**Supplementary material for 'Gene conversion rapidly generates MHC diversity in recently founded bird populations'**

## Appendix S1 – Mutual exclusivity of MHC haplotypes within lineages

We tested whether MHC haplotypes within lineages were mutually exclusive in populations using a binomial mass function. For each haplotype lineage *i*, we calculated the binomial probability that the ancestral haplotype was absent in the populations where a derived haplotype was present. The probability that the ancestral haplotype was present in a population ($A_i$) was calculated as the number of populations in which the ancestral haplotype was observed ($k$) divided by the total number of populations ($A_i = k/13$). We then tallied the number of populations where the derived haplotype was present ($n$) and where ancestral and derived haplotypes were both present ($m$), and calculated the binomial probability of finding $m$ or less populations with the ancestral haplotype:

$$P(x_i \leq m) = \sum_{i=1}^{m} \binom{n}{m} A_i^{\,m} (1 - A_i)^{n-m}$$

These 11 p-values, one for each lineage, were then analysed using Fisher's combined probability test with 22 degrees of freedom.

## Appendix S2 – Assessing convergence on empirical and simulated data

To test whether there was a significant convergence in the pipit MHC sequences, we calculated the level of shared polymorphism across unrelated sequences using the macro described in the main text, and compared the results with a simulated dataset. The simulated sequences were generated in the program Seq-Gen v. 1.3.2 (http://tree.bio.ed.ac.uk/software/seqgen). This program simulates sequences along a given tree topology according to a user-specified model of sequence evolution. We generated a neighbour joining tree using the 41 pipit MHC class I exon 3 haplotypes. This tree was used to simulate nucleotide divergence assuming the HKY model of substitution. We generated five simulated sets, each containing 41 240 bp sequences. As these sequences conform to the same tree as the real MHC sequences, levels of divergence prior to running the macro were identical (Figure S3). Differences between average levels of sequence divergence in the observed and simulated datasets

were tested using Wilcoxon tests. For these, each of the haplotypes (41 empirical and 41 x 5 = 205 simulated) was a data point. The value represents the percentage similarity between that haplotype and its ancestor after simulating convergence (i.e. the 6th column in Table S2).

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

As forward and reverse sequences had to be assembled into contigs, there is a possibility that chimeric sequences could arise if overlapping sequences are incorrectly assembled. However, it is important to remember that the observed differences within lineages can be explained by variation across lineages (i.e. 'donor' sequences in Table S2). This difference between "ancestral" and "derived" sequence (which is also the similarity between the "derived" and "donor" sequence) we suggest arises by gene conversion across lineages. In order for this same result to arise from incorrect assembly would require sequences from across lineages to be assembled together into contigs. We are confident that our bioinformatics procedures eliminated the possibility of this happening. In order for a contig to assemble, the forward and reverse reads must have 200bp overlap, and 99% sequence identity within that overlap. Yet the haplotype lineages observed are much more divergent than this (10% on average across the sequence). Indeed, no two sequences from different lineages meet the levels of similarity required from our assembly settings.

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

| Lineage | Haplotype | DG | FV | GOM | GRAC | GC | HIER | LZ | PAL | M | PS | SG | TEIDE | TF | Number pops |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **L1** | **ANBE10** | 0.402 | 0.272 | 0.225 | | 0.178 | 0.297 | 0.375 | 0.315 | 0.407 | 0.441 | 0.271 | 0.282 | 0.285 | 12 |
| | **ANBE13** | | 0.026 | 0.025 | | 0.013 | | | 0.021 | | | | | | 4 |
| | **ANBE28** | | | | | | 0.012 | | | | | | | 0.011 | 2 |
| | **ANBE18** | | | | | 0.020 | 0.009 | | | | | | | | 2 |
| **L2** | **ANBE9** | 0.078 | 0.030 | | 0.478 | | | | | 0.050 | | | | 0.053 | 5 |
| | **ANBE30** | | | | | | 0.024 | | | | | 0.043 | | | 2 |
| | **ANBE35** | | | | | | | | | | 0.005 | | | | 1 |
| **L3** | **ANBE8** | 0.179 | 0.216 | 0.206 | | 0.197 | 0.186 | 0.205 | 0.223 | 0.243 | 0.193 | 0.174 | 0.260 | 0.191 | 12 |
| | **ANBE6** | 0.025 | | 0.029 | 0.079 | 0.012 | 0.007 | | | | 0.046 | 0.072 | | 0.017 | 8 |
| | **ANBE21** | 0.014 | | | | | | | | 0.032 | 0.035 | | | | 3 |
| | **ANBE42** | | | | 0.059 | | | | | | | | | | 1 |
| | **ANBE24** | | | 0.009 | | | | 0.031 | | | | | | 0.036 | 3 |
| | **ANBE14** | | 0.069 | | | | | | | | | | | | 1 |
| | **ANBE22** | | | 0.050 | | | | | 0.014 | | | | | | 2 |
| | **ANBE15** | | | | | 0.026 | | | | | | | | | 1 |
| | **ANBE25** | | | | 0.026 | | | | | | | | | | 1 |
| | **ANBE34** | | | | | | | | 0.035 | | | | | | 1 |
| | **ANBE12** | | 0.029 | | | | | | | | | | 0.015 | | 2 |
| | **ANBE17** | | | | | 0.009 | | | | | | | | | 1 |
| **L4** | **ANBE2** | 0.013 | 0.159 | 0.117 | | 0.215 | 0.172 | 0.114 | 0.110 | 0.029 | 0.030 | 0.143 | 0.148 | 0.126 | 12 |
| **L5** | **ANBE1** | 0.069 | 0.033 | 0.021 | 0.045 | 0.007 | 0.016 | 0.030 | 0.076 | 0.053 | 0.062 | | | | 10 |
| | **ANBE41** | | | | | | | | | | | | | 0.037 | 1 |
| | **ANBE39** | | | | | | | | | | | | 0.029 | | 1 |
| | **ANBE27** | | | | | | | | | | | 0.007 | | | 1 |
| **L6** | **ANBE7** | 0.083 | 0.080 | 0.118 | 0.196 | 0.122 | 0.076 | 0.076 | 0.141 | 0.098 | | | | | 9 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **ANBE19** | | | | | | | | | | | 0.094 | 0.100 | 2 |
| | **ANBE37** | | | | | | | | | | 0.073 | | | 1 |
| | **ANBE40** | | | | | | | | | 0.068 | | | | 1 |
| **L7** | **ANBE26** | 0.009 | | | | | | | 0.009 | | | | | 2 |
| | **ANBE32** | | | | | | | | | 0.021 | | | | 1 |
| **L8** | **ANBE4** | 0.028 | | 0.117 | 0.077 | 0.120 | 0.148 | 0.088 | | | 0.153 | 0.080 | 0.076 | 9 |
| | **ANBE36** | | | | | | | | | 0.027 | | | | 1 |
| | **ANBE33** | | | | | | | | 0.042 | | | | | 1 |
| **L9** | **ANBE3** | 0.033 | 0.032 | 0.031 | 0.028 | 0.040 | 0.036 | | 0.034 | 0.015 | 0.021 | 0.046 | 0.048 | 0.034 | 12 |
| | **ANBE31** | | | | | | | 0.034 | | | | | | | 1 |
| **L10** | **ANBE11** | | 0.053 | 0.021 | | 0.038 | | 0.029 | | | | 0.008 | 0.025 | 0.020 | 7 |
| | **ANBE23** | | | | | | | | 0.054 | 0.050 | | | | | 2 |
| | **ANBE5** | 0.065 | | | | | | | | | | | | | 1 |
| **L11** | **ANBE16** | | | 0.03 | 0.013 | 0.02 | | 0.02 | | | | | 0.019 | 0.01 | 6 |
| | **ANBE38** | | | | | | | | | | | 0.01 | | | 1 |
| | **ANBE29** | | | | | | 0.02 | | | | | | | | 1 |

**Table S2.** Output from a sliding window analysis created to detect gene conversion in Berthelot's pipit MHC haplotypes (see text for details). Donor haplotypes identified by the macro that we had previously identified as ancestral are highlighted in bold.

| Haplotype | Similarity to CMH (%) | 1st donor | Similarity to 1st donor (%) | Minimum insert size 1st donor (bp) | Similarity to 1st construct sequence (%) | 2nd donor | Similarity to 2nd donor (%) | Insert size 2nd donor (bp) | Similarity to 2nd construct sequence (%) |
|---|---|---|---|---|---|---|---|---|---|
| **Putative ancestral** | | | | | | | | | |
| ANBE10 | 99.16 | **ANBE2** | 87.25 | 6 | 100 | | 87.25 | | 100 |
| ANBE11 | 97.90 | **ANBE16** | 96.36 | 7 | 99.16 | **ANBE9** | 89.78 | 3 | 99.58 |
| ANBE16 | 99.58 | **ANBE11** | 96.36 | 3 | 100 | | 96.36 | | 100 |
| ANBE9 | 99.58 | **ANBE10** | 96.64 | 3 | 100 | | 96.64 | | 100 |
| ANBE6 | 99.86 | **ANBE16** | 91.18 | 1 | 100 | | 91.18 | | 100 |
| ANBE8 | 95.80 | ANBE35 | 87.68 | 17 | 98.46 | ANBE13 | 90.90 | 5 | 99.16 |
| ANBE2 | 93.28 | ANBE42 | 92.44 | 35 | 98.74 | **ANBE1** | 91.18 | 8 | 99.86 |
| ANBE1 | 99.58 | ANBE39 | 97.34 | 3 | 100 | | 97.34 | | 100 |
| ANBE7 | 99.58 | **ANBE6** | 90.76 | 3 | 100 | | 90.76 | | 100 |
| ANBE26 | 99.58 | **ANBE3** | 83.61 | 3 | 100 | | 83.61 | | 100 |
| ANBE4 | 97.76 | **ANBE1** | 81.93 | 12 | 99.58 | | 81.93 | | 99.58 |
| ANBE3 | 99.58 | **ANBE1** | 81.51 | 3 | 100 | | 81.51 | | 100 |
| **Putative derived** | | | | | | | | | |
| ANBE13 | 98.32 | **ANBE8** | 90.90 | 10 | 100 | | 90.90 | | 100 |
| ANBE18 | 99.16 | **ANBE6** | 90.48 | 6 | 100 | | 90.48 | | 100 |
| ANBE28 | 99.16 | **ANBE2** | 88.10 | 5 | 100 | | 88.10 | | 100 |
| ANBE23 | 99.72 | **ANBE11** | 97.90 | 2 | 100 | | 97.90 | | 100 |
| ANBE5 | 99.72 | **ANBE16** | 96.22 | 2 | 100 | | 96.22 | | 100 |
| ANBE29 | 97.76 | **ANBE2** | 93.28 | 12 | 100 | | 93.28 | | 100 |
| ANBE38 | 99.58 | **ANBE3** | 90.20 | 3 | 100 | | 90.20 | | 100 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| ANBE30 | 99.58 | **ANBE6** | 91.74 | 3 | 100 | | 91.74 | | 100 |
| ANBE35 | 99.58 | **ANBE10** | 96.64 | 3 | 100 | | 96.64 | | 100 |
| ANBE12 | 99.86 | **ANBE10** | 89.78 | 1 | 100 | | 89.78 | | 100 |
| ANBE14 | 98.74 | **ANBE6** | 96.78 | 7 | 100 | | 96.78 | | 100 |
| ANBE15 | 98.46 | **ANBE8** | 94.12 | 8 | 99.58 | **ANBE38** | 90.34 | 3 | 100 |
| ANBE17 | 99.16 | **ANBE1** | 85.29 | 6 | 100 | | 85.29 | | 100 |
| ANBE21 | 99.58 | **ANBE8** | 93.42 | 3 | 100 | | 93.42 | | 100 |
| ANBE22 | 98.74 | ANBE39 | 82.49 | 5 | 99.72 | **ANBE2** | 90.76 | 2 | 100 |
| ANBE24 | 99.58 | **ANBE11** | 92.86 | 3 | 100 | | 92.86 | | 100 |
| ANBE25 | 99.58 | **ANBE10** | 90.06 | 3 | 100 | | 90.06 | | 100 |
| ANBE34 | 99.72 | **ANBE11** | 94.12 | 2 | 100 | | 94.12 | | 100 |
| ANBE42 | 99.58 | **ANBE2** | 92.44 | 3 | 100 | | 92.44 | | 100 |
| ANBE27 | 99.58 | | | | 99.58 | | | | 99.58 |
| ANBE39 | 97.34 | **ANBE10** | 86.13 | 16 | 100 | | 86.13 | | 100 |
| ANBE41 | 98.32 | **ANBE2** | 92.86 | 5 | 100 | | 92.86 | | 100 |
| ANBE19 | 98.60 | **ANBE10** | 90.20 | 9 | 100 | | 90.20 | | 100 |
| ANBE37 | 98.60 | **ANBE2** | 89.78 | 9 | 100 | | 89.78 | | 100 |
| ANBE40 | 99.58 | | | | 99.58 | | | | 99.58 |
| ANBE32 | 99.58 | **ANBE10** | 88.52 | 3 | 100 | | 88.52 | | 100 |
| ANBE33 | 97.76 | **ANBE10** | 84.03 | 15 | 100 | | 84.03 | | 100 |
| ANBE36 | 97.06 | **ANBE8** | 87.68 | 17 | 99.58 | **ANBE16** | 90.20 | 2 | 100 |
| ANBE31 | 99.58 | **ANBE10** | 84.31 | 3 | 100 | | 84.31 | | 100 |
| **Mean (all haplotypes)** | **98.85** | | **90.33** | **6** | **99.85** | | **90.39** | **4** | **99.93** |
| **Mean (derived haplotypes)** | **99.04** | | **90.07** | **6** | **99.93** | | **90.30** | **2** | **99.97** |

**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). All sequences were confirmed by subsequent 454 sequencing (see main text).

| **TF1** | **TF2** | **TF3** | **TF4** | **TF5** | **M1** |
|---|---|---|---|---|---|
| ANBE 10 | ANBE 10 | ANBE 28 | ANBE 10 | ANBE 10 | ANBE 10 |
| ANBE 2 | ANBE 9 | ANBE 8 | ANBE 8 | ANBE 8 | ANBE 9 |
| ANBE 8 | ANBE 8 | ANBE 24 | ANBE 6 | ANBE 4 | ANBE 8 |
| ANBE 24 | ANBE 6 | ANBE 2 | ANBE 2 | ANBE 2 | ANBE 23 |
| | ANBE 19 | | | | ANBE 2 |
| | ANBE 2 | | | | |
| | ANBE 4 | | | | |

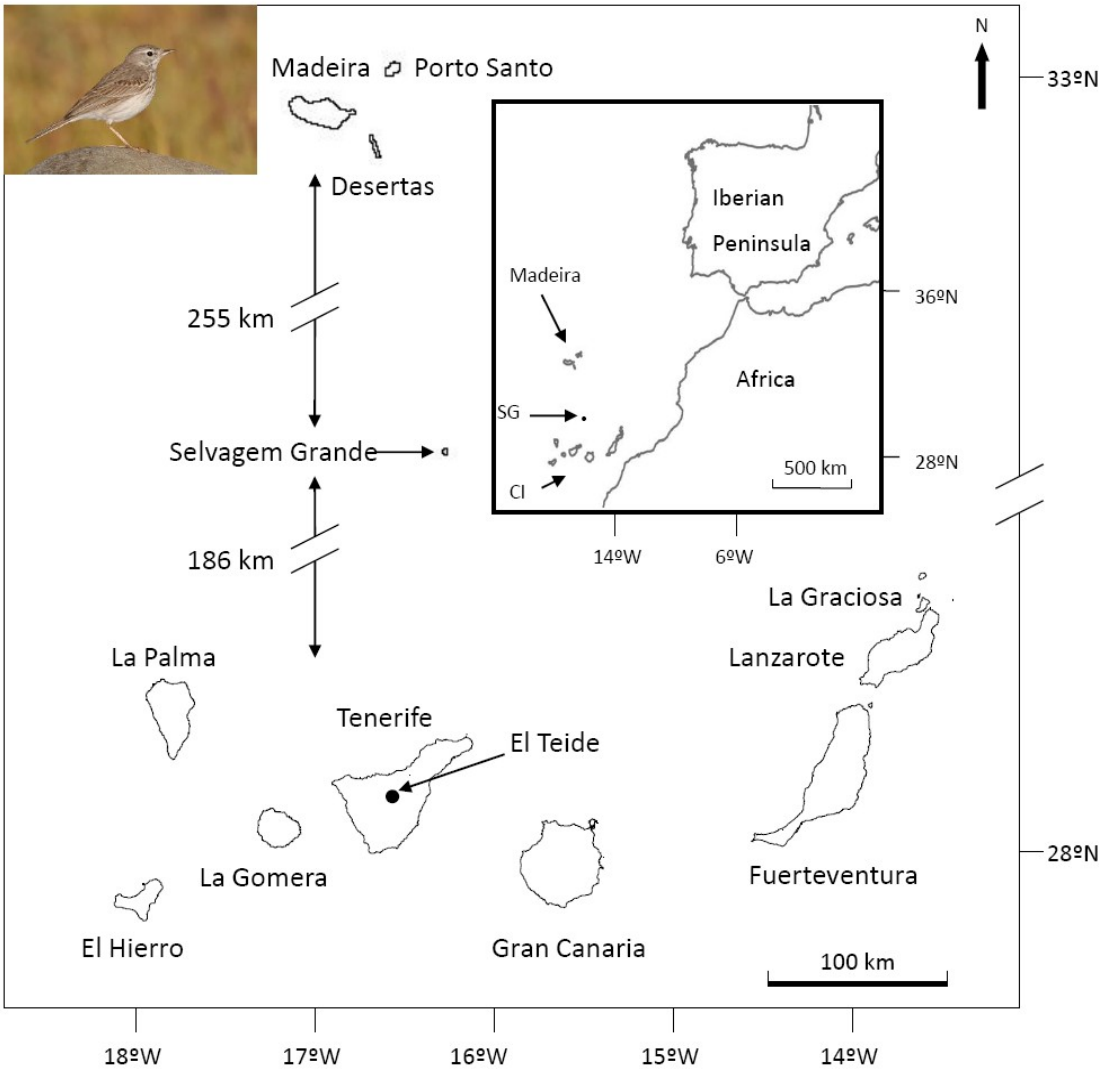



**Figure S1.** Distribution and sampling locations of Berthelot's pipits (inset) in the North Atlantic. SG, Selvagem Grande; CI, Canary Islands. Adapted from Illera *et al.* (2007).

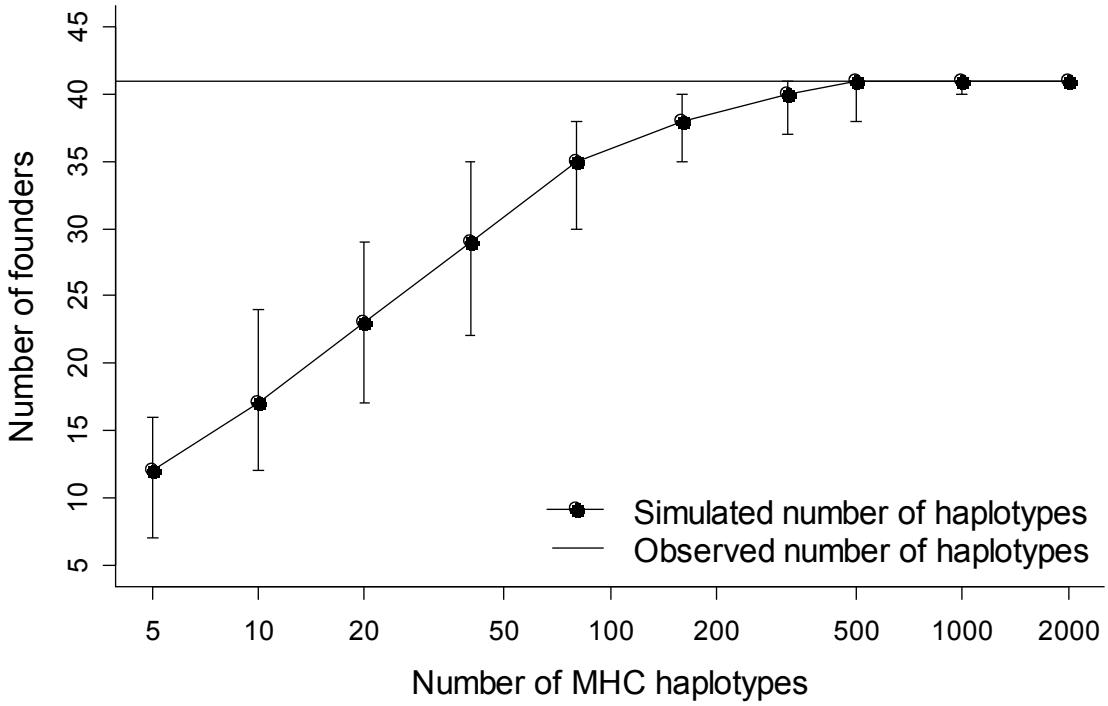



**Figure S2.** Mean (± 99% CI) number of MHC haplotypes in the pipit metapopulation as a function of the number of founders. The estimates are based on a simulation of 4 MHC genes (8 haplotypes) per individual drawn from the contemporary gene pool with the haplotype frequencies based on those observed across the entire range of pipits (see Table S1). The simulations make the following conservative assumptions: equal sex ratio, no sperm storage and no genetic drift after the founder event.

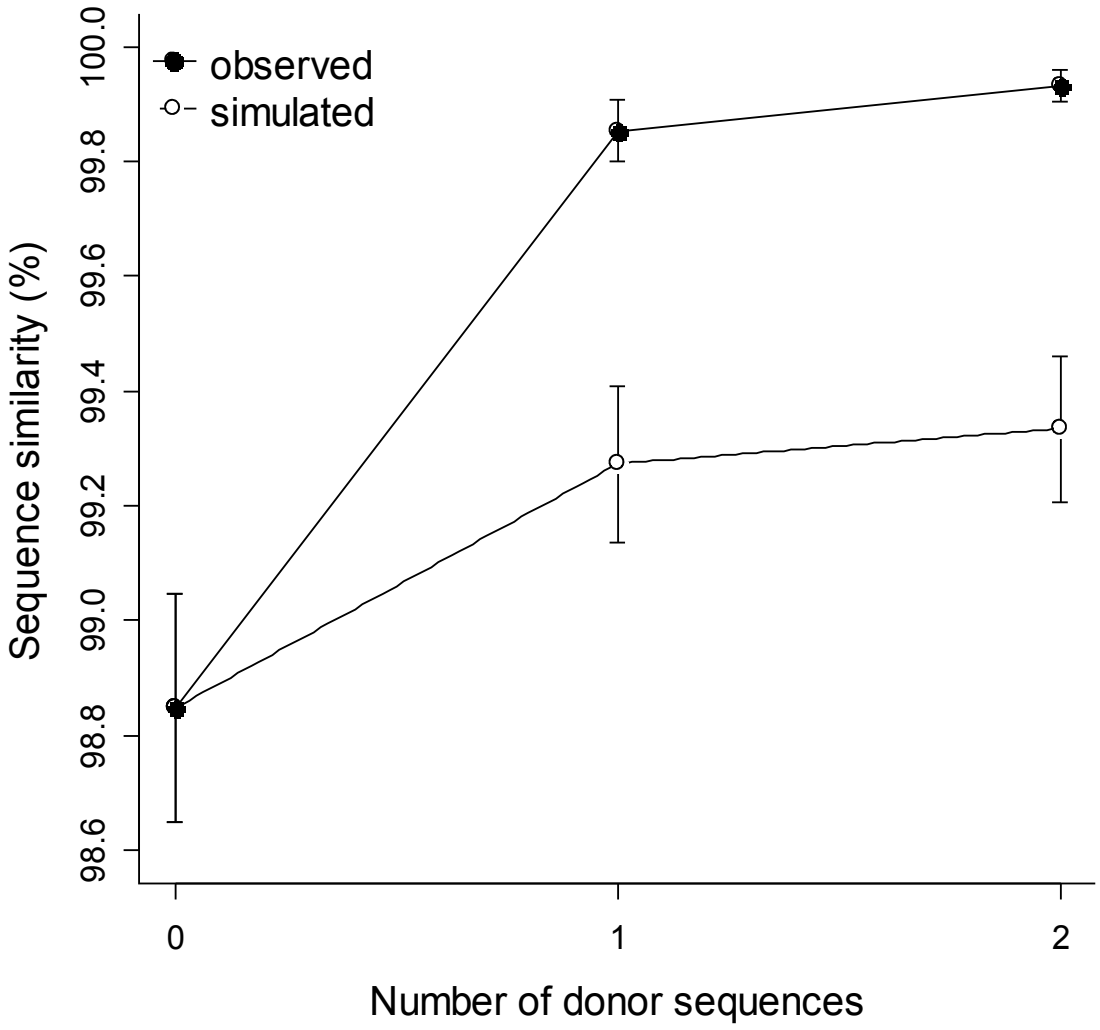
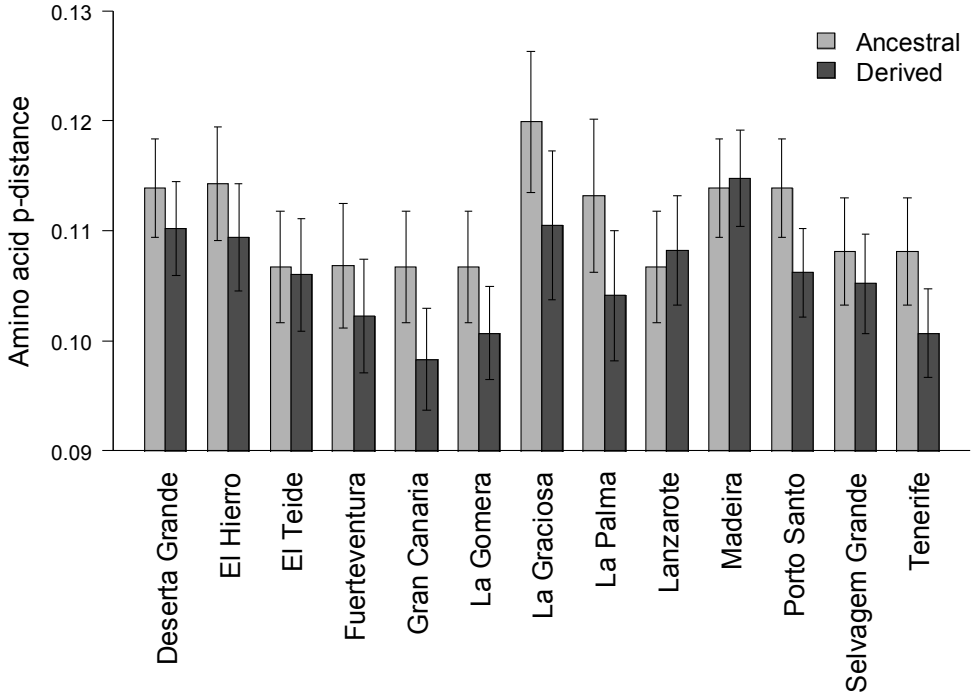



**Figure S3.** Convergence and mean (± s.e.) sequence similarity between MHC haplotypes. Each sequence was matched to its closest relative, and percentage similarity was calculated using a sliding window analysis (see methods). This value was then re-calculated allowing for the transfer of sections of DNA from one or two donor haplotypes from within the dataset. The analysis was performed on Berthelot's pipit MHC class I exon 3 haplotypes (observed), and sequences that were simulated to evolve via point mutation (see appendix S2). Improvement in sequence similarity was greater for the empirical dataset after convergence allowing for both one and two donor haplotypes (Wilcoxon tests: one donor, $P = 2.3 \times 10^{-4}$; two donors, $P = 3.8 \times 10^{-5}$).

**a)**




**b)**

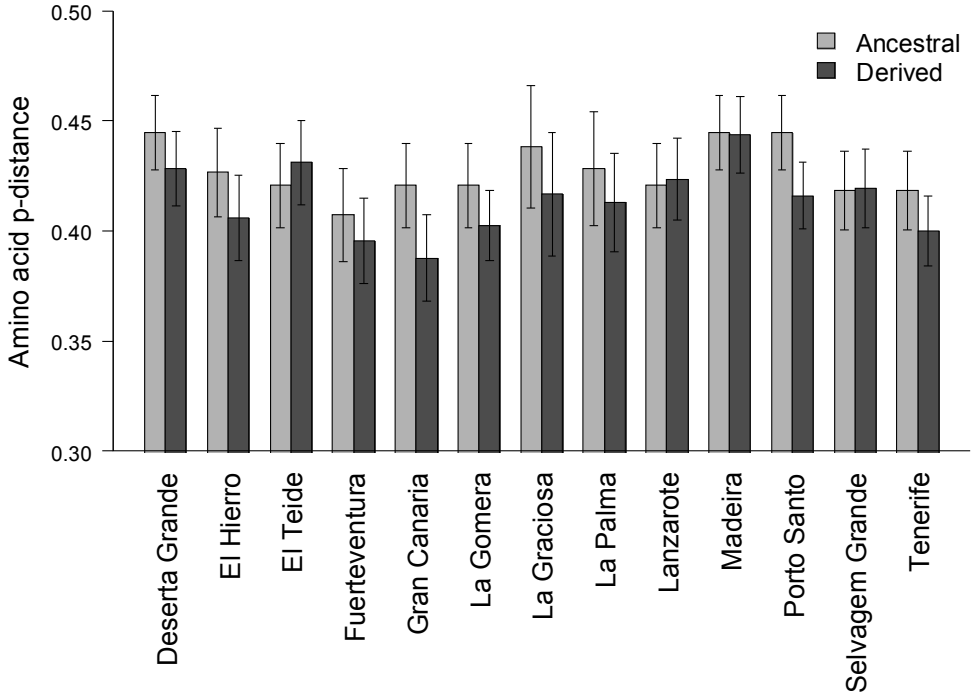



**Figure S4.** Changes in mean amino acid p-distance between all pairwise combinations of haplotypes in populations for amino acids coded by the **a)** non-PBR and **b)** PBR codons. The pale bars show the p-distances of populations in which all derived haplotypes have been replaced by their ancestral form. This removes the effect of micro-recombination. The dark bars show the p-distances based on all observed haplotypes. The rate of loss is lower for the PBR sites (3.48%) than for the non-PBR sites (4.50%), (paired t-test: $t = 2.48$, $P = 0.024$).