

Low MHC variation in the endangered Galápagos penguin (*Spheniscus mendiculus*)

Jennifer L. Bollmer · F. Hernán Vargas ·
Patricia G. Parker

Received: 30 January 2007 / Accepted: 31 March 2007 / Published online: 25 April 2007
© Springer-Verlag 2007

Abstract The major histocompatibility complex (MHC) is one of the most polymorphic regions of the genome, likely due to balancing selection acting to maintain alleles over time. Lack of MHC variability has been attributed to factors such as genetic drift in small populations and relaxed selection pressure. The Galápagos penguin (*Spheniscus mendiculus*), endemic to the Galápagos Islands, is the only penguin that occurs on the equator. It relies upon cold, nutrient-rich upwellings and experiences severe population declines when ocean temperatures rise during El Niño events. These bottlenecks, occurring in an already small population, have likely resulted in reduced genetic diversity in this species. In this study, we used MHC class II exon 2 sequence data from a *DRB1*-like gene to characterize the amount of genetic variation at the MHC in 30 Galápagos penguins, as well as one Magellanic penguin (*S. magellanicus*) and two king penguins (*Aptenodytes patagonicus*), and compared it to that in five other penguin species for which published data exist. We found that the Galápagos penguin had the lowest

MHC diversity (as measured by number of polymorphic sites and average divergence among alleles) of the eight penguin species studied. A phylogenetic analysis showed that Galápagos penguin MHC sequences are most closely related to Humboldt penguin (*Spheniscus humboldti*) sequences, its putative sister species based on other loci. An excess of non-synonymous mutations and a pattern of trans-specific evolution in the neighbor-joining tree suggest that selection is acting on the penguin MHC.

Keywords Galápagos penguin · *Spheniscus mendiculus* · Major histocompatibility complex · Genetic bottleneck · Trans-species evolution

Introduction

The genes at the major histocompatibility complex (MHC) are among the most polymorphic known, having unusually large numbers of alleles as well as higher nucleotide diversity than other loci (Parham and Ohta 1996; Gaudieri et al. 2000; Garrigan and Hedrick 2003). MHC molecules play a central role in the immune system by recognizing foreign peptides, binding to them, and presenting them to T cells, thus, initiating the immune response (Klein 1986). It is generally believed that MHC variability is the result of alleles being maintained in populations by some form of balancing selection, with the two most likely mechanisms being selection for resistance to parasites (either through overdominance or negative frequency-dependent selection) or sexual selection via mate choice (Doherty and Zinkernagel 1975; Takahata and Nei 1990; Penn and Potts 1999; reviewed in Bernatchez and Landry 2003; Piertney and Oliver 2006). The long-term maintenance of allelic lineages in populations due to balancing selection may result in trans-specific

J. L. Bollmer (✉) · P. G. Parker
Department of Biology, University of Missouri-St. Louis,
St. Louis, MO, USA
e-mail: jlbollmer@umsl.edu

F. H. Vargas
Charles Darwin Foundation,
Isla Santa Cruz, Galápagos, Ecuador

F. H. Vargas
Department of Zoology, Wildlife Conservation Research Unit,
University of Oxford,
Oxford, UK

Present address:

F. H. Vargas
The Peregrine Fund/Fondo Peregrino Panamá Apdo,
0844-00230 Panamá City, República de Panamá

evolution, where the coalescent times of MHC alleles found in different species predate speciation events (Takahata 1990; Klein et al. 1993).

In birds, there is a growing body of data describing MHC structure in non-model species. Earlier work on chickens found a very simplified MHC structure compared to that in mammals (Kaufman et al. 1999), but subsequent research on other species has shown that the simple chicken MHC is not representative of all birds, and there is much variation among species in the number and organization of MHC genes (Hess and Edwards 2002). Many studies of natural populations of birds have found the large numbers of divergent alleles expected at these loci (e.g., Ekblom et al. 2003; Bonneaud et al. 2004; Westerdahl et al. 2004). Although in some cases, the effects of genetic drift appear to outweigh balancing selection, resulting in reduced MHC polymorphism (e.g., Richardson and Westerdahl 2003; Miller and Lambert 2004a, b).

Galápagos penguins (*Spheniscus mendiculus*) are endemic to the Galápagos Islands (Fig. 1) and are the only tropical penguin species. They are able to persist at the equator due to the cold, nutrient-rich upwellings from the Cromwell current (Boersma 1977, 1978). About 95% of Galápagos penguins are distributed around the westernmost islands of Fernandina and Isabela where the upwelling is greatest, while the other 5% occur in small, isolated populations around three other islands (Bartolomé, Santiago, and Floreana; Boersma 1977, 1978). The Galápagos penguin undergoes dramatic population fluctuations in response to El Niño events when warmer water temperatures reduce food available to penguins and other species dependent on the normally cold, productive waters (Boersma 1998; Vargas et al. 2005a, 2006). Using a capture-mark-resight method, Vargas et al. (2005a) estimated that the population has fluctuated between 699 and 3,386 penguins since the first penguin census in 1970, and the population was estimated to have dropped by 77% during the 1982–1983 El Niño (Valle and Coulter 1987). In 2005, there were approximately 1,900 penguins (Vargas et al. 2005b). Because of its limited distribution and population crashes, the Galápagos penguin is listed as endangered (BirdLife International 2005).

While censuses indicate that the Galápagos penguin population has been fluctuating since at least the 1970s (Vargas et al. 2005a, 2006), it is likely that the penguin population has fluctuated for much longer. Riedinger et al. (2002) estimated that El Niño events have been occurring in Galápagos for at least the last 6,000 years, and the penguins are estimated to have arrived in the archipelago probably much earlier than that, about 4 million years ago (Baker et al. 2006). Small populations experience increased genetic drift and are thus expected to lose genetic variation more quickly than larger ones; furthermore, populations

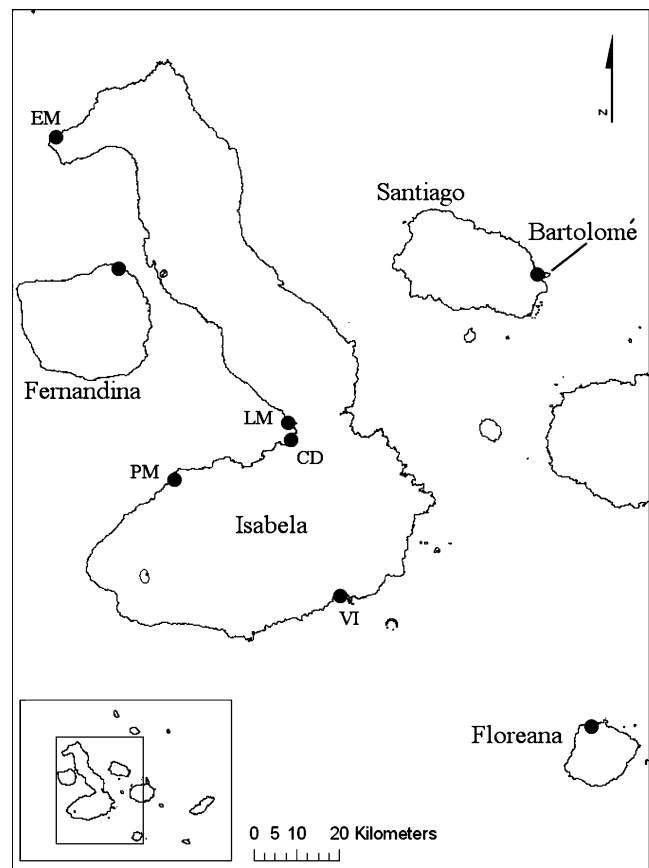


Fig. 1 Map of the western islands of the Galápagos archipelago (the inset shows the entire archipelago). The Galápagos Islands are located 1,000 km off the west coast of South America and overlap the equator. The black dots represent sites where Galápagos penguins (*Spheniscus mendiculus*) used in this study were sampled. EM El Muñeco, LM Las Marielas, CD Caleta Derek, PM Punta Moreno, CI Caleta Iguana, and VI Villamil

that undergo fluctuations in size are also expected to lose variation to drift (Wright 1931; Nei et al. 1975; Frankham 1996). Because of its small population size and repeated bottlenecks, the Galápagos penguin likely has less genetic variability than other penguin species and other outbred mainland species. In the only published genetic study of the Galápagos penguin, Akst et al. (2002) found a low level of heterozygosity (3%) at five microsatellite loci in the Galápagos penguin, which contrasted sharply with the 46% heterozygosity present in the Magellanic penguin (*S. magellanicus*), a species numbering over 1 million individuals (Gandini et al. 1996).

In this study, we present the first description of the MHC in the endangered Galápagos penguin. While the Galápagos penguin appears to have low genetic variability at neutral microsatellite loci, MHC genes are under balancing selection, so they may show variability equivalent to that in more outbred species, unless the effect of genetic drift has been too strong or selection has been relaxed. MHC variation has been

characterized in several penguin species (Tsuda et al. 2001), but it has been particularly well described in the Humboldt penguin (*Spheniscus humboldti*; Kikkawa et al. 2005), a temperate species that is sister to the Galápagos penguin (Baker et al. 2006). It is also affected by El Niño events; however, its population is much larger than the Galápagos penguin's (at least 10,000 individuals; Ellis et al. 1998; Luna-Jorquera et al. 2000). We compared MHC variability in the Galápagos penguin to that in its sister the Humboldt penguin and other penguin species for which published data exist. In addition, we incorporated our Galápagos penguin class II sequences into a phylogeny of previously published sequences from other penguin species, as well as preliminary sequences from Magellanic (*S. magellanicus*) and king (*Aptenodytes patagonicus*) penguins to compare interspecific relationships based on selected MHC genes with those based on nuclear and mitochondrial genes (Baker et al. 2006).

Materials and methods

Sampling

Blood samples were collected from Galápagos penguins during four separate trips from 2003 to 2005. Penguins were sampled from multiple sites on the islands of Floreana, Fernandina, Santiago, Bartolomé, and Isabela, covering the entire range of this species. To characterize Galápagos penguin MHC, we chose a random subset of 30 individuals from eight sites spread throughout the species' range (Fig. 1): Islote Las Bayas Pequeña by Floreana ($N=2$), Santiago/Bartolomé ($N=4$), Punta Espinosa on Fernandina ($N=4$), and five sites on Isabela (El Muñeco, Villamil, Punta Moreno, Las Marielas, and Caleta Iguana; $N=4$ individuals each for a total of 20). All birds used were adults.

We also used three blood samples taken by the Saint Louis Zoo from their penguin collection in Saint Louis, Missouri: one Magellanic and two king penguins. MHC sequences from these two species have not been previously published, and the primary purpose of these samples was for incorporation of more species (including a new genus, *Aptenodytes*) into our phylogeny.

MHC genotyping

We used the primers pen1 and pen4 (Tsuda et al. 2001; Kikkawa et al. 2005) to amplify a 198-bp fragment (primers included) of exon 2 of a class II MHC *DRB1*-like gene. Tsuda et al. (2001) found that this primer set amplified no more than two alleles per individual in the four penguin species they screened, suggesting that it was amplifying only one locus.

We genotyped 12 Galápagos penguin individuals using a combination of cloning and sequencing. First, the MHC was amplified using the pen1/pen4 primer set in 40 μ l reactions: 1 mM $MgCl_2$, 0.7X polymerase chain reaction (PCR) buffer, 0.2 μ M dNTPs, 0.4 μ M each primer, 0.3 units of *Taq* polymerase, and 80 ng genomic DNA. The PCR was run for 30 cycles of 94°C for 1 min, 57°C for 1 min, and 72°C for 2 min. Fragments were gel-purified using QIAquick gel extraction kits (QIAGEN) and then cloned using the pGEM-T easy vector cloning kit (Promega). Positive colonies were picked and suspended in 10 μ l dH_2O . They were screened for inserts of the correct size using M13 primers and then sequenced on an ABI 3100 using the primers SP6 and T7. We amplified and cloned each individual at least twice, and we sequenced at least ten positive clones from each individual. All sequences were double-stranded with 100% overlap.

We confirmed the genotypes of these 12 individuals and screened the other 18 Galápagos penguins using denaturing gradient gel electrophoresis (DGGE; Westerdahl et al. 2004; Knapp 2005). We used the same primers to amplify the MHC fragment; however, we added a GC-clamp to the 5' end of pen4 (Sheffield et al. 1989). We ran the reactions using the same temperature cycle as above, but reactions were in a volume of 48 μ l: 1.25 mM $MgCl_2$, 0.85X PCR buffer, 0.25 μ M dNTPs, 0.3 μ M each primer, 1.0 units of *Taq* polymerase, and 70 ng genomic DNA. PCR amplicons were run on 8% 19:1 acrylamide/bisacrylamide gels using a 40 to 60% denaturing gradient of formamide and urea. Gels ran at 160 V for 5 h at 60°C, and then we stained them using SYBR® gold (Promega) and visualized them on a Kodak IS440CF imaging system.

The Magellanic and king penguins were also genotyped using a combination of cloning, sequencing, and DGGE. All sequences were deposited into GenBank (accession numbers EF212007 to EF212014).

Data analysis

We assembled and edited sequences using Seqman v. 6.1 (DNASTAR) and aligned them manually in BioEdit v. 7.0.5.2 (Hall 1999). We calculated nucleotide diversity (π) using the program DnaSP v. 4.0 (Rozas et al. 2003), and we measured the rates of synonymous (d_S) and non-synonymous (d_N) substitutions using the Nei and Gojobori (1986) method with the Jukes–Cantor correction in MEGA v. 2.1 (Kumar et al. 2001). We calculated d_S and d_N separately for peptide-binding codons and non-peptide-binding codons as determined by Brown et al. (1993). We tested for positive selection ($d_N > d_S$) using a Z test in MEGA. To study the phylogenetic relationships among the MHC alleles, we constructed a neighbor-joining tree (Saitou and Nei 1987) from Jukes–Cantor distances in MEGA. It is important to

note that while we refer to our sequences as alleles, they do not encompass the entire length of exon 2 and are, thus, only partial allelic sequences.

Results

Identification of alleles

In the 30 Galápagos penguins screened, we confirmed the presence of three sequences (Spme1, 2, and 3). Two of them (Spme1 and 2) occurred in homozygous and heterozygous form in multiple individuals, while Spme3 was present in only one individual (a heterozygote). During the cloning and sequencing process, we obtained other apparent alleles as well. Those differing by a single mutation from the confirmed ones were attributed to *Taq* error, although two alleles arose multiple times in the Spme1/2 heterozygotes. When amplifying multiple sequences in one reaction, spurious alleles can form either through *in vitro* recombination when an incompletely amplified sequence pairs with the template of another or as heteroduplexes that form during the last PCR cycle when two completed alleles with different sequences anneal to each other (Jansen and Ledley 1990; L'Abbe et al. 1992; Longeri et al. 2002). The sequences of both of the suspect alleles could be explained by Spme1 and 2; one was identical to the 5' end of Spme1 and 3' end of Spme2, and the other was the reverse. When these individuals were run on the DGGE gels, they clearly had only the Spme1 and 2 alleles.

Spme3 is a combination of Spme1 and 2, matching the 5' end of Spme2 and the 3' end of Spme1; however, we confirmed it as a true allele in one individual. That individual was run multiple times on DGGE gels, and it consistently showed the Spme1 allele and another unique allele that ran slightly differently from Spme2. Cloning and sequencing of the individual produced the Spme3 allele.

We also ran a clone of the Spme3 allele adjacent to a direct PCR of the individual on a DGGE gel to verify that the clone comigrated with the unique allele.

The Magellanic penguin yielded two alleles after cloning and sequencing (Spma1 and 2), and we confirmed the sequences of three alleles from the two king penguins (Appa1, 2, and 3), although there was at least one other allele we did not confirm.

Because we only sequenced from genomic DNA, we cannot be sure that these confirmed alleles are expressed. However, we did not find any frameshift mutations or stop codons within them. Also, Tsuda et al. (2001) found that these primers amplified the same alleles from both genomic DNA and DNA from RT-PCR in an Adelie and chinstrap penguin, indicating that they amplified expressed alleles in those species.

Sequence variability

Among the three Galápagos penguin alleles, there were only three polymorphic sites in the 157 bp sequenced (after removing the primers). Spme1 and Spme2 differed from each other at three sites, while Spme3 differed from Spme2 at only one site and from Spme1 at two. MHC diversity was low in the Galápagos penguin compared to the other penguin species studied in terms of both number of alleles and degree of divergence among alleles (Table 1). In the other penguin species, there were many more polymorphic sites and consequently greater nucleotide diversity than what was present in the Galápagos penguin (Table 1, Fig. 2). While we sampled only one Magellanic and two king penguins, it appears that their variability (Table 1) may be comparable to that of the species studied by Tsuda et al. (2001) and Kikkawa et al. (2005).

All three Galápagos penguin substitutions were non-synonymous, and two of them occurred at probable antigen-binding sites (ABS; Fig. 2). We compared the

Table 1 Galápagos penguin MHC class II exon 2 sequence polymorphism compared to that in seven other penguin species (data are based on a 157-bp fragment)

Species	<i>n</i>	No. of alleles	No. of variable sites	Average no differences (\pm SE)	π
Adelie (<i>Pygoscelis adeliae</i>) ^a	4	4	20	10.7 \pm 2.3	0.068
Chinstrap (<i>P. antarctica</i>) ^a	3	3	19	12.7 \pm 2.7	0.081
Gentoo (<i>P. papua</i>) ^a	6	8	23	10.8 \pm 2.1	0.068
Little blue (<i>Eudyptula minor</i>) ^a	4	4	21	12.7 \pm 2.6	0.081
Humboldt (<i>Spheniscus humboldti</i>) ^b	20	6	20	9.7 \pm 2.1	0.062
Galápagos (<i>S. mendiculus</i>)	30	3	3	2.0 \pm 1.2	0.013
Magellanic (<i>S. magellanicus</i>)	1	2	16	16.0 \pm 3.9	0.102
King (<i>Aptenodytes patagonicus</i>)	2	3	19	12.7 \pm 2.7	0.081

The number of individuals sampled (*n*), the number of alleles found, the number of polymorphic sites, the average number of differences between alleles, and nucleotide diversity (π) are presented.

^a Data from Tsuda et al. (2001)

^b Data from Kikkawa et al. (2005)

Fig. 2 Alignment of partial MHC class II exon 2 amino acid sequences from eight species of penguin. The asterisks indicate likely antigen-binding sites based on Brown et al. (1993). *Spme* Galápagos penguin (*Spheniscus mendiculus*), *Sphu* Humboldt penguin (*S. humboldti*), *Spma* Magellanic penguin (*S. magellanicus*), *Eumi* little blue penguin (*Eudyptula minor*), *Pyad* Adelie penguin (*Pygoscelis adeliae*), *Pyan* chinstrap penguin (*P. antarctica*), *Pypa* gentoo penguin (*P. papua*), and *Appa* king penguin (*Aptenodytes patagonicus*)

		* * *	**	*	*	**	*	* * *	*	*
Spme1	(EF212007)	FVVRDIYNRQ	QDVHFDSDVG	YYVADTPLGE	PDAKYWNSQT	DILEQRRAEV	DTY			
Spme2	(EF212008)	Y.E.....A			
Spme3	(EF212009)	Y.E.....			
Sphu1	(AB154393)	Y.E.....	QF.....L			
Sphu2	(AB162144)N.....	QF.....DEA			
Sphu3	(AB154395)	Y.E.....	QF.....FK			..V
Sphu4	(AB154397)	QF.....FS			..A
Sphu5	(AB154398)	.E.Y.....	.N.....FK		
Sphu6	(AB154399)	L.E.....	.Y.....SA		
Spma1	(EF212010)	QF.....FS			..A
Spma2	(EF212011)	.D.Y.....	EY.....	QF.....SDEA		
Eumi1	(AB060946)N.....	.YA.....	H.....SRK			..V
Eumi2	(AB060949)	H.....RK			..V
Eumi3	(AB060948)K.....	EYA.....ILV			..DA
Eumi4	(AB060947)K.....	EYA.....	H.....SHL
Pyad1	(AB043601)	Y.....	.F.....	RH.....IDR		A
Pyad2	(AB043605)	Y.....	.F.....	R.....DF		A
Pyad3	(AB029998)	Y..A.....	.L.....	LFE.....F		K
Pyad4	(AB029994)	Y.....	.L.....	R.....DF		
Pyan1	(AB043556)	Y.E.Y.....	.YA.....	L.....PLR		V
Pyan2	(AB043559)	.L..N.H...	.L.....	L.....IDFR		V
Pyan3	(AB043558)	.L..N.H...	.F.....	L.....PLR		Q
Pypa1	(AB043590)	Y.E..H...	.Y.....	H.....LR		
Pypa2	(AB043600)	Y.E.Y.....	.Y...I.M.	H.....SLK		A
Pypa3	(AB043599)Y.....	H.....DR		L
Pypa4	(AB043598)	Y.E.Y.H...	.Y...I.M.	H.....LA		F
Pypa5	(AB043597)Y.....	H.....DR		
Pypa6	(AB043595)	Y.E.Y.H...	.Y...I.M.	H.....SLK		A
Pypa7	(AB043594)	.E.Y.H...	.F.....	H.....DRL		A
Pypa8	(AB043591)	.E.Y.....	MF.....	H.....DRL		A
Appa1	(EF212012)	.E.YF....	ELL.....	HF.....IDA
Appa2	(EF212013)	Y..N.....	EY.....	HH.....ID
Appa3	(EF212014)	.D.YF....	EY.....	HF.....SF		K

non-synonymous and synonymous substitution rates at probable antigen-binding codons and at the codons that are non-antigen-binding sites (non-ABS) for the penguin species studied (Table 2). Non-synonymous rates were higher at the ABS than at the non-ABS for all the species. At the non-ABS, d_N was not significantly greater than d_S for any species, while at the ABS, d_N was significantly greater than d_S for six of the eight species. The results for the Galápagos penguin ($Z=1.47, p=0.07$) and the gentoo penguin ($Z=1.24, p=0.11$) were not significant.

Phylogenetic analysis

A neighbor-joining analysis showed that all of the penguin sequences formed a monophyletic group separate from the fowl and passerine outgroups (Fig. 3). Within the penguin group, alleles from the Adelie, chinstrap, and Galápagos penguins formed single clusters by species. The little blue penguins separated into two clusters, one of which was closely related to the king penguins. The gentoo penguins fell into two clusters, while the Humboldt penguins fell into

Table 2 Comparison of rates of synonymous (d_S) and non-synonymous (d_N) substitutions among eight penguin species

Species	No. of alleles	Non-ABS		ABS			
		$d_N \pm SE$	$d_S \pm SE$	d_N/d_S	$d_N \pm SE$	$d_S \pm SE$	d_N/d_S
Adelie (<i>Pygoscelis adeliae</i>) ^a	4	0.035±0.017	0.020±0.020	1.75	0.244±0.068	0.023±0.026	10.61 ^c
Chinstrap (<i>P. antarctica</i>) ^a	3	0.052±0.025	0.039±0.030	1.33	0.270±0.079	0.000±0.000	n/a*
Gentoo (<i>P. papua</i>) ^a	8	0.051±0.018	0.031±0.027	1.65	0.168±0.043	0.076±0.066	2.21
Little blue (<i>Eudyptula minor</i>) ^a	4	0.039±0.021	0.027±0.027	1.44	0.327±0.117	0.034±0.024	9.62 ^c
Humboldt (<i>Spheniscus humboldti</i>) ^b	6	0.041±0.020	0.030±0.026	1.37	0.195±0.065	0.001±0.001	195.00 ^c
Galápagos (<i>S. mendiculus</i>)	3	0.008±0.008	0.000±0.000	n/a	0.041±0.027	0.000±0.000	n/a
Magellanic (<i>S. magellanicus</i>)	2	0.052±0.031	0.020±0.022	2.6	0.392±0.198	0.062±0.053	6.32 ^c
King (<i>Aptenodytes patagonicus</i>)	3	0.039±0.020	0.027±0.028	1.44	0.289±0.089	0.047±0.039	6.15 ^c

Rates were calculated separately for the 38 codons making up the non-antigen binding sites (ABS) and the 15 codons making up the ABS.

^a Calculated using sequences from Tsuda et al. (2001)
^b Calculated using sequences from Kikkawa et al. (2005)
^c One-tailed test indicated $d_N > d_S$ with p value <0.03

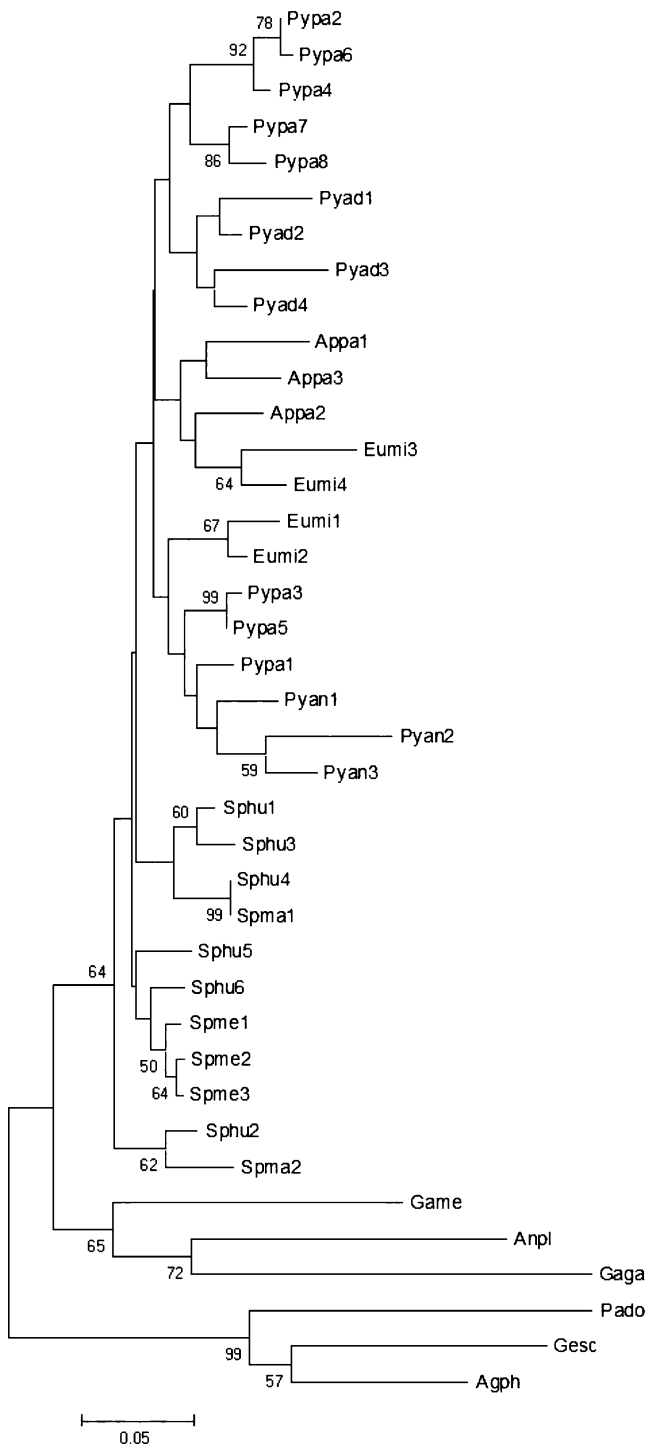


Fig. 3 Neighbor-joining tree constructed from Jukes–Cantor distances of partial MHC class II exon 2 sequences from eight penguin species. Bootstrap percentages (based on 1,000 repetitions) below 50 are not shown. Sequences from a snipe (Game, *Gallinago media*; AF485407), duck (Anpl, *Anas platyrhynchos*; AF390589), chicken (Gaga, *Gallus gallus*; M29763), sparrow (Pado, *Passer domesticus*; AY518182), blackbird (Agph, *Agelaius phoeniceus*; AF328737), and finch (Gesc, *Geospiza scandens*; Z74412) were used as outgroups. *Spme* Galápagos penguin (*Spheniscus mendiculus*), *Sphu* Humboldt penguin (*S. humboldti*), *Spma* Magellanic penguin (*S. magellanicus*), *Eumi* little blue penguin (*Eudyptula minor*), *Pyad* Adelie penguin (*Pygoscelis adeliae*), *Pyan* chinstrap penguin (*P. antarctica*), *Pypa* gentoo penguin (*P. papua*), and *Appa* king penguin (*Aptenodytes patagonicus*)

considered cautiously, as the bootstrap support for many of the nodes is very low.

Discussion

MHC loci are known for their large numbers of divergent alleles. Contrary to what has been found in most other species, the Galápagos penguin had very little genetic diversity at the *DRB1*-like MHC class II locus we studied. We found only three alleles in the 30 individuals we genotyped, and one of those alleles was present in only one individual. Because these are only partial allelic sequences from exon 2, there could be a greater number of alleles when the entire length of the exon is taken into account. While the number of sequences we found in the Galápagos penguin at these 157 bp is comparable to the number found in some other penguin species, the sample sizes of individuals genotyped in those species were small, likely missing other alleles. In addition, the Galápagos penguin was less variable than all the other species in number of polymorphic sites and divergence among alleles. The other well-sampled species is the Humboldt penguin, the Galápagos penguin's closest living relative. The Humboldt showed much higher diversity than the Galápagos penguin at this gene (20 polymorphic sites versus three and a nucleotide diversity of 0.06 versus 0.01 in 157 bp of the exon).

Genetic drift is the most likely explanation for the reduced genetic diversity in the Galápagos penguin, given its demographic history in which there was probably an initial founder effect when the population established itself in the archipelago followed by repeated population bottlenecks caused by El Niño events over thousands of years (in an already small population). There are other cases of low MHC diversity, especially in bottlenecked or naturally small populations (e.g., island populations). For example, Miller and Lambert (2004a, b) found that the Chatham Island black robin (*Petroica traversi*) of New Zealand was fixed for three of four MHC class II alleles (probably from four loci), and the endemic Seychelles warbler (*Acrocephalus sechellensis*) has substantially reduced MHC diversity

three. One of the Magellanic alleles was identical to a previously published Humboldt allele, at least at the 157 bp for which we have data, while the other fell within a cluster of Humboldt alleles. At the generic level, the three *Pygoscelis* species grouped into a large cluster, although the one *Eudyptula* and one *Aptenodytes* species fell within them. The three *Spheniscus* species also formed clusters together, although most of these relationships have to be

compared to the widespread great reed warbler (*A. arundinaceus*; Richardson and Westerdahl 2003). In mammals, small island populations of the Australian bush rat (*Rattus fuscipes greyii*) were mostly fixed for different MHC alleles (Seddon and Baverstock 1999), as were populations of the bottlenecked Eurasian beaver (*Castor fiber*; Babik et al. 2005). In these cases, it is thought that the effect of genetic drift was particularly strong and overwhelmed the effect of balancing selection. Hedrick et al. (2000) noted that in some species with reduced numbers of MHC alleles, the remaining alleles are highly divergent, which might allow for the recognition of a wider range of pathogens. The amino acid sequences of the four black robin alleles differed by an average of 25% (Miller and Lambert 2004b), whereas in the Galápagos penguin, there was an average of only 4% amino acid divergence among alleles.

An alternative explanation for lower MHC diversity is reduced selection on the MHC due to reduced exposure to parasites. Slade (1992) hypothesized that the lower MHC variation in some whales (Trowsdale et al. 1989; Murray and White 1998) and seals (Slade 1992; Lehman et al. 2004) compared to terrestrial mammals is a result of their exposure to a more limited suite of pathogens. Penguins evolved in cold, marine habitats where there are few other bird species to act as disease reservoirs and where pathogen and vector diversity is low (Jones and Shellam 1999; Clarke and Kerry 2000). There have been no records of haematozoa from Antarctic or sub-Antarctic penguin populations, and haematozoa are found at low prevalences and intensities in those more northern populations that are infected (Jones and Shellam 1999). Parasite and vector diversity should be higher in the tropics, so it might be expected that the Galápagos penguin should be under greater selection pressure than other penguin species; however, the Galápagos Islands are isolated and likely have reduced parasite diversity compared to tropical mainland habitats. In a baseline health survey of the Galápagos penguin, Travis et al. (2006) found that while 89% of the tested penguins were seropositive for *Chlamydophila psittaci*, all were seronegative for 14 common avian viruses (e.g., paramyxovirus, infectious bursal disease, Marek's disease, and adenovirus). Also, blood smears revealed the presence of microfilarid nematodes, although at low intensities and at an overall prevalence of 13.8% (Merkel et al. 2007). While relaxed selection on the MHC might be partially responsible for the reduced genetic diversity, it is likely that demographic factors are having a stronger effect.

There are species where MHC variation exists despite reduced neutral genetic variation due to genetic drift. Aguilar et al. (2004) found variation at the DRB MHC locus and three microsatellite loci linked to the MHC in a population of Channel Island foxes (*Urocyon littoralis dicheyi*) that was previously shown to be completely

monomorphic at neutral minisatellite (Gilbert et al. 1990) and microsatellite (Goldstein et al. 1999) loci. Aguilar et al. (2004) concluded that periodic balancing selection may have preserved this variation despite genetic drift (but see Hedrick 2004). Jarvi et al. (2004) suggested that balancing selection might explain why a species of honeycreeper (*Vestiaria coccinea*) that was monomorphic at the mitochondrial control region had MHC variability comparable to that of more outbred species, although in the Galápagos penguin, the level of neutral genetic variability appears to be consistent with MHC variability. Akst et al. (2002) found reduced heterozygosity at microsatellite loci in the Galápagos penguin compared to Magellanic penguins, and a microsatellite study performed in our lab shows that Galápagos penguins have a small number of alleles per locus compared to other species (Nims et al., unpublished data).

This primer set amplified transcribed alleles in two other penguin species (Tsuda et al. 2001), making it unlikely that the low MHC variability seen here was due to the amplification of a pseudogene or nonclassical locus (Hess et al. 2000; Aguilar et al. 2006). Furthermore, we found higher non-synonymous substitution rates at antigen-binding sites than at non-antigen-binding sites in all penguins, and higher non-synonymous substitution rates than synonymous substitution rates at antigen-binding sites in six of eight species, which also suggests that this locus is under selection.

Trans-specific evolution of MHC alleles

The penguin phylogeny of Baker et al. (2006), based on 5,691 bp of nuclear and mitochondrial DNA, indicated that the *Aptenodytes* penguins (king and emperor) were the most basal, followed by the *Pygoscelis* species, chinstraps, and gentoos being more closely related to each other than either was to the Adelies. *Eudyptula* and *Spheniscus* were sister genera, and within *Spheniscus*, Galápagos and Humboldt penguins were sister, as were Magellanic and African penguins.

Our neighbor-joining tree based on 157 bp of MHC sequence data did not match these relationships. Instead of the king sequences being most basal followed by *Pygoscelis* and *Spheniscus*, it was the reverse, with *Spheniscus* sequences being most basal. The king penguin sequences clustered with two *Eudyptula* sequences within a larger cluster of *Pygoscelis* species. Gentoo sequences clustered with both Adelie and chinstrap sequences rather than more closely with the chinstraps. The *Eudyptula* sequences were more closely related to *Pygoscelis* and *Aptenodytes* than *Spheniscus*. Based on the data of Baker et al. (2006), we had expected the close relationship we found between the Humboldt and Galápagos sequences, as they are likely sister

species. We had expected the Magellanic sequences, however, to be more divergent from the Humboldt and Galápagos sequences, but they clustered with the Humboldts, one of them being identical to a Humboldt allele at the 157-bp fragment sequenced. It will be interesting to see if this relationship holds true with a larger sampling of Magellanic penguins. The Magellanic and Humboldt penguin distributions overlap in the wild, and our sampled Magellanic penguin's close relationship with the Humboldt penguins in the phylogenetic tree suggests a hybridization event in its ancestry.

This lack of concordance between the two phylogenies could be due to sampling error, either from few individuals being genotyped for some species or a lack of resolution due to the relatively short fragment size sequenced. There was low bootstrap support for many of the nodes. Alternatively, the differences in topology and lack of stronger structure could be due to the effect of selection acting on the MHC. Balancing selection on MHC alleles may result in trans-species evolution, where alleles have long coalescent times (often predating speciation events) and show less divergence among species than what is found at neutral markers. Alleles from related species are sometimes interdigitated on trees as has been found within honeycreepers (Jarvi et al. 2004), Darwin's finches (Vincek et al. 1997), and warblers (Richardson and Westerdahl 2003). There was evidence of this within the penguins as well. Only three of the eight species formed monophyletic clusters, and the Magellanic penguin shared an allele with the Humboldt penguins.

There is growing concern about the threat introduced diseases may pose to native bird species in the Galápagos Islands (Wikelski et al. 2004; Parker et al. 2006). Mosquito vectors are present in the archipelago, including the species known to vector avian malaria (*Plasmodium*) elsewhere (Whiteman et al. 2005). Both in the wild and in captivity, penguins are highly susceptible to exotic pathogens (Clarke and Kerry 1993). Other *Spheniscus* species have been found to be very susceptible to *Plasmodium* in captivity (Fix et al. 1988; Cranfield et al. 1991). There has not yet been any evidence of *Plasmodium* in Galápagos penguins (Miller et al. 2001), although other blood parasites (*Haemoproteus* sp. and microfilariae) have been reported (Parker et al. 2006; Travis et al. 2006). The Galápagos penguin is already at risk due to demographic factors (small population size and periodic bottlenecks), but it appears that genetic monomorphism at these immunological loci that are instrumental in disease resistance may put this species at even further risk.

Acknowledgments We would like to thank the people who collected the samples used in this study: C. Duffie, N. Gottdenker, G. Jimenez Uzcátegui, J. Merkel, C. Larrea, and E. Travis. We would

also like to thank the Galápagos National Park and the Charles Darwin Foundation for allowing us to do this work and for their support in the field. Michael Macek and Anne Tieber at the Saint Louis Zoo provided the Magellanic and king penguin samples. Funds for this project were provided by the Saint Louis Zoo Field Research for Conservation Program, the Des Lee Collaborative Vision in Zoological Research, D. Swarovski & Co., and the American Ornithologists' Union. For help with the laboratory work, we are grateful to B. Nims, K. Matson, and especially K. Halbert.

References

- Aguilar A, Roemer G, Debenham S, Binns M, Garcelon D, Wayne RK (2004) High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. *Proc Natl Acad Sci USA* 101:3490–3494
- Aguilar A, Edwards SV, Smith TB, Wayne RK (2006) Patterns of variation in MHC class II β loci of the little greenbul (*Andropadus virens*) with comments on MHC evolution in birds. *J Heredity* 97:133–142
- Akst EP, Boersma PD, Fleischer RC (2002) A comparison of genetic diversity between the Galápagos penguin and the Magellanic penguin. *Conservation Genetics* 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 AJ, Pereira SL, Haddrath OP, Edge KA (2006) Multiple gene evidence for expansion of extant penguins out of Antarctica due to global cooling. *Proc R Soc Lond B Biol Sci* 273:11–17
- Bernatchez L, Landry C (2003) MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *J Evol Biol* 16:363–377
- BirdLife International (2005) *Spheniscus mendiculus*. In: IUCN 2006. 2006 IUCN red list of threatened species (<http://www.iucnredlist.org>) downloaded on 21 January 2007
- Boersma PD (1977) An ecological and behavioral study of the Galápagos Penguin. *Living Bird* 15:43–93
- Boersma PD (1978) Breeding patterns of Galápagos penguins as an indicator of oceanographic conditions. *Science* 200:1481–1483
- Boersma PD (1998) Population trends of the Galápagos penguin: impacts of El Niño and La Niña. *Condor* 100:245–253
- Bonneaud C, Sorci G, Morin V, Westerdahl H, Zoorob R, Wittzell H (2004) Diversity of *Mhc* class I and IIB genes in house sparrows (*Passer domesticus*). *Immunogenetics* 55:855–865
- Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL, Wiley DC (1993) Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* 364:33–39
- Clarke JR, Kerry KR (1993) Diseases and parasites of penguins. *Kor J Polar Res* 4:79–96
- Clarke JR, Kerry KR (2000) Diseases and parasites of penguins. *Penguin Conserv* 13:5–24
- Cranfield MR, Beall FB, Skjoldager MT, Ialeggio DM (1991) Avian malaria. *Spheniscus Penguin Newsletter* 4:5–7
- Doherty PC, Zinkernagel RM (1975) Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature* 256:50–52
- Eklom R, Grahn M, Höglund J (2003) Patterns of polymorphism in the MHC class II of a non-passerine bird, the great snipe (*Gallinago media*). *Immunogenetics* 54:734–741
- Ellis S, Croxall JP, Cooper J (eds) (1998) Penguin conservation assessment and management plan. IUCN/SSC Conservation Breeding Specialist Group, Apple Valley, Minnesota

- Fix AS, Waterhouse C, Greiner EC, Stoskopf MK (1988) *Plasmodium relictum* as a cause of avian malaria in wild-caught Magellanic penguins (*Spheniscus magellanicus*). *J Wildl Dis* 24:610–619
- Frankham R (1996) Relationship of genetic variation to population size in wildlife. *Conserv Biol* 10:1500–1508
- Gandini P, Frere E, Boersma PD (1996) Status and conservation of Magellanic penguins *Spheniscus magellanicus* in Patagonia, Argentina. *Bird Conserv Int* 6:307–316
- Garrigan D, Hedrick PW (2003) Detecting adaptive molecular polymorphism: lessons from the MHC. *Evolution* 57:1707–1722
- Gaudieri S, Dawkins RL, Habara K, Kulski JK, Gojobori T (2000) SNP profile within the human major histocompatibility complex reveals an extreme interrupted level of nucleotide diversity. *Genome Res* 10:1579–1586
- Gilbert DA, Lehman N, O'Brien SJ, Wayne RK (1990) Genetic fingerprinting reflects population differentiation in the California Channel Island fox. *Nature* 344:764–767
- Goldstein DB, Roemer GW, Smith DA, Reich DE, Bergman A, Wayne RK (1999) The use of microsatellite variation to infer population structure and demographic history in a natural model system. *Genetics* 151:797–801
- Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 41:95–98
- Hedrick P (2004) Foxy MHC selection story. *Heredity* 93:237–238
- Hedrick PW, Parker KM, Gutiérrez-Espeleta GA, Rattink A, Liewers K (2000) Major histocompatibility complex variation in the Arabian oryx. *Evolution* 54:2145–2151
- Hess CM, Edwards SV (2002) The evolution of the major histocompatibility complex in birds. *Bioscience* 52:423–431
- Hess CM, Gasper J, Hoekstra HE, Hill CE, Edwards SV (2000) MHC class II pseudogene and genomic signature of a 32-kb cosmid in the house finch (*Carpodacus mexicanus*). *Genome Res* 10:613–623
- Jansen R, Ledley FD (1990) Disruption of phase during PCR amplification and cloning of heterozygous target sequences. *Nucleic Acids Res* 18:5153–5156
- Jarvi SI, Tarr CL, McIntosh CE, Atkinson CT, Fleischer RC (2004) Natural selection of the major histocompatibility complex (*Mhc*) in Hawaiian honeycreepers (Drepanidinae). *Mol Ecol* 13:2157–2168
- Jones HI, Shellam GR (1999) The occurrence of blood-inhabiting protozoa in captive and free-living penguins. *Polar Biol* 21:5–10
- Kaufman J, Milne S, Göbel TWF, Walker BA, Jacob JP, Auffray C, Zoorob R, Beck S (1999) The chicken B locus is a minimal-essential major histocompatibility complex. *Nature* 401:923–925
- Kikkawa EF, Tsuda TT, Naruse TK, Sumiyama D, Fukuda M, Kurita M, Murata K, Wilson RP, LeMaho Y, Tsuda M, Kulski JK, Inoko H (2005) Analysis of the sequence variations in the *Mhc DRB1*-like gene of the endangered Humboldt penguin (*Spheniscus humboldti*). *Immunogenetics* 57:99–107
- Klein J (1986) Natural history of the histocompatibility complex. Wiley, New York
- 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
- Knapp LA (2005) Denaturing gradient gel electrophoresis and its use in the detection of major histocompatibility complex polymorphism. *Tissue Antigens* 65:211–219
- Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17:1244–1245
- L'Abbe D, Belmaaza A, Decary F, Chartrand P (1992) Elimination of heteroduplex artifacts when sequencing HLA genes amplified by polymerase chain reaction (PCR). *Immunogenetics* 35:395–397
- Lehman N, Decker DJ, Stewart BS (2004) Divergent patterns of variation in major histocompatibility complex class II alleles among Antarctic phocid pinnipeds. *J Mammal* 85:1215–1224
- Longeri M, Zanotti M, Damiani G (2002) Recombinant DRB sequences produced by mismatch repair of heteroduplexes during cloning in *Escherichia coli*. *Eur J Immunogenet* 29:517–523
- Luna-Jorquera G, Garthe S, Sepulveda FG, Weichler T, Vásquez JA (2000) Population size of Humboldt penguins assessed by combined terrestrial and at-sea counts. *Waterbirds* 23:506–510
- Merkel J, Jones HI, Whiteman NK, Gottdenker N, Vargas H, Travis EK, Miller RE, Parker PG (2007) Microfilariae in Galápagos penguins (*Spheniscus mendiculus*) and flightless cormorants (*Phalacrocorax harrisi*): genetics, morphology and prevalence. *J Parasitol* (in press)
- Miller HC, Lambert DM (2004a) Gene duplication and gene conversion in class II MHC genes of New Zealand robins (Petroicidae). *Immunogenetics* 56:178–191
- Miller HC, Lambert DM (2004b) Genetic drift outweighs balancing selection in shaping post-bottleneck major histocompatibility complex variation in New Zealand robins (Petroicidae). *Mol Ecol* 13:3709–3721
- Miller GD, Hofkin BV, Snell H, Hahn A, Miller RD (2001) Avian malaria and Marek's disease: potential threats to Galapagos penguins *Spheniscus mendiculus*. *Mar Ornithol* 29:43–46
- Murray BW, White BN (1998) Sequence variation at the major histocompatibility complex *DRB* loci in beluga (*Delphinapterus leucas*) and narwhal (*Monodon monoceros*). *Immunogenetics* 48:242–252
- Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3:418–426
- Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. *Evolution* 29:1–10
- Parham P, Ohta T (1996) Population biology of antigen presentation by MHC class I molecules. *Science* 272:67–74
- Parker PG, Whiteman NK, Miller RE (2006) Conservation medicine on the Galápagos Islands: partnerships among behavioral, population, and veterinary scientists. *Auk* 123:625–638
- Penn DJ, Potts WK (1999) The evolution of mating preferences and major histocompatibility complex genes. *Am Nat* 153:145–164
- Piertney SB, Oliver MK (2006) The evolutionary ecology of the major histocompatibility complex. *Heredity* 96:7–21
- Richardson DS, Westerdahl H (2003) MHC diversity in two *Acrocephalus* species: the outbred great reed warbler and the inbred Seychelles warbler. *Mol Ecol* 12:3523–3529
- Riedinger MA, Steinitz-Kannan M, Last WM, Brenner M (2002) A ~6100 ¹⁴C yr record of El Niño activity from the Galápagos Islands. *J Paleolimnol* 27:1–7
- Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496–2497
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425
- Seddon JM, Baverstock PR (1999) Variation on islands: major histocompatibility complex (*Mhc*) polymorphism in populations of the Australian bush rat. *Mol Ecol* 8:2071–2079
- Sheffield VC, Cox DR, Lerman LS, Myers RM (1989) Attachment of a 40-base-pair G+C-rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes. *Proc Natl Acad Sci USA* 86:232–236
- Slade RW (1992) Limited MHC polymorphism in the southern elephant seal: implications for MHC evolution and marine mammal population biology. *Proc R Soc Lond B Biol Sci* 249:163–171

- Takahata N (1990) A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. *Proc Natl Acad Sci USA* 87:2419–2423
- Takahata N, Nei M (1990) Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics* 124:967–978
- Travis EK, Vargas FH, Merkel J, Gottdenker N, Miller RE, Parker PG (2006) Hematology, serum chemistry, and serology of Galápagos penguins in the Galápagos Islands, Ecuador. *J Wildl Dis* 42: 625–632
- Trowsdale J, Groves V, Arnason A (1989) Limited MHC polymorphism in whales. *Immunogenetics* 29:19–24
- Tsuda TT, Tsuda M, Naruse T, Kawata H, Ando A, Shiina T, Fukuda M, Kurita M, LeMaho I, Kulski JK, Inoko H (2001) Phylogenetic analysis of penguin (*Spheniscidae*) species based on sequence variation in MHC class II genes. *Immunogenetics* 53:712–716
- Valle CA, Coulter MC (1987) Present status of the flightless cormorant, Galapagos penguin and greater flamingo populations in the Galapagos Islands, Ecuador, after the 1982–83 El Niño. *Condor* 89:276–281
- Vargas H, Loughheed C, Snell H (2005a) Population size and trends of the Galápagos penguin *Spheniscus mendiculus*. *Ibis* 147:367–374
- Vargas H, Steinfurth A, Larrea C, Jiménez G, Llerena W (2005b) Penguin and cormorant census 2005. Report to the Charles Darwin Research Station and the Galápagos National Park Service, Oxford
- Vargas FH, Harrison S, Rea S, Macdonald DW (2006) Biological effects of El Niño on the Galápagos penguin. *Biol Conserv* 127:107–114
- Vincek V, O’uigin C, Satta Y, Takahata N, Boag PT, Grant PR, Grant BR, Klein J (1997) How large was the founding population of Darwin’s finches? *Proc R Soc Lond B Biol Sci* 264:111–118
- Westerdahl H, Wittzell H, von Schantz T, Bensch S (2004) MHC class I typing in a songbird with numerous loci and high polymorphism using motif-specific PCR and DGGE. *Heredity* 92: 534–542
- Whiteman NK, Goodman SJ, Sinclair BJ, Walsh T, Cunningham AA, Kramer LD, Parker PG (2005) Establishment of the avian disease vector *Culex quinquefasciatus* Say, 1823 (Diptera: Culicidae) on the Galápagos Islands, Ecuador. *Ibis* 147:844–847
- Wikelski M, Foufopoulos J, Vargas H, Snell H (2004) Galápagos birds and diseases: invasive pathogens as threats for island species. *Ecol Soc* 9:5, available online <http://www.ecologyandsociety.org/vol9/iss1/art5>
- Wright S (1931) Evolution in Mendelian populations. *Genetics* 16: 97–159