Major-Histocompatibility-Complex Variation in Two Species of Cichlid Fishes from Lake Malawi[1]

Hideki Ono, Colm O'hUigin,* Herbert Tichy,* and Jan Klein*,†*

*Max-Planck-Institut für Biologie, Abteilung Immungenetik, Tübingen; and †Department of Microbiology and Immunology, University of Miami School of Medicine

Lake Malawi in eastern Africa harbors >500 endemic species of cichlid fishes, all of which are believed to have emerged from a single founding population in the past 2 Myr. Molecular characterization of differences among the species could provide important information about the nature of speciation in the period of adaptive radiation. Because of the close relationship, however, molecular variation among the species has been difficult to ascertain. In this communication, we provide evidence for extensive differences, in major-histocompatibility-complex (*Mhc*) class II genes, between two related species, *Pseudotropheus zebra* and *Melanochromis auratus.* We used specific primers to amplify and sequence intron 1 and exon 2 of the class II genes from 18 individuals. Although we found 20 different sequences among the 42 that we produced, there was not a single sequence shared by the two species. Thus the study suggests that different cichlid species of Lake Malawi have different profiles of class II alleles, presumably because the polymorphism present in the ancestral founding population segregated differentially into the various species. These results make *Mhc* genes an important tool for elucidating speciation.

Introduction

Cichlids are bony fishes (teleosts) of the order Perciformes (perchlike fishes), family Cichlidae. They are characterized by a bilaterally compressed body, a single pair of nostrils, a single well-developed dorsal fin whose anterior rays protrude as spines, and a pair of bones (lower pharyngeals) in the floor of the throat (Fryer and Iles 1972; Keenleyside 1991). They occur naturally in Africa, Madagascar, South and Central America, Syria and the Jordan Valley, India, and Ceylon; they have been introduced by humans into other tropical regions, and they are maintained by aquarists all over the world. In several regions, cichlids have undergone recent adaptive radiations resulting in richness of species. Nowhere, however, have the radiations been as spectacular as in the Great Lakes of eastern Africa—Lake Malawi, Lake Tanganyika, and Lake Victoria. In Lake Malawi alone, there are >500 cichlid species, all except a few endemic (Ribbink 1984; Lewis et al. 1986; Eccles and Trewavas 1989). Most of the species of Lake Malawi are believed to have arisen in the past 2 Myr (the estimated age of the lake; Beadle 1981), possibly from a single, geographically isolated founding population (Eccles and Trewavas 1989; Meyer et al. 1990). Hence Lake Malawi, as well as other African lakes, can be viewed as natural laboratories for the study of speciation.

Until recently, most of the work describing speciation of African cichlids has

1. Key words: major histocompatibility complex, cichlid fishes, adaptive radiation, speciation.

Address for correspondence and reprints: Jan Klein, Max-Planck-Institut für Biologie, Abteilung Immungenetik, Correnstrasse 42, D-7400 Tübingen, Germany.

Mol. Biol. Evol. 10(5):1060–1072. 1993.
© 1993 by The University of Chicago. All rights reserved.
0737-4038/93/1005-0010$02.00

focused on morphological and behavioral characters (Greenwood 1981; Eccles and Trewavas 1989). In the past few years, however, a few studies have appeared in which biochemical traits, such as allozyme variation (Kornfield 1991) and mitochondrial DNA polymorphism (Meyer et al. 1990; Sturmbauer and Meyer 1992, and accepted), have been exploited. A serious limitation of the biochemical studies is proving to be the generally low variability of the loci, because of short divergence times; one exception to this is the major histocompatibility complex (*Mhc*).

The genes of the *Mhc* code for proteins that bind short peptides derived from other proteins inside the cell and that then display them on the cell surface (Klein 1986; Rothbard and Gefter 1991). In an uninfected cell or organism, the displayed peptides are derived from degraded self proteins; in an infection, some of the self peptides can be replaced by peptides derived from proteins of the invading parasite. The displayed nonself peptides are recognized by specialized receptors on thymus-derived lymphocytes, and the recognition initiates the specific arm of the immune response. In all vertebrates studied thus far, the *Mhc* consists of a large set of loci that structurally and functionally can be divided into two classes, I and II. In each class, the loci can further be divided into two subclasses, *A* and *B*, coding for the α and β polypeptide chains, respectively. The functional *Mhc* loci are highly polymorphic: A large number of alleles occurring at appreciable frequencies can be discerned at many of the loci, and the nucleotide diversity of the alleles is often considerable [it is not uncommon for two alleles to differ by >50 substitutions in the coding regions alone (Klein and Figueroa 1986)]. The *Mhc* polymorphism is known to evolve transspecifically: In mammals, some of the allelic lineages have been estimated to be >30 Myr old and are believed to have been passed through many speciation events (Klein 1980, 1987; Klein et al. 1993). This last feature makes the *Mhc* particularly valuable for the study of speciation and adaptive radiation. If the founding population of adaptively radiated species could be shown to be highly polymorphic at its *Mhc* loci, and if the polymorphism could be demonstrated to have segregated into these species, the *Mhc* could become a useful tool for elucidating the mechanism of speciation. In an earlier publication (Klein et al., accepted), we provided evidence for *Mhc* class II *B* gene polymorphism of the founding populations of African lake cichlids. Here we take the first step toward demonstrating segregation of the polymorphism in different species.

Material and Methods

Two cichlid species—*Pseudotropheus zebra* Boulenger, 1897 ("blue" or "red dorsal") and *Melanochromis auratus* Boulenger, 1897—were obtained from local dealers (Aquaristik H. Pelz, Bondorf, Germany, and Kölle Zoo, Stuttgart). DNA was isolated from the spleen and hepatopancreas of individual fishes by the standard method in an automated nucleic acid extractor (Applied Biosystems, Weiterstadt, Germany). Fragments of the isolated genomic DNA encompassing intron 1 and exon 2 were amplified by the polymerase chain reaction (PCR); the amplification products were cloned; and the clones were sequenced. The conditions of the PCR were as follows: denaturation for 2 min at 94°C, followed by 40 cycles of denaturation for 1 min at 94°C, annealing for 30 s at 55°C, and primer extension for 2 min at 72°C; the final extension was for 10 min at 72°C. The sequences of the primers were as follows: Tu383 (3′ end of exon 1)—5′-CTCTTCATCAGCCTCAGCACA-3′; Tu377 (degenerate primer, 3′ end of exon 2)—5′-TGATTTAGACAGA(G/A)(T/G)G(T/G)-(T/C)GCTGTA-3′; Tu422 (3′ end of exon 2, based on the *Auha*-M-231a sequence;

see Klein et al., accepted)—5′-TGATTTAGACAGAATGGCGCTGTA-3′; and Tu423 (3′ end of exon 2, based on the *Auha*-M-231b sequence)—5′-TGATTTAGACA-GAGGGTTGCTGTA-3′. The primers were used in the combination Tu383-Tu377 or Tu383-Tu422+Tu423.

The 335–457-bp-long PCR products were purified in low-melting-point agarose (GIBCO BRL, Eggenstein, Germany) and were cloned in pBluescript II plasmid vector (Stratagene, Heidelberg). The clones were sequenced by the dideoxy chain-termination method (Sanger et al. 1977) using the version 2.0 Sequenase Kit (U.S. Biochemical, Cleveland) and [^{35}S]dATPαS (Amersham Buchler, Braunschweig). Evolutionary relationships among the sequences were inferred from genetic distances estimated by using Kimura's (1980) two-parameter method and from a dendrogram constructed by using the neighbor-joining method (Saitou and Nei 1987). The sequences reported in this communication have been deposited in the GenBank under the accession numbers L17439–L17484.

Results

The assumption underlying the present study was that the large founding cichlid population in Lake Malawi contained a high number of alleles at its *Mhc* loci and that the emergence of individual species from this population was accompanied by segregation of these alleles. According to this assumption, cichlid species from Lake Malawi should be found to differ in the spectrum of *Mhc* genes that they possess. The differences could be either quantitative (in frequencies of shared *Mhc* genes), qualitative (in the presence or absence of certain *Mhc* alleles), or both. To test this assumption, we selected two widely distributed species from the same trophic group (epilithic algae feeders) and from a similar habitat (the rocky shores of Lake Malawi)—*Pseudotropheus zebra* and *Melanochromis auratus*—positing that, if differences can be found between them, differences between other species in different trophic groups and habitats would be even more likely to occur. We obtained six individuals of *P. zebra* ("blue"—individuals 1–3; "red dorsal"—individuals 189, 192, and 193) and 12 individuals of *M. auratus,* isolated genomic DNA from them, PCR-amplified segments of their *Mhc* class II *B* genes, and sequenced the cloned amplified products. The amplified segment encompasses intron 1 and almost the entire exon 2 (the primers used in the amplification were specific for sequences at the 3′ end of exons 1 and 2; see Klein et al., accepted). Its length varied from 335 bp to 457 bp, depending on the length of intron 1 in the different *Mhc* class II *B* genes. This segment was chosen because exon 2 was shown to be the most polymorphic part of coding class II *B* sequences (Ono et al. 1992, 1993; Klein et al., accepted) and because intron 1 of cichlid fishes consists of a 12-nucleotide element tandemly repeated a different number of times in different genes—a feature that serves as a convenient marker for the classification of genes into groups (Klein et al., accepted).

Altogether, we obtained 18 and 28 nucleotide sequences from *P. zebra* and *M. auratus,* respectively. The different intron 1 sequences are given in figure 1; the translation of the exon 2 sequences into amino acid sequences appears in figure 2. A dendrogram showing the relationships among the exon 2 sequences described in this study is given in figure 3.

Some of the sequences are undoubtedly those of alleles at the same locus. At the same time, however, some of the sequences must be derived from genes at different loci, because more than two different sequences were obtained from most of the tested individuals. [Unlike some other fishes that are believed to be of polyploid origin (Ohno

```
GROUP 1      1                                                                                                            110
CONSENSUS    GTATGATCGATATACTCCTCACTGATCAATACACTCATCAATACACTCCTTAATATACTGCAGAGATACACTGATCACCAATCACTGTACTGATCACTGGTCAATACACT
Psze-M-193g  --------------*******************-----------------------------------------------------------------------------
Meau-M-194d  --------------*******************-----------------------------------------------------------------------------
Meau-M-227a  --------------*******************-----------------------------------------------------------------------------

           111                                                                                                                   227
CONSENSUS    CATCACTAATCAGTATAGTGATCAATACATACAAATCCACTCATCAGCACACACATACACTCTCTCACTGGTTGTGATCAATACAGTGATCAGTCCTGATG TTCTGTTTGTTCTCAG
Psze-M-193g  -------------------------------------------------------------------------A--------------------------- ----------------
Meau-M-194d  -----------------------------G----------------------------------------------------------------------- ----------------
Meau-M-227a  -----------------------------G----------------------------------------------------------------------- ----------------

GROUP 2      1                                                                                                            110
CONSENSUS    GTACTATCAATATACACTGATTAATACACTGATCACTGATCAATACACTGCAGAGATACACTGATCACTTAACAGTACAGTGATCAATAAATGCATCACTGATCAAGTTA
Meau-M-194a  ---------**-------T-------------------------------------------------------------------------------------------
Meau-M-233b  ---------**-------T-------------------------------------------------------------------------------------------
Meau-M-234a  ---------**-------T-------------------------------------------------------------------------------------------
Meau-M-231a  ---------**-------T-------------------------------------------------------------------------------------------
Meau-M-230b  ---------**-------T-------------------------------------------------------------------------------------------
Meau-M-229e  ---------**-------T-------------------------------------------------------------------------------------------
Meau-M-233a  ---------**-------T-------------------------------------------------------------------------------------------
Meau-M-228a  ---------**-------T-------------------------------------------------------------------------------------------
Meau-M-227g  ---------**-------T-------------------------------------------------------------------------------------------
Psze-M-2d    ---------**---------------*******----A------------------------------------------------------------------------
Psze-M-1a    ----C-------------C------------------C-----------------------C-----------------------------------------------*
Psze-M-1c    ----C-------------C------------------C-----------------------C-----------------------------------------------*

           111                                                                  172
CONSENSUS    TTTGTACATCTGTACACAGCAGGTGTGATCAGCTCACAGTCTGATG TTGTGTGTGGTTCTGTTTGTCCTCAG
Meau-M-194a  ---------------------------------------------- ------**********----------
Meau-M-233b  ---------------T------------------------------ ------**********----------
Meau-M-234a  ---------------------------------------------- ------**********----------
Meau-M-231a  ---------------------------------------------- ------**********----------
Meau-M-230b  ---------------------------------------------- ------**********----------
Meau-M-229e  ---------------------------------------------- ------**********----------
Meau-M-233a  ---------------------------------------------- ------**********----------
Meau-M-228a  ---------------------------------------------- ------**********----------
Meau-M-227g  ---------------------------------------------- ------**********----------
Psze-M-2d    ---------------T------------------------------ ------**********----------
Psze-M-1a    ---------------T------------------------------ ------**********----------
Psze-M-1c    ---------------T----------------TG------------ --C-----------------T-----

GROUP 4      1                                                                                                            110
CONSENSUS    GTATGATCAATACACTCATTCATACACTGATCACTCATCAATACCCTAAGCAATGATGTGGCTGAAGATACTGATCAATACTCTGATCTCAGATCAGTCCACACTTTGAC
Psze-M-189a  ---CT-----------T---A-----------------------------------------------------------------------------------------
Psze-M-192a  ---CT-----------T---A-----------------------------------------------------------------------------------------
Meau-M-233f  ---------------------------------------------------------------------------------------G----------------------
Meau-M-230d  ---------------------------------------------------------------------------------------G----------------------
Meau-M-229a  ---------------------------------------------------------------------------------------G----------------------
Psze-M-193b  ------------------------------------------------------------A--------------------------G----------------------
Psze-M-3e    ------------------------------------------------------------A--------------------------G----------------------
Psze-M-2a    ------------------------------------------------------------A--------------------------G----------------------
Meau-M-235e  ------------------------------------------------------------A--------------------------G----------------------
Meau-M-227d  ------------------------------------------------------------A--------------------------G----------------------
```

FIG. 1.—Nucleotide sequences of intron 1 from *Meau* and *Psze*, arranged into homology groups. Dashes (-) indicate identity with the top sequence, and asterisks (*) indicate gaps introduced to improve the alignment. The sequences are arranged into groups according to their length and sequence similarity. The simple majority consensus was obtained from sequences reported elsewhere (Klein et al., accepted). The 12-nucleotide tandem repeats are underlined. Nucleotides are numbered in each group separately. Note that introns of different groups have similar sequences at their 3′ ends.

```
              111                    137
CONSENSUS     AGTGAATGTGA TTCTGTTTGTTCTCAG
Psze-M-189a   ----------- ----------------
Psze-M-192a   ----------- ----------------
Meau-M-233f   ----------- ----------------
Meau-M-230d   ----------- ----------------
Meau-M-229a   ----------- ----------------
Psze-M-193b   ----------- ----------------
Psze-M-3e     ----------- ----------------
Psze-M-2a     ----------- ----------------
Meau-M-235e   ----------- ----------------
Meau-M-227d   ----------- ----------------

GROUP 5       1                                                                                                  103
CONSENSUS     GTACCATCAATATACACTGATCAATACACTGCAGAGATACACTGATCAATAGTCTGACCAATGATGTGATCAGTTCACAGTCTGATG TTCTGTTTGTTCTCAG
Psze-M-189c   ------------------C------------------C------------------------------------------------- ----------------
Meau-M-228g   --------------------------------------------------------------------------------------- ----------------
Meau-M-226f   --------------------------------------------------------------------------------------- ----------------
Meau-M-236a   --------------------------------------------------------------------------------------- ----------------
Meau-M-226c   --------------------------------------------------------------------------------------- ----------------
Psze-M-192k   ------GG--C-**---------C-GAT-AAT-CT---*------------CA----GTG--ACA--------GC------------ ----------------

GROUP 6       1                                                                                                       108
CONSENSUS     GTACGATCAATACACTCATAAATGATCATTAATACACTGATCAACACTCTGATCAATGATGTGATCAGTGCACAGTCTGATGTTCTGTGTGG TTCTGTTTGTCCTCAG
Psze-M-3a     ----C-----------------------A--------------------------------------------------------------- ----------------
Psze-M-193e   ----C-----------------------------------****************************************------------ ----------------
Psze-M-189b   ---T----G---T---------------------------****************************************------------ ----------------
Psze-M-192c   ---T----G---T---------------------------****************************************------------ ----------------

GROUP 9       1                                                                                      88
CONSENSUS     GTATGATCGATATACTCATCACTGATCAATACTCTGATCAGTGATCAGGCCACAGTCTGATGTTCTGTGTGG TTCTGTTTGTTCTCAG
Meau-M-227c   --------A-C-C---G---------------A------------------T-------------------- --------A-------
Meau-M-194h   --------A-C-C---G---------------A------------------T-------------------- --------A-------
Psze-M-193d   ------------------------------------------------------------------------ ----------------
Psze-M-3d     ------------------------------------------------------------------------ ----------------
Psze-M-2c     ------------------------------------------------------------------------ ----------------

GROUP 10      1                                                                                                                    110
CONSENSUS     GTATGATCAATATACTCATCACCGATCAACACATTGATCATTACTGATCACTGATCATTAATACATTCGTCACTGATTAATACTCTGATCTATGATGTGATCAGTCCACAGTCTGATGTT
Meau-M-236c   ------------------------------------------------------------------------------------------------------------------------
Meau-M-234c   ------------------------------------------------------------------------------------------------------------------------
Meau-M-232c   ------------------------------------------------------------------------------------------------------------------------
Meau-M-231d   ------------------------------------------------------------------------------------------------------------------------
Meau-M-228b   ------------------------------------------------------------------------------------------------------------------------
Meau-M-226e   ------------------------------------------------------------------------------------------------------------------------

              111                 134
CONSENSUS     TTGTGTGC TTCTGTGTTTTCTCAG
Meau-M-236c   -------- ----------------
Meau-M-234c   -------- ----------------
Meau-M-232c   -------- ----------------
Meau-M-231d   -------- ----------------
Meau-M-228b   -------- ----------------
Meau-M-226e   -------- ----------------
```

```
                    5          15         25         35         45         55         65         75
Consensus           .GFLMYMVDR CDFNSTELKD IEYIRSYYYN KIEYVRFSSS VGKFVGYTEY GVKNAEYWNK DPGQLAAWRA QKETYCQHNI GVW
Psze-M-3d           .--ME-Y-W- ---------- ---TV----- ----A----- --EY--F-K- --H--DI--- -QA-M----- ------LP-- -ID
Psze-M-193d         .--ME-Y-W- ---------- ---TA----- ----A----- --EY--F-K- --H--DI--- -QA-M----- ------LP-- -ID
Meau-M-227c         .--ME-C-W- ---------- --FTL----- ----A----- --EY--F-KF --H--DI--- -Q-D---M-- ---------- --E
Meau-M-228g         .---E-RM-- -V-------- --F-----F- -V-D------ ---Y--F-KH --Y--DR--N --AE-NNR-- ---R--V--- -ID
Meau-M-236c         .--MD-Q-T- -V----DP-- ----L----- ----A----- ---Y--F-KH --Y--DR--N --AE-NNR-- ---R--V--- -ID
Meau-M-236a         .---E-RM-- -V-------- --F-----F- -V-D------ ---Y-----P -L-Y-AD--- -Q-----MK- -------QH- ---
Psze-M-192k         .---E-A-I- -V-------G --F-----F- -V-D------ ---Y-----P -L-Y--N--- -QALM-SEK- ---------- DID
Psze-M-189c         .---E-S-A- ---------- --F-I---F- -V-D------ ---------P -L-Y-AD--- -Q-D---M-- ------L--- --D
Meau-M-227a         .--MS---G- ------DP-- -QF-L----- MM--I--D-- ------F-QL -----DRL-- -Q-------- --G---V--- --D
Psze-M-193g         .--MS-I-N- -V-------- ----Q----- TM-IL--D-N --EY----QL -----DR--- ---V--VR-- ----V-HN-- ---
Psze-M-3e           .---S-R-S- -Q-------- ----Y-Q--- -L-IY----- ---------- ---Q-K-F-- -TAYVSSLN- ---------- ---
Meau-M-233f         .---S-G--- -Q-------- ----Y-E--- -L-IY----- L--------- ---Q-K-F-- -TAYVSSLN- ---------- -I-
Psze-M-189a         .---S-S-E- -Q-------- --L-L-Q--- -L-IY----- L--------- ---Q-K-F-D L--EV-R-T- E--R------ -N-
Psze-M-192a         .---S-S-E- -Q-------- --L-L-Q--- -L-IY----- L--------- ---Q-K-FSD L--EV-R-T- E--R------ -N-
Meau-M-235e         .---S-S--- -Q-------- ------V--- -L-IF----- ---------- ---Q-D-R-N -KAI-SSMK- ------HN-- --D
Psze-M-189b         .--K--W-N- -N-------- ----V---F- ----T--D-- --HY------ ---------N L--V--R--- A--------- -N-
Psze-M-1c           .--E---M-- -N------Q- --F-----FR -V-FT--D-D ---Y------ --R---E--- -Q-D---MK- ---------- --Y
Psze-M-3a           .--E--R-S- -N---S--Q- ----K---FR -V-FT--D-- L--Y------ --R---A--N --AI--SD-- R-----L--- -N-
Psze-M-193e         .--K---M-- -N-----PQ- ---------- ---VT--D-D --HY------ ---T-----N --AM---D-- A--R-----V -N-
Meau-M-234a         .--K--T--- -N-------- -Q-VL----- ---VN--D-D ---------- ------L--* SGAY-TQVKT E-G---L--- ---
Psze-M-2d           .--K--T--- -N-------- -Q-VL----- ---VN--D-D ---------- ------L--* SGAY-TQVKT E-G---L--- ---
Psze-M-1a           .--K--T--- -N-------- -Q-VL----- ---VN--D-N ---------- ------L--* SGAY-TQVKT E--R------ -I-
Meau-M-233b         .--K-FV-S- ---------- ---M------ ---V---D-D ------F-DF ---Q-KHF-S Q--R--TL-- E--R--HN-- ---
```

FIG. 2.—Amino acid sequences translated from exon 2 sequences of *Meau* and *Psze*. Dashes (–) indicate identity with the top sequence, asterisks (*) indicate gaps introduced to improve the alignment, and dots (.) indicate unavailability of sequence information. The simple majority consensus was obtained from sequences that included those reported elsewhere (Klein et al., accepted). The residues are numbered starting with the N-terminal residue of the mature protein (the first four residues are encoded in exon 1 and hence are not shown). The amino acids are abbreviated according to the international single-letter code.

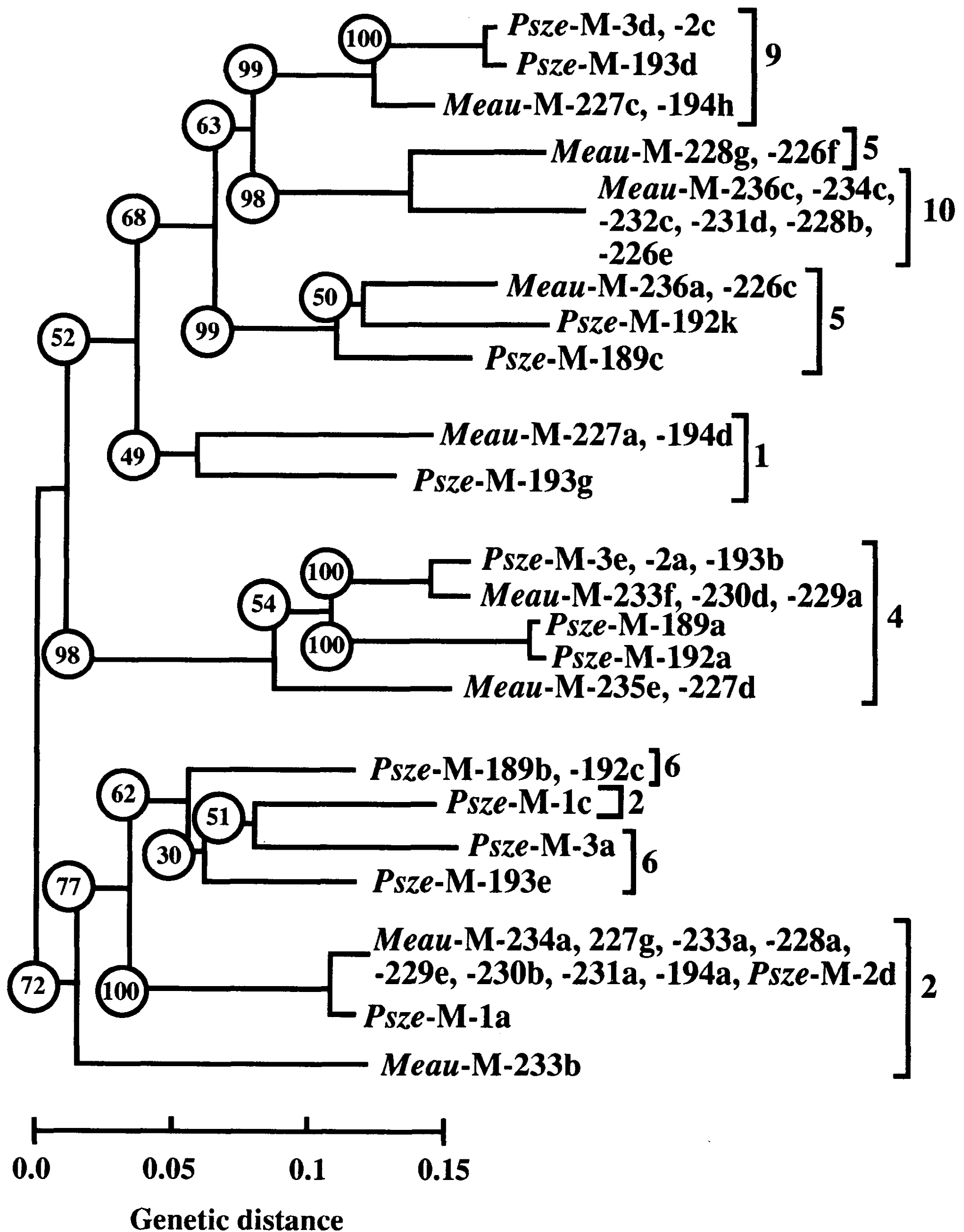

FIG. 3.—Genetic-distance dendrogram showing the relationships among the *Meau* and *Psze* exon 2 nucleotide sequences. In batches of identical sequences only the first is listed by its full name; the remaining sequences in the batch are listed by their numbers only. The scale shows genetic distances calculated by using Kimura's (1980) two-parameter method. The dendrogram is constructed according to the neighbor-joining principle (Saitou and Nei 1987). Numbering on nodes indicates the number of times that a particular branch is recovered per 100 bootstrap replications. Classification of the genes, based on intron 1 sequences, is indicated by brackets. Note that genes *Meau*-M-228g, *Meau*-M-236c, *Psze*-M-189a, *Psze*-M-192a, *Meau*-M-235e, *Meau*-M-227d, *Psze*-M-1c, *Psze*-M-2d, and *Meau*-M-233b may be recombinants, and this fact may have influenced the topology of the dendrogram.

1970), cichlids appear to be diploid (Kornfield 1984), and hence a maximum of two alleles is expected to occur at each locus per individual.] Since in the present study we are unable to distinguish alleles at a locus from genes at different loci (this distinction will be possible only when the genomic organization of the cichlid *Mhc* is worked out), we refrain from assigning definitive genetic symbols to the described sequences. Instead, we use tentative designations consisting of the genus and species abbreviation ("*Meau*" for *Melanochromis auratus* and "*Psze*" for *Pseudotropheus zebra;* see Klein et al. 1990), the lake abbreviation ("M" for Lake Malawi), the serial number of the individual, and the serial letter of the sequence from each individual. Our inability, at this stage of the analysis, to distinguish alleles from genes at different loci has little bearing on the aim of the present study, which is to determine the representation of the different genes in the two cichlid species.

The sequences are divided into groups on the basis of the length and sequence similarity of intron 1 (fig. 1; see Klein et al., accepted). The variation in length stems from the fact that different genes vary in the number of 12-nucleotide repeats. The consensus sequence of the repeat is ATCAATACACTG, but variations on this theme are common, both in sequence and in length (Klein et al., accepted). Intron 1 sequences classified as belonging to the same group (fig. 1) are clustered also in the dendrogram constructed from exon 2 sequences (fig. 3).

Our results reveal that *M. auratus* and *P. zebra* do indeed differ in the spectra of their *Mhc* genes (figs. 1–3). Among the 46 sequenced genes, there are no two that are identical in the two species, while repeated occurrences of the same sequence in different individuals of the same species are common. For example, 50% of *M. auratus* individuals have a sequence identical to that of *Meau*-M-236c, and another 17% have a recombinant sequence related to it (see below), while none of these sequences has been found in *P. zebra.* Similarly, 67% of *M. auratus* individuals have a sequence identical to that of *Meau*-M-227g, while no *P. zebra* individuals do (although, in this case, the latter species has two related sequences—*Psze*-M-2d and *Psze*-M-1a). Were there a complete overlap in alleles present in the two species, it would have been detected even in the relatively small sample of individuals tested. That no overlap has been detected in the samples suggests that the overlap in the entire population is either small or nonexistent. Absence of overlap, in turn, indicates that segregation of alleles occurred during the formation of the two species.

In addition to this main conclusion, our study also provides additional information about cichlid *Mhc* class II *B* genes. In an earlier study (Klein et al., accepted), we defined nine groups of cichlid class II *B* genes, based mainly on intron 1 sequences. Here we define a tenth group, found in *Meau*-M-236c and in corresponding genes of other *M. auratus* individuals. Like all other groups, it is distinguished by its length determined by the number of 12-nucleotide repeats and by sequence similarity. Of the two species, *M. auratus* has been found to possess genes of groups 1, 2, 4, 5, 9, and 10, whereas *P. zebra* has groups 1, 2, 4–6, and 9. Hence groups 3, 7, and 8 are missing in both species, while group 6 is missing in *M. auratus* and group 10 is missing in *P. zebra.* A few notes on the genes, in the order in which they appear (from top to bottom) in figure 3, follow.

Genes *Psze*-M-3d, *Psze*-M-2c, and *Psze*-M-193d have identical intron 1 sequences (group 9); the first two also have identical exon 2 sequences, while the third differs from the two by a single, nonsynonymous substitution. The identical intron 1 sequences of *Meau*-M-227c and *Meau*-M-194h differ from the three *Psze* sequences by 7 substitutions, whereas the *Meau* and *Psze* exon 2 sequences differ by 17.9 (*Psze*-M-193d)

and 16.8 (*Psze*-M-3d) substitutions, of which 15.8 and 14.7, respectively, are nonsynonymous [the substitutions are estimated by using the method of Li et al. (1985)]. The dominance of nonsynonymous over synonymous substitutions and of exon 2 over intron substitutions, in these and other genes discussed below, suggests that balancing selection operates on these genes (Hughes and Nei 1989) and hence that they are functional.

Genes (*Meau*-M-228g and *Meau*-M-226f), (*Meau*-M-236c, *Meau*-M-234c, *Meau*-M-232c, *Meau*-M-231d, *Meau*-M-228b, and *Meau*-M-226e), and (*Meau*-M-236a and *Meau*-M-226c) fall into three batches indicated by the parentheses. Genes in parentheses are identical in both intron 1 and exon 2 sequences, and hence, for further discussion, we shall consider only the first member of each batch. The *Meau*-M-228g and *Meau*-M-236a genes have identical intron 1 sequences (group 5), and their exon 2 sequences are also identical, but only up to codon 50; the 3′ parts of their exon 2 sequences are very different from each other (differing by 41.7 substitutions, of which 33.3 are nonsynonymous). The *Meau*-M-236c gene (group 10), on the other hand, is very different from *Meau*-M-228g (group 5) and *Meau*-M-236a (group 5) in intron 1 and in codons 5–40 of exon 2 (differing from *Meau*-M-228g by 21.4 substitutions, of which 15.9 are nonsynonymous) but is identical to *Meau*-M-228g in the remaining exon 2 codons. It appears, therefore, that *Meau*-M-228g and *Meau*-M-236c are reciprocal recombinants, the former being derived from *Meau*-M-236a and an unidentified gene X and the latter being derived from X and an unidentified gene Y. In both recombinants, the exchange took place in approximately the same position in the middle of exon 2.

Genes *Psze*-M-192k and *Psze*-M-189c differ from each other in both intron 1 (group 5) and exon 2. Genes *Meau*-M-227a and *Meau*-M-194d are identical in intron 1 and exon 2, and gene *Psze*-M-193g has an intron 1 similar to the two (only two substitutions differentiate the two sets of intron sequences) but a different exon 2.

Genes *Psze*-M-3e, *Psze*-M-2a, and *Psze*-M-193b are identical in their intron 1 (group 4) and exon 2 sequences. Similarly, *Meau*-M-233f, *Meau*-M-230d, and *Meau*-M-229a are also identical. The intron 1 sequences of the *Psze* and *Meau* batches differ by a single nucleotide substitution; the exon 2 sequences differ by 6.1 substitutions, all of which are nonsynonymous. Genes *Meau*-M-235e and *Meau*-M-227d have identical intron 1 and exon 2 sequences, and the intron 1 sequence is identical with that of the above *Psze* group 4 sequences; the exon 2 sequence is, however, completely different from exon 2 sequences of that group. Moreover, the *Psze*-M-189a and *Psze*-M-192a genes, which have identical intron 1 sequence and differ by a single, nonsynonymous substitution in their exon 2 sequences, differ, in 5–6 nucleotides only, from the above intron 1 group 4 sequences, but they differ from the *Meau*-M-235e gene by 53.1 substitutions, of which 45.6 are nonsynonymous. Most of the latter differences are in the 3′ part of exon 2, while the 5′ parts of these genes are rather similar. Hence, all these genes appear to be related to one another, but some of them may have diversified by recombination with genes from outside this group.

Genes *Psze*-M-189b and *Psze*-M-192c are identical in their intron 1 (group 6) and exon 2 sequences. They appear to be distantly related to the *Psze*-M-3a and *Psze*-M-193e genes, both of which possess group 6 introns. The *Psze*-M-1c gene has group 2 intron 1, except for the 3′ end, where it resembles group 6 introns. Since its exon 2 sequence is related to the exon 2 sequences of *Psze*-M-189b, *Psze*-M-3a, and *Psze*-M-193e, the *Psze*-M-1c gene is probably the result of recombination between group 2 and group 6 genes.

Genes *Meau*-M-234a, *Meau*-M-227g, *Meau*-M-233a, *Meau*-M-228a, *Meau*-M-229e, *Meau*-M-230b, *Meau*-M-231a, and *Meau*-M-194a are all identical in both their intron 1 (group 2) and exon 2 sequences. Gene *Psze*-M-2d is identical with this batch in exon 2 but differs from it by three substitutions and one 7-nucleotide indel in intron 1. Since exon 2 normally accumulates substitutions more rapidly than does intron 1, the *Psze*-M-2d may also be derived by recombination. Gene *Meau*-M-233b may be yet another recombinant, since its intron 1 differs from intron 1 of the *Meau*-M-227g batch by a single substitution, but exon 2 sequences of the two genes are very different from each other. Intron 1 of *Psze*-M-2d and the *Psze*-M-1a gene may be distantly related to these sequence batches.

Discussion

The study of molecular evolution in cichlid fishes of the eastern African lakes is hampered by a paucity of suitable markers. Variability of nuclear genes among different species of a single lake is very limited (McKaye et al. 1982; Sage et al. 1984; Verheyen et al. 1985; Kornfield 1991), and mitochondrial DNA analysis, while useful for defining main phylogenetic lineages within Lakes Malawi and Tanganyika, is powerless as far as most recent radiations are concerned, particularly in Lake Victoria (Meyer et al. 1990; Sturmbauer and Meyer 1992, and accepted). While the *Mhc* genes may not evolve faster than many other nuclear genes (Klein et al. 1993), they have the advantage of retaining their polymorphism over periods of time far exceeding the species' life spans. Utilization of *Mhc* polymorphism in the analysis of adaptive radiation, however, presupposes the fulfillment of two requirements: (*a*) the polymorphism of the entire species flock must be large, and (*b*) the species—or at least the species lineages—must differ in their *Mhc* genes. The differences cannot be expected to arise by generation of new alleles during or since the speciation event (for that to occur, the time interval is too short, just as it is for other nuclear genes); rather, they may be generated by differential apportioning (segregation) of individual alleles into different species because of genetic drift in the founding populations, if these latter are relatively small.

Extensive *Mhc* variability in the species flocks has now been amply documented (Klein et al. 1993; present paper). In *Melanochromis auratus* and *Pseudotropheus zebra* alone, we were able to demonstrate the presence of ≥23 different genes, and many more are present in other species (Klein et al., accepted). While some of these genes probably occupy different loci, many are likely to be alleles and thus to constitute true polymorphisms.

In a test of the second requirement, we have demonstrated in this communication that two species from the same flock have different gene profiles at their *Mhc* loci. We are aware of the potential limitations of the present study: the sample tested may have been too small to reveal the entire range of variability in natural populations of the two species, and the stocks maintained by dealers may have been subjected to an artificial bottleneck, thus restricting and perhaps skewing the actual variability. We have attempted to offset these limitations by obtaining individuals of each species from different dealers and by choosing different morphs ("blue" and "red") of one species (*P. zebra*). The fact that no two individuals of the sample have the same composition of *Mhc* genes indicates, however, that the stocks maintained by the dealers are not inbred. Hence, while it is possible that more extensive testing will reveal partial overlap between the two species, it is highly unlikely that the differences observed are solely due to sampling. Nonetheless, we are in the process of acquiring individuals from natural populations and also expanding the study to other cichlid species.

For this initial study, we selected intentionally two species with an intermediate degree of relatedness between them, compared with other species pairs. Both *M. auratus* and *P. zebra* are members of the same species flock in Lake Malawi. They were once considered to belong to the same genus (*Pseudotropheus*), but more recently taxonomists regard the two species as divergent enough to assign them to different genera (Lewis et al. 1986). Both species belong to the same trophic group: they feed on algae that grow on the surface of rocks (the so-called *Aufwuchs*). They have evolved highly specialized food-collecting equipment, which enables them to scrape the *Aufwuchs* from the rock surfaces. Both species are widely distributed along the western coast of Lake Malawi, *P. zebra* in the northern part and *M. auratus* in the southern part. They are strictly bound to rocks in shallow waters along the shore. Other members of the Lake Malawi species flock, particularly those belonging to different trophic groups, may be less related to one another and hence can be expected to show even greater differences in their *Mhc* genes than *P. zebra* and *M. auratus* do. Hence, analysis of *Mhc* variability in cichlid fishes, though time-consuming, should prove to be rewarding. It may provide means of sorting out relationships among the species in the flock and thus of learning about the nature of the speciation process.

The high representation of apparent recombinants among the sampled genes is interesting. Belich et al. (1992), as well as Watkins et al. (1992), reported a similar common occurrence of recombinant *Mhc* (*HLA-B*) genes in South American Indian tribes and argued that recombination may provide rapid means of *Mhc* diversification in populations colonizing new environments. Our findings can be interpreted as being consistent with this interpretation.

Acknowledgments

We thank Lynne Yakes for editorial assistance, Anica Milosev for the preparation of the computer graphic, and Holger Sültmann for critical reading of the manuscript. This work was supported in part by National Institutes of Health grant AI 23667.

LITERATURE CITED

BEADLE, L. C. 1981. The inland waters of tropical Africa, 2d ed. Longman, London.

BELICH, M. P., J. A. MADRIGAL, W. H. HILDEBRAND, J. ZEMMOUR, R. C. WILLIAMS, R. LUZ, M. L. PETZL-ERLER, and P. PARHAM. 1992. Unusual *HLA-B* alleles in two tribes of Brazilian Indians. Nature **357**:326–329.

ECCLES, D., and E. TREWAVAS. 1989. Malawian cichlid fishes: the classification of some haplochromine genera. Lake Fish Movies, Herten, Germany.

FRYER, G., and T. D. ILES. 1972. The cichlid fishes of the great lakes of Africa. TFH Publications, Neptune City, N.J.

GREENWOOD, P. H. 1981. The haplochromine fishes of the east African lakes. Cornell University Press, Ithaca, N.Y.

HUGHES, A. L., and M. NEI. 1989. Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. Proc. Natl. Acad. Sci. USA **86**:958–962.

KEENLEYSIDE, M. H. A., ed. 1991. Cichlid fishes: behavior, ecology and evolution. Chapman & Hall, London.

KIMURA, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16**:111–120.

KLEIN, D., H. ONO, C. O'HUIGIN, V. VINCEK, T. GOLDSCHMIDT, and J. KLEIN. Extensive *Mhc* variability in cichlid fishes of Lake Malawi. Nature (accepted).

KLEIN, J. 1980. Generation of diversity at MHC loci: implications for T-cell receptor repertoires.

Pp. 239–253 *in* M. FOUGEREAU and J. DAUSSET, eds. Immunology 80. Academic Press, London.

———. 1986. Natural history of the major histocompatibility complex. John Wiley & Sons, New York.

———. 1987. Origin of the major histocompatibility complex polymorphism: the trans-species hypothesis. Hum. Immunol. **19:**155–162.

KLEIN, J., R. E. BONTROP, R. L. DAWKINS, H. A. ERLICH, U. B. GYLLENSTEN, E. R. HEISE, P. P. JONES, P. PARHAM, E. K. WAKELAND, and D. I. WATKINS. 1990. Nomenclature for the major histocompatibility complexes of different species: a proposal. Immunogenetics **31:** 217–219.

KLEIN, J., and F. FIGUEROA. 1986. Evolution of the major histocompatibility complex. CRC Crit. Rev. Immunol. **6:**295–386.

KLEIN, J., Y. SATTA, C. O'HUIGIN, and N. TAKAHATA. 1993. The molecular descent of the major histocompatibility complex. Annu. Rev. Immunol. **11:**269–295.

KORNFIELD, I. L. 1984. Descriptive genetics of cichlid fishes. Pp. 591–616 *in* B. J. Turner, ed. Evolutionary genetics of fishes. Plenum, New York.

———. 1991. Genetics. Pp. 103–128 *in* M. H. A. KEENLEYSIDE, ed. Cichlid fishes: behavior, ecology and evolution. Chapman & Hall, London.

LEWIS, D., P. REINTHAL, and J. TRENDALL. 1986. A guide to the fishes of Lake Malawi National Park. WWF World Conservation Center, Gland, Switzerland.

LI, W.-H., C.-I. WU, and C.-C. LUO. 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol. Biol. Evol. **2:**150–174.

MCKAYE, K. R., T. KOCHER, P. REINTHAL, and I. KORNFIELD. 1982. A sympatric sibling species complex of *Petrotilapia* Trewavas from Lake Malawi analyzed by enzyme electrophoresis (Pisces: Cichlidae). Zool. J. Linnean Soc. **76:**91–96.

MEYER, A., T. D. KOCHER, P. BASASIBWAKI, and A. C. WILSON. 1990. Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. Nature **347:**550–553.

OHNO, S. 1970. Evolution by gene duplication. Springer, New York.

ONO, H., F. FIGUEROA, C. O'hUIGIN, and J. KLEIN. 1993. Cloning of the β_2-microglobulin gene in the zebrafish. Immunogenetics **38:**1–10.

ONO, H., D. KLEIN, V. VINCEK, F. FIGUEROA, C. O'HUIGIN, H. TICHY, and J. KLEIN. 1992. *Mhc* class II genes of zebrafish: origin and polymorphism. Proc. Natl. Acad. Sci. USA **89:** 11886–11890.

RIBBINK, A. J. 1984. Is the species flock concept tenable? Pp. 21–25 *in* A. A. ECHELLE and I. KORNFIELD, eds. Evolution of fish species flocks. University of Maine at Orono Press, Orono.

ROTHBARD, J. B., and M. L. GEFTER. 1991. Interactions between immunogenetic peptides and Mhc proteins. Annu. Rev. Immunol. **9:**527–565.

SAGE, R. D., P. V. LOISELLE, P. BASASIBWAKI, and A. C. WILSON. 1984. Molecular versus morphological change among cichlid fishes of Lake Victoria. Pp. 185–201 *in* A. A. ECHELLE and I. KORNFIELD, eds. Evolution of fish species flocks. University of Maine at Orono Press, Orono.

SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:**406–425.

SANGER, F., S. NICKLEN, and A. R. COULSON. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.

STURMBAUER, C., and A. MEYER. 1992. Genetic divergence, speciation and morphological stasis in a lineage of African cichlid fishes. Nature **358:**578–581.

———. Mitochondrial phylogeny of the endemic mouthbrooding lineages of cichlid fishes from Lake Tanganyika, East Africa. Mol. Biol. Evol. (accepted).

VERHEYEN, E., J. VAN ROMPAEY, and M. SELENS. 1985. Enzyme variations in haplochromine cichlid fishes from Lake Victoria. Neth. J. Zool. **35:**469–478.

WATKINS, D. I., S. N. MCADAM, X. LIU, C. R. STRANG, E. L. MILFORD, C. G. LEVINE, T. L. GARBER, A. L. DOGON, C. I. LORD, S. H. GHIM, G. M. TROUP, A. L. HUGHES, and N. L. LETVIN. 1992. New recombinant *HLA-B* alleles in a tribe of South Amerindians indicate rapid evolution of MHC class I loci. Nature **357**:329–333.

PAUL M. SHARP, reviewing editor

Received November 30, 1992; revision received March 24, 1993

Accepted March 24, 1993