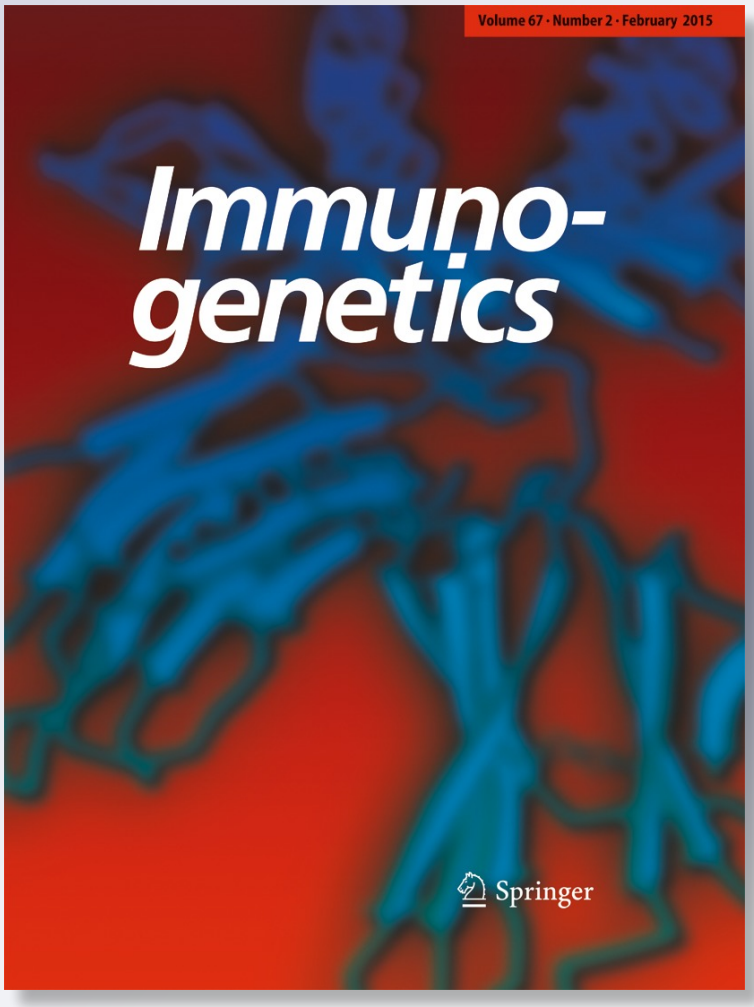
Non-neutral evolution and reciprocal monophyly of two expressed Mhc class II B genes in Leach's storm-petrel

Donald C. Dearborn, Andrea B. Gager, Morgan E. Gilmour, Andrew G. McArthur, Douglas A. Hinerfeld & Robert A. Mauck

Immunogenetics

ISSN 0093-7711
Volume 67
Number 2

Immunogenetics (2015) 67:111-123
DOI 10.1007/s00251-014-0813-2



Springer

Immunogenetics (2015) 67:111–123
DOI 10.1007/s00251-014-0813-2

ORIGINAL PAPER

Non-neutral evolution and reciprocal monophyly of two expressed Mhc class II B genes in Leach's storm-petrel

Donald C. Dearborn · Andrea B. Gager · Morgan E. Gilmour · Andrew G. McArthur · Douglas A. Hinerfeld · Robert A. Mauck

Received: 20 August 2014 / Accepted: 30 October 2014 / Published online: 23 November 2014

Abstract The major histocompatibility complex (Mhc) is subject to pathogen-mediated balancing selection and can link natural selection with mate choice. We characterized two Mhc class II B loci in Leach's storm-petrel, *Oceanodroma leucorhoa*, focusing on exon 2 which encodes the portion of the protein that binds pathogen peptides. We amplified and sequenced exon 2 with locus-specific nested PCR and Illumina MiSeq using individually barcoded primers. Repeat genotyping of 78 single-locus genotypes produced identical results in 77 cases (98.7 %). Sequencing of messenger RNA (mRNA) from three birds confirmed expression of both loci, consistent with the observed absence of stop codons or frameshifts in all alleles. In 48 birds, we found 9 and 12 alleles at the two loci, respectively, and all 21 alleles translated to unique amino acid sequences. Unlike many studies of duplicated Mhc genes, alleles of the two loci clustered into monophyletic groups. Consistent with this phylogenetic result, interlocus gene conversion appears to have affected only two short fragments of the exon. As predicted under a paradigm of pathogen-mediated selection, comparison of synonymous and non-synonymous substitution rates found evidence of a history of positive selection at putative peptide binding sites. Overall, the results suggest that the gene duplication event leading to these two loci is not recent and that point mutations and positive selection on the peptide binding sites may be the predominant forces acting on these genes. Characterization of these loci sets the stage for population-level work on the evolutionary ecology of Mhc in this species.

Keywords Major histocompatibility complex · Seabirds · Positive selection · Gene conversion · Gene duplication · Peptide binding site

Introduction

The major histocompatibility complex (Mhc) provides a system for exploring host-pathogen interactions (Carrington et al. 1999; Hill et al. 1991; Sommer 2005; Thursz et al. 1997; Wegner et al. 2003), balancing selection (Doherty and Zinkernagel 1975; Hughes and Nei 1989; Spurgin and Richardson 2010), and the link between natural selection and mate choice (Ekblom et al. 2010; Milinski 2006; Ober et al. 1997; Potts et al. 1991; Wedekind et al. 1995). Mate choice is expected to target Mhc because of fitness differences between genotypes (Kaufman and Wallny 1996; Kurtz et al. 2004; Paterson et al. 1998; Worley et al. 2010), but the action of sexual selection or natural selection on Mhc may depend on the type and amount and variation in a population. Like members of other multi-gene families, Mhc genes follow a birth-and-death model of evolution, with genes arising by duplication and departing by inactivation or deletion (Nei and Rooney 2005; Sato et al. 2001). Additional evolutionary processes create or eliminate Mhc diversity—point mutations,

Electronic supplementary material The online version of this article (doi:10.1007/s00251-014-0813-2) contains supplementary material, which is available to authorized users.

D. C. Dearborn (✉) · A. B. Gager
Department of Biology, Bates College, 44 Campus Ave, Lewiston, ME, USA
e-mail: ddearbor@bates.edu

M. E. Gilmour
Ocean Sciences Department, University of California Santa Cruz, Santa Cruz, CA, USA

A. G. McArthur
Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton L8S 4K1, ON, Canada

D. A. Hinerfeld
Jackson Laboratories, Bar Harbor, ME, USA

R. A. Mauck
Biology Department, Kenyon College, Gambier, OH, USA

gene conversion, recombination, and parallel or convergent evolution (Chaves et al. 2010; Hess and Edwards 2002; Hughes and Nei 1989; Nei and Rooney 2005; Reusch and Langefors 2005; Richman et al. 2003; Zangenberg et al. 1995). The relative importance of these processes is debated and is likely to vary across taxa.

Understanding the drivers of Mhc diversity requires data from many species, but such efforts are constrained by the difficulty of characterizing Mhc genes in non-model animals (Bernatchez and Landry 2003). One problem is that gene duplication has led to wide variation in the number of loci between species (Balakrishnan et al. 2010; Bollmer et al. 2010; Jaratlerdsiri et al. 2014; Kaufman et al. 1999). This gene duplication, along with concerted evolution (Alcaide et al. 2007; Bahr and Wilson 2012; Zangenberg et al. 1995), can make single-locus genotyping difficult (Strand et al. 2013).

Here, we make an initial characterization of Mhc class II B in Leach's storm-petrel, *Oceanodroma leucorhoa*. Seabirds have been suggested as good systems for studying the interplay between natural selection and sexual selection on Mhc (Zelano and Edwards 2002) because of their long lifespan and monogamy, and because they have an olfactory capability which is thought to be the key to assessing the Mhc genotypes of potential mates (Boehm and Zufall 2006; Spehr et al. 2006). Leach's storm-petrels fit all of these characteristics: They can live more than 25 years (Mauck et al. 2004), have long-term, monogamous pair-bonds (Mauck et al. 1995), and use olfactory cues to find their nesting burrows and to find food (Grubb Jr 1974; Nevitt and Haberman 2003).

To lay the groundwork for studies of Mhc-targeted mate choice in a framework of pathogen-mediated selection, three key questions must be answered about the focal population. First, are the loci under consideration being expressed? If Mhc-based mate assessment occurs via olfactory detection of Mhc peptides—as has been shown beautifully in fish (e.g., Milinski et al. 2005)—mate-choosing individuals will detect only loci that are expressed. Mhc pseudogenes may provide an opportunity to study evolution after the relaxation of selection (Eirín-López et al. 2012), but they would not be plausible candidates for Mhc-targeted mate choice.

Second, how much diversity exists at exon 2, a segment encoding the extracellular domain that presents pathogen peptides to helper T cells? Imagine, for example, that exon 2 diversity within the focal population is extremely high, with a large number of alleles of roughly equal frequency. In a scenario of selection favoring Mhc-disassortative mating (Juola and Dearborn 2012), even random mate choice would be likely to produce heterozygous offspring. In contrast, more targeted mate choice for Mhc complementarity would be needed to achieve the same outcome in a population with lower allelic diversity.

Third, is there evidence that natural selection has shaped Mhc alleles in a manner consistent with pathogen-mediated selection? Within exon 2, particular codons translate to amino acids that interact with pathogen peptides (Brown et al. 1993), and variation at these peptide binding sites (PBSs) governs the particular peptides that can be displayed to the immune system (Schad et al. 2005). Under pathogen-mediated selection, Mhc alleles are predicted to show functional differences at the PBS (i.e., a history of positive selection) while showing purifying selection at non-PBS codons (Bernatchez and Landry 2003).

Therefore, we had three aims in the present study of Leach's storm-petrels: (1) to determine the number of Mhc class II B loci and whether each is expressed; (2) to characterize diversity of exon 2, in terms of the number of alleles, their frequencies, and their phylogenetic relationships to one another; and (3) to test for evidence of a history of positive selection on peptide binding sites, as predicted under a paradigm of pathogen-mediated selection. We sampled storm-petrels at the Bowdoin Scientific Station on Kent Island (44° 35′ N, 66° 45′ W), a long-term study site in Canada's Bay of Fundy (Haussmann and Mauck 2008; Mauck et al. 2012; Mauck et al. 2004; Ricklefs et al. 1985).

Materials and methods

Sampling methods

We characterized Mhc class II B loci using 46 reproductively active adults and two unrelated nestlings, all captured by hand from nest burrows on Kent Island. Blood samples were taken in 2010 and 2013, by brachial venipuncture. For PCR of genomic DNA, blood was either stored whole in lysis buffer at ambient temperature or was separated into plasma and packed cells and then frozen at −20 °C. DNA was extracted with Qiagen DNEasy kits or alcohol precipitation. For RNA analysis, ~150 μl whole blood was suspended in 500 μl RNA*later* buffer (Qiagen), stored <30 min at 4 °C, then frozen at −20 °C at the Bowdoin Scientific Station and later stored at −80 °C in the lab at Bates College.

Initial Mhc exploration and primer development

We used GenBank Mhc sequences and previously published primers from a variety of species (Aguilar et al. 2006; Alcaide et al. 2007; Canal et al. 2010; Ekblom et al. 2003; Juola and Dearborn 2012; Miller and Lambert 2004) to develop and screen preliminary primers for Leach's storm-petrels. When examining initial Mhc class II B sequences for Leach's storm-petrels that ranged from approximately the end of exon 1 to the start of exon 3 (Online Resource 1), we found conserved regions primarily in introns 1 and 2. In intron 1 in particular, there was a dichotomy of pronounced, fixed differences suggestive of two different loci. We designed primers for these

areas with the goal of developing a nested PCR that would completely amplify exon 2 in a locus-specific manner while providing a short enough amplicon for sequencing with 2× 250 paired-end reads on the Illumina MiSeq platform.

Our resulting nested PCR protocol (see below for reaction details) used locus-specific outer primers that amplified 836 bp of locus 1 and 747 bp of locus 2, followed by inner primers that annealed in the introns immediately flanking exon 2 (Table 1). Because of a fully conserved sequence in intron 2, we used the same reverse inner primer for both loci, with locus specificity being provided by the forward inner primer and by both of the outer primers. This nested PCR yielded locus-specific amplicons that were short enough to be sequenced with some overlap of the 2×250 paired-end reads on the Illumina MiSeq platform.

Exon 2 amplification and Illumina sequencing

For Mhc genotyping, we used this locus-specific, nested PCR (Table 1) to amplify the complete exon 2 of two Mhc class II B genes. The inner PCR for each individual bird was conducted with a unique combination of an 8-nt barcode attached to the forward and reverse primers, to allow subsequent demultiplexing of the sequence data. The barcode sequences were designed so that any two barcodes differed in at least three positions, minimizing the chance that a sequencing error would cause an amplicon sequence to be associated with the incorrect bird. The barcoded primers were tailed at the 5′ end with a string of five random nucleotides to generate initial sequence diversity and thereby reduce the need for PhiX spike-in during Illumina sequencing. Thus, the composite inner PCR primers were 5′–NNNNN-8 nt barcode-19 nt template complement–3′. Including this composite primer, the inner PCR amplicon size was 353 bp for locus 1 and 395 bp for locus 2, spanning all 270 bp of exon 2.

Reaction components were 1.5 μl 10× GeneAmp Gold Buffer, 1.5 μl 25 mM $MgCl_2$, 1.5 μl 2 mM dNTP, 0.5 μl 10 μM primer F, 0.5 μl 10 μM primer R, 8.925 μl water, 0.075 μl AmpliTaq Gold DNA Polymerase, and 0.5 μl template DNA. Template for outer PCR was 20 ng/μl genomic DNA. Template for inner PCR was the undiluted product of the outer PCR. For outer and inner PCR reactions at both loci, the cycling parameters were 95 °C for 10 min, followed by 25 cycles of 95 °C for 30 s, T_A for 45 s, and 72 °C for 60 s, ending with 72 °C for 7 min and a 12 °C hold. Annealing temperature was 62 °C for locus 1 outer and inner, 61 °C for locus 2 outer, and 59 °C for locus 2 inner.

We took precautions to minimize the risk of chimera formation (Lenz and Becker 2008) by using a minimum number of PCR cycles in both the outer and inner PCR reactions, a long extension time to avoid incomplete synthesis and subsequent heteroduplex formation, and a hot-start polymerase that lacks proofreading capability because some studies have found that proofreading polymerase enzymes increase the risk of chimeras (Judo et al. 1998; Zylstra et al. 1998). We also repeat-genotyped 39 individuals by independent amplification and sequencing.

An aliquot of each reaction was visualized with agarose electrophoresis to assess the presence and relative concentration of the appropriately sized amplicon. Amplicon volumes were adjusted prior to pooling in an effort to obtain similar read depth for all birds during Illumina sequencing.

A PCR-free Illumina sequencing library was prepared from 1 μg of pooled PCR amplicons using the KAPA HTP Library Preparation Kit (Kapa Biosystems, Wilmington, MA) and an Illumina TruSeq style adapter. The “no size selection” protocol was used for the post-adapter ligation cleanup.

Quality control was performed on the resulting library using an Agilent Technologies 2100 Bioanalyzer to characterize DNA fragment size and concentration. The concentration of DNA fragments with the correct adapters on both sides was determined using a quantitative PCR strategy, following the kit manufacturer’s protocol (Kapa Biosystems, Wilmington, MA), and the library was subjected to 2×250 bp paired-end sequencing on the Illumina MiSeq.

To assess the quality of the genotyping procedure, we repeat-genotyped 39 birds at both loci by independent PCR with a new pair of barcodes. Three negative controls were also subjected to amplification and sequencing protocols for each locus. Each negative control contained all PCR reagents

Table 1 Primers used in nested PCR for locus 1 and locus 2

Primer	Sequence 5′ to 3′	Location	T_M (°C)
Locus 1 outer F	CGTGGGGAGGAGACCTCA	Exon 1, 293 bp upstream from exon 2	64.5
Locus 1 outer R	CTCCACCTCGGGTTGAACT	Spans boundary of intron 2 and exon 3	62.3
Locus 1 inner F	AGCCCTGACCTCTCCATGT	Intron 1, 32 bp upstream from exon 2	62.3
Loci 1 and 2 inner R	AGGGAAATGCTCTGCCAAG	Intron 2, 26 bp downstream from exon 2	60.2
Locus 2 outer F	GTCCCCAGGCTGAGAAATTG	Intron 1, 205 bp upstream from exon 2	62.4
Locus 2 outer R	CCTCACCTTGGGCTGAACT	Spans boundary of intron 2 and exon 3	62.3
Locus 2 inner F	GTGCTGAGAGCACCTTGAG	Intron 1, 64 bp upstream from exon 2	62.3

except template, and each included its own unique pair of barcoded primers to allow any contaminant amplicons to be sequenced and recognized.

Processing of DNA sequence data

Prior to demultiplexing of the Illumina sequence data and assembling the two ends of the paired-end reads, sequences were trimmed to a length of 200 bp using the FASTX package (http://hannonlab.cshl.edu/fastx_toolkit/) based on the average quality curve: sequences averaged Q25 at 300 bp and Q30 at 200 bp. As the sequence data included both forward and reserve primer sequences for each amplicon, custom software developed by AGM sorted the data by amplicon and determined the barcode embedded in both the forward and the reverse primers to demultiplex the data by both individual bird and MHC II B locus, discarding any amplicons that could not be resolved due to missing primer sequence or unrecognized barcode. The resulting sample of amplicon sequences was then collated into a catalog of unique amplicon sequences for each locus, flagging those containing sequence ambiguities (i.e., sequencing error) or those with read length less than 200 bp for either primer (i.e., possible PCR contamination). To recover the complete sequence of each amplicon, the trimmed forward and reverse 200 bp from the paired-end reads were then assembled, with 111-bp overlap for locus 1 and 73-bp overlap for locus 2 (all with perfect consensus in the overlapping sequence).

As allele coverage within bird and locus should vary since the sequencing results are the product of a pooling of many PCR products, using a coverage cutoff to identify real alleles (versus sequencing noise) was not feasible as coverage of individual birds for individual loci varied widely. Instead, when predicting alleles, we examined the most abundant amplicon sequences (i.e., peaks) within a bird/locus: There should be one (homozygous) or two (heterozygous) amplicons with the highest abundance (i.e., coverage), the remainder being sequencing error and low abundance amplicons. Using a custom genotyping algorithm developed by AGM that examined peak abundance relative to sequencing error, genotype was predicted for all birds/loci for those having at least one amplicon with coverage greater than 10 (based on examination of amplicon frequencies in the negative controls). Any sample with all amplicons with coverage of 10 or less was manually inspected to determine genotype, if possible. Similarly, any sample unresolved by the genotyping algorithm was manually inspected to determine genotype. Lastly, a random sample of 20 algorithm-determined genotypes had their sequencing peaks manually inspected for confirmation.

Sequences were trimmed to exon 2 based on alignment with Mhc sequences of seabirds in the orders Procellariiformes, Sphenisciformes, and Charadriiformes: blue petrel, *Halobaena caerulea*, JF276893.1; thin-billed prion, *Pachyptila belcheri*, FJ588549.1; Galápagos penguin, *Spheniscus mendiculus*, AB302191.1; little penguin, *Eudyptula minor*, AB302093.2; Magellanic penguin, *Spheniscus magellanicus*, AB303945.1; Humboldt penguin, *Spheniscus humboldti*, AB302087.1; African penguin, *Spheniscus demersus*, AB301944.1; Atlantic puffin, *Fratercula arctica*, HQ822509.1; common murre, *Uria aalge*, EU326276.1; razorbill , *Alca torda*, HQ822527.1; and black-legged kittiwake, *Rissa tridactyla*, HQ822434.1. Putative peptide binding residues of exon 2 at both loci were determined by comparison with human HLA molecules (Brown et al. 1993) and with similarly aligned sequences from blue petrel (Strandh et al. 2011; JF276893.1).

RNA extraction and sequencing

To determine whether these loci are expressed, we isolated and sequenced messenger RNA (mRNA) from three individuals—two adults and one nestling. Whole blood that had been preserved in RNA*later* (Qiagen, Valencia, CA) was thawed, combined with RNA Stat-60 (Tel-Test, Friendswood, TX), physically homogenized, and mixed with chloroform. Purified RNA was isolated from the resulting supernatant with an RNA Clean & Concentrator-5 kit (Zymo Research, Irvine, CA) according to the manufacturer's protocol. We digested any potential gDNA in the extracted RNA with TURBO DNase (Ambion/Life Technologies, Grand Island, NY).

Because the PCR primers used to amplify gDNA were partially or entirely located in introns, we used our gDNA sequences to design new exon 2 primers for RNA amplification. Conserved regions in exon 2 were uncommon and were not amenable to primer design; therefore, we designed primers that were allele-specific (Online Resource 2). Total mRNA was reverse-transcribed to complementary DNA (cDNA) and then selectively amplified using the Access RT-PCR kit (Promega, Madison, WI), with the following reaction conditions: 10 μl 5× *AMV/Tfl* buffer, 5 μl 2 mM dNTP, 5 μl 10 μM primer F, 5 μl 10 μM primer R, 2 μl 25 mM $MgSO_4$, 16 μl water, 1 μl *AMV* reverse transcriptase, 1 μl *Tfl* polymerase, and 5 μl mRNA. Cycling parameters were 45 °C for 45 min, 94 °C for 2 min, followed by 45 cycles of 94 °C for 30 s, T_A for 30 s, and 72 °C for 30 s, ending with 72 °C for 7 min and a 4 °C hold. Depending on the allele-specific primer, T_A was 65 or 67 °C, and the total region sequenced per allele was 224 to 246 bp of the 270-bp exon 2 (Online Resource 2).

Amplicons were gel-purified with a Zymoclean Gel DNA Recovery kit (Zymo Research, Irvine, CA) and then Sanger-sequenced in both directions by Operon SimpleSeq (Operon/Eurofins Genomics, Huntsville, AL). Cloning was not necessary because primers were allele specific. Consensus of

forward and reverse sequences of each allele from each bird was compared to gDNA genotype sequences from the Illumina MiSeq run.

As a last step, we sought confirmation that our cDNA sequences were truly derived from mRNA rather than from any possible gDNA contamination that had somehow escaped the DNase treatment. Toward this end, we attempted to use intron-based primers (Table 1) to amplify locus 2 via the same nested PCR reaction that had successfully generated the DNA amplicons for Illumina genotyping. If our mRNA/cDNA sequences are valid, this intron-based PCR should fail. As a positive control, we simultaneously ran the PCR reaction on gDNA samples from the same three birds.

Characterizing allelic diversity

We used Sequencher 5.2.2 (Gene Codes Corporation, Ann Arbor, MI) to edit and align nucleotide sequences, Clustal2W (Larkin et al. 2007) to align and compare amino acid sequences, and DNAsp (Rozas et al. 2003) to measure nucleotide diversity. To view the relationships among alleles within a locus while simultaneously exploring allele frequencies, we built haplotype networks for each locus with Haploviewer (Salzburger et al. 2011) based on the dnapars module in Phylip 3.695 (Felsenstein 2005). We also wished to confirm that allele frequencies were not markedly different between males and females. Because sexes of this species are not distinguishable by plumage or external morphology, we used a PCR-based sexing technique that amplifies differentially sized introns in the sex-linked CHD1W and CHD1Z genes with primers 2550F and 2718R (Fridolfsson and Ellegren 1999).

To assess whether alleles of the two loci are reciprocally monophyletic or, instead, are phylogenetically intermingled as would be expected under a scenario of extensive gene conversion or convergent evolution, we built a maximum likelihood phylogeny of all alleles from both loci with MEGA 6.0.2 (Tamura et al. 2007) using a GTR+I+G model with 1000 bootstrap replicates.

We looked for additional evidence of gene conversion or recombination with GENECONV 1.81 (Sawyer 1999). GENECONV uses patterns of variation between aligned sequences to find pairs of alleles with longer-than-expected regions of sequence identity. We repeated this analysis using a range of gscale values that influence the extent to which sequences must be identical. Specifically, we tested for gene conversion under gscale values of 0 (no mismatches allowed), 1 (small mismatch penalty), 2, and 3 (large mismatch penalty). Allowing no mismatches means searching only for fully identical sequences, which might represent recent gene conversion events, while the use of non-zero gscale values allows the detection of sequences that could be the result of older gene conversion events followed by subsequent point mutations.

Testing for positive selection

We tested for evidence of historical positive selection with MEGA 6.0.2 (Tamura et al. 2007). The difference in rates of non-synonymous and synonymous substitutions ($d_N - d_S$) was assessed with a Z-test using the modified Nei-Gojobori method with pairwise deletion of gaps, a transition/transversion bias of 2, and 1000 bootstraps for estimating variance.

Results

Illumina sequencing quality and data validation

The Mhc sequences presented here were obtained as part of a larger Illumina sequencing run. In composite, that run totaled 30,232,256 sequencing reads. The valid, barcoded sequences resolved to 9,241,320 amplicons (i.e., forward+reverse sequence with a diagnostic barcode combination) with detection of amplicons for both Mhc II B loci from all storm-petrel DNA samples. Resolved alleles had average amplicon coverage of 1921 and 1218 for locus 1 and locus 2, respectively. For locus 1, only one resolved allele among all birds had less than 50-fold coverage, while nine resolved alleles among all birds had less than 50-fold coverage for locus 2. All of these low-coverage alleles received manual inspection. None of the negative controls yielded amplicons that were visible on agarose gels and all negative control sequencing peaks had coverage ≤5 (mean coverage of most abundant two peaks for the six negative controls was 1.42).

A manual inspection of a random sample of 20 algorithm-predicted genotypes did not find any erroneous results. For additional assessment of genotyping validity, we repeat-genotyped 39 birds at both loci by independent PCR with different barcodes. Of those 78 single-locus genotypes which were run in duplicate, 77 (98.7 %) produced identical results. In the remaining case, one of the duplicate genotypes had a peak of sequencing error as its second most abundant "allele" at locus 1, causing a false homozygote call; the true genotype, captured by the repeat sample, was L1A4/L1A60.

For each locus-specific primer set, all birds yielded Illumina genotypes that consisted of either one or two alleles, as expected if we were separately amplifying two loci.

Gene expression

We reverse-transcribed mRNA and amplified and sequenced the resulting cDNA from both loci for three individuals. For all three birds, the mRNA sequences showed complete consensus with their DNA genotypes: L1A4/L1A4, L2A50/L2A50 for the first bird; L1A28/L1A28, L2A50/L2A74 for the second bird; and L1A4/L1A4, L2A131/L2A50 for the

Table 2 Exon 2 nucleotide sequences of locus 1 and locus 2

```
   LOCUS 1
Allele   Freq        ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|70
 L1A4     0.427     GGTTTTTCCA GGATATGTTT AAGGCCGAAT GTTACTTCAC CAACGGCACC GAGCGGGTGA GGCTTCTGGC
 L1A28    0.177     ...A...... ...A...... .......... .......... .......... .......... ..T......A
 L1A55    0.167     .......... ...A...... .......... .......... .......... .......... ..T......A
 L1A80    0.052     ...A...... ...A...... .......... .......... .......... .......... ..T......A
 L1A113   0.052     ...A...... ...A...... .......... .......... .......... A......... ..T......A
 L1A90    0.042     ...A...... .......... .......... .......... .......... .......... ..T......A
 L1A79    0.031     ...A...... ...A...... ...AG..... .......... .......... ....A..... ...A.G..A.
 L1A60    0.031     ...A...... ...A...... .......... .......... .......... .......... ..T......A
 L1A149   0.021     .......... .......... .......... .......... .......... .......... ..........

           ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|150
 L1A4     GAGGTACATC TACAACCGGC AGCAGGACGT GCACTTCGAC AGCGATGTGG GGTTCTTTGT GGCTGACACC CCCCTGGGCG
 L1A28    C......... .......... .....CG..C .......... .......... ......A... .......... ..........
 L1A55    C......... .......... .....CG..C .......... .......... ......A... ...C...... ..........
 L1A80    C......... .......... .....TT... .......... .......... ......A... ...C...... ..........
 L1A113   C......... .......... .....CG..C .......... .......... ......A... .......... ..........
 L1A90    C......... .......... .....TT... .......... .......... ......A... ...C...... ..........
 L1A79    .......... C......... .....TT... .......... .......... ......A... .......... ..........
 L1A60    C......... .......... .....CG..C .......... .......... ......A... .......... ..........
 L1A149   .......... .......... .......... .......... .......... .......... .......... ..........

           ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|230
 L1A4     AGCCTGATGC CAAGTACTGG AACAGCCAGC CGGACCTCCT GGAGGACAGA CGGGCTTCGG TGGACACCTT CTGCCGGCAC
 L1A28    .....AG... .......... .......... .....A.A.. ....CGG.A. .......... .......... ..........
 L1A55    .......... .......... .......... .....A.A.. ......TGA. ......G... .......... ..........
 L1A80    .....AG... .......... .......... .....A.A.. ....CGG.A. .......... .......... ..........
 L1A113   .....AG... .......... .......... .....A.A.. ....CGG.A. .......... .......... ..........
 L1A90    .......... .......... .......... .....A.A.. ......TGA. ......G... .......... ..........
 L1A79    .....AT... ....G..... .......... ...C.A.A.. ....C.G.C. ......GA.A .......... ..........
 L1A60    .....AG... .......... .......... .....A.A.. ....CGG.A. .......... ........-- -.........
 L1A149   .....AT... ....G..... .......... ...C.A.A.. ....C.G.C. ......GA.A .......... ..........

           ....|....| ....|....| ....|....| ....|....|270
 L1A4     AACTACGGGG TGTGGACCCC TTTCACCGTG GAGAGGAGGG
 L1A28    .......... .......... .......... .......A..
 L1A55    .......... .......... .......... ..........
 L1A80    .......... .......... .......... .......A..
 L1A113   .......... .......... .......... .......A..
 L1A90    .......... .......... .......... .......A..
 L1A79    .......... .......... .......... ..........
 L1A60    .......... .......... .......... .......A..
 L1A149   .......... .......... .......... ..........

   LOCUS 2
Allele   Freq        ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|70
 L2A50    0.479     GGTTTTTCCA GTGGATAGGA AAGGCTGAGT GTCAGTACCT CAACGGCACC GAGCGGGTGA GGCTTCTGGT
 L2A54    0.156     ..G....... .C.A..GTTT .....C.... .......... .......... .......... .........C
 L2A131   0.125     ..G....... .C.A..GTTT .....C.... .......... .......... .......... ..TA.....C
 L2A74    0.052     .A........ .GAA..GTTT .....C.... .......... .......... .......... .........C
 L2A193   0.042     ..G....... .C.A..GTTT .....C.... .......... .......... .......... .........C
 L2A539   0.031     .......... .......... .......... .......... .......... .......... ..........
 L2A176   0.031     ..G....... .GA...G... .......... ....T...T. .......... .......... .........C
 L2A46    0.021     ..G....... .......... .......... .......... .......... .......... ..........
 L2A1158  0.021     ..G....... .GA...G... .......... ....T...T. .........G .......... ..T......A
 L2A249   0.021     .A........ .GAA..GTTT .....C.... .......... .......... .......... .........C
 L2A791   0.010     .A........ .GAA..GTTT .....C.... .......... .......... .......... .........C
 L2A1553  0.010     .......... .......... .......... .......... .......... .......... ..TA.....A

           ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|150
 L2A50    GAGGTACATC CACAACCGGC AGCAGTTCGT GCACTTCGAC AGCGATGTGG GGTTCTATGT GGCCGACACC CCCCTGGGCG
 L2A54    .....C.... T......... .......... .......... .......... .......... .......... ..........
 L2A131   .....C.... T......... .......... .......... .......... .......... .......... ..........
 L2A74    .....C.... T......... .......... .......... .......... .......... ...T...... ..........
 L2A193   .....C.... T......... .......... .......... .......... .......... .......... ..........
 L2A539   .......... .......... .......... .......... .......... ......T... .......... ..........
 L2A176   .....C.... T......... .......... .......... .......... .......... .......... ..........
 L2A46    .......... .......... .....GA... .......... .......... ......T... .......... ..........
 L2A1158  C......... T......... .....GA... .......... .......... .......... ...T...... ..........
 L2A249   .....C.... T......... .......... .......... .......... .......... .......... ..........
 L2A791   .....C.... T......... .......... .......... .......... .......... ...T...... ..........
 L2A1553  A......... .......... .......... .......... .......... .......... .......... ..........

           ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|230
 L2A50    AGCCTGATGC CAAGTACTGG AACAGCCAGC CGGACCTCCT GGAGCAGAGA CGGGCTGAGG TGGACACCTA CTGCCGACAC
 L2A54    .......... .......... .......... .......... .......... .......... .......... ..........
 L2A131   .....AT... ....G..... .......... .....A.A.. .......... .......C.. .......... .....G...
 L2A74    .....AT... ....G..... .......... .......... .......... ......TC.. .........-- -.....G...
 L2A193   .....AT... ....G..... .......... .....A.A.. .......... .......C.. .......... .....G...
 L2A539   .......... .......... .......... .....A.A.. .........A .......... .........- -.....G...
 L2A176   .....AT... ....G..... .......... .....A.A.. .......... .......C.. .......... .....G...
 L2A46    .....AT... ....G..... .......... .....A.A.. .......... .......... .......... .....G...
 L2A1158  .......... .......... .......... .....A.A.. .........A .......... .........- -.....G...
 L2A249   .......... .......... .......... .......... .......... .......... .......... ..........
 L2A791   .....AT... ....G..... .......... .......... .......... .......... .......... ..........
 L2A1553  .....AT... ....G..... .......... .....A.A.. .......... .......... .......... .....G...

           ....|....| ....|....| ....|....| ....|....|270
 L2A50    AACTACGGGG TGTCGACCCC TTTCATCGTG GAGAGGAGGG
 L2A54    .......... .......... .......... ..........
 L2A131   .......... ..GT...... .....C.... ..........
 L2A74    .......... .......... .......... .......A..
 L2A193   .......... ..G....... .....C.... ..........
 L2A539   .......... ..G....... .....C.... ..........
 L2A176   .......... ..GT...... .....C.... ..........
 L2A46    .......... ..GT...... .....C.... ..........
 L2A1158  .......... ..G....... .....C.... ..........
 L2A249   .......... .......... .......... ..........
 L2A791   .......... .......... .......... ..........
 L2A1553  .......... ..GT...... .....C.... ..........
```

Note that in both loci, there are alleles with a 3-bp deletion at sites 219–221. Reading frame is +3

Table 3 Amino acid alignment of translated exon 2 sequences of locus 1 and locus 2

```
   LOCUS 1            5    10   15   20   25   30   35   40   45   50   55   60   65   70   75   80   85
Allele  Freq     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....
 L1A4    0.427   FFQDMFKAECYFTNGTERVRLLARYIYNRQQDVHFDSDVGFFVADTPLGEPDAKYWNSQPDLLEDRRASVDTFCRHNYGVWTPFTVERR
 L1A28   0.177   Y..E..............F.D........RA........Y.........S.........I..RK.....................K
 L1A55   0.167   ...E..............F.D........RA........Y.....................I...E..A...................
 L1A80   0.052   Y..E..............F.D........F.........Y.........S.........I..RK.....................K
 L1A113  0.052   Y..E...........K...F.D........RA........Y.........S.........I..RK.....................K
 L1A90   0.042   Y.................F.D........F.........Y.....................I...E..A...................K
 L1A79   0.031   Y..E...S.........Q..HVT...H....F.........Y.........I..D.....AI..QT..EM.................
 L1A60   0.031   Y..E..............F.D.........RA........Y.........S.........I..RK......-..............K
 L1A149  0.021   ...D..............................................I..D.....AI..QT..EM.................

   LOCUS 2            5    10   15   20   25   30   35   40   45   50   55   60   65   70   75   80   85
Allele  Freq     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....
 L2A50   0.479   FFQWIGKAECQYLNGTERVRLLVRYIHNRQQFVHFDSDVGFYVADTPLGEPDAKYWNSQPDLLEQRRAEVDTYCRHNYGVSTPFIVERR
 L2A54   0.156   V..RMF......L.........A.S.Y.....................................................................
 L2A131  0.125   V..RMF......L.......Y.A.S.Y.....................I..D.....I.....A.....................V...T....
 L2A74   0.052   ...EMF......L.........A.S.Y.....................I..D.....I.....S...-.....................K
 L2A193  0.042   V..RMF......L.........A.S.Y.....................I..D.....I.....A.....................A...T....
 L2A539  0.031   ............L.....................F.....................I...K.....-.......A...T....
 L2A176  0.031   V..EM.....H.F.........A.S.Y.....................I..D.....I.....A.....................V...T....
 L2A46   0.021   V...........L.....................D.........F.........I..D.....I.....................V...T....
 L2A1158 0.021   V..EM.....H.F.......F.D...Y....D.....................................I...K.....-.......A...T....
 L2A249  0.021   ...EMF......L.........A.S.Y.....................................................................
 L2A791  0.010   ...EMF......L.........A.S.Y.....................I..D.....................................
 L2A1553 0.010   ............L.......Y.E.....................I..D.....I.....................V...T....
```

Note that in both loci, there are alleles with a 3-bp deletion at codon 73. Putative peptide binding sites are shaded in gray

third bird (sequence data in Online Resource 3). Confirming that the cDNA sequences were truly derived from mRNA and not from gDNA contamination, PCR with these samples failed when using intron-based primers.

As expected of genes that are expressed and functionally important, all alleles found in the Illumina genotyping of gDNA (see below) had an open reading frame throughout exon 2, and the only indel was in-frame (Tables 2 and 3).

Genetic diversity and relationship among alleles

Among the 48 birds genotyped at exon 2 of both loci, we found 9 alleles at locus 1 (GenBank KP090142-KP090150) and 12 alleles at locus 2 (GenBank KP090151-KP090162), all of which produced unique amino acid sequences (Tables 2, 3, and 4). All but two alleles were found in multiple individuals, and the two unique alleles were confirmed in duplicate PCR and sequencing. Genotypes of all birds are listed in Online Resource 4.

Alleles ranged in frequency from 0.010 to 0.479, and both loci had several alleles that were fairly common (Fig. 1 and Table 2). Amplification of CHD1W and CHD1Z genes revealed that our sample comprised 25 males and 23 females, and there were no marked differences in allele frequencies between the sexes (Fig. 1).

A maximum likelihood tree of all alleles combined showed that each locus was monophyletic (Fig. 2). The

Table 4 Diversity at loci 1 and 2 (n=48 birds) and tests for positive selection

Region	S, no. (%) of sites polymorphic	Nucleotide diversity π±SD	No. of nucleotide haplotypes	No. of amino acid haplotypes	d_N-d_S	p value[a] against H_0: $d_N=d_S$
Locus 1						
PBS only (99 bp)	26 (26.3)	0.10014±0.00382	8	8	3.749	0.0001
Non-PBS only (171 bp)	7 (4.1)	0.00983±0.00075	7	5	−0.620	1.0000
All of exon 2 (270 bp)	33 (12.2)	0.04219±0.00179	9	9	3.145	0.0010
Locus 2						
PBS only (99 bp)	27 (27.3)	0.09785±0.00595	12	12	1.759	0.0406
Non-PBS only (171 bp)	9 (5.3)	0.01218±0.00112	10	6	−0.476	1.0000
All of exon 2 (270 bp)	36 (13.3)	0.04298±0.00266	12	12	1.249	0.1069

[a] p value is from a test for positive selection (H_0: d_N=d_S vs H_A: d_N>d_S) in MEGA 6.0.2, using a modified Nei-Gojobori method with 1000 bootstraps, transition/transversion bias=2, and pairwise deletion of gaps

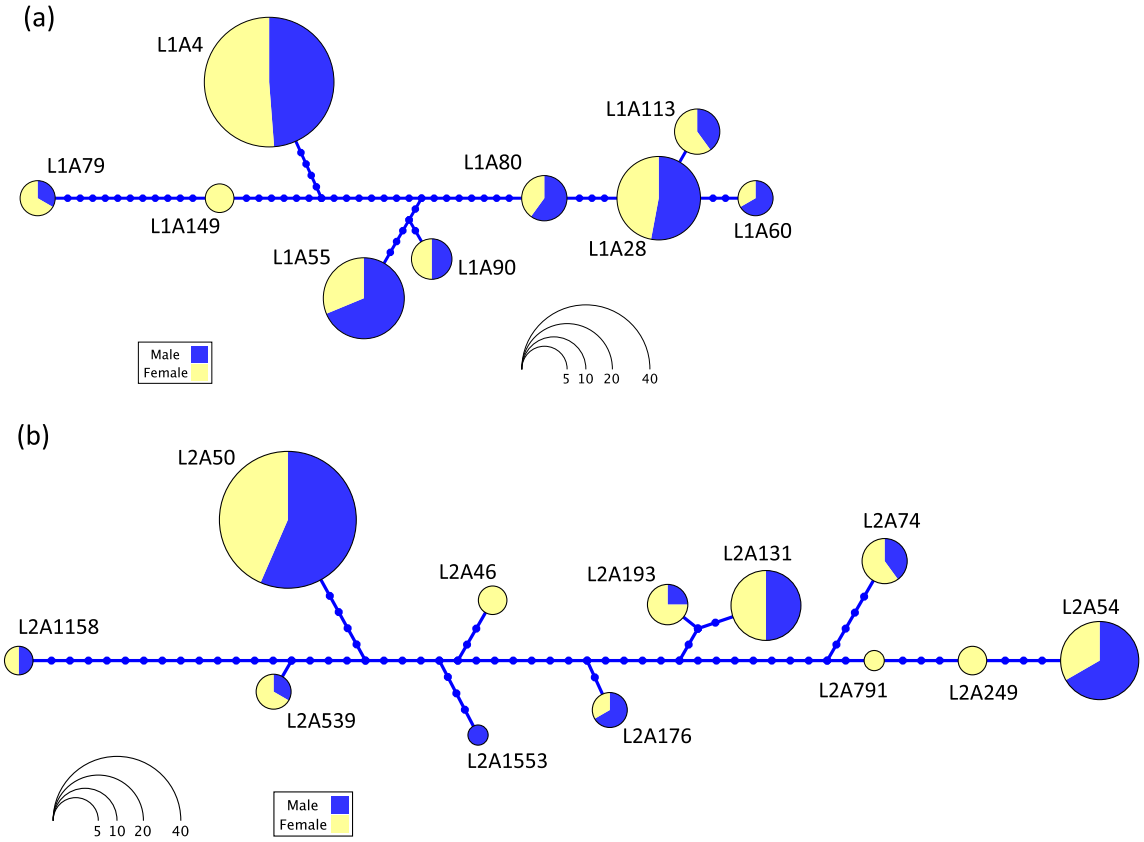


Fig. 1 Parsimony-based haplotype network for 270-bp nucleotide sequences of exon 2 of two Mhc class II B genes: **a** nine alleles of locus 1, and **b** 12 alleles of locus 2. Size of *circles* is proportional to number of copies of each haplotype in a sample of 48 diploid animals (i.e., 96 alleles total)

tree in Fig. 2 was based on all 270 nucleotides of exon 2, but similar results were obtained when using only third-position bases or only first- and second-position bases, and also when using methods other than maximum likelihood (Online Resource 5).

GENECONV found evidence for up to four cases of a fragment shared between two alleles, depending on the gscale value used to penalize mismatches (Table 5); the shared sequences were most frequently detected under a model that allows no mismatches, consistent with recent enough gene conversion events that no subsequent point mutations have occurred. Two of the putative gene conversion events were intralocus and two were interlocus. All shared sequences were located near the start of the exon, and the two interlocus shared fragments were short relative to the intralocus shared fragments and overlapped only 4 of 33 codons for putative peptide binding sites.

Evidence for selection

At both loci, there was evidence for positive selection at the peptide binding sites, with $d_N - d_S > 0$, but not at the non-peptide binding sites (Table 4).

Discussion

We found clear evidence for a minimum of two expressed Mhc class II B loci in Leach's storm-petrel, with moderate allelic diversity, monophyletic clustering of alleles by locus, and a signature of positive selection on peptide binding sites.

Because of varying histories of gene duplication, the number of Mhc loci appears to differ widely between taxa, ranging from only 1 (Hughes et al. 2008; Zeisset and Beebee 2009) to perhaps 20 or more (Bollmer et al. 2010), although the methods used in some studies provide only a minimum estimate of loci rather than an absolute number. Among bird species, current data suggest the possibility of only one locus in several taxonomic groups: Psittaciformes (Hughes et al. 2008), Ciconiiformes (Zhang et al. 2006), Falconiformes (Alcaide et al. 2008), and Sphenisciformes (Bollmer et al. 2007; Kikkawa et al. 2009). Other avian orders have more loci, but the limited sampling coverage means that we should generalize with caution. There is documentation of at least two loci in a member of Pelecaniformes (Juola and Dearborn 2012) and at least two to four loci in Charadriiformes (Ekblom et al. 2003; Walsh and Friesen 2003), in a member of Gruiformes (Alcaide et al. 2014), and in most Galliformes

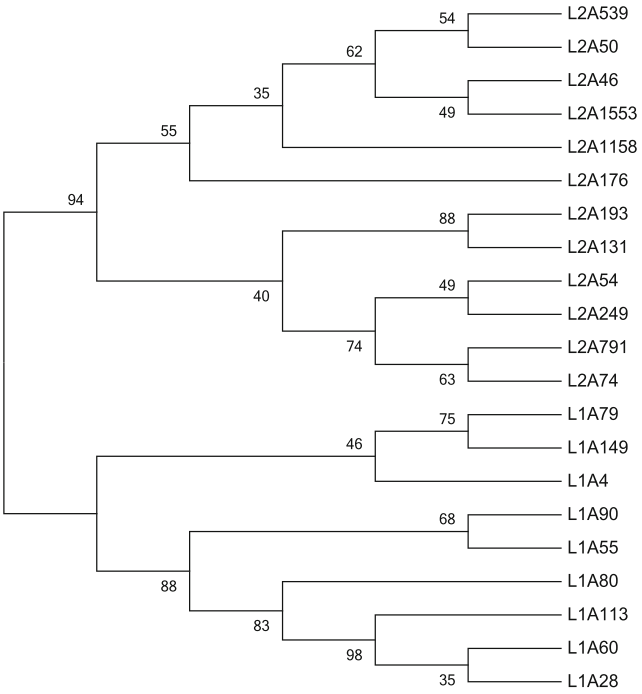


Fig. 2 Bootstrap consensus tree of a maximum likelihood analysis of all alleles from both loci, implemented in MEGA 6.0.2 (Tamura et al. 2007) using a GTR+I+G model with 1000 bootstraps. Percentage bootstrap support is shown for each node. See Online Resource 5 for trees built with other methods

(Jacob et al. 2000; Piertney 2003; Strand et al. 2007; Wittzell et al. 1999). In contrast, there seems to be a highly variable but much larger number of loci in songbirds, the Passeriformes (Aguilar et al. 2006; Bollmer et al. 2010; Sato et al. 2001; Westerdahl et al. 2000; Zagalska-Neubauer et al. 2010). The detection of two loci in our study is similar to what has been seen in other Procellariiformes (one or two loci (Silva and Edwards 2009; Strandh et al. 2011)), but we emphasize that additional loci might exist that were not amplified by the PCR primers we used. A further caveat about cross-species comparisons is that many studies have been forced to infer the number of loci based on the maximum number of alleles amplified in one individual by a primer pair that was not locus specific. In our study, the use of locus-specific primers provides strong confirmation that the allelic sequences in our analyses were drawn from exactly two loci.

A related issue is whether multiple Mhc II B loci in a species are functional. Copies of Mhc genes arise by gene duplication, after which they may or may not continue to be expressed (Bollmer et al. 2010; Edwards et al. 2000). In Leach's storm-petrel, we found clear evidence that both of these loci are expressed, based on sequencing of mRNA from three genotyped individuals. This RNA-based evidence for gene expression is consistent with additional indirect evidence: the absence of stop codons or frame-shift mutations in the alleles found in our sample. At both of our loci, we found only one type of indel: a 3-bp deletion beginning in the 219th nucleotide of exon 2. This deletion was found in one allele at locus 1 and three alleles at locus 2 and was the deletion of a complete codon at a non-peptide binding site. Overall, the nature of the evidence for gene expression at these two loci is suitable for determining whether the loci are expressed at all, but it cannot address whether the two loci are expressed equally. Expression levels can vary widely between duplicated Mhc genes, as has been shown, for example, in chickens at both class I (Wallny et al. 2006) and class II loci (Jacob et al. 2000; Wallny et al. 2006). It would not be surprising if the two storm-petrel loci in our study are expressed at different levels, but this remains unknown, as does the relative importance of the two loci in pathogen defense.

In total, our sample of 48 birds exhibited nine alleles at locus 1 and 12 alleles at locus 2, all of which were verified and all of which translated to unique amino acid sequences. Compared to some studies (Alcaide et al. 2008; Alcaide et al. 2014; Juola and Dearborn 2012), this is slightly low allelic diversity, particularly given that the most common three alleles at locus 1 and locus 2 in our sample had combined frequencies of 0.77 and 0.76, respectively. The observed uneven frequency distribution of alleles is not unusual under the complex range of balancing selection scenarios that are possible for Mhc (Ejsmond et al. 2010), but the low number of

Table 5 Putative gene conversion events identified by GENECONV

Type of shared fragment	Pairs of alleles	Starting position[a]	Length (bp)	gscale value[b]	Sim *p* value[c]
Within-locus	L1A4 and L1A149	1	155	0, 2, 3	0.0130
	L2A249 and L2A193	14	142	0	0.0439
Between-loci	L1A55 and L2A791	3	26	0, 2, 3	0.0341
	L1A55 and L2A74	3	26	0, 2, 3	0.0341
Outside[d]	None		–	None	0.3190

[a] Relative to start of exon 2

[b] Values of gscale (0, 1, 2, 3) at which there was significant evidence for this particular gene conversion event

[c] Global *p* values from analyses with gscale=0; *p* values are simulated via 10,000 permutations and are adjusted for multiple comparisons

[d] Gene conversion between an allele in the data set and an unknown allele not in the data set

alleles in our sample surprised us somewhat. Although Kent Island is small, at 80 ha, two factors led us to expect higher genetic diversity: The breeding population of the island is estimated at 15,000 pairs (L. Minich and R. Mauck, unpublished), and natal dispersal precedes the first breeding effort (Bicknell et al. 2014; Mauck and Huntington, unpublished data) resulting in extensive gene flow within the Atlantic and Pacific population clusters (Bicknell et al. 2012).

At the scale of genes, Mhc diversity is shaped by the gains of gene duplication and the losses of inactivation or deletion (i.e., a birth-and-death model) (Nei and Rooney 2005; Sato et al. 2001). At the scale of alleles, Mhc diversity can be shaped by an array of processes, including point mutations and gene conversion (Hess and Edwards 2002). In phylogenetic analyses of all exon 2 sequences in our study, alleles clustered by locus (Fig. 2). Trees built with other methods showed the same fundamental pattern of both loci being monophyletic (Online Resource 5). Some studies of other species have found a similar pattern (Hablützel et al. 2013; Michel et al. 2009; Worley et al. 2008), but others have found alleles of multiple Mhc class II loci within a species to be phylogenetically intermingled (Reusch and Langefors 2005; Strand et al. 2013), which is interpreted as evidence for concerted evolution via extensive interlocus recombination (Hess and Edwards 2002; Strand et al. 2013). The pattern of reciprocal monophyly in our study suggests that the gene duplication event giving rise to the two loci in Leach's storm-petrels may be a relatively old one that was followed predominantly by point mutations (and perhaps intralocus gene conversion) rather than by widespread interlocus gene conversion. This idea is consistent with the GENECONV results which indicated that exon 2 gene conversion events were few and that the two interlocus gene conversions involved a relatively short fragment. It is also interesting to note that all four fragments identified as being involved in gene conversion were detected under the GENECONV model of gscale=0 that disallows mismatches, suggesting that these gene conversion events may be recent (i.e., without time for subsequent point mutations) or that selection is imposing constraints on these particular shared fragments.

Although nothing is currently known about pathogen-mediated selection pressures in the Kent Island population, both loci showed evidence of a history of positive selection on the amino acids that bind to pathogen peptides. Specifically, there was a higher frequency of non-synonymous than synonymous substitutions at the putative peptide binding sites (PBSs; Table 4) but no difference between d_N and d_S at the non-PBS codons. This pattern has been seen in other taxa (Hughes and Nei 1989; Xi et al. 2014) and is consistent with the idea that pathogen-mediated balancing selection acts on the peptide binding sites while purifying selection acts on the remainder of the protein structure (Hughes and Nei 1989; Sommer 2005).

Moving forward, two population-level aspects of Mhc may be interesting in Leach's storm-petrels. First, our view of Mhc evolution in this system would be greatly enhanced by knowledge of pathogens along with a comparison of Mhc diversity and neutral genetic diversity across disjunct populations. Leach's storm-petrel has a very wide geographic range and there is evidence for gene flow, based on microsatellite markers and mtDNA control region (Bicknell et al. 2012). If regional variation in pathogens drives Mhc evolution, a geographic signal of Mhc variation would stand in stark contrast to existing data from non-Mhc sequences. Second, the relatively low allelic diversity in our population—along with a moderately high frequency of a handful of alleles—means that mate choice for Mhc complementarity could be important. This would be in contrast to a population with extraordinarily high Mhc diversity, in which individuals could mate at random and stand a good chance of producing heterozygous offspring. The allelic composition of our study population joins other factors that make Leach's storm-petrel a likely candidate for Mhc-based mate choice: long lifespan, long-term monogamy, and demonstrated olfactory capacity (Huntington et al. 1996).

In sum, we have characterized two Mhc class II B loci that are expressed, are reciprocally monophyletic, and have experienced positive selection at peptide binding sites. This sets the stage for future work on population-level questions about ecology and evolution of Mhc in this species.

Acknowledgments This is contribution number 251 from the Bowdoin Scientific Station on Kent Island and was supported by funds from the A.V. Davis Foundation. For logistical support or assistance in the lab or field, we thank Mary Barter, Lucy Rowe, Sandy Daigle, Larissa Williams, Haley Acker, Colin McMahon, Damon Gannon, Russell Ingalls, and Mark Murray, and we thank Frank Hailer and anonymous reviewers for feedback on the manuscript.

Ethical standards All animal work was conducted in accordance with permits from the Canadian Wildlife Service, Bowdoin College's Institutional Animal Care and Use Committee, and the USDA Animal and Plant Health Inspection Service.

Conflict of interest The authors declare that they have no conflict of interest.

References

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 Hered 97:133–142

Alcaide M, Edwards SV, Negro JJ (2007) Characterization, polymorphism, and evolution of MHC class II B genes in birds of prey. J Mol Evol 65:541–554

Alcaide M, Edwards SV, Negro JJ, Serrano D, Tella JL (2008) Extensive polymorphism and geographical variation at a positively selected

MHC class II B gene of the lesser kestrel (*Falco naumanni*). Mol Ecol 17:2652–2665

Alcaide M, Muñoz J, Martínez-de la Puente J, Soriguer R, Figuerola J (2014) Extraordinary MHC class II B diversity in a non-passerine, wild bird: the eurasian coot *Fulica atra* (Aves: Rallidae). Ecol Evol 4:688–698

Bahr A, Wilson AB (2012) The evolution of MHC diversity: evidence of intralocus gene conversion and recombination in a single-locus system. Gene 497:52–57

Balakrishnan CN, Ekblom R, Völker M, Westerdahl H, Godinez R, Kotkiewicz H, Burt DW, Graves T, Griffin DK, Warren WC, Edwards SV (2010) Gene duplication and fragmentation in the zebra finch major histocompatibility complex. BMC Biol 8:29

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

Bicknell AWJ, Knight ME, Bilton D, Reid JB, Burke T, Votier SC (2012) Population genetic structure and long-distance dispersal among seabird populations: Implications for colony persistence. Mol Ecol 21:2863–2876

Bicknell AWJ, Knight ME, Bilton DT, Campbell M, Reid JB, Newton J, Votier SC (2014) Intercolony movement of pre-breeding seabirds over oceanic scales: implications of cryptic age-classes for conservation and metapopulation dynamics. Divers Distrib 20:160–168

Boehm T, Zufall F (2006) MHC peptides and the sensory evaluation of genotype. Trends Neurosci 29:100–107

Bollmer JL, Vargas FH, Parker PG (2007) Low MHC variation in the endangered Galápagos penguin (*Spheniscus mendiculus*). Immunogenet 59:593–602

Bollmer JL, Dunn PO, Whittingham LA, Wimpee C (2010) Extensive MHC class II B gene duplication in a passerine, the common yellowthroat (*Geothlypis trichas*). J Hered 101:448–460

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. Nat 364:33–39

Canal D, Alcaide M, Anmarkrud JA, Potti J (2010) Towards the simplification of MHC typing protocols: targeting classical MHC class II genes in a passerine, the pied flycatcher Ficedula hypoleuca. BMC Res Notes 3:236

Carrington M, Nelson GW, Martin MP, Kissner T, Vlahov D, Goedert JJ, Kaslow R, Buchbinder S, Hoots K, O'Brien SJ (1999) HLA and HIV-1: Heterozygote advantage and B*35-Cw*04 disadvantage. Sci 283:1748–1752

Chaves LD, Faile GM, Krueth SB, Hendrickson JA, Reed KM (2010) Haplotype variation, recombination, and gene conversion within the turkey MHC-B locus. Immunogenet 62:465–477

Doherty PC, Zinkernagel RM (1975) Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. Nat 256:50–52

Edwards SV, Gasper J, Garrigan D, Martindale D, Koop BF (2000) A 39-kb sequence around a blackbird Mhc class II gene: ghost of selection past and songbird genome architecture. Mol Biol Evol 17:1384–1395

Eirín-López JM, Rebordinos L, Rooney AP, Rozas J (2012) The birth-and-death evolution of multigene families revisited. Genome Dyn 7: 170–196

Ejsmond MJ, Babik W, Radwan J (2010) MHC allele frequency distributions under parasite-driven selection: a simulation model. BMC Evol Biol 10:332

Ekblom R, Grahn M, Höglund J (2003) Patterns of polymorphism in the MHC class II of a non-passerine bird, the great snipe (*Gallinago media*). Immunogenet 54:734–741

Ekblom R, Saether SA, Fiske P, Käläs JA, Höglund J (2010) Balancing selection, sexual selection and geographic structure in MHC genes of Great Snipe. Genet 138:453–461

Felsenstein J (2005) PHYLIP (Phylogeny Inference Package). 3.6.5 edn. Distributed by the author., Department of Genome Sciences, University of Washington, Seattle

Fridolfsson A-K, Ellegren H (1999) A simple and universal method for molecular sexing of non-ratite birds. J Avian Biol 30:116–121

Grubb TC Jr (1974) Olfactory navigation to the nesting burrow in Leach's petrel (*Oceanodroma leucorrhoa*). Anim Behav 22:192–202

Hablützel PI, Volckaert FAM, Hellemans B, Raeymaekers JAM (2013) Differential modes of MHC class IIB gene evolution in cichlid fishes. Immunogenet 65:795–809

Haussmann MF, Mauck RA (2008) Telomeres and longevity: testing an evolutionary hypothesis. Mol Biol Evol 25:220–228

Hess CM, Edwards SV (2002) The evolution of the major histocompatibility complex in birds. Biosci 52:423–431

Hill AVS, Allsopp CEM, Kwiatkowski D, Anstey NM, Twumasi P, Rowe PA, Bennett S, Brewster D, McMichael AJ, Greenwood BM (1991) Common West African HLA antigens are associated with protection from severe malaria. Nat 352:595–600

Hughes AL, Nei M (1989) Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. Proc Natl Acad Sci U S A 86:958–962

Hughes CR, Miles S, Walbroehl JM (2008) Support for the minimal essential MHC hypothesis: a parrot with a single, highly polymorphic MHC class II B gene. Immunogenet 60:219–231

Huntington CE, Butler RG, Mauck RA (1996) Leach's storm-petrel (*Oceanodroma leucorhoa*). In: Poole A, Gill F (eds) The birds of North America, no. 233. The Academy of Natural Sciences, Philadelphia, and The American Ornithologists' Union, Washington

Jacob JP, Milne S, Beck S, Kaufman J (2000) The major and a minor class II β chain (B-LB) gene flank the Tapasin gene in the B-F/B-L region of the chicken major histocompatibility complex. Immunogenet 51: 138–147

Jaratlerdsiri W, Isberg SR, Higgins DP, Ho SYW, Salomonsen J, Skjodt K, Miles LG, Gongora J (2014) Evolution of MHC class I in the order Crocodylia. Immunogenet 66:53–65

Judo MSB, Wedel AB, Wilson C (1998) Stimulation and suppression of PCR-mediated recombination. Nucleic Acids Res 26:1819–1825

Juola FA, Dearborn DC (2012) Sequence-based evidence for major histocompatibility complex-disassortative mating in a colonial seabird. Proc R Soc B Biol Sci 279:153–162

Kaufman J, Wallny HJ (1996) Chicken MHC molecules, disease resistance and the evolutionary origin of birds. Curr Top Microbiol Immunol 212:129–141

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. Nat 401:923–925

Kikkawa EF, Tsuda TT, Sumiyama D, Naruse TK, Fukuda M, Kurita M, Wilson RP, Lemaho Y, Miller GD, Tsuda M, Murata K, Kulski JK, Inoko H (2009) Trans-species polymorphism of the Mhc class II DRB-like gene in banded penguins (genus *Spheniscus*). Immunogenet 61:341–352

Kurtz J, Kalbe M, Aeschlimann PB, Haberli MA, Wegner KM, Reusch TBH, Milinski M (2004) Major histocompatibility complex diversity influences parasite resistance and innate immunity in sticklebacks. Proc R Soc Lond Ser B-Biol Sci 271:197–204

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948

Lenz TL, Becker S (2008) Simple approach to reduce PCR artefact formation leads to reliable genotyping of MHC and other highly polymorphic loci—implications for evolutionary analysis. Gene 427:117–123

Mauck RA, Waite TA, Parker PG (1995) Monogamy in Leach's storm-petrel: DNA-fingerprinting evidence. Auk 112:473–482

Mauck RA, Huntington CE, Grubb TC Jr (2004) Age-specific reproductive success: evidence for the selection hypothesis. Evol 58:880–885

Mauck RA, Huntington CE, Doherty PF (2012) Experience versus effort: what explains dynamic heterogeneity with respect to age? Oikos 121:1379–1390

Michel C, Bernatchez L, Behrmann-Godel J (2009) Diversity and evolution of MHII β genes in a non-model percid species-the Eurasian perch (Perca fluviatilis L.). Mol Immunol 46:3399–3410

Milinski M (2006) The major histocompatibility complex, sexual selection, and mate choice. Annu Rev Ecol Evol Syst 37:159–186

Milinski M, Griffiths S, Wegner KM, Reusch TBH, Haas-Assenbaum A, Boehm T (2005) Mate choice decisions of stickleback females predictably modified by MHC peptide ligands. Proc Natl Acad Sci U S A 102:4414–4418

Miller HC, Lambert DM (2004) Gene duplication and gene conversion in class II MHC genes of New Zealand robins (Petroicidae). Immunogenet 56:178–191

Nei M, Rooney AP (2005) Concerted and birth-and-death evolution of multigene families. Annu Rev Genet 39:121–152

Nevitt GA, Haberman K (2003) Behavioral attraction of Leach's storm-petrels (*Oceanodroma leucorhoa*) to dimethyl sulfide. J Exp Biol 206:1497–1501

Ober C, Weitkamp LR, Cox N, Dytch H, Kostyu D, Elias S (1997) HLA and mate choice in humans. Am J Hum Genet 61:497–504

Paterson S, Wilson K, Pemberton JM (1998) Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population (*Ovis aries* L.). Proc Natl Acad Sci U S A 95:3714–3719

Piertney SB (2003) Major histocompatibility complex B-LB gene variation in red grouse *Lagopus lagopus scoticus*. Wildl Biol 9:251–259

Potts WK, Manning CJ, Wakeland EK (1991) Mating patterns in seminatural populations of mice influenced by MHC genotype. Nat 352:619–621

Reusch TBH, Langefors Å (2005) Inter- and intralocus recombination drive MHC class IIB gene diversification in a teleost, the three-spined stickleback *Gasterosteus aculeatus*. J Mol Evol 61:531–541

Richman AD, Herrera LG, Nash D, Schierup MH (2003) Relative roles of mutation and recombination in generating allelic polymorphism at an MHC class II locus in Peromyscus maniculatus. Genet Res 82:89–99

Ricklefs RE, Day CH, Huntington CE, Williams JB (1985) Variability in feeding rate and meal size of Leach's storm-petrel at Kent Island, New Brunswick. J Anim Ecol 54:883–898

Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497

Salzburger W, Ewing GB, Von Haeseler A (2011) The performance of phylogenetic algorithms in estimating haplotype genealogies with migration. Mol Ecol 20:1952–1963

Sato A, Mayer WE, Tichy H, Grant PR, Grant RB, Klein J (2001) Evolution of Mhc class II B genes in Darwin's finches and their closest relatives: birth of a new gene. Immunogenet 53: 792–801

Sawyer SA (1999) GENECONV: a computer package for the statistical detection of gene conversion. 1.8.1 edn. Distributed by the author, Department of Mathematics, Washington University in St. Louis, available at http://www.math.wustl.edu/~sawyer

Schad J, Ganzhorn JU, Sommer S (2005) Parasite burden and constitution of major histocompatibility complex in the Malagasy mouse lemur, *Microcebus murinus*. Evol 59:439–450

Silva MC, Edwards SV (2009) Structure and evolution of a new avian MHC class II B gene in a sub-antarctic seabird, the thin-billed prion (Procellariiformes: *Pachyptila belcheri*). J Mol Evol 68:279–291

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

Spehr M, Kelliher KR, Li XH, Boehm T, Leinders-Zufall T, Zufall F (2006) Essential role of the main olfactory system in social recognition of major histocompatibility complex peptide ligands. J Neurosci 26:1961–1970

Spurgin LG, Richardson DS (2010) How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. Proc R Soc B Biol Sci 277:979–988

Strand T, Westerdahl H, Höglund J, V. Alatalo R, Siitari H (2007) The Mhc class II of the Black grouse (*Tetrao tetrix*) consists of low numbers of B and Y genes with variable diversity and expression. Immunogenet 59:725–734

Strand T, Wang B, Meyer-Lucht Y, Höglund J (2013) Evolutionary history of black grouse major histocompatibility complex class IIB genes revealed through single locus sequence-based genotyping. BMC Genet 14(1):29

Strandh M, Lannefors M, Bonadonna F, Westerdahl H (2011) Characterization of MHC class I and II genes in a subantarctic seabird, the blue petrel, *Halobaena caerulea* (Procellariiformes). Immunogenet 63:653–666

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

Thursz MR, Thomas HC, Greenwood BM, Hill AV (1997) Heterozygote advantage for HLA class-II type in hepatitis B virus infection. Nat Genet 17:11–12

Wallny HJ, Avila D, Hunt LG, Powell TJ, Riegert P, Salomonsen J, Skødt K, Vainio O, Vilbois F, Wiles MV, Kaufman J (2006) Peptide motifs of the single dominantly expressed class I molecule explain the striking MHC-determined response to Rous sarcoma virus in chickens. Proc Natl Acad Sci U S A 103:1434–1439

Walsh HE, Friesen VL (2003) A comparison of intraspecific patterns of DNA sequence variation in mitochondrial DNA, α-enolase, and MHC class II B loci in auklets (Charadriiformes: Alcidae). J Mol Evol 57:681–693

Wedekind C, Seebeck T, Bettens F, Paepke AJ (1995) MHC-dependent mate preferences in humans. Proc R Soc B Biol Sci 260:245–249

Wegner KM, Kalbe M, Kurtz J, Reusch TBH, Milinski M (2003) Parasite selection for immunogenetic optimality. Sci 301:1343

Westerdahl H, Wittzell H, Von Schantz T (2000) Mhc diversity in two passerine birds: no evidence for a minimal essential Mhc. Immunogenet 52:92–100

Wittzell H, Bernot A, Auffray C, Zoorob R (1999) Concerted evolution of two Mhc class II B loci in pheasants and domestic chickens. Mol Biol Evol 16:479–490

Worley K, Gillingham M, Jensen P, Kennedy LJ, Pizzari T, Kaufman J, Richardson DS (2008) Single locus typing of MHC class I and class II B loci in a population of red jungle fowl. Immunogenet 60:233–247

Worley K, Collet J, Spurgin LG, Cornwallis C, Pizzari T, Richardson DS (2010) MHC heterozygosity and survival in red junglefowl. Mol Ecol 19:3064–3075

Xi D, Hao T, He Y, Leng J, Sun Y, Yang Y, Mao H, Deng W (2014) Nucleotide sequence and polymorphism of MHC class II DQB exon 2 alleles in Chinese yakow (*Bos grunniens*×*Bos taurus*). Int J Immunogenet 41:269–275

Zagalska-Neubauer M, Babik W, Stuglik M, Gustafsson L, Cichón M, Radwan J (2010) 454 sequencing reveals extreme complexity of the class II major histocompatibility complex in the collared flycatcher. BMC Evol Biol 10:395

Zangenberg G, Huang MM, Arnheim N, Erlich H (1995) New HLA-DPB1 alleles generated by interallelic gene conversion detected by analysis of sperm. Nat Genet 10:407–414

Zeisset I, Beebee TJC (2009) Molecular characterization of major histocompatibility complex class II alleles in the common frog, *Rana*

temporaria. Mol Ecol Resour 9:738–745

Zelano B, Edwards SV (2002) An Mhc component to kin recognition and mate choice in birds: predictions, progress, and prospects. Am Nat 160:S225–S237

Zhang B, Fang SG, Xi YM (2006) Major histocompatibility complex variation in the endangered crested ibis *Nipponia nippon* and implications for reintroduction. Biochem Genet 44:113–123

Zylstra P, Rothenfluh HS, Weiller GF, Blanden RV, Steele EJ (1998) PCR amplification of murine immunoglobulin germline V genes: Strategies for minimization of recombination artefacts. Immunol Cell Biol 76:395–405