

# Gene conversion rapidly generates major histocompatibility complex diversity in recently founded bird populations

LEWIS G. SPURGIN,\* COCK VAN OOSTERHOUT,† JUAN CARLOS ILLERA,\*‡§ STEPHEN BRIDGETT,¶ KARIM GHARBI,¶ BRENT C. EMERSON\*‡ and DAVID S. RICHARDSON\*

\*School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK, †School of Environmental Sciences, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK, ‡Island Ecology and Evolution Research Group, IPNA-CSIC, C/Astrofísico Francisco Sánchez 3, 38206 La Laguna, Tenerife, Canary Islands, Spain, § Research Unit of Biodiversity (UO-CSIC-PA), C/Catedrático Uría, s/n, Oviedo University, Campus del Cristo, 33006 Oviedo, Spain, ¶ The GenePool Genomics Facility, University of Edinburgh, Edinburgh, Lothian and Borders, UK

## Abstract

Population bottlenecks can restrict variation at functional genes, reducing the ability of populations to adapt to new and changing environments. Understanding how populations generate adaptive genetic variation following bottlenecks is therefore central to evolutionary biology. Genes of the major histocompatibility complex (MHC) are ideal models for studying adaptive genetic variation due to their central role in pathogen recognition. While *de novo* MHC sequence variation is generated by point mutation, gene conversion can generate new haplotypes by transferring sections of DNA within and across duplicated MHC loci. However, the extent to which gene conversion generates new MHC haplotypes in wild populations is poorly understood. We developed a 454 sequencing protocol to screen MHC class I exon 3 variation across all 13 island populations of Berthelot's pipit (*Anthus berthelotii*). We reveal that just 11–15 MHC haplotypes were retained when the Berthelot's pipit dispersed across its island range in the North Atlantic ca. 75 000 years ago. Since then, at least 26 new haplotypes have been generated *in situ* across populations. We show that most of these haplotypes were generated by gene conversion across divergent lineages, and that the rate of gene conversion exceeded that of point mutation by an order of magnitude. Gene conversion resulted in significantly more changes at nucleotide sites directly involved with pathogen recognition, indicating selection for functional variants. We suggest that the creation of new variants by gene conversion is the predominant mechanism generating MHC variation in genetically depauperate populations, thus allowing them to respond to pathogenic challenges.

**Keywords:** Berthelot's pipit, major histocompatibility complex, micro-recombination, mutation, selection

Received 4 August 2011; revision received 29 September 2011; accepted 2 October 2011

## Introduction

Reductions in population size can result in loss of diversity at functional genes, compromising the ability of populations to adapt (Mayr 1965; Dlugosch & Parker

2008). While populations with low levels of genetic variation may be able to persist in stable environments, loss of adaptive genetic diversity is likely to be especially important in scenarios where changes in habitat, climate or pathogens occur (Lande & Shannon 1996). Following a population bottleneck, novel genetic diversity may be introduced by migration or mutation. In isolated populations where migration rates are low,

Correspondence: David S. Richardson, Fax: 0044 1603 592250; E-mail: david.richardson@uea.ac.uk

mutation may be the only mechanism for generating new variation. Understanding how mutation generates adaptive genetic diversity in isolated populations is therefore central to evolution, and given the role of such diversity in the persistence of populations, is key from a long-term conservation perspective.

Vertebrate major histocompatibility complex (MHC) genes are arguably the best studied examples of how selection operates at the genetic level. The MHC genes code for cell surface glycoproteins that present pathogenic peptides to cells of the immune system, and so play a direct role in pathogen defence (Klein 1986). These genes display extraordinary levels of polymorphism, often with hundreds of alleles at individual loci (Garrigan & Hedrick 2003). Patterns of MHC variation are directly associated with pathogen resistance (Doherty & Zinkernagel 1975; Ilmonen *et al.* 2007; Oliver *et al.* 2009) and survival (Brouwer *et al.* 2010; Worley *et al.* 2010), suggesting that pathogen-mediated natural selection is the predominant force driving MHC diversity (reviewed in Jeffery & Bangham 2000; Spurgin & Richardson 2010). MHC-based mate choice (Penn 2002; Richardson *et al.* 2005) and selection against deleterious mutations at MHC-linked regions (e.g. van Oosterhout 2009) can operate alongside pathogen-mediated selection to maintain MHC variation, and a large body of research has been devoted to establishing the relative roles of these processes (Apanius *et al.* 1997; Piertney & Oliver 2005).

Demographic processes can also play an important role in regulating genetic diversity at the MHC (Alcaide 2010). In small, isolated populations, the effect of genetic drift can override the effects of selection at MHC loci, leading to a depletion in diversity (e.g. Hedrick *et al.* 2001; Miller & Lambert 2004; Seddon & Ellegren 2004). Such reductions in MHC variation may compromise the ability of populations to adapt to new and rapidly evolving pathogens, although establishing direct links between reduced MHC variation, and population viability has proved difficult (Edwards & Potts 1996; Radwan *et al.* 2010). Even less well understood is how, and at what rate, populations are able to recover from the reduced MHC variation caused by reductions in population size over evolutionary timescales. These are questions of fundamental importance for conservation biologists, and of intrinsic interest to molecular biologists, but are difficult to address, and so have received comparatively little attention.

New MHC sequence variation is generated by point mutation, with positive selection resulting in an excess of amino acid changing substitutions in the peptide-binding region (PBR)—the part of the molecule directly involved with pathogen recognition (Hughes & Nei 1988). Gene conversion-like processes (sometimes referred to as ‘micro-recombination’; Geliebter & Nathenson 1988;

hereafter referred to as ‘gene conversion’) can also generate new haplotypes, by shuffling existing MHC variation within or across duplicated loci (Ohta 1991). Gene conversion may be particularly important for regenerating MHC haplotype variation in bottlenecked populations, as the rate at which it can do so may greatly exceed that of point mutation (Parham & Ohta 1996). The occurrence, and to some extent, the mechanics of gene conversion, have been documented (Baltimore 1981; Schulze *et al.* 1983; Reusch & Langefors 2005; Chen *et al.* 2007), including in the avian MHC (Witzell *et al.* 1999; Burri *et al.* 2010). However, the ability of gene conversion to generate MHC variation relative to point mutation is still highly controversial, primarily due to a lack of empirical data from natural populations (Martinson *et al.* 1999; Klein *et al.* 2007). Moreover, statistical methods for detecting gene conversion lack power (Mansai & Innan 2010), and it can be difficult to separate gene conversion from convergent accumulation of point mutations due to selection (Kriener *et al.* 2000; Sato *et al.* 2011).

Berthelot’s pipit (*Anthus berthelotii*) is a passerine bird endemic to 13 island populations in three North Atlantic archipelagos (Fig. S1, Supporting information). A previous study revealed that this species has extremely low levels of genetic variation, with just one and four haplotypes at the mitochondrial control region and cytochrome *b* genes respectively, across all 13 populations (Illera *et al.* 2007). The pipit was also found to have lower levels of genetic variation compared to other passerine species at a set of conserved microsatellite markers (Dawson *et al.* 2010). Significant genetic structure was observed across the 13 populations, particularly across, but also within archipelagos, and significant isolation by distance was detected. These data are consistent with a small founding population of pipits having recently colonized these islands, with limited subsequent dispersal (Illera *et al.* 2007).

Here, we use the Berthelot’s pipit as a model to study how MHC genes evolve in the early stages of differentiation. We first test whether MHC diversity in Berthelot’s pipit has been reduced to a similar extent as the neutral diversity as a result of the founder event. We then test whether novel MHC haplotype variation has been generated *in situ* post-colonization. Finally, we quantify the relative roles of point mutation and gene conversion in generating new MHC haplotypes in the pipit populations.

## Methods

### Sampling

Thirty Berthelot’s pipits were caught in each of the 13 main populations in which this species occurs (Fig. S1,

Supporting information) between August 2005 and September 2006. Birds were caught across multiple locations in each island, using spring traps baited with *Tenebrio molitor* larvae. Blood samples (c. 40 µL) were taken by brachial venipuncture and stored in absolute ethanol.

### *Molecular procedures*

Genomic DNA was extracted from blood using a salt extraction method (Richardson *et al.* 2001), and diluted to 10–50 ng/µL. Degenerate fusion primers based on those designed by Westerdahl *et al.* (2004; DG2, TTGCGCTCYAGCTCYTTCTGCT; GENDG, TCCCCAC AGGTCTCCACAC), incorporating the Roche 454 FLX adapter sequences were used to amplify 240 bp (out of 270 bp) of the MHC class I exon 3 locus of each individual. Exon 3 contains sites which code for the PBR of the class I molecule, and variation at these sites is likely to be directly associated with immune defence (Hughes & Yeager 1998).

Initial amplification and cloning was performed on five individuals from Tenerife and one from Madeira, using the amplification and cloning protocols outlined in Westerdahl *et al.* (1999). Twelve alleles were found, all of which were confirmed by subsequent 454 sequencing (Table S3, Supporting information). Between four and seven alleles were detected per individual, indicating that a minimum of four loci were present (Table S3, Supporting information). This is concordant with the number of class I loci found in other passerine birds (Richardson & Westerdahl 2003; Pomerová *et al.* 2009).

We used population-level tagged primers and 454 sequencing to screen population-level MHC variation. Thirteen variants of the MHC primers were designed, each with a 7-bp population-specific tag at the 5' end. To reduce the risk of sequencing errors, tags were preceded with a 'CCD' (D = G, A or T) motif, stretches of the same nucleotide longer than two were avoided, and all tags differed by at least 2 bp (Valentini *et al.* 2009). With primers and tags included, the amplicon was longer than the average read length of the 454 sequencer (the total target amplicon length including all primers and tags was 322 bp). Both forward and reverse primers were tagged, so that, amplicons could be sequenced using bi-directional beads and the exon could be re-assembled using the read overlap. Each individual within a population was amplified in a separate PCR using the same population tagged primer. The PCRs were performed in 15 µL reactions, containing 7.5-µL 2× Thermoprime PCR mastermix (ABgene), 2-µL each primer, 1.5-µL H<sub>2</sub>O and 2-µL template DNA, under the following conditions: initial denaturation at

96 °C for 3 min, followed by 27 cycles of 94 °C for 30 s, 61 °C for 30 s, and 72 °C for 60 s, with a final extension stage of 72 °C for 10 min. The reduced number of cycles and long extension time were used to reduce the formation of chimeric sequences (Lenz & Becker 2008). The PCR products were purified using MinElute PCR Purification Kits (Qiagen). Concentrations were assessed using 1.2% agarose gels and a NanoDrop ND-1000 spectrophotometer.

Amplicons from all individuals from a given population were equalized and pooled. The mixed amplicons from each separate population were then equalized and pooled into one overall sample containing all populations and individuals. The final sample was sequenced using bi-directional sequencing on half a 454 FLX plate.

### *Identification of MHC haplotypes*

Forward and reverse reads corresponding to each of the 13 population-specific tags were extracted using an in-house Perl script, and grouped into the populations. The reads from each population were assembled into contigs with the Roche 454 Newbler assembler version 2.3, using highly stringent settings (99% minimum overlap identity and 200 bp minimum overlap length). Most short and low frequency reads, which may represent PCR and sequencing artefacts (Babik *et al.* 2009; Galan *et al.* 2010), would not group into contigs using this procedure. Nonetheless, some may, and so contigs shorter than 220 bp and low frequency contigs (see below) were excluded. If each individual in a population (30 individuals) had eight different haplotypes (based on the estimate of four duplicated loci), the frequency of a single haplotype copy in the population would be ~0.4%. To minimize the possibility of missing very rare haplotypes, the identified haplotypes from each population were mapped back to the discarded contigs from all other populations, with the aim of identifying haplotypes that had been verified in some populations, but discarded in others.

### *Founder population size and drift*

We developed a simple simulation to establish the minimum number of founder individuals required to introduce the observed number of MHC haplotypes, using a macro developed in Minitab version 12.1 (Minitab Inc.; available from the authors on request). The rationale behind this simulation was that if the number of founder individuals required to introduce the total observed number of haplotypes is much larger than is plausible based on the mitochondrial diversity, then it is likely that some MHC variation has evolved *in situ*, post-colonization.

The simulation re-sampled eight haplotypes per individual (based on four duplicated loci) from the contemporary gene pool 1000 times, increasing the number of individuals until the observed number of MHC haplotypes occurred in 1%, 50% and 99% of simulations. The simulations assumed equal sex ratio and no genetic drift after the founder event. Although the second assumption is unrealistic, it will result in an underestimation of the number of founding individuals required. As our aim was to simulate the *minimum* number of founding individuals required to introduce all MHC variation, it can be considered conservative.

The effect of genetic drift was examined by comparing population size with allelic richness at both the MHC and at the five microsatellite loci amplified by Illera *et al.* (2007) using linear regression.

### Sequence analyses

To visualize the time frame over which MHC loci have been diverging across populations, it is necessary to obtain an estimate of how long the populations have been separated. To do so, time to most recent common ancestor (TMRCA) of the pipit mitochondrial cytochrome *b* sequences from Illera *et al.* (2007) was estimated using BEAST version 1.6.1 (Drummond & Rambaut 2007). Two independent MCMC analyses of 10 000 000 steps, each with a burn-in of 1 000 000 steps, were performed. We used the HKY model of nucleotide substitution (Hasegawa *et al.* 1985) and a rate of 0.01 substitutions per site per million years, corresponding to sequence a divergence rate of 2% per million years. Convergence of the chains to a stationary distribution was assessed with Tracer version 1.4. Mutation rates at passerine mitochondrial genes can vary, and 2% per million years is likely to be an underestimate (Nabholz *et al.* 2009). The true date of dispersal may therefore be more recent than our conservative estimate of TMRCA.

At the MHC, nucleotide sites known to code for the PBR in humans (Brown *et al.* 1993) were used to estimate PBR and non-PBR sites in the pipit sequences. We calculated  $d_N$  and  $d_S$  separately for PBR and non-PBR regions (Nei & Gojobori 1986) in MEGA (Tamura *et al.* 2007). Differences between  $d_N$  and  $d_S$ , as well as differences across PBR and non-PBR regions, were assessed using Mann–Whitney *U*-tests.

We tested for individual gene conversion events using GENECONV (Sawyer 1999), and recombination using the RDP (Martin & Rybicki 2000), MaxChi (Smith 1992) and methods implemented in the RDP3 package. However, GENECONV, the most commonly used statistical method to detect gene conversion, has been shown to suffer from low power (Mansai & Innan 2010), and other methods are not designed to detect the

transfer of small sequence fragments. We also calculated rate of recombination ( $\rho$ ) relative to point mutation ( $\theta$ ) across the global alignment using LDHat (McVean *et al.* 2002). This program uses a coalescent-based approach, and therefore while estimates of  $\rho/\theta > 1$  are likely to be indicative of recombination, absolute estimates from data across populations need to be interpreted with caution.

We developed a macro to quantify the level of shared polymorphism across unrelated haplotypes (hereafter 'convergence'), using Minitab version 12.1 (available from the authors on request). Each haplotype (the 'derived') is first paired with its closest matching haplotype (the 'ancestor') using a 3-bp sliding window, and the percentage sequence similarity between the two is calculated. The macro then searches for a 'donor' haplotype within the genepool that can explain 100% of sequence similarity at the regions where the ancestor and derived haplotypes differ. A 'construct' haplotype is then created by recombining the derived haplotype with the donor at the region of mismatch. The percentage sequence similarity between the construct haplotype and the ancestor haplotype is then calculated to give the level of sequence similarity after allowing 'convergence' with one donor. The average improvement in sequence similarity was compared to a data set simulated to evolve neutrally via point mutation only, using Wilcoxon tests (see Appendix S2, Supporting information). A significant increase in sequence similarity is predicted to arise from convergent evolution via either point mutation or by gene conversion. We therefore stress that the above method quantifies the level of convergence in haplotypes, rather than gene conversion *per se*. We assessed the likelihood of the convergent evolution occurring via point mutation by looking for independent point mutations at synonymous sites (see Results).

## Results

### 454 sequencing and identification of MHC haplotypes

The 454 run yielded 268 267 reads, 257 708 (96%) of which exactly matched one of the 13 population-level tags. Thus, the rate of error was relatively low. The number of reads obtained per population ranged from 7763 to 33 102. The Newbler assembler grouped these sequences into between 12 and 91 contigs per population, using 227 056 (88%) of the reads. After removing short contigs, and those with <0.4% of the reads (this corresponded to 38 reads in the population with the lowest read number), between 9 and 14 haplotypes per population were identified (Table 1) making use of 208 199 (81%) of the reads. None of the remaining

**Table 1** Haplotype and lineage diversity at major histocompatibility complex (MHC) class I genes across Berthelot's pipit populations

Island	Number of MHC haplotypes	Lineages										
		L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11
Madeira												
Deserta Grande	12	10	9	8,6,21	2	1	7	26	4	3	5	
Madeira	11	10	9	8,21	2	1	7	26	33	3	23	
Porto Santo	12	10	35	8,6,21	2	1	40	32	36	3	23	
Selvagens												
Selvagem Grande	11	10	30	8,6	2	27	37		4	3	11	38
Canary Islands												
La Graciosa	9		9	6,42,25		1	7		4	3		16
Lanzarote	10	10		8,24	2	1	7		4	31	11	16
Fuerteventura	11	10,13	9	8,12,14	2	1	7			3	11	
Gran Canaria	14	10,13,28		8,6,15,17	2	1	7		4	3	11	16
El Teide	10	10		8,12	2	39	19		4	3	11	16
Tenerife	13	10,28	9	8,6,24	2	41	19		4	3	11	16
La Gomera	13	10,13		8,6,24,22	2	1	7		4	3	11	16
La Palma	9	10,13		8,22,34	2	1	7			3		
El Hierro	12	10,18,28	30	8,6	2	1	7		4	3		29
Probability		1	0.2330	1	1	0.0123	0.0090	0.8462	0.0947	0.0769	0.0123	0.2899

Columns L1–11 represent haplotype lineages, the numbers inside cells are the labels of individual MHC haplotypes. The colour of cells corresponds to whether, within each population, we found only the most widespread haplotype within a lineage (blue), a rare haplotype found in just a few populations (red), or cases where multiple haplotypes of the same lineage were present (yellow). Furthermore, included are the probabilities from binomial mass functions that haplotypes within lineages are mutually exclusive across populations (see Appendix S1, Supporting information).

haplotypes contained indels or stop codons. Some haplotypes occurred at high frequencies across most populations, whereas others were rare, and only present in one, or a subset of populations, making a total of 41 unique haplotypes (Table S1, Supporting information; Fig. 2). These assembled haplotype sequences have been deposited in GenBank (accession numbers).

A phylogenetic network of the MHC haplotypes revealed eleven divergent lineages, each containing between one and twelve closely related haplotypes (Fig. 1). Lineages typically consist of one geographically widespread haplotype and a few rarer haplotypes, unique to just one or two populations (Table 1). Not all populations contain all 11 lineages (range 6–10), but with the exception of the more complicated L1 and L3 lineages, all populations possess only one haplotype from each of the lineages they do contain (Table 1). In other words, haplotypes within lineages are mutually exclusive across populations (Fisher's combined probability test:  $\chi^2 = 53.8$ , d.f. = 20,  $P = 6.19 \times 10^{-5}$ , Table 1).

*Colonization time, founder population size and drift*

Analyses of the pipit mitochondrial DNA sequences using BEAST yielded a mean age estimate for the MRCA of extant mtDNA haplotype diversity of 75 000 years (HPD 20 000–170 000), which we take as a conservative

estimate of when the Berthelot's pipit dispersed to form its current range. This assumes a single founding mtDNA lineage from which all extant lineages are derived, and it should be pointed out that if colonization had involved more than one of the four mtDNA haplotypes from which the single extant haplotype diversity is derived, or if our mutation rates are too low, colonization time would be closer to the present.

Simulations revealed that a minimum founder population size of 320 (outbred individuals) is required to introduce the observed number of 41 MHC haplotypes. A founder size of ca. 500 birds will introduce a median of 41 MHC haplotypes, and more than 2000 birds are required to explain the presence of 41 haplotypes with 99% confidence (Fig. S2, Supporting information). Given that no drift was allowed after the founder event, these simulations are conservative (i.e. they are underestimating the 'true' founder population size if all extant MHC variation has to be introduced during founding). As a single ancestral mitochondrial haplotype was involved in colonization (Illera *et al.* 2007), such a large number of colonizing individuals seems highly unlikely, given the levels of mitochondrial haplotype diversity seen in other outbred bird populations (Grapputo *et al.* 1998; Questiau *et al.* 1998; Kimura *et al.* 2002). Furthermore, if all 41 haplotypes were present in the founding population, and the extant

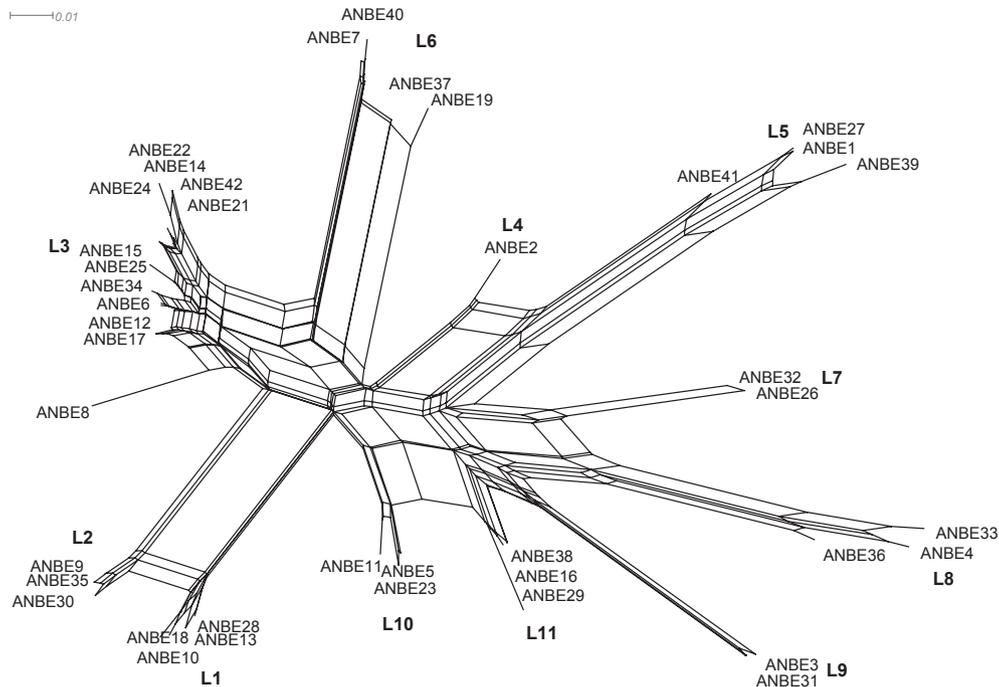


Fig. 1 Neighbour-net constructed from 41 Berthelot's pipit MHC class I exon 3 haplotypes using Jukes–Cantor distances.

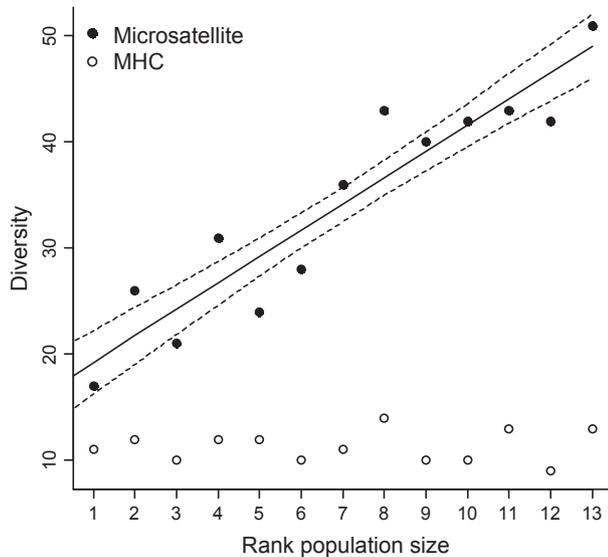
haplotype distribution was caused by stochastic patterns of loss across populations due to drift, there should be a correlation between population size and number of MHC haplotypes. We found strong evidence for drift at neutral loci, indicated by a positive relationship between microsatellite variation and population size; however, this was not the case for the MHC, which showed a strikingly different pattern (Fig. 2). Consequently, we suggest that the geographically widespread haplotypes of each lineage are ancestral, that novel haplotype variation has been generated *in situ* within each population, and that new haplotypes normally replaced the haplotype from which they were derived. Depending on the number of ancestral haplotypes in the complicated L1 and L3 lineages, there are between 11 and 15 founder MHC haplotypes, with between 26 and 30 haplotypes having evolved *in situ* after the initial founding event. In subsequent analyses, we assume 12 ancestral alleles—one from each lineage, and the two most widespread haplotypes from lineage L3 (ANBE6 and ANBE8). Fewer than 20 individuals are required to introduce 11–15 haplotypes to each island with 99% confidence (Fig. S2, Supporting information). This founder population size appears to be more consistent with the low mtDNA variation of the species.

#### Sequence analyses

Putative derived MHC haplotypes differed from their ancestral forms by distinctive nucleotide clusters that

were present in ancestral haplotypes from other lineages (Fig. 3). The effect of these recurring motifs was apparent in the phylogenetic network (Fig. 1), which reveals widespread conflicting splits, indicating ambiguous relationships among the MHC haplotypes. The RDP, GENCONV and MaxChi methods of recombination detection found 3, 11 and 19 recombination events respectively, strongly suggesting that micro-recombination or gene conversion has occurred at the pipit MHC. However, these methods failed to detect instances of gene conversion that were very clear when inspected by eye (the event shown in Fig. 3B, for example, was not detected). Considering all sequences together, the estimated population recombination rate ( $\rho$ ) was 26.94 (5% and 95% bounds were 18.34 and 38.22 respectively), compared to a mutation rate ( $\theta$ ) of 15.89, giving a recombination–mutation ratio of 1.695. This indicates that recombination is more prevalent than point mutation.

Our macro developed to detect convergence among haplotypes revealed that MHC haplotypes were highly similar to their closest relative, differing on average by just over 1% (Table S2, Supporting information). In contrast, putative donor haplotypes were on average 10% different to the derived haplotypes across the sequence as a whole, but usually matched perfectly in the donor regions. The size of donor regions varied, with minimum insert size ranging from 2 to 35 bp (Table S2, Supporting information). There was an increase in sequence similarity between haplotypes and their closest relatives after allowing a single donor



**Fig. 2** Variation at major histocompatibility complex (MHC) and microsatellite loci in relation to rank population size (Illera 2007) across island populations. For the MHC, this is the number of haplotypes found in each population. For microsatellites, it is the total number of haplotypes detected across five loci (Illera *et al.* 2007). Dotted lines represent 95% confidence intervals. There is a significant positive relation for microsatellites ( $R^2 = 0.87$ , d.f. = 11,  $P = 2.92 \times 10^{-6}$ ), but not for the MHC ( $R^2 = 0.001$ , d.f. = 11,  $P = 0.89$ ). This relationship also holds for island size (microsatellite:  $R^2 = 0.71$ , d.f. = 11,  $P = 3.14 \times 10^{-4}$ ; MHC:  $R^2 = 0.015$ , d.f. = 11,  $P = 0.69$ ).

haplotype, and after allowing two donors, 100% of the observed nucleotide polymorphisms could be explained in all but two putative derived haplotypes, which both differed from their closest relative by a single point mutation (Table S2, Supporting information). Sequence variation in the putative ancestral haplotypes could be less well explained by donor haplotypes (Table S2, Supporting information). Moreover, putative ancestral haplotypes were identified by the macro as donor sequences in 26 of 27 cases. Thus, these analyses strongly support our identification of ancestral and derived haplotypes. The average level of convergence in the observed MHC sequences was much greater than that at sequences simulated to evolve neutrally via point mutation (Fig. S3, Supporting information). In other words, to explain the shared polymorphisms observed across lineages, either gene conversion or convergent evolution by point mutation must be operating. Assuming 12 ancestral and 29 derived sequences, a combination of 27 gene conversion events and two point mutations could potentially explain all the MHC sequence variation in the pipit populations.

To assess the likelihood that convergent accumulation of point mutations, rather than gene conversion,

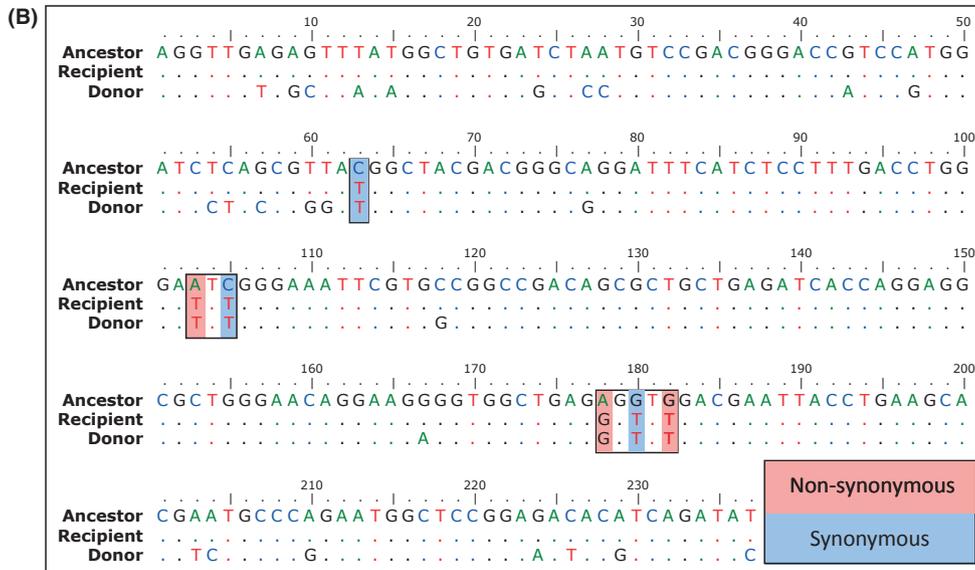
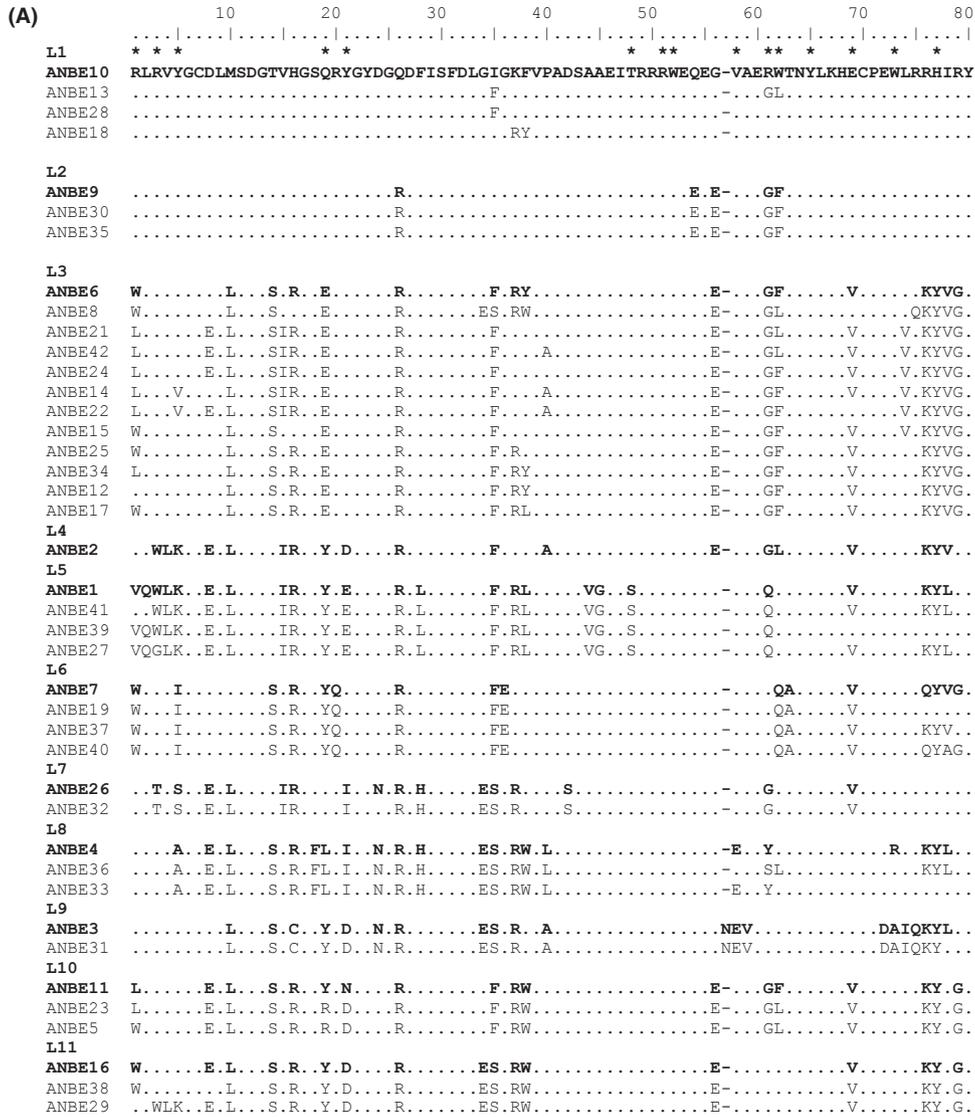
explained our observed patterns, patterns of variation at (mostly neutral) third codon positions were examined. If all within-lineage variation could be explained by point mutation, we would expect to see, by chance, some evidence of independent neutral mutations within lineages. We did not see this: of 80 third codon sites, 50 were conserved across all haplotypes, 14 sites were variable across (but not within) lineages and 15 sites exhibited within-lineage variation. Importantly, all within-lineage variation could be explained by polymorphism found across lineages, with no evidence of independent mutations within lineages. The probability of 15 neutral mutations occurring at the same nucleotide positions in different lineages by chance given the number of variable sites is extremely low ( $\chi^2 = 29.17$ , d.f. = 1,  $P = 6.7 \times 10^{-8}$ ). Considering all 240 sites, only two unique (nonsynonymous) within-lineage mutations occurred, which were also identified by the macro and appear to represent point mutations.

Gene conversion resulted in more changes at putative PBR sites than at non-PBR sites ( $\chi^2 = 10.89$ , d.f. = 1,  $P = 9.6 \times 10^{-4}$ ; Fig. 4). This is reflected in rates of nonsynonymous ( $d_N$ ) and synonymous ( $d_S$ ) substitutions across the pipit sequences. Both  $d_N$  and  $d_S$  were significantly higher at PBR compared to non-PBR sites (Fig. 4), but  $d_N$  was not significantly higher than  $d_S$  in the PBR. However,  $d_S$  was significantly higher than  $d_N$  at non-PBR sites (Fig. 4) indicating purifying selection in these areas.

## Discussion

Here, we show that, in the Berthelot's pipit, only a limited number of divergent MHC haplotypes persisted when this species dispersed across its range ca. 75 000 years ago. Importantly, we were able to characterize the generation of MHC diversity subsequent to the founding event. We found that gene conversion rapidly generated functional MHC haplotype diversity across populations, with conversion events outstripping point mutations by as much as an order of magnitude.

The argument could be made that the recombinant haplotypes we observe are chimeras—artefacts of the PCR process known to occur when amplifying multiple loci. We believe that this is highly unlikely. First, we used a PCR protocol designed to minimize the formation of chimeras (Lenz & Becker 2008). Second, chimeras are expected to be rare in comparison to true haplotypes (Babik *et al.* 2009), whereas we find the same recombinants occurring at high frequencies both within a population, and most importantly, across multiple independently amplified populations. To invoke an explanation of chimeras for our data would require independent nonrandom generation of the same



chimeric sequence in independent PCR reactions, from independent populations. Third, by virtue of the dynamics of their formation, chimeras should co-occur with the parental sequences, from which they formed in a PCR reaction (Lenz & Becker 2008). In contrast, we find that recombinant and parental haplotypes do not typically co-occur within populations (Table 1). Fourth, there would be no reason to expect chimeras to form preferentially at the PBR, as we have found. Finally, we would generally expect chimeras to resemble single breakpoint recombination events within our short fragment, and many (but not all) of the recombination events we observe would require multiple breakpoints (Fig. 3A). We can thus exclude any explanation of chimeric origin for our recombinant sequences.

Using population-level tagged primers and 454 sequencing allowed us to MHC genotype a large number of individuals rapidly and efficiently. With this approach, however, individual-level variation could not be resolved. While initial cloning indicated that four duplicated class I loci are present in Berthelot's pipit—a similar number to other studies on passerines (Richardson & Westerdahl 2003; Promerová *et al.* 2009)—we cannot exclude the possibility that more loci are present. However, given the number of haplotypes we detected per population (between 9 and 14), it is unlikely that we have drastically underestimated the number of loci present. A second challenge that our approach presented was that our amplicon exceeded the available read length, meaning that we had to assemble forward and reverse reads into contigs. We are confident that the use of highly stringent assembly criteria to reconstruct haplotypes allows us to exclude the presence of chimeric alleles assembled from overlapping haplotypes for all but the most similar sequences (<1% nucleotide divergence). The longer read-lengths now available with 454 sequencing are likely to eliminate the need for this bioinformatics phase for amplicons up to ~300 bp (see for example Galan *et al.* 2010), although our paired-end approach offers the opportunity for longer amplicons (up to ~700 bp) to be genotyped.

As derived haplotypes appear to have replaced their ancestors within the pipit populations, gene conversion has not increased haplotype diversity within each independent population. This same pattern has been found

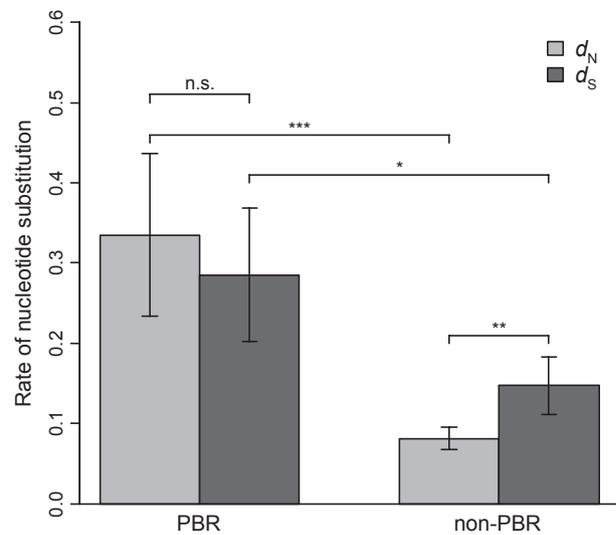


Fig. 4 Rates of nonsynonymous ( $d_N$ ) and synonymous ( $d_S$ ) substitutions in peptide-binding (PBR) and non-PBR regions in Berthelot's pipit major histocompatibility complex (MHC) class I exon 3 sequences. Significance levels were evaluated with Mann-Whitney *U*-tests (\* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ ). Error bars represent 95% confidence intervals.

in humans (Parham & Ohta 1996), suggesting that it may be a widespread phenomenon in MHC evolution. The replacement of ancestral MHC haplotypes by new haplotypes within populations may be because of direct positive selection of the derived haplotype, or because derived haplotypes share the same 'sheltered load' of deleterious mutations at MHC-linked regions as their ancestors (van Oosterhout 2009). Importantly, gene conversion generated different MHC haplotypes in eight of the 13 populations. Consequently, these results show that while haplotype diversity did not increase within populations, gene conversion can rapidly increase MHC haplotype diversity at the metapopulation level.

Although there is a large body of evidence suggesting that gene conversion operates within the MHC, the evolutionary significance of this has remained both poorly understood and controversial (Martinson *et al.* 1999; Klein *et al.* 2007). This is largely due to the difficulty of detecting gene conversion in wild populations. In populations that have evolved independently for long periods of time, homoplasy arising from recurrent gene

Fig. 3 (A) Amino acid alignment of 41 Berthelot's pipit major histocompatibility complex (MHC) class I exon 3 haplotypes, grouped by lineage according to a neighbour-net (Fig. 1). Putative ancestral haplotypes are at the top of each lineage, and highlighted in bold. Sites marked with a \* indicate those that encode for the human peptide-binding region. (B) Nucleotide alignment showing an example of gene conversion in three MHC sequences. This example consists of a widespread ancestral haplotype (ANBE10), a closely related haplotype derived from this (derived, ANBE13), and another distantly related ancestral haplotype (donor, ANBE2). Areas in which the ancestral and derived haplotypes differ are outlined with shaded boxes, and nonsynonymous and synonymous substitutions are highlighted in red and blue respectively.

conversion events and/or point mutations will make the two difficult to distinguish. In Berthelot's pipit, the combination of multiple populations and limited migration has allowed us to separate ancestral and derived variation, and as a result, clearly observe gene conversion. Moreover, the recent divergence among populations has provided a 'window of opportunity' to characterize how MHC genes evolve in the first stages of diversification, without the effects of homoplasy. Additional research is now required to determine how selection and/or drift operate to maintain new haplotypes within populations, and how migration may affect how patterns of MHC variation and newly evolved haplotypes are distributed across populations.

An additional factor underpinning the difficulty of quantifying gene conversion in natural populations is that detecting individual gene conversion events using statistical methods is highly problematic, although all three methods implemented in the present study yielded positive results. This is a particular problem when gene conversion does not result in large changes—for example, when small sequence tracts are transferred. We present a simple method for detecting the overall degree of convergence between sequences. Again, we stress that this method does not distinguish between gene conversion and convergence of point mutations; in our study, this was only achieved by looking at patterns of synonymous variation. The probability that point mutation can explain the high levels of convergence seen in the pipit MHC is exceedingly small ( $P = 6.7 \times 10^{-8}$ ). We acknowledge, however, that we were not able to statistically identify all of the gene conversion events, and cannot completely rule out the possibility that some of the convergent substitutions observed were the result of point mutations.

No evidence was found for elevated  $d_N/d_S$  in the PBR, which is widely regarded as the classic sign of selection at MHC genes (Hughes & Nei 1988). Instead, we found that both  $d_N$  and  $d_S$  were higher in PBR compared to non-PBR regions. This pattern is to be expected under a scenario of gene conversion combined with positive selection (Ohta 1995), whereby conversion events that contain beneficial non-synonymous mutations at the PBR are positively selected, and synonymous mutations hitchhike across with them (Fig. 3B). While a number of other population genetic studies have found elevated  $d_S$  at PBR compared to non-PBR sites (e.g. Richardson & Westerdahl 2003; Ekblom *et al.* 2007), to our knowledge, none has considered the possibility that gene conversion may be the cause.

Gene conversion can occur across alleles or across loci when genes have been duplicated (Chen *et al.* 2007). Thus, gene conversion across loci has the potential to generate new haplotypes, even if the population is

entirely homozygous at each of the paralogous MHC loci. Nonetheless, gene conversion can only re-shuffle existing (intra- or inter-locus) polymorphisms to create new variants. Therefore, although it can rapidly generate variation at the haplotype level, variation at the sequence level is expected to be homogenized over time, especially when larger fragments are being transferred (Takuno *et al.* 2008). In populations of Berthelot's pipit, gene conversion did cause a reduction in amino acid variation, although to a lesser extent in the PBR than non-PBR regions, and not in all populations (Fig. S4, Supporting information). It therefore appears that in these populations, the beneficial effect of novel haplotype variation from gene conversion outweighed the effect of sequence variation becoming marginally homogenized. This is to be expected when sequences within a population are divergent—a common observation at the MHC following population bottlenecks (Hedrick *et al.* 2000, 2002; Richardson & Westerdahl 2003; van Oosterhout *et al.* 2006). In such scenarios, we suggest that gene conversion may be the primary mechanism generating new functional haplotypes.

## Acknowledgements

We are grateful to the friends who provided assistance in the field. Iain Barr performed initial cloning and sequencing of the Berthelot's pipit MHC. Local governments in the Canary Islands and Madeira also provided accommodation. The Spanish Environment Ministry and the Natural Park of Madeira gave permission to work, and the Portuguese Navy helped with transport. The NERC Biomolecular Analysis Facility nodes in Liverpool and Edinburgh carried out 454 sequencing and bioinformatics respectively. We thank T. Burke, G. Hewitt and J. Kaufman for discussions. This work was funded by PhD and MGF small project grants awarded to DSR from the Natural Environment Research Council, and a Spanish postdoctoral fellowship (Ramón y Cajal) to JCI.

## References

- Alcaide M (2010) On the relative roles of selection and genetic drift in shaping MHC variation. *Molecular Ecology*, **19**, 3842–3844.
- Apanius V, Penn DJ, Slev PR, Ruff LR, Potts WK (1997) The nature of selection on the major histocompatibility complex. *Critical Reviews in Immunology*, **17**, 179–224.
- Babik W, Taberlet P, Ejsmond MJ, Radwan J (2009) New generation sequencers as a tool for genotyping of highly polymorphic multilocus MHC system. *Molecular Ecology Resources*, **9**, 713–719.
- Baltimore D (1981) Gene conversion: some implications for immunoglobulin genes. *Cell*, **24**, 592.
- Brouwer L, Barr I, Van de Pol M *et al.* (2010) MHC-dependent survival in a wild population: evidence for hidden genetic benefits gained through extra-pair fertilizations. *Molecular Ecology*, **19**, 3444–3455.

- Brown JH, Jardetzky TS, Gorga JC *et al.* (1993) Three-dimensional structure of the human class-II histocompatibility antigen HLA-DR1. *Nature*, **364**, 33–39.
- Burri R, Salamin N, Studer RA, Roulin A, Fumagalli L (2010) Adaptive divergence of ancient gene duplicates in the avian MHC class II. *Molecular Biology and Evolution*, **27**, 2360.
- Chen JM, Cooper DN, Chuzhanova N, Férec C, Patrinos GP (2007) Gene conversion: mechanisms, evolution and human disease. *Nature Reviews Genetics*, **8**, 762–775.
- Dawson DA, Horsburgh GJ, Küpper C *et al.* (2010) New methods to identify conserved microsatellite loci and develop primer sets of high cross-species utility; as demonstrated for birds. *Molecular Ecology Resources*, **10**, 475–494.
- Dlugosch KM, Parker IM (2008) Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology*, **17**, 431–449.
- Doherty PC, Zinkernagel RM (1975) Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature*, **256**, 50–52.
- Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.
- Edwards SV, Potts WK (1996) Polymorphism of genes in the major histocompatibility complex (MHC): implications for conservation genetics of vertebrates. In: *Molecular Genetic Approaches in Conservation* (eds Smith TB and Wayne RK), pp. 214–237. Oxford University Press, Oxford.
- Eklblom R, Saether SA, Jacobsson P *et al.* (2007) Spatial pattern of MHC class II variation in the great snipe (*Gallinago media*). *Molecular Ecology*, **16**, 1439–1451.
- Galan M, Guivier E, Caraux G, Charbonnel N, Cosson JF (2010) A 454 multiplex sequencing method for rapid and reliable genotyping of highly polymorphic genes in large-scale studies. *BMC Genomics*, **11**, 296.
- Garrigan D, Hedrick PW (2003) Detecting adaptive molecular polymorphism: lessons from the MHC. *Evolution*, **57**, 1707–1722.
- Geliebter J, Nathenson SG (1988) Microrecombinations generate sequence diversity in the murine major histocompatibility complex: analysis of the Kbm3, Kbm4, Kbm10, and Kbm11 mutants. *Molecular and Cellular Biology*, **8**, 4342.
- Grapputo A, Pilastro A, Marin G (1998) Genetic variation and bill size dimorphism in a passerine bird, the reed bunting *Emberiza schoeniclus*. *Molecular Ecology*, **7**, 1173–1182.
- Hasegawa M, Kishino H, Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, **22**, 160–174.
- Hedrick PW, Parker KM, Gutierrez-Espeleta GA, Rattink A, Lievers K (2000) Major histocompatibility complex variation in the Arabian oryx. *Evolution*, **54**, 2145–2151.
- Hedrick PW, Gutierrez-Espeleta GA, Lee RN (2001) Founder effect in an island population of bighorn sheep. *Molecular Ecology*, **10**, 851–857.
- Hedrick PW, Lee RN, Garrigan D (2002) Major histocompatibility complex variation in red wolves: evidence for common ancestry with coyotes and balancing selection. *Molecular Ecology*, **11**, 1905–1913.
- Hughes AL, Nei M (1988) Patterns of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature*, **335**, 167–170.
- Hughes AL, Yeager M (1998) Natural selection at major histocompatibility complex loci of vertebrates. *Annual Review of Genetics*, **32**, 415–435.
- Illera JC (2007) Bisbita Caminero *Anthus berthelotii*. In: *Atlas de las aves nidificantes en el archipiélago Canario* (ed. Lorenzo JA), pp. 344–347. Dirección General de Conservación de la Naturaleza-Sociedad Española de Ornitología, Madrid.
- Illera JC, Emerson BC, Richardson DS (2007) Population history of Berthelot's pipit: colonization, gene flow and morphological divergence in Macaronesia. *Molecular Ecology*, **16**, 4599–4612.
- Ilmonen P, Penn DJ, Damjanovich K *et al.* (2007) Major histocompatibility complex heterozygosity reduces fitness in experimentally infected mice. *Genetics*, **176**, 2501–2508.
- Jeffery KJM, Bangham CRM (2000) Do infectious diseases drive MHC diversity? *Microbes and Infection*, **2**, 1335–1341.
- Kimura M, Clegg SM, Lovette IJ *et al.* (2002) Phylogeographical approaches to assessing demographic connectivity between breeding and overwintering regions in a Nearctic-Neotropical warbler (*Wilsonia pusilla*). *Molecular Ecology*, **11**, 1605–1616.
- Klein J (1986) *Natural History of the Major Histocompatibility Complex*. John Wiley and Sons, New York, NY.
- Klein J, Sato A, Nikolaidis N (2007) MHC, TSP, and the origin of species: from immunogenetics to evolutionary genetics. *Annual Review of Genetics*, **41**, 281–304.
- Kriener K, O'HUigin C, Tichy H, Klein J (2000) Convergent evolution of major histocompatibility complex molecules in humans and New World monkeys. *Immunogenetics*, **51**, 169–178.
- Lande R, Shannon S (1996) The role of genetic variation in adaptation and population persistence in a changing environment. *Evolution*, **50**, 434–437.
- 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.
- Mansai SP, Innan H (2010) The power of the methods for detecting interlocus gene conversion. *Genetics*, **184**, 517.
- Martin D, Rybicki E (2000) RDP: detection of recombination amongst aligned sequences. *Bioinformatics*, **16**, 562–563.
- Martinsohn JT, Sousa AB, Guethlein LA, Howard JC (1999) The gene conversion hypothesis of MHC evolution: a review. *Immunogenetics*, **50**, 168–200.
- Mayr E (1965) The nature of colonizations in birds. In: *The Genetics of Colonizing Species* (eds Baker HG and Stebbins GL), pp. 29–43. Academic Press, London.
- McVean G, Awadalla P, Fearnhead P (2002) A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics*, **160**, 1231.
- Miller HC, Lambert DM (2004) Genetic drift outweighs balancing selection in shaping post-bottleneck major histocompatibility complex variation in New Zealand robins (Petroicidae). *Molecular Ecology*, **13**, 3709–3721.
- Nabholz B, Glémin S, Galtier N (2009) The erratic mitochondrial clock: variations of mutation rate, not population size, affect mtDNA diversity across birds and mammals. *BMC Evolutionary Biology*, **9**, 54.
- Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution*, **3**, 418–426.

- Ohta T (1991) Role of diversifying selection and gene conversion in evolution of major histocompatibility complex loci. *Proceedings of the National Academy of Sciences of the United States of America*, **88**, 6716.
- Ohta T (1995) Gene conversion vs point mutation in generating variability at the antigen recognition site of major histocompatibility complex loci. *Journal of Molecular Evolution*, **41**, 115–119.
- Oliver MK, Telfer S, Piertney SB (2009) Major histocompatibility complex (MHC) heterozygote superiority to natural multi-parasite infections in the water vole (*Arvicola terrestris*). *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **276**, 1119–1128.
- van Oosterhout C (2009) A new theory of MHC evolution: beyond selection on the immune genes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **276**, 657.
- van Oosterhout C, Joyce DA, Cummings SM *et al.* (2006) Balancing selection, random genetic drift, and genetic variation at the major histocompatibility complex in two wild populations of guppies (*Poecilia reticulata*). *Evolution*, **60**, 2562–2574.
- Parham P, Ohta T (1996) Population biology of antigen presentation by MHC class I molecules. *Science*, **272**, 67–74.
- Penn DJ (2002) The scent of genetic compatibility: sexual selection and the major histocompatibility complex. *Ethology*, **108**, 1–21.
- Piertney SB, Oliver MK (2005) The evolutionary ecology of the major histocompatibility complex. *Heredity*, **96**, 7–21.
- Promerová M, Albrecht T, Bryja J (2009) Extremely high MHC class I variation in a population of a long-distance migrant, the Scarlet Rosefinch (*Carpodacus erythrinus*). *Immunogenetics*, **61**, 451–461.
- Questiau S, Eybert MC, Gaginskaya AR, Gielly L, Taberlet P (1998) Recent divergence between two morphologically differentiated subspecies of bluethroat (Aves: Muscicapidae: *Luscinia svecica*) inferred from mitochondrial DNA sequence variation. *Molecular Ecology*, **7**, 239–245.
- Radwan J, Biedrzycka A, Babik W (2010) Does reduced MHC diversity decrease viability of vertebrate populations? *Biological Conservation*, **143**, 537–544.
- Reusch TBH, Langefors Å (2005) Inter- and intralocus recombination drive MHC class IIB gene diversification in a teleost, the three-spined stickleback *Gasterosteus aculeatus*. *Journal of Molecular Evolution*, **61**, 531–541.
- Richardson DS, Westerdahl H (2003) MHC diversity in two *Acrocephalus* species: the outbred Great reed warbler and the inbred Seychelles warbler. *Molecular Ecology*, **12**, 3523–3529.
- Richardson DS, Jury FL, Blaakmeer K, Komdeur J, Burke T (2001) Parentage assignment and extra-group paternity in a cooperative breeder: the Seychelles warbler (*Acrocephalus sechellensis*). *Molecular Ecology*, **10**, 2263–2273.
- Richardson DS, Komdeur J, Burke T, von Schantz T (2005) MHC-based patterns of social and extra-pair mate choice in the Seychelles warbler. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **272**, 759–767.
- Sato A, Tichy H, Grant PR *et al.* (2011) Spectrum of MHC class II variability in Darwin finches and their close relatives. *Molecular Biology and Evolution*, **28**, 1943–1956.
- Sawyer SA (1999) *GENECONV: A Computer Package for the Statistical Detection of Gene Conversion*. Department of Mathematics, Washington University, St. Louis, MO.
- Schulze DH, Pease LR, Geier SS *et al.* (1983) Comparison of the cloned H-2Kbm1 variant gene with the H-2Kb gene shows a cluster of seven nucleotide differences. *Proceedings of the National Academy of Sciences of the United States of America*, **80**, 2007.
- Seddon JM, Ellegren H (2004) A temporal analysis shows major histocompatibility complex loci in the Scandinavian wolf population are consistent with neutral evolution. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **271**, 2283–2291.
- Smith JM (1992) Analyzing the mosaic structure of genes. *Journal of Molecular Evolution*, **34**, 126–129.
- Spurgin LG, Richardson DS (2010) How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **277**, 979–988.
- Takuno S, Nishio T, Satta Y, Innan H (2008) Preservation of a pseudogene by gene conversion and diversifying selection. *Genetics*, **180**, 517–531.
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, **24**, 1596–1599.
- Valentini A, Miquel C, Ali Nawaz M *et al.* (2009) New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the *trnL* approach. *Molecular Ecology Resources*, **9**, 51–60.
- Westerdahl H, Wittzell H, von Schantz T (1999) Polymorphism and transcription of MHC class I genes in a passerine bird, the great reed warbler. *Immunogenetics*, **49**, 158–170.
- 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.
- Wittzell H, Bernot A, Auffray C, Zoorob R (1999) Concerted evolution of two MHC class II B loci in pheasants and domestic chickens. *Molecular Biology and Evolution*, **16**, 479.
- Worley K, Collet J, Spurgin LG *et al.* (2010) MHC heterozygosity and survival in red junglefowl. *Molecular Ecology*, **19**, 3064–3075.

---

L.G.S. is a PhD student studying the neutral and adaptive processes driving genetic differentiation across island bird populations. C.v.O. is focused on the evolution of the MHC, particularly the role of gene conversion and the effect of mutational load in this multigene family. J.C.I. studies how different processes (colonisation, adaptation) influence genetic variation within and among species inhabiting oceanic islands. B.C.E. uses molecular data to interpret phylogenetic history and population dynamics, particularly within arthropods and island ecosystems. K.G. is interested in genome biology, with a focus on using next-generation genomics to understand the basis of pathogen susceptibility in fish, while S.B. has broad interests in next-generation sequencing data analysis. Finally, D.S.R. is interested in understanding the causes and consequences of the patterns of genetic variation observed within natural populations and how this impacts upon individual survival and population persistence.

---

### Data accessibility

DNA Sequences: GenBank accession numbers JN799601–JN799641.

For MHC, microsatellite and mtDNA data, please see the Supporting information. For *usat* data, please refer to Illera *et al.* (2007).

### Supporting information

Additional supporting information may be found in the online version of this article.

**Fig. S1** Distribution and sampling locations of Berthelot's pipits (inset) in the North Atlantic.

**Fig. S2** Mean ( $\pm 99\%$  CI) number of MHC haplotypes in the pipit metapopulation as a function of the number of founders.

**Fig. S3** Convergence and mean ( $\pm$ SE) sequence similarity between MHC haplotypes.

**Fig. S4** Changes in mean amino acid p-distance between all pairwise combinations of haplotypes in populations for amino acids coded by the (a) non-PBR and (b) PBR codons.

**Table S1** Frequencies of MHC class I exon 3 haplotypes across 13 island populations of Berthelot's pipit, grouped by lineage.

**Table S2** Output from a sliding window analysis created to detect gene conversion in Berthelot's pipit MHC haplotypes (see text for details).

**Table S3** Sequences obtained from amplifying and cloning duplicated MHC class I exon 3 loci from individual Berthelot's pipits from Tenerife (TF) and Madeira (M).

**Appendix S1** Mutual exclusivity of MHC haplotypes within lineages.

**Appendix S2** Assessing convergence on empirical and simulated data.

**Appendix S3** Assembly settings and likelihood of chimeric sequences arising *in silico*.

**Data S1** Microsatellite genotypes and mitochondrial cytochrome b haplotypes in Berthelot's pipit individuals.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.