Immunogenetics (2001) 53:792–801
DOI 10.1007/s00251-001-0393-9

ORIGINAL PAPER

**Akie Sato · Werner E. Mayer · Herbert Tichy**
**Peter R. Grant · B. Rosemary Grant · Jan Klein**

# Evolution of *Mhc* class II *B* genes in Darwin's finches and their closest relatives: birth of a new gene

Received: 17 September 2001 / Revised: 26 October 2001 / Published online: 19 December 2001


**Abstract** The 15 extant species of Darwin's finches on the Galápagos and Cocos Islands are the products of an unfinished adaptive radiation from a founder flock of birds related to the South American species *Tiaris obscura*. Molecular characterization of their major histocompatibility complex (*Mhc*) class II *B* genes has revealed the existence of several related groups of sequences (presumably encoded in distinct loci) from which one (group 5) stands out because of its low divergence over extended time periods. Analysis of group 5 exon 2 and intron 2 sequences has revealed that the encoding locus apparently arose 2–3 million years ago in the *Tiaris* group of South and Central American Thraupini. The locus shows no evidence of inactivation, but displays a very low degree of polymorphism, both in terms of number of alleles and genetic distances between alleles. Some of the polymorphism, however, appears to be trans-specific. All the observed intergenic differences can be explained by point mutations and most of the exon 2 changes represent nonsynonymous substitutions, although the rate of nonsynonymous and synonymous substitutions appears to be the same. The origin of the new locus is explained by the birth-and-death model of *Mhc* evolution with two important extensions. First, the ancestor of the group 5 genes may have arisen without new gene duplication and second, the birth of the new group may have been brought about by a switch from balancing to directional selection. The ancestor of the group 5 genes may have been a classical class II *B* allele (one of many) which directional selection fixed in the ancestral population and drove into the category of nonclassical genes.

A. Sato (✉) · W.E. Mayer · H. Tichy · J. Klein
Max-Planck-Institut für Biologie, Abteilung Immungenetik,
Corrensstrasse 42, 72076 Tübingen, Germany
e-mail: akie.sato@tuebingen.mpg.de
Tel.: +49-7071-601290, Fax: +49-7071-600437

P.R. Grant · B.R. Grant
Department of Ecology and Evolutionary Biology,
Princeton University, Princeton, NJ 08544-1003, USA

**Keywords** Darwin's finches · *Mhc* · Birth-and-death model · Nonclassical class II gene · Evolution

## Introduction

As a rule, the major histocompatibility complex, *Mhc*, of a species consists of multiple loci which are of different ages (Klein and Figueroa 1986). Thus, for example, in the human *Mhc*, the *HLA* complex, there is a group of classical, class Ia loci whose origin postdates the divergence of Platyrrhini and Catarrhini some 30 million years (my) ago, a group of nonclassical, class Ib loci which may have arisen following the divergence of primates from other eutherian orders 70 my ago, and a group of class II loci which emerged close to the origin of the eutherian mammals >120 my ago (Hughes and Nei 1989; Klein et al. 1993a). Within each of these three groups, different loci arose at different times after the appearance of the first gene of that group.

The age differences among the *Mhc* loci are explained by the accordion (Klein et al. 1993b) or the birth-and-death (Nei and Hughes 1992; Nei et al. 1997) models of *Mhc* evolution. In both models, the *Mhc* is viewed as a highly unstable chromosomal region (or regions) subject to expansions by gene duplications and contractions by gene deletions. This instability leads to the creation (birth) of new genes and the demise (death) of old genes. Some of the new genes assume (or retain) functions associated with the classical *Mhc* loci – they acquire the ability to present a wide spectrum of peptides, including important foreign, parasite-derived peptides, to the T-cell receptor of the lymphocytes, an ability normally conditioned on their polymorphism. The two characteristic features of the polymorphism are, first, a large number of alleles and second, great genetic distances between some of the alleles (Klein et al. 1993c). The time needed for the development of these two features is generally much longer than the life span of a species (Klein 1980). Allelic lineages are thus generated which are passed from ancestral to descendant species along a phylogene-

tic line of descent and are maintained by balancing selection acting on the peptide-binding region (PBR) sites (Hughes and Nei 1988).

The specifics of the birth-and-death process are not well understood. What, for example, are the reasons for the turnover of the classical loci? From which category of genes are the new classical loci recruited? From particular alleles of old classical loci or from nonclassical loci? What are the dynamics of the replacement of the old by the new loci? What forces the replacement? Do the old and the new loci coexist functionally over a time interval and if so, for how long? On what time scale does the replacement take place?

These and related questions cannot be answered by studying the traditional *Mhc* model systems, if for no other reason than because the evolutionary distances between the *Mhcs* of the different taxa are simply too large to reveal the replacement mechanics. More than 2×70 my of evolution separate the human and the mouse *Mhc* systems and more than 2×300 my of evolution separate both from the *Mhc* of the domestic fowl. The *Mhc* systems of the two nearest living relatives of *Homo sapiens*, the chimpanzee and the gorilla, are separated from the *HLA* complex by 2×5–8 my of evolution and the rat *Mhc* from the mouse *H2* system by a time gap of 2×20 my. Over such long time intervals, the clues to the events underlying the "change of guards" can be expected to be all but obliterated.

Gaining an insight into the birth of new functional loci requires turning to groups of species generated by recent adaptive radiations and examining the newborns for telltale signs of participation in the replacement process. In our laboratories, we have focused on two classical examples of adaptive radiation in vertebrates, the cichlid fishes of the Great African Lakes (reviewed by Klein et al. 1997) and Darwin's finches of the Galápagos/Cocos Island Archipelago (Sato et al. 2000; Vincek et al. 1997). The 15 living species of Darwin's finches (Grant 1986; Lack 1947) are the descendants of a flock of >30 birds (Vincek et al. 1997) that reached the Galápagos Islands ~2.3 my ago (Sato et al. 2001) on a nonstop flight from the west coast of South America, covering a distance of at least 960 km. The closest living relatives of Darwin's finches on the South American continent and the Caribbean islands are members of the tribe Thraupini (Sato et al. 2001), subfamily Emberizinae, family Fringillidae (Sibley and Ahlquist 1990). Mitochondrial (mt) DNA analysis identifies the genus *Tiaris*, the grassquits, and specifically the dull-colored grassquit, *Tiaris obscura*, as the nearest extant relative of Darwin's finches (Sato et al. 2001). Upon reaching the Galápagos Archipelago, the ancestors of Darwin's finches began to adapt to the different ecological niches available on the volcanic islands, and in the process diverged into different groups currently represented by the ground finches (genus *Geospiza*), tree finches (genera *Camarhynchus* and *Cactospiza*), the vegetarian finch (*Platyspiza*), and warbler finches (*Certhidea*). Subsequent to the colonization, a group of the ancestors reached the Cocos Island at a distance of 720 km northeast of the Galápagos Archipelago to evolve into the Cocos finch, *Pinaroloxias inornata*. mtDNA (Freeland and Boag 1999a, 1999b; Sato et al. 1999, 2001) and microsatellite DNA studies (Petren et al. 1999) support the monophyly of the 15 extant species of Darwin's finches, relative to the continental representatives of the Thraupini. The main adaptations of the radiating ancestral populations were to different types of food and resulted in a divergence in body and beak size and shape. In earlier publications (Sato et al. 2000; Vincek et al. 1997) we described some of the class II *B* genes coding for the β chain of the different class II αβ dimers. In the present study, we focus on one of the loci identified in the analysis, provide evidence for its recent origin, and relate its emergence to the birth-and-death model of *Mhc* evolution.

## Materials and methods

### Birds, blood samples, and DNA extraction

The sources of blood samples and the manner in which they were collected have been described in previous publications (Sato et al. 1999, 2000, 2001). Generally, 10–20 µl of blood was obtained from wing veins, the blood was stored in an AS-Buffer (Qiagen Blood Kit; Qiagen, Hilden, Germany), and the Qiagen kit was also used for DNA extraction. Genomic DNA was then amplified by using appropriate oligonucleotide primers (Table 1) and the amplification products cloned and sequenced.

### Polymerase chain reaction

The polymerase chain reaction (PCR) conditions were as follows: one cycle of denaturation for 30 s at 94°C, annealing for 15 s at the annealing temperature, and extension for 7 min at 72°C, followed by 34 cycles of denaturation for 15 s at 94°C, annealing for 15 s at the annealing temperature, and extension for 1 min at 72°C, and a final extension for 7 min at 72°C. In each reaction, 2 µl of genomic DNA, 0.2 mM of each of the four deoxyribonucleotides, 0.5 µM of each of the sense and antisense primers, 2.5 units of *Taq* polymerase (Amersham Pharmacia Biotech, Freiburg, Germany), and 0.4 units *Pfu* DNA polymerase (Stratagene, Heidelberg, Germany) were added to 10 µl of 5×PCR buffer. Hot-start PCR was carried out using HotWax 3.5 mM $Mg^{2+}$ beads (Invitrogen, Leek, The Netherlands). The DNA was amplified in the GeneAmp PCR System 9700 (AB Applied Biosystems, Weiterstadt, Germany) or the PTC-200 Thermal Controller (Biozym Diagnostik, Hess. Oldendorf, Germany).

### Cloning of PCR products

Forty microliters of PCR product was purified with the aid of the QIAquick Gel Extraction Kit (Qiagen). The eluted DNA was blunt-ended, phosphorylated, ligated to *Sma*I-digested pUC18 plasmid vector with the help of the SureClone ligation kit (Amersham Pharmacia Biotech), and used to transform *Escherichia coli* XL-1 blue competent bacteria.

### DNA sequencing

Double-stranded DNA was prepared with the help of the QIAGEN Plasmid Kit and 0.5 µg of the DNA was then used in the dideoxy chain-termination cycle sequencing reaction. The Thermo Sequenase Fluorescent Labeled Primer Cycle Sequencing Kit was used

**Table 1** Finch *Mhc* class II *B* primers. The primer position is specified from the border between exon 2 and intron 2 in either the 5′ (negative values) or the 3′ (positive numbers) direction (*E* exon, *I* intron, *S* sense, *AS* antisense)

| Name | Specificity (class II) | Position | Sequence | Orientation | Annealing temperature (°C) |
|---|---|---|---|---|---|
| HOPE1 | E2 | –252 to –224 | 5′-GAAAGCTCGAGTGTCACTTCACGAACGGC-3′ | S | 58 |
| HOPE3 | E2 | –241 to –215 | 5′-TGTCACTTCACGAACGGCACGGAGAAG-3′ | S | 58 |
| HOPE2 | E2 | –50 to –22 | 5′-GGGTGACAATCCGGTAGTTGTGCCGGCAG-3′ | AS | 58 |
| H4 | I2 | 132 to 160 | 5′-GTTTGGATTGGGAATGGTTTGGGACGAAC-3′ | AS | 58 |
| FA | E3 | 84 to 111 | 5′-ACCTCACCTGGATCGGGGCAGGGTAGAA-3′ | AS | 58 |
| HOPE4 | E2 | –50 to –22 | 5′-GGCGGAACACCTCGTAGTTGTGCCGGCAG-3′ | AS | 58 |
| HOPE6 | E2 | –43 to –15 | 5′-CTGAACGGGCGGAACACCTCGTAGTTGTG-3′ | AS | 58 |
| HOPE10 | E2 | –34 to –9 | 5′-TCCACGCTGAACGGGCGGAACACCTC-3′ | AS | 58 |
| HOPE50 | E2 | –53 to –36 | 5′-TAGTTGTGCCGGCAGTAC-3′ | AS | 58 |
| HOPE72 | I2 | 525 to 550 | 5′-GAGTGGGTGTAGGTTTGAGAGTGAAC-3′ | AS | 58 |
| HOPE103 | E2 | –238 to –215 | 5′-CACTTCACGAACGGCACGGAGAAG-3′ | S | 58 |
| HOPE102 | E2 | –42 to –19 | 5′-ACGGGCGGAACACCTCGTAGTTGT-3′ | AS | 58 |
| HOPE111 | E2 | –186 to –170 | 5′-CCGGCAGCAGTACGTGA-3′ | S | 50 |
| HOPE112 | E2 | –85 to –69 | 5′-GCATGTACTTCAGTCTG-3′ | AS | 50 |
| HOPE24 | I2 | 70 to 97 | 5′-AGGGAATATACTGGAAGCGACTGGTAGG-3′ | AS | 58 |
| HOPE20 | I2 | 81 to 107 | 5′-TCGGAGGGTCAGGGAATATACTGGAAG-3′ | AS | 58 |
| HOPE22 | I2 | 88 to 115 | 5′-GAATGGACTCGGAGGGTCAGGGAATATA-3′ | AS | 58 |

with Thermo Sequenase DNA polymerase to carry out the reaction, which was then processed by the LI-COR 4200 Long ReadIR DNA sequencer (MWG Biotech, Ebersberg, Germany). To confirm the sequences, 5–10 µg of DNA was used for the dideoxy chain-termination sequencing reaction carried out with the Auto-Read Sequencing Kit (Amersham Pharmacia Biotech), which contained 5′ fluorescent-labeled sequencing primers and T7 DNA polymerase. The reactions were processed by the Automated Laser Fluorescent (A.L.F.) Sequencer (Amersham Pharmacia Biotech). Although the possibility that some of the sequenced products may have been derived from heteroduplex DNA could not be excluded, this was unlikely because most of the sequences were reproduced from independent cloning.

### Phylogenetic analysis

The sequences were aligned with the help of the SeqPup software, version 0.6f (Gilbert 1995). Numbers and rates of synonymous and nonsynonymous substitutions were estimated using MEGA version 2.1 (Kumar et al. 2001) and the Nei-Gojobori method with Jukes-Cantor correction. Phylogenetic trees were drawn with the help of the MEGA software or the PAUP*4.0b8 software for the Macintosh (Swofford 2001), using the neighbor-joining algorithm (Saitou and Nei 1987) and *p*-distances for amino acid sequences. The reliability of the branching order was determined by 500 bootstrap replications.

## Results

In our earlier studies (Sato et al. 2000; Vincek et al. 1997), which were concentrated on the Darwin's finches themselves, we identified four groups of *Mhc* class II *B* sequences (groups 1–4), each group presumably representing a distinct locus and each displaying a high degree of polymorphism. In an effort to trace the origin of this polymorphism, we subsequently extended the study to the closest relatives of Darwin's finches living in South America and on the Caribbean islands. The results of this extended study will be reported in a separate publication (A. Sato, W.E. Mayer, H. Tichy, P.R. Grant, B.R. Grant, J. Klein, unpublished data). Here, we describe an additional group of sequences (group 5) distinguished from all the other groups by certain interesting properties.

Initially, the group 5 sequences, like most of the other sequences in this study, were obtained by PCR amplification of genomic DNA using the pair of primers HOPE1 and HOPE10 (Table 1). The product of this amplification encompassed 189 bp of sequence covering the middle, most variable part of exon 2 (Figs. 1, 2). Later, we used other pairs of primers to characterize group 5 more closely. The HOPE1-HOPE10 pair amplified not only group 5 sequences, but also sequences of other groups in an unpredictable and uncontrollable manner. To restrict the specificity of the amplification to group 5 sequences, we used the primer combinations HOPE1-HOPE20, HOPE1-HOPE22, HOPE1-HOPE24, and HOPE111-HOPE112 (Table 1). In the first three combinations, the second primer of each pair was complementary to a portion of the group 5 intron 2 sequence – a different portion in each pair. The three pairs did indeed amplify group 5 sequences of Darwin's finches and of five of the related Caribbean species (*Melanospiza richardsoni*, *T. bicolor*, *T. canora*, *Loxigilla noctis*, and *Poospiza hispaniolensis*). DNA of *T. obscura* could not be amplified even though it could be shown to contain group 5 sequences by the use of other primer pairs. Apparently, the intron 2 sequences of this species are too divergent from those of the Darwin's finches to allow amplification under the conditions used. Primers HOPE111 and HOPE112 (Table 1) were complementary to group 5-specific stretches of exon 2 sequences and hence primed amplification of group 5 genes specifically. However, the amplification product was only 81 bp long, so that the primers could be used only to ascertain the presence or absence of group 5 genes in the different species and individuals, but not for the characterization of the genes.

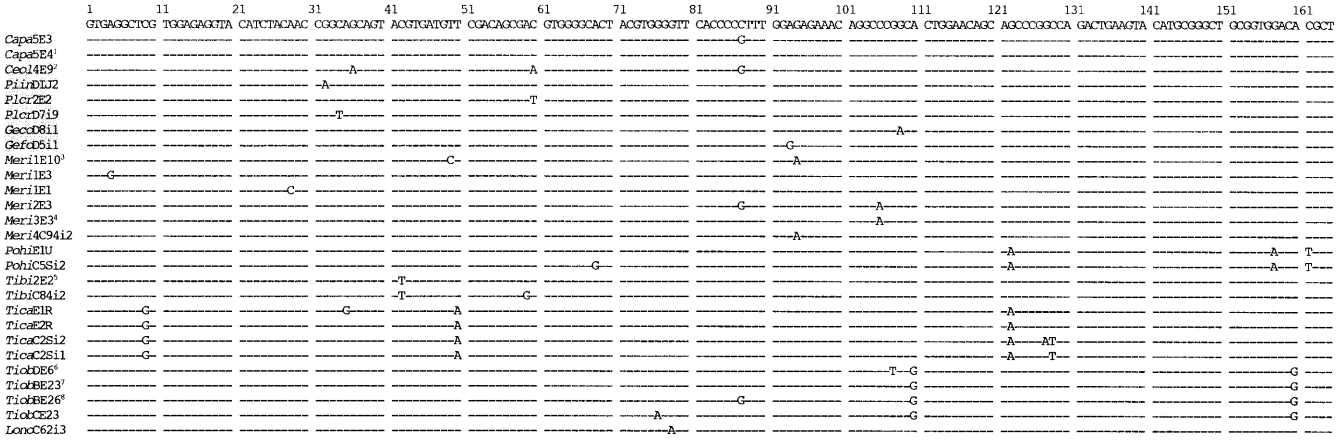

**Fig. 1** Nucleotide sequences of group 5 exon 2 genes. Site 1 in the numbering system used corresponds to site 57 from the 5′ end of the exon. The sequence at the top is a simple majority consensus of all the group 5 sequences. Identity with the consensus is indicated by a dash (–). The distinct sequences are numbered. Identical sequences were found in the following species and individuals: [1]*Capa*4E1, *Capr*HE3, *Capr*E1, *Capr*GE14, *Capu*E1, *Gesc*KE26, *Gesc*KE4, *Gesc*KE27, *Piin*DIJ1, *Piin*DIK10, *Piin*DIK11, *Piin*DIJ3, *Lono*E4, *Meri*1E2, *Meri*2E1, *Plcr*2E8, *Tibi*2E1, *Plcr*D7i8, *Plcr*D7i6, *Gefu*D6i1, *Gema*D1i1, *Gedi*D12i1, *Capu*D10Ai1, *Meri*1C9i4, *Meri*1C9i2, *Meri*4C94i1, *Lono*C62i1; [2]*Ceol*D44i1; [3]*Meri*3E1, *Meri*1C9i9, *Meri*1C9i3; [4]*Meri*1C9i6; [5]*Tibi*1E4, *Tibi*C84i1; [6]*Tiob*AE23 *Tiob*FE12; [7]*Tiob*FE23, *Tiob*DE24, *Tiob*FE19, *Tiob*GE25, *Tiob*CE25; [8]*Tiob*GE22, *Tiob*GE18. Species abbreviations are as follows: *Capa*, *Cactospiza pallida*; *Capr*, *Camarhynchus parvulus*; *Capu*, *Camarhynchus pauper*; *Ceol*, *Certhidea olivacea*; *Geco*, *Geospiza conirostris*; *Gedi*, *Geospiza difficilis*; *Gefo*, *Geospiza fortis*; *Gefu*, *Geospiza fuliginosa*; *Gema*, *Geospiza magnirostris*; *Gesc*, *Geospiza scandens*; *Lono*, *Loxigilla noctis*; *Meri*, *Melanospiza richardsoni*; *Piin*, *Pinaroloxias inornata*; *Plcr*, *Platyspiza crassirostris*; *Pohi*, *Poospiza hispaniolensis*; *Tibi*, *Tiaris bicolor*; *Tica*, *Tiaris canora*; *Tiob*, *Tiaris obscura*. (Darwin's finches are *highlighted*.) The sequences are deposited in the GenBank database under the following accession codes: AY064421–AY064463

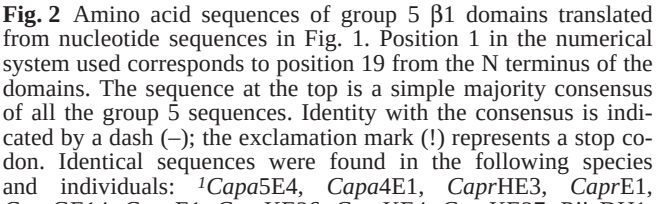

**Fig. 2** Amino acid sequences of group 5 β1 domains translated from nucleotide sequences in Fig. 1. Position 1 in the numerical system used corresponds to position 19 from the N terminus of the domains. The sequence at the top is a simple majority consensus of all the group 5 sequences. Identity with the consensus is indicated by a dash (–); the exclamation mark (!) represents a stop codon. Identical sequences were found in the following species and individuals: [1]*Capa*5E4, *Capa*4E1, *Capr*HE3, *Capr*E1, *Capr*GE14, *Capu*E1, *Gesc*KE26, *Gesc*KE4, *Gesc*KE27, *Piin*DIJ1, *Piin*DIK10, *Piin*DIK11, *Piin*DIJ3, *Lono*E4, *Meri*1E2, *Meri*2E1, *Meri*2E3, *Meri*3E3, *Plcr*2E2, *Plcr*2E8, *Tibi*2E2, *Tibi*1E4, *Tibi*2E1, *Geco*D8i1, *Plcr*D7i8, *Plcr*D7i6, *Gefu*D6i1, *Gefo*D5i1, *Gema*D1i1, *Gedi*D12i1, *Capu*D10Ai1, *Meri*1C9i6, *Meri*1C9i4, *Meri*1C9i2, *Meri*4C94i1, *Tibi*C84i1, *Lono*C62i3, *Lono*C62i1; [2]*Meri*3E1, *Meri*1C9i3, *Meri*1C9i9; [3]*Tiob*AE23, *Tiob*FE12; [4]*Tiob*BE26, *Tiob*GE22, *Tiob*GE18, *Tiob*FE23, *Tiob*DE24, *Tiob*FE19, *Tiob*GE25, *Tiob*CE25. (Darwin's finches are *highlighted*.) The species abbreviations are as in Fig. 1. Amino acid residues are given in the IUPAC-IUB single-letter code. Asterisks (*) indicate putative PBR sites according to Westerdahl and co-workers (2000)

Finally, to confirm the distinctiveness of group 5 sequences from the sequences of the other groups, we used the primer combination HOPE1-FA (Table 1), in which the FA primer is complementary to a stretch of sequence in the 5′ part of exon 3, so that the amplification products encompassed most of exon 2, the entire intron 2, and part of exon 3. The reason for including intron 2 in the amplification was that it could be shown to be of different length in the distinct loci (Sato et al., unpublished data). By this approach, group 5 sequences could indeed be shown to be distinguished by an intron 2 of a characteristic length and composition (Fig. 3). However, the HOPE1-FA primer combination primed the amplification of not only group 5 sequences, but also sequences of other groups, so that obtaining the former turned out to be largely a matter of persistence and luck.

The distinctiveness of group 5 sequences from those of the other groups is indicated by three observations. First, on phylogenetic trees based on 359 *Mhc* class II *B* exon 2 sequences from 15 Thraupini species (Sato et al. 2000, Sato et al., unpublished data; Vincek et al. 1997), the group 5 sequences form a single compact clade, well separated from all the other sequences and supported by a bootstrap value of 100% (not shown, but a tree including group 5 sequences and representative sequences of the four other groups is presented in Fig. 4). The separateness of the clade is a reflection of a series of nucleotide substitutions and amino acid replacements (L3, V15, K33, Q34, R36, H37, S41, R44, K46, Y47, M48, and A50 in Fig. 2) which distinguish this group from other groups. Second, group 5 genes possess a distinctive intron 2 which is short compared to the introns of genes in the other groups (267 bp, in contrast to >2 kb of the longest introns in some of the other groups) and is

**A**

```
                1          11         21         31         41         51         61         71         81         91
 Consensus* GTGAGCGCGG GGCAGAGCGT GTGCCCTCGG GCCCTGCCCC TGCCAGTGAC CATCCCAGTC CAATCCCAGT CAATTCCCAG TCCCTCCCAG TCCATTCCCA
GemaD1i1    ------A--- ---------- --C------- ---------* ---------- --------C- *--------- -CC-*A---- -----T---- G.........
GecoD8i1[1] ---------- ---------- --C------- ---------* ---------- --------C- *--------- -CC-*A---- -----T---- ..........
GecoD8i6    ---------- ---------- --C------- ---------* ---------- --------C- *--------- -CC-*A---- -----T---- -AT------T
CapaD3i5    ---------- ---------- --C------- ---------* ---------- --------C- *--------- -CC-*A---- -----T---- -AT------T
CapuD10Ai1[2] ------A--- ---------- --C------- ---------* ---------- --------C- *--------- TCC-*A---- --G--T---- ..........
Piin1DIJ1   ------A--- ---------- --C------- ---------* ---------- --------C- *--------- -CC-*A---- --G--T---- -AT------T
Piin1DIJ2[3] ------A--- ---------- --C------- ---------* ---------- --------C- *--------- -CC-*A---- -----T---- -AT------T
Piin2DIK10  ---------- ---------- --C------- ---------* ---------- --------C- *--------- -CC-*A---- --G--T---- -AT------T
PlcrD7i8    ---------- ---------- --C------- ---------* ---------- --------C- *--------- -.......... .......... ..........
PlcrD7i9    ------A--- ---------- --C------- ---------* ---------- --------C- *--------- -.......... .......... ..........
CeolD44i6   ------A--- ---------- --C------- ---------* ---------- --------C- *--------- -CC-*A---- -----T---- -AT------T
TicaC2Si8   ------A-A- -------A-- --C------A- ---------* ---------- --------C- *--------- -CC-*A---- -----T---- -AT------T
TicaC2Si1[4] ------A-A- -------A-- --C------- ---------* ---------- --------C- *--------- -CC-*A---- -----T---- ..........
TibiC84i1[5] ---------- ---------- --C------- ---------* ---------- --------C- *--------- -CC-*A---- -----T---- ..........
Meri1C9i4   -----T---- ---------- --C------- ---------* ---------- --------C- *--------- -CC-*A---- --G--T---- ..........
Meri1C9i6[6] ---------- ---------- --C------- ---------* ---------- --------C- *--------- -CC-*A---- --G--T---- ..........
LonoC62i1   ---------- ---------- --C------- ---------* ---------- ---------- *--------- -CC-*A---- -----T---- ..........
LonoC62i3   ---------- ---------- --C------- ---------* ---------- ---------- *--------- -CC-*A---- .......... ..........
PohiC5Si2   ---------- ---------- --C------- ---------* ---------- --C------- *--------- -CCC*A---- -----T---- ..........
```

**B**

Here follows a sequence of ~2 kb present in the consensus but absent in group 5

```
          101        111        121        131        141        151        161        171        181        191
 Consensus* GTCCCTTCCC AGTCCATTCC CTGACCCTTC CAGTCCATTC CCTGACCCTC CCAGTCCATT CCCTGACCCT CCCAGTCCAT TCCCAGTCCC TCCCAGTCC
GecoD8i6    -A----*--G ---------- -A-T--A--- -CAGT-CAC- --AAT--A-- ------T--- ---A-T--T- --------** ********** *********
CapaD3i5    -A----*--G ---------- -A-T--A--- -CAGT-CAC- --AAT--A-- -------C-C --AGTC-AT- --------** ********** *********
Piin1DIJ1   -A----*--G ---------- -A-T--A--- -CAGT-CAC- --AAT--A-- ------T--- ---A-T--T- --------** ********** *********
Piin1DIJ2[7] -A----*--* ---------- -A-T--A--- -CAGT-CAC- --AAT--A-- ------T--- ---A-T--T- --------** ********** *********
CeolD44i6   -A----*--A ---------- -A-T--A--- -CAGT-CAC- --AAT--A-- ------T--- ---A-T--T- --------C- *--------T ---------
TicaC2Si8   -A----*--G ---------- -A-T--A--- -CAGT-CAC- --AAT--A-T ------T--- ---A-T--T- --------C- C--AGTC-AT ---------
```

**C**

```
          2241       2251       2261       2271       2281       2291       2301       2311       2321       2331
 Consensus* CCCTCCCCAG CGCACCCCAG CCCTCCCCTC TCTCTCCCAG TGCCCCCCAG CTGATCCCAG TGTGTCCCCG CTCTCTCTCT CTCTGTCCCC CAG
GecoD8i6    ------**-- T-----T--- ---------- ---------- ---------- ---------- ---------- ---------- -----**---- ---
CapaD3i5    ------**-- ---------- --------*- ---------- ---------- ---------- ---------- ------T--- -------T-- ---
Piin1DIJ1   ------**-- ---------- ---------- ---------- ---------- ---------- ---------- ---------- -------T-- ---
Piin1DIJ2   ---C--**-- ------T--- ---------- ---------- --T------- ---------- --------T- ------T--- -------T-- ---
Piin1DIJ3   ------**-- ------T--- ---------- ---------- --T------- ---------- --------T- ------T--- -------T-- ---
Piin2DIK10  ------**-- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ****---T-- ---
CeolD44i6   ------**-- ------T--- ---------- ---------- ---------- ---------- ---------- ------T--- -------T-- ---
TicaC2Si8   -T----**-- ------T--- ---------- ---------- ---------- ---------- ------T--- ---C------ ****C--T-- ---
```

*Consensus of groups 1-4.

**Fig. 3** Nucleotide sequences of group 5 intron 2 obtained from different species of Darwin's finches (highlighted) and their close relations: 5′ part of intron 2 (**A**), continuation of 5′ part (**B**), 3′ part of intron 2 (**C**). In groups 1–4, an extra sequence is present between the parts depicted in **B** and **C** (indicated by a *box*); this sequence is not shown here. The single majority consensus sequence at the top is based on intron 2 sequences of other groups which will be reported elsewhere (Sato et al.,unpublished data). Numbering starts with the first site of intron 2. Identity with the consensus sequence is indicated by a dash (–); asterisks (*) stand for insertions/deletions introduced to optimize the alignment, and dots (.) for unavailability of sequence information. Identical sequences were found in the following species and individuals: [1]*Gefu*D6i1, *Gefo*D5i1, *Gedi*D12i1; [2]*Plcr*D7i6; [3]*Piin*1DIJ3; [4]*Tica*C2Si2; [5]*Meri*4C94i1, *Meri*4C94i2, *Tibi*C84i2; [6]*Meri*1C9i2, *Meri*1C9i3, *Meri*1C9i9; [7]*Piin*1DIJ3, *Piin*2DIK10. Species abbreviations are explained in Fig. 1

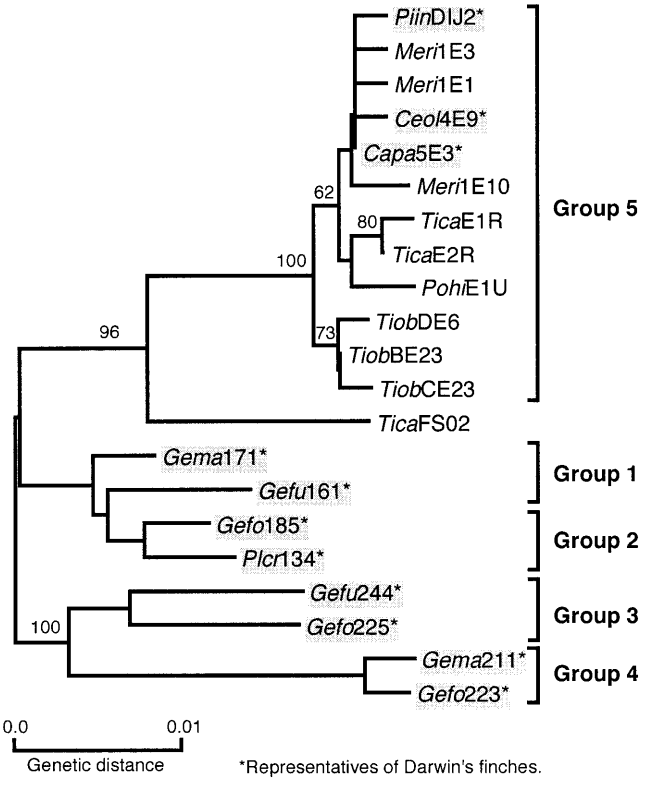


**Fig. 4** Neighbor-joining tree based on group 5 amino acid sequences from Fig. 2, and of representative sequences from groups 1–4, as well as the sequence (*Tica* FS02) that is closest to group 5 in the tree based on the complete collection (Sato et al., unpublished data). Darwin's finches are *highlighted*. Numbers placed at nodes are bootstrap values. Species abbreviations are explained in Fig. 1

marked by characteristic substitutions (Fig. 3). The 5′ part of the intron in group 5 genes (~170 bp in length) is alignable with the 5′ part of intron 2 sequences of other genes and similarly, the 3′ part (~80 bp) is alignable with the 3′ parts of genes belonging to other groups, but the middle part present in other groups is missing in group 5 sequences. Third, the group 5 sequences are distinguished from those of other genes by the characteristic features described below.

Group 5 sequences were obtained from 12 species of Darwin's finches (*Certhidea olivacea*, *Platyspiza crassirostris*, *Geospiza scandens*, *G. magnirostris*, *G. difficilis*, *G. conirostris*, *G. fortis*, *G. fuliginosa*, *Pinaroloxias inornata*, *Camarhynchus pauper*, *C. parvulus*, and *Cactospiza pallida*), as well as from six species of Central/South American Thraupini (*P. hispaniolensis*, *L. noctis*, *M. richardsoni*, *T. canora*, *T. bicolor*, and *T. obscura*). Since the 12 tested species of Darwin's finches represent the different groups and genera, we assume that all the Darwin's finches possess the group 5 locus. The six positively testing Central/South American species cluster together on a phylogenetic tree based on mtDNA cytochrome *b* sequences (Fig. 5; see Sato et al. 2001), five of them in a clade that includes all of the Darwin's finches, and one (*P. hispaniolensis*) outside of the clade, although its outlier position is supported by low bootstrap values. An additional nine species traditionally classified as Thraupini (*Sporophila nigricollis*, *Ory-zoborus angolensis*, *Diglossa humeralis*, *Sicalis flaveola*, *Volatinia jacarina*, *Coereba flaveola*, *Euphonia musica*, *Coryphospingus cucullatus*, and *Rhamphocelus carbo*; but see Fig. 5), as well as representatives of Cardinalini (*Cyanocompsa parellina*, *Pheucticus aureoventris*), Parulini (*Dendroica adelaidae*), Emberizini (*Zonotrichia capensis*, *Atlapetes rufinucha*), Icterini (*Sturnella bellicosa*), Tyrannidae (*Capsiempis flaveola*, *Elaenia martinica*), and Carduelini (*Carduelis magellanica*) tested negative for the group 5 locus (Fig. 5). Also negative were species representing more distant taxa: *Columba squamosa*, *Geotrygon montanum*, *Phaethornis yaruqui*, *Piaya minuta*, *Pachyrhamphus cinnamoneum*, and *Stelgidopteryx ruficollis* (not shown). Since the tests that yielded negative results were quite extensive and employed a variety of primer combinations, we conclude that the group 5 sequences represent a new locus which arose in the ancestors of the clade that gave rise to Darwin's finches and their closest relatives in Central and South America. Although only a small number of individuals could be tested from the various species, the entire collection as a whole gives no indication that there are group 5-negative individuals in any of the positive species. In other words, the group 5 locus appears to be fixed in all the group 5-positive species.

In the collection of birds tested and the class II *B* sequences obtained, the one closest to the group 5 sequences (but well separated from them) appears to be the *T. canora* sequence FS02 (Fig. 4), which shares four diagnostic amino acid residues with group 5 (V15, Q34, R44, and Y47; Fig. 2). The FS02 sequence is not a member of any of the defined groups but rather part of a cluster of solitary sequences characterized by long branch lengths (Sato and co-workers, unpublished data). Whether this or some other sequence in the loose cluster can be considered to be related directly to the immediate ancestor of the group 5 sequences is uncertain, however.

The most conspicuous feature of the group 5 sequences is their relative invariance compared to the sequences of the other groups. The group 5 amino acid sequences of Darwin's finches are virtually identical (Fig. 2) and the exon 2 nucleotide sequences show only three substitutions differentiating the warbler finch from the other Darwin's finch sequences (Fig. 1). This invariability is in stark contrast to the high variability of the *Mhc* class II *B* exon 2 sequences in other groups. In the initial surveys, we found 21, 18, 13, and 8 different sequences in groups 1, 2, 3, and 4, respectively, all of them differentiated by multiple substitutions (Vincek et al. 1997). Although some of these groups may encompass more than one locus, the loci redefined by subsequent surveys all proved to be highly polymorphic, both in terms of the number of alleles and the genetic distances between the alleles (Sato et al., unpublished data). Even more remarkable is the observation that some of the group 5 exon 2 sequences of the South/Central American species (e.g., *L. noctis*, *M. richardsoni*, and *T. bicolor*) are identical with the Darwin's finch sequences (Fig. 1), even though the latter diverged from the former >2.3 my ago (Sato et al. 2001). The invariance is reflected in the shallowness of the branching pattern in the phylogenic tree based on the group 5 exon 2 sequences (Fig. 4).

The low variability observed among the group 5 exon 2 sequences is of two kinds: apparently fixed differences between species and polymorphisms within species. The former is exemplified by the 110 A→G and the 160 A→G substitutions in *T. obscura* which appear to be fixed in this species (Fig. 1), although the number of individuals tested is too small to place high confidence in this conclusion. Hints of species-specific substitutions are also seen in the *T. bicolor* and *T. canora* sequences. Whether fixed or on the way to fixation, the variability observed in these three species suggests that the substitutions arose a sufficiently long time ago to attain high frequencies in the populations.

The remaining exon 2 variability observed in the group 5 sequences appears to represent polymorphism, some of which is shared between species and other which is restricted to a given species. Of the two observed cases of trans-species polymorphism – G at site 88 in *C. pallida*, *C. olivacea*, *M. richardsoni*, and *T. obscura*, as well as A at site 122 in *P. hispaniolensis* and *T. canora* (Fig. 1) – the former represents a synonymous and the latter a nonsynonymous substitution. If the synonymous substitution is regarded as representing a neutral change, its persistence as a polymorphism for a time period in excess of 2.3 my is surprising, especially in view of the conclusion that the size of the founding population of Darwin's finches was probably <100 individuals. Alternative explanations for this polymorphism

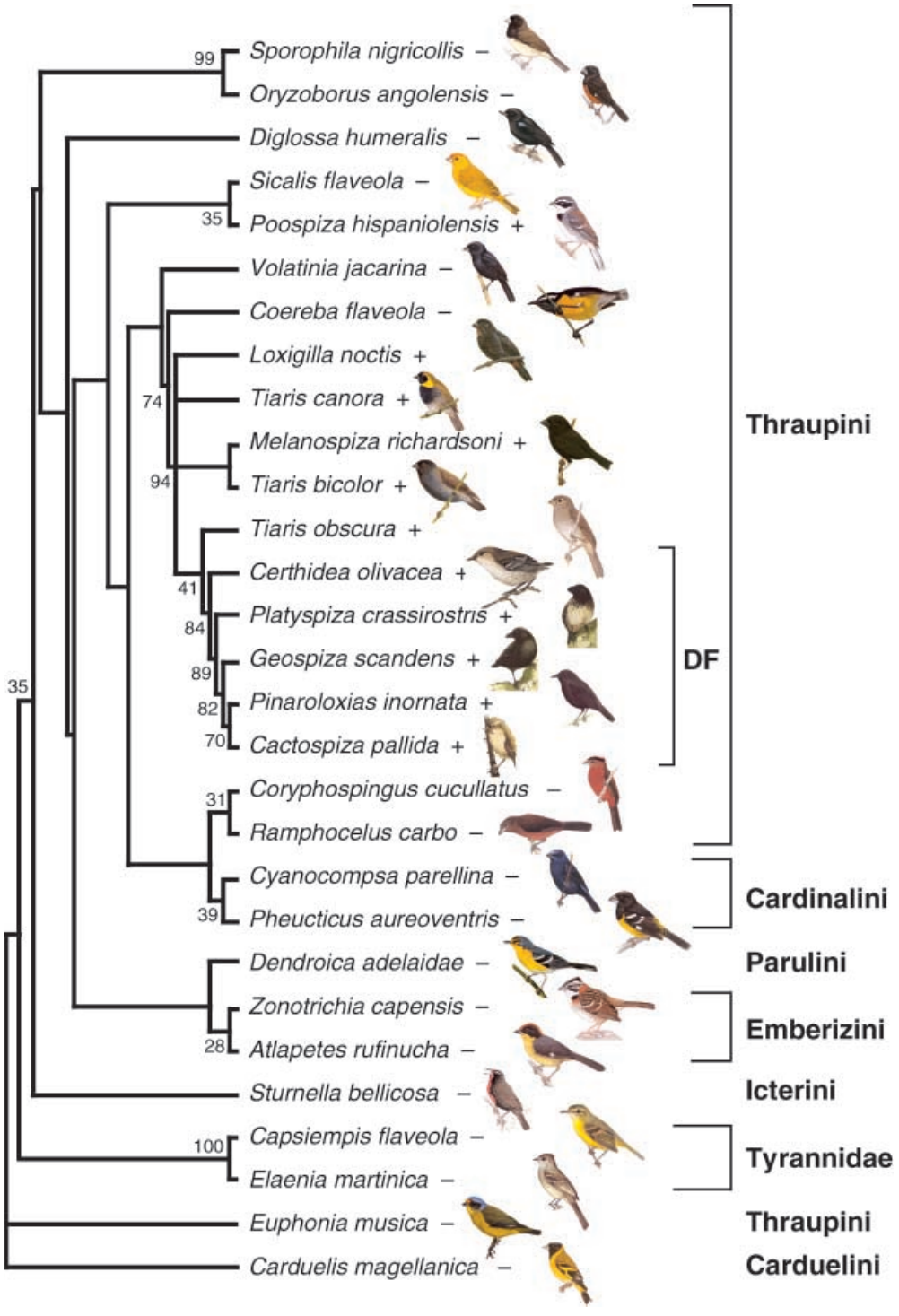


**Fig. 5** Presence (+) or absence (–) of the group 5 locus among the Thraupini and related taxa. The maximum-parsimony tree is based on mtDNA control region sequences reported by Sato and co-workers (2001). The bird icons were taken from Steadman and Zousmer (1989), Stiles and Skutch (1995), Howell and Webb (1995), Bond (1993), and Ridgely and Tudor (1989) (*DF* Darwin's finches)

are either that the substitution is neutral but linked to a locus under selection; that despite its synonymity, the substitution is not neutral; or that the substitution occurred independently in different species. We find no support for the possibility that the observed changes are the result of systematic amplification or sequencing errors.

The majority (75%) of the observed variation in the group 5 exon 2 sequences is of the nonsynonymous type. Since, however, the synonymous sites in this part of exon 2 are in a minority (23%), the calculated overall synonymous distance (0.0136) is approximately the same as the nonsynonymous distance (0.0128), and both are similar to the mean distance between intron 2 sequences (0.0155). However,, because of the overall sparsity of substitutions, these estimates have a large standard error and therefore do not provide a reliable basis for drawing conclusions about the involvement of selec-

tion in the evolution of group 5 sequences. Similarly, the distribution of the substitutions shows no clearcut preference for the putative PBR sites, although some of the changes do occur in the otherwise highly variable parts of exon 2. Here again, however, it must be pointed out that the location of the PBR sites in bird class II genes is uncertain and that the most polymorphic sites do not always coincide with predicted PBR sites. Interestingly, the substitutions that are shared by all group 5 sequences and that distinguish them from other groups of sequences are all at sites that are highly variable in these other groups. This observation suggests that the group 5 locus might be derived from a gene with a function of "classical" gene.

Whether the group 5 sequences themselves are functional could not be established with certainty because the restrictions placed on the collection of material on the Galápagos Islands and in South America and the circumstances under which the material was collected did not allow us to obtain tissue samples suitable for RNA isolation or for more extensive genomic analysis. All we can say at this stage is that none of the sequences gave any indication of gene inactivation (the putative stop codon in the *Plcr*D7i9 sequence has been found in a single clone only) and that some of the characteristics of the sequences suggest that the group 5 locus has, at least until recently, been active.

## Discussion

The group 5 sequences described in the present study represent *Mhc* class II *B* genes at a single locus which originated 2–3 my ago during the adaptive radiations that produced, on the one hand, the *Tiaris* clade of Thraupini in South and Central America and on the other, the Darwin's finches of the Galápagos and Cocos Islands. The locus arose from a single gene that had the functional properties of a classical class II *B* gene in that its variable part encompassing the PBR had accumulated characteristic substitutions distinguishing it from other genes. This accumulation presumably occurred under the influence of balancing selection. The ancestral gene itself was apparently a member of a distinct group of functional, polymorphic class II *B* loci.

The point in the adaptive radiation at which the group 5 locus arose remains uncertain. Its presence in *P. hispaniolensis* indicates either that the phylogeny shown in Fig. 5 is unreliable or that the locus was lost in several lineages independently. Neither of these two possibilities can be excluded at present. The branching pattern of the clades in Fig. 5 is supported by low bootstrap values (Sato et al. 2001) and hence cannot be regarded as firmly established. On the other hand, the selection pressure behind the emergence of locus 5 (see below) may have existed in some lineages but not in others, and the loss of the locus (e.g., by unequal crossing over) in the latter lineages could therefore have occurred repeatedly.

The origin of the group 5 locus can be envisioned in one of two ways. First, one of the genes in the ancestral group duplicated and the copy then became fixed in the ancestral species or its descendants (=the classical birth-and-death scenario). Second, the selection on one of the genes in the ancestral group switched from the balancing to the directional type and as a consequence, the gene swept through the population replacing all other alleles. The outcomes of the two scenarios could be expected to differ. In the first scenario, the ancestral group might be expected to persist with the new copy and the new group derived from this gene would appear as a subset of the old group. In the second scenario, the ancestral group would be ablated and the new group would not appear to be directly related to any of the other groups. Since the phylogenetic analysis shows that group 5 stands alone as a separate clade in the collection of 359 class II *B* sequences obtained from the sampled Thraupini birds, the second scenario seems to provide a more plausible explanation for the origin of the founder of the group 5 sequences.

Assuming that the fixation of one of the genes in the ancestral group occurred by selection rather than by random genetic drift in a bottleneck population, what might have been responsible for the shift from balancing to directional selection? Balancing selection at *Mhc* loci is believed to be driven by the need to provide an organism with the ability to recognize and respond immunologically to an assortment of parasites, each product of a series of allelic genes being able to accommodate a certain range of peptides and all the products together covering the entire spectrum of parasites a species may encounter (Doherty and Zinkernagel 1975; Takahata 1995). Directional selection, on the other hand, is expected to be driven by a small group of parasites or even a single parasite which may have become particularly abundant and threatened the survival of the species. This is the explanation commonly offered to account for the existence of nonclassical genes, some of whose products have been shown to specialize in the presentation of a narrow range of peptides, including those derived from a particular group of parasites (Fischer-Lindahl et al. 1997). The next question therefore is: are the group 5 sequences derived from a nonclassical class II locus?

The term "nonclassical" in this context is usually applied to class I loci and specifically to the mammalian class Ib loci characterized by low polymorphism and an unorthodox mode of expression (Fischer-Lindahl et al. 1997). Nonclassical class II loci have not been described, to our knowledge, but we find no compelling reason why they should not exist. Superficially, the group 5 class II *B* locus seems to resemble a nonclassical class I locus in its low polymorphism. The nature of the polymorphism of the group 5 locus, too, is similar to that of the mammalian class Ib loci in that the genetic distances between alleles are small and the substitutions are scattered over the exon 2 sequences. Interpreting the group 5 locus as a nonclassical (class IIb) locus is therefore an obvious possibility.

Other possibilities, however, must also be considered. One is that the group 5 locus has assumed or is in the process of assuming a new function, either entirely or partially different from peptide (antigen) presentation. Several examples of such a change in function have been reported for both the class I and the class II loci (reviewed in Klein and Hořejší 1997). In the latter, the mammalian *DM* and *DO* genes, for example, have been shown to participate in protein processing rather than peptide binding and presentation (reviewed in van Ham et al. 2000; Alfonso and Karlsson 2000). Attempts to find any evolutionary connection between the group 5 locus and the *DM* or *DO* genes have failed (not shown), but the possibility that the group 5 locus is evolving toward a new function cannot be discounted entirely.

Another possibility is that the emergence of the group 5 locus represents a functional shift from one classical locus to another. Although the birth-and-death process is usually associated with gene duplication and deletion, we see no reason why it could not also occur without gene duplication, by one classical gene replacing all its alleles and becoming the ancestor of a new group of alleles. Arguing against this explanation of the origin of the group 5 sequences is the absence of any evidence for balancing selection acting on the presumptive PBR sites in these sequences. But in the early stages of evolution of a new classical locus, the signs of balancing selection may not be recognizable, particularly in view of the observation that the development of polymorphism characterizing the classical *Mhc* loci is a slow process (Satta et al. 1993) extending over long phylogenetic lineages (Klein et al. 1993c). Nevertheless, at the present state of knowledge, the interpretation that the group 5 sequences represent a nonclassical class II locus newly derived from a classical locus appears to be the most likely of the three possibilities considered.

**Acknowledgements** We thank Ms. Sabine Rosner for technical and Ms. Jane Kraushaar for editorial assistance; the Galápagos Park Services and the Charles Darwin Research Station at Galápagos, and the Smithsonian Tropical Research Institute for administrative and logistical support; as well as the National Science Foundation for financial support to B.R.G. and P.R.G.

## References

Alfonso C, Karlsson L (2000) Nonclassical *MHC* class II molecules. Annu Rev Immunol 18:113–142

Bond J (1993) Collins field guide: Birds of the West Indies, 5th edn. Harper Collins, Hong Kong

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

Fischer-Lindahl K, Byers DE, Dabhi VM, Hovik R, Jones EP, Smith GP, Wang CR, Xiao H, Yoshino M (1997) H2-M3, a full-service class Ib histocompatibility antigen. Annu Rev Immunol 15:851–879

Freeland JR, Boag PT (1999a) Phylogenetics of Darwin's finches: paraphyly in the tree finches, and two divergent lineages in the warbler finch. Auk 116:577–587

Freeland JR, Boag PT (1999b) The mitochondrial and nuclear genetic homogeneity of the phenotypically diverse Darwin's ground finches. Evolution 53:1553–1563

Gilbert DG (1995) SeqPup version 0.6f: a biosequence editor and analysis application. http://iubio.bio.indiana.edu/soft/molbio

Grant PR (1986) Ecology and evolution of Darwin's finches. Princeton University Press, Princeton, NJ

Ham M van, Lith M van, Griekspoor A, Neefjes J (2000) What to do with HLA-DO? Immunogenetics 51:765–770

Howell SNG, Webb S (1995) A guide to the birds of Mexico and northern Central America. Oxford University Press, Oxford

Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 355:167–170

Hughes AL, Nei M (1989) Evolution of the major histocompatibility complex: independent origin of nonclassical class I genes in different groups of mammals. Mol Biol Evol 6:559–579

Klein J (1980) Generation of diversity at *MHC* loci: implications for T-cell receptor repertoires. In: Fougereau M, Dausset J (eds) Immunology 80. Academic Press, London, pp 239–253

Klein J, Figueroa F (1986) Evolution of the major histocompatibility complex. CRC Crit Rev Immunol 6:295–386

Klein J, Hořejší V (1997) Immunology, 2nd edn. Blackwell, Oxford, pp 146–149

Klein J, O'hUigin C, Figueroa F, Mayer WE, Klein D (1993a) Different modes of *Mhc* evolution in primates. Mol Biol Evol 10:48–59

Klein J, Ono H, Klein D, O'hUigin C (1993b) The accordion model of *Mhc* evolution. In: Gergely J, Petranyi G (eds) Progress in immunology. Springer, Berlin Heidelberg New York, pp 137–143

Klein J, Satta Y, O'hUigin C, Takahata N (1993c) The molecular descent of the major histocompatibility complex. Annu Rev Immunol 11:269–295

Klein J, Klein D, Figueroa F, Sato A, O'hUigin C (1997) Major histocompatibility complex genes in the study of fish phylogeny. In: Kocher TD, Stepien CA (eds) Molecular systematics of fishes. Academic Press, New York, pp 271–283

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA 2: molecular evolutionary genetic analysis software. Arizona State University, Tempe, Ariz

Lack D (1947) Darwin's finches. Cambridge University Press, Cambridge, UK

Nei M, Hughes A (1992) Balanced polymorphism and evolution by the birth-and-death process in the *MHC* loci. In: Tsuji K, Aizawa M, Sasazuki T (eds) HLA 1991: Proceedings of the eleventh international histocompatibility workshop and conference, vol 2. Oxford University Press, Oxford, pp 27–38

Nei M, Gu X, Sitnikova T (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune system. Proc Natl Acad Sci USA 94:7799–7806

Petren K, Grant BR, Grant PR (1999) A phylogeny of Darwin's finches based on microsatellite DNA variation. Proc R Soc Lond B 266:321–329

Ridgely RS, Tudor G (1989) The birds of South America, vol 1. The oscine passerines. Oxford University Press, Oxford

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sato A, O'hUigin C, Figueroa F, Grant PR, Grant BR, Tichy H, Klein J (1999) Phylogeny of Darwin's finches as revealed by mtDNA sequences. Proc Natl Acad Sci USA 96:5101–5106

Sato A, Figueroa F, Mayer WE, Grant PR, Grant BR, Klein J (2000) *Mhc* class II genes of Darwin's finches: divergence by point mutations and reciprocal recombination. In: Kasahara M (ed) Major histocompatibility complex: evolution, structure, and function. Springer, Tokyo, pp 518–541

Sato A, Tichy H, O'hUigin C, Grant PR, Grant BR, Klein J (2001) On the origin of Darwin's finches. Mol Biol Evol 18:299–311

Satta Y, O'hUigin C, Takahata N, Klein J (1993) The synonymous substitution rate of the major histocompatibility complex loci in primates. Proc Natl Acad Sci USA 90:7480–7484

Sibley CG, Ahlquist JE (1990) Phylogeny and classification of birds: a study in molecular evolution. Yale University Press, New Haven, Conn

Steadman DW, Zousmer S (1989) Galápagos: discovery on Darwin's islands. Smithsonian Institution Press, Washington, DC

Stiles FG, Skutch AF (1995) A guide to the birds of Costa Rica. Helm, London

Swofford DL (2001) PAUP: phylogenetic analysis using parsimony (and other methods), version 4.0b8. Sinauer, Sunderland, Mass

Takahata N (1995) *Mhc* diversities and selection. Immunol Rev 143:225–247

Vincek V, O'hUigin C, Satta Y, Takahata N, Boag PT, Grant PR, Grant BR, Klein J (1997) How large was the founding population of Darwin's finches? Proc R Soc Lond B 264: 111–118

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