (n), the number of alleles detected (A), number of variable sites (V), average number of nucleotide differences (N-Diff), average proportion of differences (p-distance), nucleotide diversity (π), and haplotype diversity (*h*). Diversity estimates for Magellanic and Humboldt penguins are based on the 420 bp region of the *DRß1* exon 2 amplified in this study whereas estimates for the Galapagos penguin are based on Bollmer's et al. (2007) analysis of a 157 bp portion of exon 2. SE represents standard errors.

Species	n	A	V	N-Diff (±SE)	p-distance	π	*h*
Magellanic penguin	100	45	54	14.74±2.1	0.036±0.005	0.030	0.930
Humboldt penguin [a]	20	7	29	12.8±2.3	0.031±0.006	0.022	0.671
Galápagos penguin [b]	30	3	3	2.0±1.3	0.013±0.007	-	-

[a] data from Kikkawa et al. (2009).

[b] data from Bollmer et al. (2007); π and *h* could not be estimated because allele frequencies were not reported.

Table 3. Proportion of non-synonymous (d_N) and synonymous (d_S) substitutions for the complete Exon 2 sequence of the Magellanic penguin *DRβ1* gene, the peptide binding region (PBR), and the non-peptide binding region (non-PBR). d_N/d_S ratios were all significant at $p<0.05$. SE represents standard errors.

	d_N ±SE	d_S ±SE	d_N/d_S
Exon 2	0.0490±0.011	0.007±0.003	7.006
PBR	0.168 0±0.059	0.014 ± 0.009	11.20
Non-PBR	0.043±0.007	0.015±0.010	2.86

Figure 1. See page 24. Neighbor-joining gene tree of MHC class II DRB1 exon 2 sequences from three species of penguin in the *Spheniscus* genus. The domestic chicken (*Gallus gallus*) and the common cactus-finch (*Geospiza scandens*) class II β sequences 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).

Figure 1 (continuation)

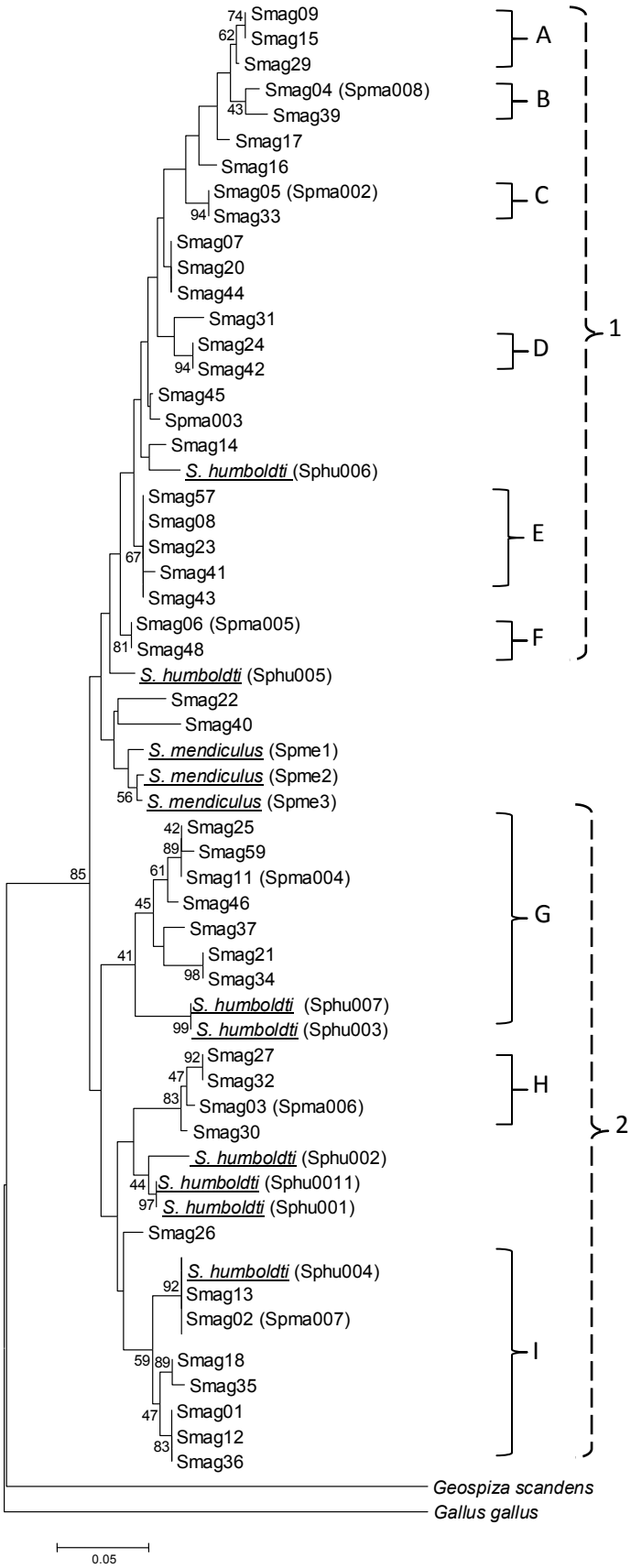
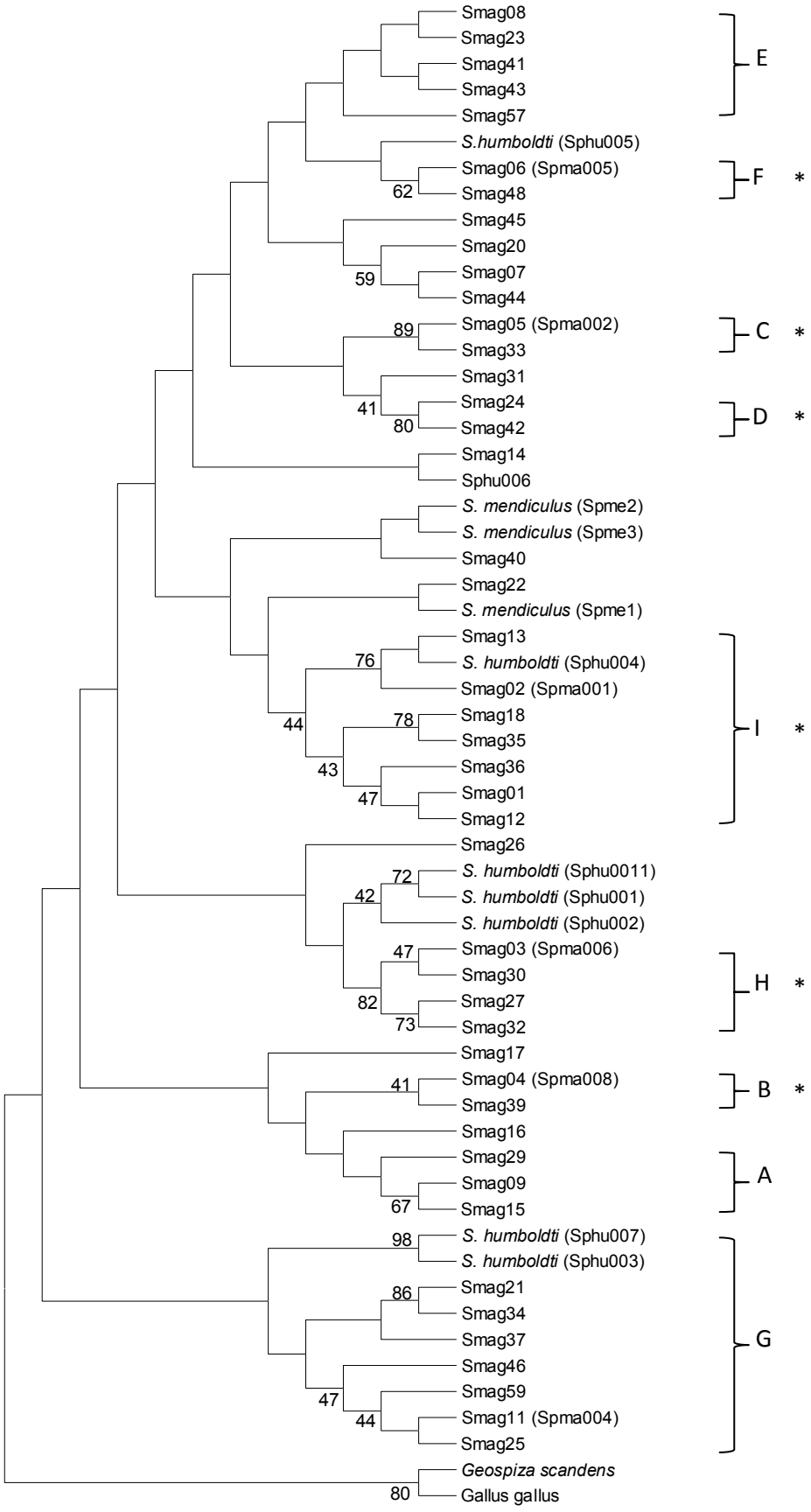


Figure 2. See page 26. Maximum parsimony gene tree of MHC class II DRB1 exon 2 sequences from three species of penguin in the *Spheniscus* genus. The domestic chicken (*Gallus gallus*) and the common cactus-finch (*Geospiza scandens*) class II β sequences were used as outgroups. Numbers at nodes are bootstrap values based on 500 replicates. Brackets A through I represent allelic lineages defined by Neighbor-joining methods. Asterisks represent those allelic lineage defined by Neighbor-joining that retained confidence of 41% or greater when analyzed by maximum parsimony. Refer to Figure 1 for abbreviations regarding sequence designations.

Figure 2 (continuation)



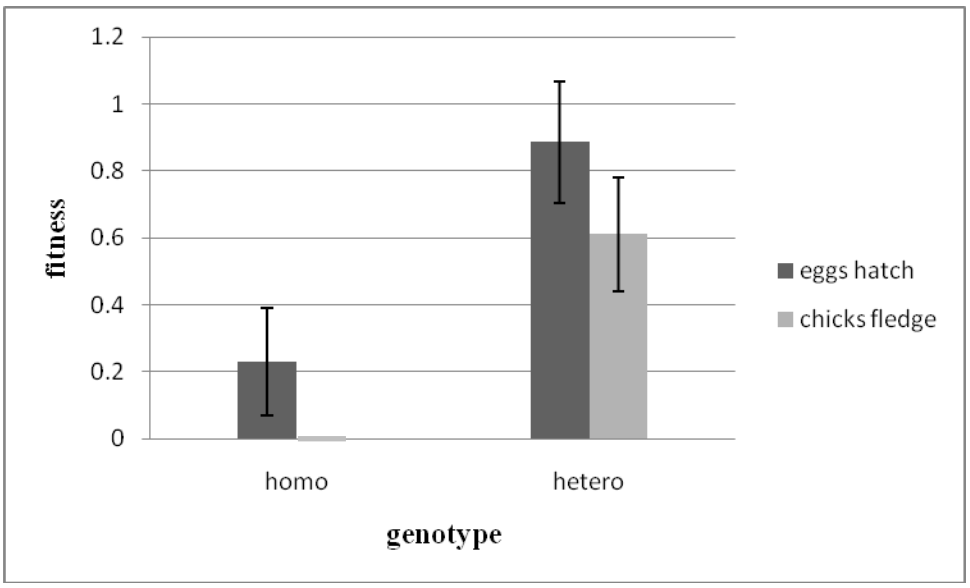
```
                            * * *    **       *       *   **   * * ** *
Smag01 YFQEMLKAECHFLNGTERVRFVVRDIYNRQQDVHFDSDVGYYVADTPLGEPDAKYWNSQTDFLEQSRAAV
Smag02 YFQEMLKAECHFLNGTERVRFVVRDIYNRQQDVHFDSDVGQFVADTPLGEPDAKYWNSQTDFLEQSRAAV
Smag03 YFQEMLKAECHFLNGTERVRYVVRDIYNRQQNVHFDSDVGQFVADTPLGELIAKYWNSQTDLLEQRRAEV
Smag07 YFQEMGKFECHFLNGTERVRFVDRYIYNRQQYVHFDSDVGYYVADTPLGEPSAKYLNSQTDLLEQRRAEV
Smag08 YFQEMGKAECHFLNGTERVRFVERYLYNRQQYVHFDSDVGYYVADTPLGEPSAKYWNSQTDLLEQKRAEV
Smag04 YFQEMGKAECHFLNGTERVRFVDRYIYNRQEYVHFDSDVGQFVADTPLGELIAKYLNSQTDLLEQRRAEV
Smag09 YFQEMGKFECHFLNGTERVRFVDRYIYNRQQYVHFDSDVGQFVADTPLGELIAKYLNSQTDLLEQTRAAV
Smag06 YFQEMGKSECHFLNGTERVRFVERYLYNRQQYVHFDSDVGYYVADTPLGEPDAKYWNSQTDLLEQKRAEV
Smag12 YFQEMLKAECHFLNGTERVRFVVRDIYNRQQDVHFDSDVGYYVADTPLGEPDAKYWNSQTDFLEQSRAAV
Smag13 YFQEMLKAECHFLNGTERVRFVVRDIYNRQQDVHFDSDVGQFVADTPLGEPDAKYWNSQTDFLEQSRAAV
Smag14 YFQEMGKSECHFLNGTERVRFVERYLYNRQQYVHFDSDVGYYVADTPLGEPSAKYWNSQTDLLEQSRAAV
Smag15 YFQEMGKFECHFLNGTERVRFVDRYIYNRQQYVHFDSDVGQFVADTPLGELIAKYLNSQTDLLEQTRAAV
Smag16 YFQEMGKFECHFLNGTERVRFVDRYIYNRQQYVHFDSDVGQFVADTPLGEPSAKYLNSQTDLLEQRRAAV
Smag17 YFQEMHKAECHFLNGTERVRLVERYIYNRQQYVHFDSDVGQFVADTPLGELIAKYWNSQTDLLEQRRAAV
Smag18 YFQEMVKAECHFLNGTERVRFVVRDIYNRQQDVHFDSDVGYYVADTPLGEPDAKYWNSQTDILEQSRAAV
Smag20 YFQEMGKFECHFLNGTERVRFVDRYIYNRQQYVHFDSDVGYYVADTPLGEPSAKYLNSQTDLLEQRRAEV
Smag21 YFQEMLKAECHFLNGTERVRFVDRYIYNRQEYVHFDSDVGQFVADTPLGELIAKYWNSQTDILEDERAEV
Smag22 YFQEMLKAECHFLNGTERVRFVVRDIYNRQQDVHFDSDVGYYVADTPLGVPDAKYWNSQTDFLEQKRAAV
Smag23 YFQEMGKSECHFLNGTERVRFVERYLYNRQQYVHFDSDVGYYVADTPLGEPSAKYWNSQTDLLEQKRAEV
Smag11 YFQEMGKAECHFLNGTERVRFVDRYIYNRQEYVHFDSDVGQFVADTPLGEPSAKYWNSQTDILEDERAAV
Smag05 YFQEMGKFECHFLNGTERVRFVDRYIYNRQQYVHFDSDVGYYVADTPLGELIAKYLNSQTDLLEQRRAAV
Smag24 YFQEMGKFECHFLNGTERVRFVDRYIYNRQQYVHFDSDVGYYVADTPLGEPSAKYLNSQTDFLEQKRAAV
Smag25 YFQEMGKAECHFLNGTERVRFVDRYIYNRQEYVHFDSDVGQFVADTPLGEPSAKYWNSQTDILEDERAAV
Smag26 YFQRMLKFECHFLNGTERVRLVERDIYNRQQDVHFDSDVGQFVADTPLGEPDAKYWNSQTDFLEQRRAAV
Smag27 YFQEMLKFECHFLNGTERVRYVVRDIYNRQQNVHFDSDVGQFVADTPLGELIAKYWNSQTDFLEQKRAEV
Smag29 YFQEMGKFECHFLNGTERVRFVDRYIYNRQQYVHFDSDVGQFVADTPLGELIAKYLNSQTDLLEQRRAAV
Smag30 YFQEMLKAECHFLNGTERVRYVVRDIYNRQQYVHFDSDVGQFVADTPLGELIAKYWNSQTDLLEQRRAEV
Smag32 YFQEMLKFECHFLNGAERVRYVVRDIYNRQQNVHFDSDVGQFVADTPLGELIAKYWNSQTDFLEQKRAEV
Smag33 YFQEMGKFECHFLNGTERVRFVDRYIYNRQQYVHFDSDVGYYVADTPLGELIAKYLNSQTDLLEQRRAAV
Smag34 YFQEMLKAECHFLNGTERVRFVDRYIYNRQEYVHFDSDVGQFVADTPLGELIAKYWNSQTDILEDERAEV
Smag35 YFQEMVKAECHFLNGTERVRFVVRDIYNRQQDVHFDSDVGYYVADXPLGEPDAKYWNSQTDILEQSRAAV
Smag36 YFQEMLKAECHFLNGTERVRFVVRDIYNRQQDVHFDSDVGYYVADTPLGEPDAKYWNSQTDFLEQSRAAV
Smag37 YFQEMGKAECHFLNGTERVRFVERYIYNRQEYVHFDSDVGQFVADTPLGELIAKYWNSQTDFLEDERAAV
Smag39 YFQEMHKAECHFLNGTERVRFVDRYIYNRQEYVHFDSDVGQFVADTPLGELIAKYLNSQTDILEQKRAAV
Smag40 YFQEMHKSKCHFLNGTERVRYVERYIYNRQQDVHFDSDVGYYVADTPLGEPDAKYWNSQTDILERKQAAV
Smag41 YFQEMGKSECHFLNGTERVRFVERYLYNRQQYVHFDSDVGYYVADTPLGEPSAKYWNSQTDLLEQKRAEV
Smag42 YFQEMGKFECHFLNGTERVRFVDRYIYNRQQYVHFDSDVGYYVADTPLGEPSAKYLNSQTDFLEQKRAAV
Smag43 YFQEMGKAECHFLNGTERVRFVERYLYNRQQYVHFDSDVGYYVADTPLGEPSAKYWNSQTDLLEQKRAEV
Smag44 YFQEMGKFECHFLNGTERVRFVDRYIYNRQQYVHFDSDVGYYVADTPLGEPSAKYLNSQTDLLEQRRAEV
Smag45 YFQEMGKFECHFLNGTERVRFVDRYIYNRQQYVHFDSDVGYYVADTPLGEPSAKYWNSQTDLLEQKRAEV
Smag46 YFQEMGKAECHFLNGTERVRFVDRYIYNRQEYVHFDSDVGQFVADTPLGEPSAKYWNSQTDFLEDERAAV
Smag48 YFQEMLKAECHFLNGTERVRFVERYLYNRQQYVHFDSDVGYYVADTPLGEPDAKYWNSQTDLLEQKRAEV
Smag59 YFQEMGKAECHFLNGTERVRFVDRYIYNRQEYAHFDSDVGQFVADTPLGEPSAKYWNSQTDILEDERAAV
Smag57 YFQEMGKAECHFLNGTERVRFVERYLYNRQQYVHFDSDVGYYVADTPLGEPSAKYWNSQTDLLEQKRAEV
Smag31 YFQEMGKAECHFLNGTERVRYVERYIYNRQQYVHFDSDVGYYVADTPLGEPSAKYLNSQTDLLERKRAAV
```

Figure 3. MHC class II *DRβ1* exon 2 amino acid sequences of the 45 Magellanic penguin alleles reported in this study. Asterisks represent possible peptide binding regions from Brown et al. (1993).

```
          *
Smag01 DTICRHNYGVFTPFTVERRGECVAEHLPGGRAQAKPRG
Smag02 DTICRHNYGVVTPFTVERRGECVAEHLPGGRAQAKPRG
Smag03 DTVCRHNYGVVTPFTVERRGECVAEHLPGGRAQAKPRG
Smag07 DTYCRHNYGVVTPFTVERRGECVAEHLPGGRAQAKPRG
Smag08 DTYCRHNYGVVTPFTVERRGECVAEHLPGGRAQAKPRG
Smag04 DRYCRHNYGVVTPFTVERRGECVAEHLPGGRAQAKPRG
Smag09 DTYCRHNYGVVTPFTVERRGECVAEHLPGGRAQAKPRG
Smag06 DTYCRHNYGVVTPFTVERRGECVAEHLPGGRAQAKPRG
Smag12 DTICRHNYRVFTPFTVERRGECVAEHLPGGRAQAKPRG
Smag13 DTICRHNYGVVTPFTVERRGECVAEHLPGGRAQAKPRG
Smag14 DTYCRHNYGVVTPFTVERRGECVAEHLPGGRAQAKPRG
Smag15 DTYCRHNYGVVTPFTVERRGECVAEHLPGGRAQAKPRG
Smag16 DTYCRHNYGVGTPFTVERRGECVAEHLPGGRAQAKPRG
Smag17 DTYCRHNYGVVTPFTVERRGECVAEHLPGGRAQAKPRG
Smag18 DTICRHNYGVFTPFTVERRGECVAEHLPGGRAQAKPRG
Smag20 DTYCRHNYGVVTPFTVERRGECVAEHLPGGRAQAKPRG
Smag21 DRYCRHNYGVFTPFTVERRGECVAEHLPGGRAQAKPRG
Smag22 DRYCRHNYGVVTPFTVERRGECVAEHLPGGRAQAKPRG
Smag23 DTYCRHNYGVVTPFTVERRGECVAEHLPGGRAQAKPRG
Smag11 DTYCRHNYGVFTPFTVERRGECVAEHLPGGRAQAKPRG
Smag05 DRYCRHNYGVGTPFTVERRGECVAEHLPGGRAQAKPRG
Smag24 DTYCRHNYGVGTPFTVERRGECVAEHLPGGRAQAKPRG
Smag25 DTYCRHNYGVFTPFTVERRGECVAEHLPGGRTQAKPRG
Smag26 DTYCRHNYGVGTPFTVERRGECVAEHLPGGRAQAKPRG
Smag27 DTVCRHNYGVVTPFTVERRGECVAEHLPGGRAQAKPRG
Smag29 DTYCRHNYGVVTPFTVERRGECVAEHLPGGRAQAKPRG
Smag30 DTVCRHNYGVVTPFTVERRGECVAEHLPGGRAQAKPRG
Smag32 DTVCRHNYGVVTPFTVERRGECVAEHLPGGRAQAKPRG
Smag33 DTYCRHNYGVGTPFTVERRGECVAEHLPGGRAQAKPRG
Smag34 DRYCRHNYGVFTPFTVERRGECVAEHLPGGRAQAKPRG
Smag35 DTICRHNYGVFTPFTVERRGECVAEHLPGGRAQAKPRG
Smag36 DTICRHNYGVFTPFTVERRGECVAEHLPGGRAQAKPRG
Smag37 DTYCRHNYGVVTPFTVERRGECVAEHLPGGRAQAKPRG
Smag39 DRYCRHNYGVVTPFTVERRGECVAEHLPGGRAQAKPRG
Smag40 DRYCRHNYGVFTPFTVERRGECVAEHLPGGRAQAKPRG
Smag41 DTYCRHNYGVVTPFTVERRGECVAEHLPGGRAQAKPRG
Smag42 DRYCRHNYRVGTPFTVERRGECVAEHLPGGRAQAKPRG
Smag43 DTYCRHNYGVVTPFTVERRGECVAEHLPGGRAQAKPRG
Smag44 DTYCRHNYGVVTPFTVERRGECVAEHLPGGRAQAKPRG
Smag45 DTYCRHNYGVVTPFTVERRGECVAEHLSGGRAQAKPRG
Smag46 DTYCRHNYGVFTPFTVERRGECVAEHLPGGRAQAKPRG
Smag48 DTYCRHNYGVVTPFTVERRGECVAEHLPGGRAQAKPRG
Smag59 DTYCRHNYGVFTPFTVERRGECVAEHLPGGRAQAKPRG
Smag57 DTYCRHNYGVVTPFTVERRGECVAEHLPGGRAQAKPRG
Smag31 DRYCRHNYGVVTPFTVERRGECVAEHLPGGRAQAKPRG
```

Figure 3 (continuation). MHC class II *DRβ1* exon 2 amino acid sequences of the 45 Magellanic penguin alleles reported in this study. Asterisks represent possible peptide binding regions from Brown et al. (1993).

(a) Female



(b) Male

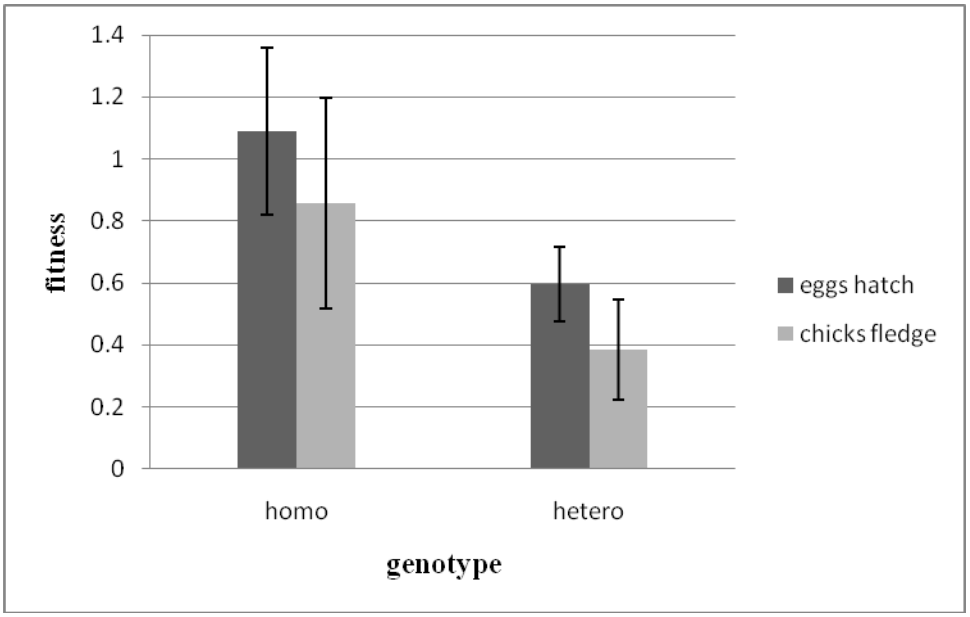
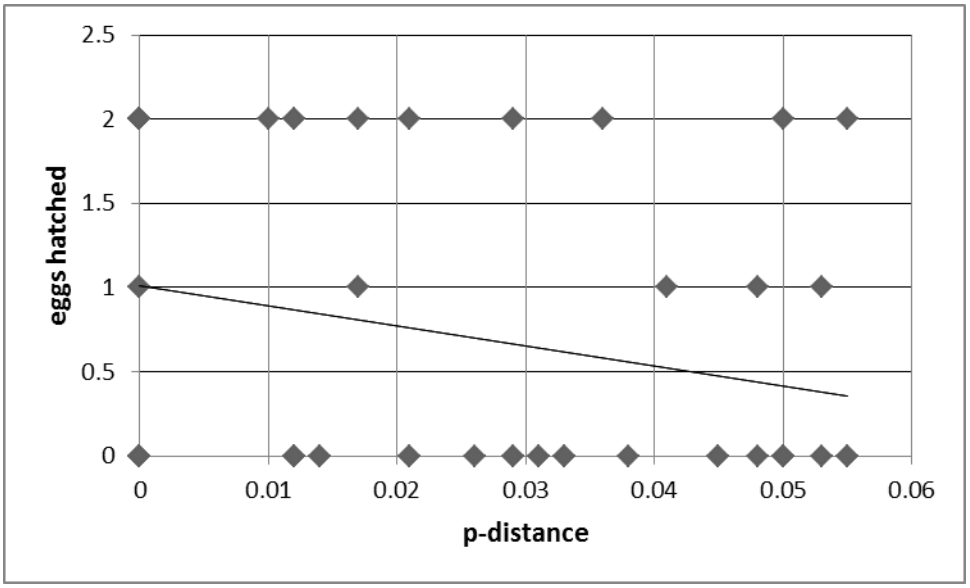


Figure 4. Mean number of eggs hatched (dark grey) and chicks fledged (light grey) for females (a) and males (b) in relation to MHC class II *DRβ1* genotype. Female heterozygotes significantly hatched more eggs than homozygotes; p=0.024. Note that there were no female homozygotes (considering those that hatched eggs) that fledged chicks. Male MHC genotype/fitness associations were nonsignificant for eggs hatched and chicks fledged; ≥0.103 .

(a) Male p-distance on eggs hatched



(b) Male p-distance on chicks fledged

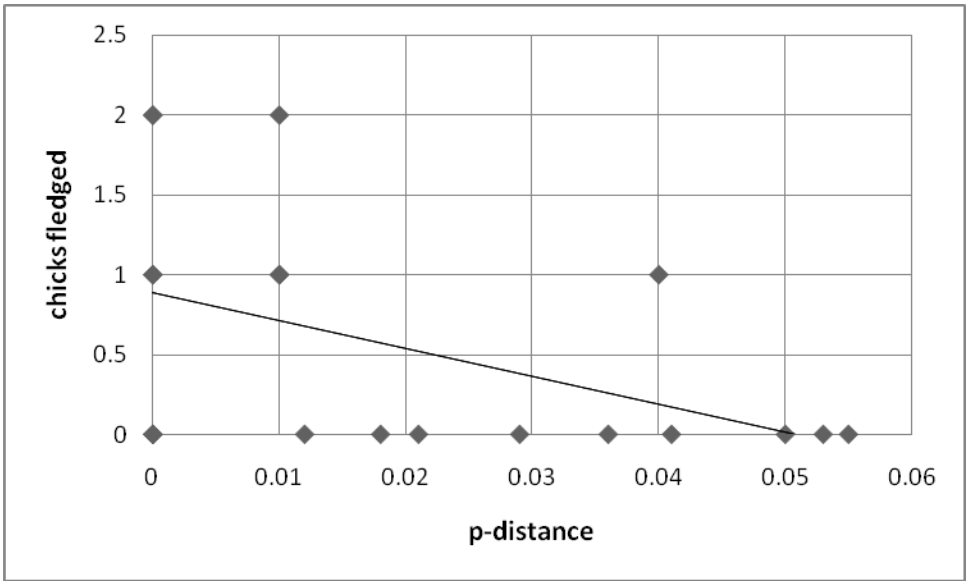
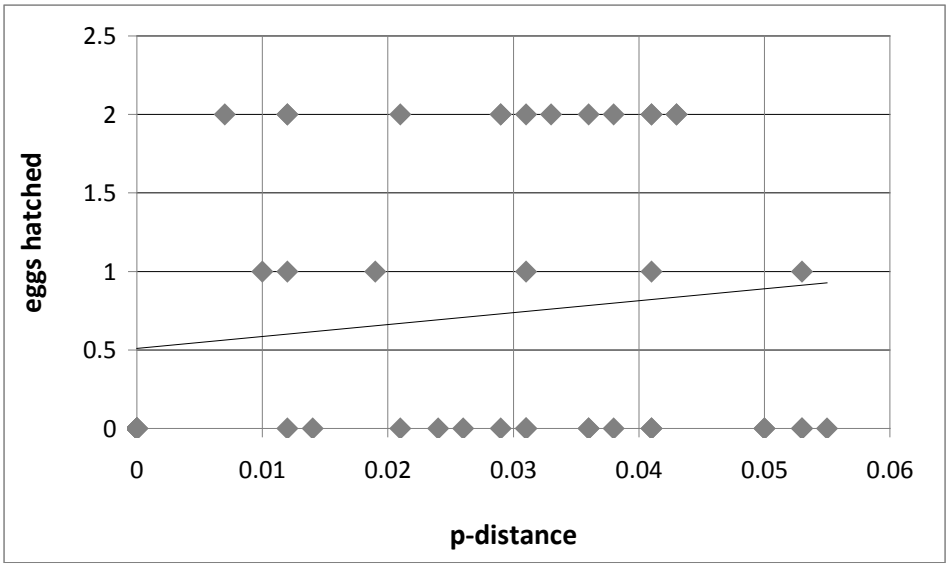


Figure 5. Regression analysis of male sequence variation (p-distance) on number of eggs hatched (a) and chicks fledged (b). Male sequence divergences among alleles in individual genotypes could not explain fitness differences in either eggs hatched or chicks fledged ($p \geq 0.064$).

(a) Female p-distance on eggs hatched



(b) Female p-distance on chicks fledged

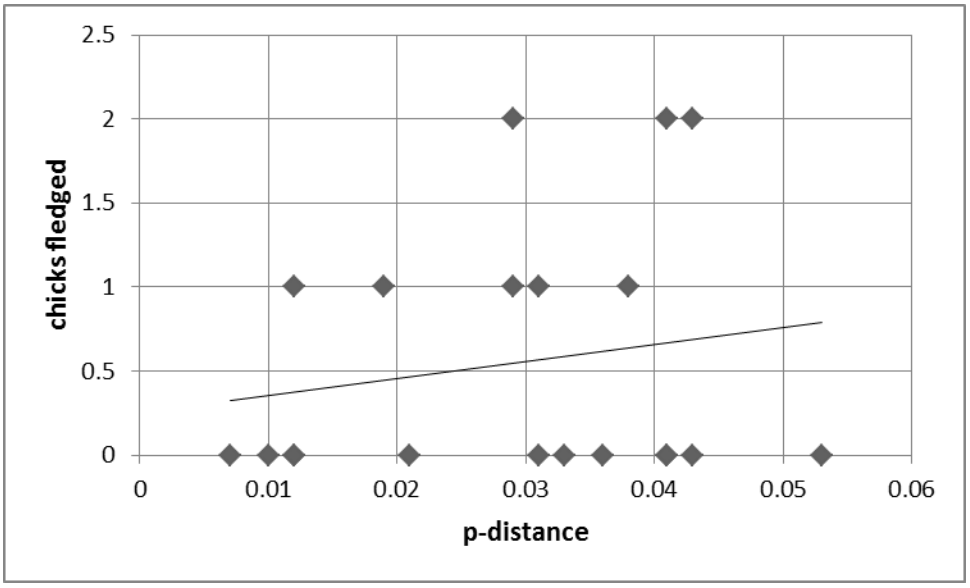


Figure 6. Regression analysis of female sequence variation (p-distance) on number of eggs hatched (a) and chicks fledged (b). Female sequence divergences among alleles in individual genotypes could not explain fitness differences in either eggs hatched or chicks fledged ($p \geq 0.076$).

APPENDIX

Genetic Diversity Estimates of the MHC class II DRβ1 gene before and after applying ISAG's criteria for defining MHC alleles

Introduction

The Characterization of MHC alleles is often problematic due to the potential amplification of multiple MHC loci and the formation of chimeric sequences during PCR (see ***MHC Genotyping*** section; pg. 5). Estimates of genetic diversity and d_N/d_S ratios commonly used to provide support for positive selection could ultimately be confounded by the presence of DNA chimeras; therefore, ISAG's conservative criteria for defining new MHC alleles was used for all analyses in this thesis. However, since alleles obtained before applying ISAG's criteria may include true alleles rather than chimeric sequences, it is important to know how using the criteria may affect levels of genetic variation and d_N/d_S ratios at the MHC class II *DRβ1* locus. In this Appendix, I provide genetic diversity estimates and rates of non-synonymous to synonymous substitutions obtained using all potential MHC *DRβ1* alleles detected in our sample of 100 Magellanic penguins.

Methods

Genetic diversity estimates of all potential alleles detected during cloning and sequencing techniques were analyzed prior to applying ISAG's conservative criteria for identifying MHC alleles. DNA sequences from these "unconfirmed" alleles were assembled and aligned using BioEdit, version 7.0.5.3 (Hall 1999). Based on the multiple sequence alignment, I 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 (h) were calculated using the program DnaSP, while observed and expected heterozygosities (H_O and H_E) as well as the total number of alleles in the

sample were calculated using Arlequin. As previously described for the confirmed alleles, I used MEGA, v.4, to measure the relative rate of non-synonymous (d_N) and synonymous (d_S) substitutions. To test whether positive selection was operating at this locus, I estimated the variances of d_N and d_S using the bootstrap method implemented in MEGA (500 replications) and compared the relative abundance of synonymous and non-synonymous substitutions using a Z-test. As described in the main body of the thesis, d_N/d_S ratios were calculated for the entire exon 2 of the *DRß1* gene as well as independently for the antigen binding regions and non-peptide binding regions within the exon.

Results

Results from DNA sequence analyses including unconfirmed alleles were compared to the genetic diversity estimates obtained using ISAG's conservative criteria (Table A1). Expected heterozygosity (H_E= 0.939), average sequence divergence (p-dist= 0.037), and nucleotide (π= 0.032) and haplotype (h= 0.940) diversity estimates of all potential alleles were similar to those estimates obtained using ISAG's criteria for defining MHC alleles. Expectedly, the number of detected alleles prior to applying these criteria (59) was greater than the number of alleles after applying the criteria (45). Accordingly, observed heterozygosity was also greater (H_O=0.900) and did not significantly deviate from heterozygosity under Hardy-Weinberg equilibrium. The majority of the 59 alleles detected (34) appeared only once in the sample whereas three alleles were present in relatively high frequency (0.09, 0.13, and 0.17). Table A2 compares non-synonomous to synonymous substitution rates for confirmed and unconfirmed alleles. Proportions of non-synonymous to synonymous substitutions were not significantly different than those reported using ISAG's criteria for defining MHC alleles, in both cases revealing evidence for positive selection at this locus.

Table A1. Genetic diversity estimates of 420 bp of the MHC class II *DRβ1* gene obtained before (Unconfirmed) and after (Confirmed) applying ISAG's criteria for defining MHC alleles. V= number of variable sites; A=number of alleles; H_O and H_E = observed and expected heterozygosities; π =nucleotide diversity; p-dist=average proportion of differences; and *h*= haplotype diversity.

	V	A	H_O	H_E	π	p-dist	*h*
Unconfirmed	71	59	0.900	0.939	0.032	0.037	0.940
Confirmed	54	45	0.770	0.930	0.030	0.036	0.930

Table A2. Proportion of non-synonymous (d_N) to synonymous (d_S) substitutions for the complete Exon 2 sequence of the Magellanic penguin *DRβ1* gene, the peptide binding region (PBR), and the non-peptide binding region (non-PBR) obtained before (Unconfirmed) and after (Confirmed) applying ISAG's criteria for defining MHC alleles. d_N/d_S ratios were all significant at $p<0.05$.

	Exon 2	PBR	Non-PBR
Unconfirmed	6.25	11.26	2.80
Confirmed	7.01	11.20	2.86

REFERENCES

Aguilar, A. 2006 Patterns of variation in MHC class II beta loci of the little greenbul (*Andropadus virens*) with comments on MHC evolution in birds. *J. Heredity* **97**, 133-42.

Aguilar, A., Roemer, G., Debenham, S., Binns, M., Garcelon, D., & Wayne, R.K. 2004 High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. *Proc. Natl. Acad. Sci. USA* **101**, 3490-3494.

Akst, E.P., Boersma, P.D., & Fleischer, R.C. 2002 A comparison of genetic diversity between the Galapagos penguin and the Magellanic penguin. *Conserv. Genet.* **3**, 375-383.

Babik, W., Durka, W., & Radwan, J. 2005 Sequence diversity of the MHC DRB gene in the Eurasian beaver (*Castor fiber*). *Mol. Ecol.* **14**, 4249–4257.

Baker, A.J., Pereira, S.L., Haddrath, O.P., & Edge, K.A. 2006 Multiple gene evidence for expansion of extant penguins out of Antarctica due to global cooling. *Proc. R. Soc. B* **273**, 11–17.

Boersma, P.D. 2008 Penguins as marine sentinels. *BioScience* **58**, 597-607.

Boersma, P.D., Stokes, D.L., & Yorio, P. 1990 Reproductive variability and historical change of Magellanic penguins (*Spheniscus magellanicus*) at Punta Tombo, Argentina. In *Biology of Penguins* (ed. L. Davis and J. Darby), pp. 15-43. New York: Macmillan.

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

Brouwer, L., Barr, I., van de Pol, M., Burke, T., Komdeur, J., & Richardson, D.S. 2010 MHC-dependent survival in a wild population: evidence for hidden genetic benefits gained through extra-pair fertilizations. *Mol. Ecol.* **19**, 3444-3455.

Brown, J.H., Jardetzky, T.S., Gorga, J.C., Stern, L.J., Urban, R.G., Strominger, .J.L, & Wiley, D.C. 1993 Three-dimesional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* **36,** 33-39.

Burri, R., Hirzel, N.H., Salamin, N., Roulin, A., & Fumagalli, L. 2008 Evolutionary patterns of MHC Class II B in owls and their implications for the understanding of avian MHC evolution. *Mol. Biol. Evol.* **25**, 1180–1191.

Consuegra, S., & de Leaniz, C.G. 2008 MHC-mediated mate choice increases parasite resistance in salmon. *Proc. R. Soc. B* **275**, 1397-1403.

Davies, C.J., Andersson, L., Ellis, S.A., Hensen, E.J., Lewin, H.A., Mikko, S., Muggli-Cockett, N.E., van der Poel, J.J., & Russell, G.C. 1997 Nomenclature for the factors of the *BoLA* system, 1996: report of the ISAG *BoLA* Nomenclature Committee. *Anim. Genet.* **28**, 159-168.

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

Edwards, S.V., Hess, C.M., Gasper, .J, & Garrigan, J. 1999 Toward an evolutionary genomics of the avian *Mhc*. *Immunol. Rev.* **167**, 119–132.

Eizaguirre, C., Yeates, S.E., Lenz, T.L., Kalbe, & Milinski, M. 2009 MHC-based mate choice combines good genes and maintenance of MHC polymorphism. *Mol. Ecol.* **18**, 3316–3329.

Ewald, P.W. 1994 *The Evolution of Infectious Disease*. New York: Oxford University Press.

Felsenstein, J. 2004 *Inferring Phylogenies*. Massachusettes: Sinauer Associates.

Freeman-Gallant, C.R., Geguerdichian, M., Wheelwright, N.T., & Sollecito, S.V. 2003 Social-pairing and female fidelity predicted by restriction fragment length polymorphism similarity at the major histocompatibility complex in a songbird. *Mol. Ecol.* **12**, 3077-3083.

Gandini, P.A., Frere, E., & Boersma, P.D. 1996 Status and conservation of Magellanic penguins (*Spheniscus magellanicus*) in Patagonia, Argentina. *Bird. Conserv. Int.* **6**, 307-316.

Hall, T.A. 1999 BioEdit: A user:friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acid Symposium Series* **41**, 95-98.

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

Hess, C.M., & Edwards, S.V. 2002 The evolution of the major histocompatibility complex in birds. *Bioscience* **52**, 423–431.

Hughes, A.L., & Yeager, M. 1998 Natural selection at major histocompatibility complex of vertebrates. *Annu. Rev. Genet.* **32**, 415-435.

Hughes, C.R., Miles, S., & Walbroehl, J.M. 2008 Support for the minimal essential MHC hypothesis: a parrot with a single, highly polymorphic MHC class II B gene. *Immunogenetics* **60**, 219-231.

Jukes, T.H., & Cantor, C.R. 1969 Evolution of protein molecules. In *Mammalian Protein Metabolism* (ed. H.N. Munro), pp. 21-123. New York: Academic Press.

Kennedy, L.J., Altet, L., Angles, J.M., Barnes, A., Carter, S.D., Francino, O., Gerlach, J.A., Happ, G.M., Ollier, W.E.R., Polvi, A., Thomson, W., & Wagner, J.L. 1999 Nomenclature for the factors of the Dog Major Histocompatibility system (DLA), 1998: first report of the ISAG DLA Nomenclature Committee. *Tissue Antigens* **54**, 312-321.

Kikkawa, E.F., Tsuda, T.T., Naruse, T.K., Sumiyama, D., Fukuda, M., Kurita, M., Murata, K., Wilson, R.P., LeMaho, Y., Tsuda, M., Kulski, J.K., & Inoko, H. 2005 Analysis of the sequence variations in the Mhc DRß1-like gene of the endangered Humboldt penguin (*Spheniscus humboldti*). *Immunogenetics* **57**, 99–107

Kikkawa, E.F., Tsuda, T.T., Sumiyama, D., Naruse, T.K., Fukuda, M., Kurita, M., Wilson, R.P., LeMaho, Y., Miller, G.D., Tsuda, M., Murata, K., Kulski, J.K., & Inoko, H. 2009 Trans-species polymorphism of the Mhc class II DRB-like gene in banded penguins (genus Spheniscus). *Immunogenetics* **61**, 341–352.

Klein, J. 1986 *Natural History of the Major Histocombatibility Complex*. New York: John Wiley and Sons.

Klein, J., Satta, Y., Takahata, N., & O'hUigin, C. 1993 Trans-specific MHC polymorphism and the origin of species in primates. *J. Med. Primatol.* **22**, 57–64.

Landry, C., Garant, D., Duchesne, P., & Bernatchez, L. 2001 "Good genes as heterozygosity": the major histocompatibility complex and mate choice in the Atlantic salmon (*Salmo salar*). *Proc. R. Soc. B* **268**, 1279-1285.

Lenz, T., & Becker, S. 2008 Simple approach to reduce PCR artifact formation leads to reliable genotyping of *MHC* and other highly polymorphic loci-Implications for evolutionary analysis. *Gene* **427**, 117-123.

Loisel, D.A., Alberts, S.C., & Ober, C. 2007 Functional significance of MHC variation in mate choice, reproductive outcome, and disease risk. In *Evolution in Health and Disease* (ed. S.C. Stearns), pp. 241-252. England: Oxford University Press.

Marsh, S.G., Bodmer, J.G., Albert, E.D., Bodmer, W.F., Bontrop, R.E., Dupont, B., Erlich, H.A., Hansen, J.A., Mach, B., Mayr, W.R., Parham ,P., Petersdorf, E.W., Sasazuki, T., Schreuder, G.M., Strominger, J.L., Svejgaard, A., & Teraski, P.I. 2001 Nomenclature for factors of the HLA system, 2000. *Tissue Antigens* **57**, 236-283.

Milinski, M. 2006 The major histocompatibility complex, sexual selection, and mate choice. *Annu. Rev. Eco.l Evol. Syst.* **37**, 159-186.

Nei, M., & Gojobori, T 1986 Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mo.l Biol. Evol.* **3**, 418-426.

Penn, D.J., & Potts, W.K. 1999 The evolution of mating preferences and major histocompatibility genes. *Am. Nat.* **153**, 145-164.

Penn, D.J., Damjanovich, K., & Potts, W.K. 2002 MHC heterozygosity confers a selective advantage against multiple-strain infections. *Proc. Nat. Acad. Sci.* **99**, 11260–11264.

Potti, J., Moreno, J., Yorio, P., Briones, V., Garcia-Borborglu, P., Villar, S., & Ballesteros, C. 2002 Bacteria divert resources from growth for Magellanic penguin chicks. *Ecol. Lett.* **5**, 709–714.

Potts, W.K., & Wakeland, E.K. 1993 The evolution of MHC genetic diversity: a tale of incest, pestilence, and sexual preference. *Trends. Genet.* **9**, 408-412.

Rozas, J., Sanchez-DelBarrio, C., Messeguer, X., & Rozas, R. 2003 DnaSP, DNApolymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496-2497.

Sambrook, J., Fritsch, E.F., & Maniatis, T. 1989 *Molecular Cloning: A Laboratory Manual,* 2nd edn. New York: Cold Spring Harbor Laboratory Press.

SAS Institute Inc., SAS OnlineDoc, Version 8, Cary, NC: SAS Institute Inc., 2000.

Schiavini, A.P., Yorio, P., Gandini, P., Rey, A., & Boersma, P.D. 2005 Los pingüinos de las costas argentinas: estado poblacional y conservación. *Hornero* **20**, 5-23.

Schneider, S.D., Roessli, D., & Excoffier, L. 2000 ARLEQUIN: A Software for Population Genetic Data Analysis. University of Geneva, Geneva Switzerland. [Online.] Available at lgb.unige.ch/arlequin/.

Seutin, G., White, B.N., & Boag, P.T. 1991 Preservation of avian blood and tissue samples for DNA analyses. *Can. J. Zool.* **69**, 82-90.

Snell, G.D. 1968 The H-2 locus of the mouse: observations and speculations concerning its comparative genetics and its polymorphism. *Folia Biologica* (Prague) **14**, 335–358.

Stokes, D.L., & Boersma, P.D. 1998 Nest-site characteristics and reproductive success in Magellanic penguins (*Spheniscus magellanicus*). *Auk* **115**, 34-49.

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

Tamura, K., Dudley, J., Nei, M., & Kumar, S. 2007 MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**, 1596-1599.

Trivers, R.L. 1972 Parental investment and sexual selection. In *Sexual selection and the descent of man* (ed. B.G. Campbell), pp. 136-179. Chicago: Aldine-Atherton, Chicago.

Trowsdale, J. 1993 Genomic structure and function in the MHC. *Trends Genet* **9**, 117-122.

Tsuda, T.T., Tsuda, M., Naruse, T., Kawata, H., Ando, A., Shiina, T., Fukuda, M., Kurita, M., LeMaho, I., Kulski, J.K., & Inoko, H. 2001 Phylogenetic analysis of penguin (Spheniscidae) species based on sequence variation in Mhc class II gene. *Immunogenetics* **53**, 712–716.

Van Buskirk, R.W., Nevitt, G.A. 2007 Evolutionary arguments for olfactory behavior in modern birds. *ChemoSense* **10**, 1-10.

von Schantz, T., Wittzell, H., Goransson, G., & Grahn, M. 1997 Mate choice, male condition-dependent ornamentation and MHC in the pheasant . *Hereditas* (Lund) **127**, 133-140.

Wakeland, E.K., Boehme, S., & She, J.X. 1990 The generation and maintenance of MHC class II gene polymorphism in rodents. *Immunol. Rev.* **113**, 39-46.

Wenzel, B. 2006 Olfactory sensation in the Kiwi and other birds. *Annals. N.Y. Acad. Sci.* **188**, 183-193.

Williams, T.D., & Boersma, P.D. 1995 Magellanic penguin. In *Bird families of the world, the penguins* (ed. T.D. Williams), pp. 249-258. England: Oxford University Press.

Worley, K., Collet, J., Spurgin, L.G., Cornwallis, C., Pizzari, T., & Richardson, D.S. 2010 MHC heterozygosity and survival in red junglefowl. *Mol. Ecol.* **19**, 3064-3075.