First publ. in: Genetics ; 149 (1998), 3. - S. 1527-1537

Linkage Relationships and Haplotype Polymorphism Among Cichlid *Mhc* Class II *B* Loci

Edward Málaga-Trillo,*,† Zofia Zaleska-Rutczynska,* Brendan McAndrew,‡ Vladimir Vincek,† Felipe Figueroa,* Holger Sültmann* and Jan Klein*,†

**Max-Planck-Institut für Biologie, D-72076 Tübingen, Germany, †University of Miami School of Medicine, Miami, Florida 33101, and ‡Institute of Aquaculture, University of Stirling, Stirling FK9 4LA, Scotland*

ABSTRACT

The species flocks of cichlid fishes in the Great East African Lakes are paradigms of adaptive radiation and hence, of great interest to evolutionary biologists. Phylogenetic studies of these fishes have, however, been hampered by the lack of suitable polymorphic markers. The genes of the major histocompatibility complex hold the promise to provide, through their extensive polymorphism, a large number of such markers, but their use has been hampered by the complexity of the genetic system and the lack of definition of the individual loci. In this study we take the first substantial step to alleviate this problem. Using a combination of methods, including the typing of single sperm cells, gyno- or androgenetic individuals, and haploid embryos, as well as sequencing of class II *B* restriction fragments isolated from gels for Southern blots, we identify the previously characterized homology groups as distinct loci. At least 17 polymorphic class II *B* loci, all of which are presumably transcribed, have been found among the different species studied. Most of these loci are shared across the various cichlid species and genera. The number of loci per haplotype varies from individual to individual, ranging from 1 to 13. A total of 21 distinct haplotypes differing in the number of loci they carry has thus far been identified. All the polymorphic loci are part of the same cluster in which, however, distances between at least some of the loci (as indicated by recombination frequencies) are relatively large. Both the individual loci and the haplotypes can now be used to study phylogenetic relationships among the members of the species flocks and the mode in which speciation occurs during adaptive radiation.

CICHLID fishes (family Cichlidae, order Perciformes) of the East African Great Lakes are often cited, along with Darwin's finches, as a classic example of adaptive radiation. Of the more than 1300 species widely distributed over the African continent, Central and South America, the Near East, South India, Madagascar and Sri Lanka, nearly 1000 are endemic to the East African lakes, especially the large Lakes Tanganyika, Victoria and Malawi (Fryer and Iles 1972). The rapid rise of hundreds of cichlid species from a limited number of ancestors and their coexistence within single lakes constitute a largely unsolved problem in evolutionary biology. Only recently have some of the major events in the history of these species flocks begun to be uncovered with the use of molecular markers (Meyer *et al.* 1990; Kocher *et al.* 1993; Moran and Kornfield 1993; Sturmbauer and Meyer 1993; Schliewen *et al.* 1994; Franck *et al.* 1994; Sültmann *et al.* 1995; Verheyen *et al.* 1996; Zardoya *et al.* 1996; Mayer *et al.* 1997). However, a common limitation to the use of these markers is their low variability in African cichlids attributable to the recency of their radiation. One exception is the set of markers encoded in the genes of the major histocompatibility complex (*Mhc*). The *Mhc* offers new possibilities for the study of cichlid evolution because its extensive polymorphism (Klein and Figueroa 1986; Klein 1990) evolves transspecifically (Klein 1980, 1987; Figueroa *et al.* 1988; J. Klein *et al.* 1993), making it possible to relate taxa through their sharing of old allelic lineages, to differentiate them through the segregation of ancestral polymorphisms, and to answer questions concerning the size of founding populations and the manner of adaptive radiation.

The *Mhc* genes code for proteins that bind short peptides derived from self and nonself molecules and display them on the surface of antigen presenting cells (Klein 1986; Srivastava *et al.* 1991). Recognition of Mhc-nonself peptide complexes by T lymphocytes initiates the specific form of the immune response. *Mhc* genes fall into two classes, I and II, whose products present intra- and extracellularly derived peptides, respectively. Class II genes are further divided into two subclasses, *A* and *B*, encoding the α and β polypeptide chains that form class II heterodimers. The class II *B* genes of several African cichlid species have been shown to be highly polymorphic (D. Klein *et al.* 1993; Ono *et al.* 1993a,b). Each of the genes consists of six exons

Corresponding author: Jan Klein, Max-Planck Institut für Biologie, Abteilung Immungenetik, Corrensstr. 42, D-72076 Tübingen, Germany. E-mail: jan.klein@tuebingen.mpg.de

separated by short introns, exon 2 being the most polymorphic segment of the coding region (Ono *et al.* 1993a). Sequencing of class II *B* intron 1 and exon 2 from a large number of African cichlids has revealed the existence of 17 homology groups of class II *B* genes that can be distinguished by their sequences and intron 1 length (D. Klein *et al.* 1993; Ono *et al.* 1993a,b; E. Málaga-Trillo, H. Tichy, P. Walsh, C. Wood, J. Maina, V. Vincek and J. Klein, unpublished data). However, no information was available as to whether the homology groups were distinct loci or old allelic lineages at one locus. Orthologous comparisons between sequences from different species could therefore not be made with certainty. To overcome this limitation to the use of *Mhc* genes in evolutionary studies, we decided to use sperm typing and segregation analysis of haploid embryos, gyno- and androgenetic fishes to determine the locus-allele relationships among the cichlid class II *B* homology groups.

MATERIALS AND METHODS

Fishes: The *Pseudotropheus zebra* (*Psze*) specimens employed for sperm typing as well as the *Aulonocara hansbaenschi* (*Auha*) and *Haplochromis nubilus* (*Hanu*) individuals used for the isolation of *Mhc* transcripts were obtained from a local dealer (Aquarium Pelz, Bondorf, Germany) and maintained in our aquaria at the Max-Planck-Institut für Biologie in Tübingen. The *Oreochromis niloticus* (*Orni*) androgenetic, gynogenetic and normal crosses were at the University of Stirling Institute of Aquaculture, and the haploid embryos were generated from the *O. niloticus* stocks provided by RWE Energie, Aktiengesellschaft, Anwendungstechnik, Essen, Germany and maintained at the Max-Planck-Institut für Biologie, Tübingen.

Single sperm isolation and lysis: Semen samples were collected by palpation from an informative *P. zebra* male and single sperm cells were sorted by micromanipulation. Tenfold serial dilutions were prepared in 96-well microtiter plates and 20 μl of each dilution was scored on Terasaki plates under an inverted stereoscope to find the well containing 1 cell/2 μl. To keep all sperms in one observation plane, the plates were centrifuged for 2 min at 1200 rpm before each screening. The correct dilution was aliquoted in 2 μl volumes into new Terasaki plates to obtain one sperm cell per well on average. After another centrifugation, the wells were visually rescored to confirm the presence of single cells. The single sperms in selected wells were then lysed by the addition of 5 μl of an alkaline solution (200 mm KOH/50 mm dithiothreitol). After 10 min incubation at 65°, 5 μl of neutralizing solution (900 mm Tris-HCl, pH 8.3/300 mm KCl/200 mm HCl) was added.

***Mhc* typing of single sperm cells:** Amplification of genomic targets from single sperm cells was achieved by using heminested PCR. Out of the 10-μl single sperm lysates, 1-μl aliquots were used as template for the first round of amplification with intron 1-exon 2 generic primers TU383 and TU377 (Figure 1). The conditions for the first round PCR were: 2 min at 94° and then 30 cycles of 1 min at 94°, 1 min at 55° and 2 min at 72°, followed by a final extension of 10 min at 72°, in a total volume of 50 μl. For the second round PCR, only the intron 1 or exon 2 regions were targeted by using generic primer pairs (TU383/TU798-TU823, and TU822-TU377/TU799) or upstream generic primer TU383 in conjunction with downstream group-specific primers (TU931, TU937, TU938, TU939, TU940 and TU941) under the same conditions as the first amplification, using 40-PCR cycles. The products were visualized in 2% agarose gels, excised from the agarose, cloned and sequenced.

Ploidy manipulations in *O. niloticus*: Two androgenetic and three gynogenetic *O. niloticus* families were generated at the University of Stirling Institute of Aquaculture according to previously established techniques (Hussain *et al.* 1993; Myers *et al.* 1995). Nine sets of *O. niloticus* haploid embryos were produced at the Max-Planck-Institut für Biologie, Tübingen: Oocytes were collected in large numbers from selected females and fertilized with UV-irradiated sperm from a *Pseudotropheus tropheops* male. The resulting embryos were incubated at 28° in aereated cylinders for 2–3 days and then frozen in liquid nitrogen and stored at −80° until needed. Additional gynogenetic families were generated in Tübingen according to the techniques mentioned above.

Isolation of genomic DNA: Muscle tissue DNA was according to Maniatis *et al.* (1989). To obtain DNA from dorsal fin samples, sections weighing approximately 50 mg and no larger than 10% of the fin's area were carefully excised from live fish. The samples were kept on ice to continue with the procedure, or immediately frozen in liquid nitrogen and kept at −80° until needed. The fin fragments were finely chopped and lysed at 55° for 6 hr in 400 μl of ATL-Buffer (Qiagen, Hilden, Germany) containing 2 mg/ml of Proteinase K. After two extractions with phenol, one with phenol-chloroform-isoamyl alcohol (25:24:1) and one with chloroform-isoamyl alcohol (24:1), the DNA was precipitated by adding 0.1 vol sodium acetate (NaOAc) and 2.5 vol 100% ethanol. The white DNA precipitate was picked up with a Pasteur pipette, rinsed in 80% ethanol, and redissolved at room temperature in 500 μl of TE buffer pH 8.0 (10 mm Tris-HCl, 1 mm ethylene diaminetetraacetic acid, EDTA). A similar protocol was also used for DNA isolation from 2–3-day-old embryos. DNA isolation from blood samples was achieved with the aid of the Qiagen Blood Kit (Qiagen).

Polymerase chain reaction (PCR) and primers: PCR amplifications were performed according to the conditions described for the GeneAmp PCR Core Reagents Kit (Perkin Elmer-Cetus, Norwalk, CT). We used 10× PCR buffer, Taq DNA polymerase and dNTPs from Perkin Elmer-Cetus (Norwalk, CT) and Pharmacia LKB (Freiburg, Germany). The following generic primers were used for the amplification of cichlid class II *B* genes: TU383 (upstream, end of exon 1: 5′-TGATTTAGACAGAG(A)T(G)GT(G)T(C)GCTGTA-3′); TU377 (downstream, end of exon 2: 5′-CTCTTCATCAGCCTCAGCACA-3′); TU957 [downstream, 56 base pairs (bp) inside of exon 2: 5′-CTTCAGCTCAGTGGAGTTAAA-3′]; Ex3F (upstream, beginning of exon 3: 5′-GCCATGTTGGTCTGCAGCGTCTAT-3′); Ex3R (downstream, end of exon 3: 5′-CTGGGTGTGTACTCCAGGTGTGAG-3′). The following group-specific primers (located in exon 2) were designed and used in conjunction with TU383 or TU377: TU931 [group 1-specific: 5′-ATATACACTC(G)ATCAATAC(T)AC-3′]; TU937 (group 3-specific: 5′-ATACACTCATTCATACACTG-3′); TU938 (group 4-specific: 5′-GAATCACTTTCACTGTCAAA-3′); TU939 [group 11-specific: 5′-CG(A)ATACACTCATAAATGAT-3′]; TU940 [group 13-specific: 5′-AGAACCACACAG(A)AACATCAGA-3′]; TU941 [group 14-specific: 5′-AGAACATCAGACTGTGAA(G)CTG-3′]; TU1209 (middle of exon 2, *Auha* group 1-specific downstream primer: 5′-GATTCAAAATGTATTGGGTGA-3′); TU1210 (middle of exon 2, *Auha* group 2-specific downstream primer: 5′-GATTCATGAATTACAGAATGA-3′); TU1211 (middle of exon 2, *Auha* group 4-specific downstream primer: 5′-GATTCAAAATGTTTTGGGTGC-3′); TU1212 (middle of exon 2, *Auha* group 6-specific downstream primer: 5′-GATTCATGGACTATTTTGTGT-3′); TU1213 (middle of exon

2, *Auha* group 10-specific downstream primer: 5′-GATTTCTGAGTTATAGTGTGG-3′); TU1247 (middle of exon 2, *Auha* group 1-specific upstream primer: 5′-TCACCCAATACATTTTGAATC3′), TU1248 (middle of exon 2, *Auha* group 2-specific upstream primer: 5′-TCATTCTGTAATTCATGAATC-3′), TU1253 (middle of exon 2, *Auha* group 4-specific upstream primer: 5′-GCACCCAAAACATTTTGAATC-3′), and TU1254 (middle of exon 2, *Auha* group 6-specific upstream primer: 5′-ACACAAAATAGTCCATGAATC-3′). *Auha* first strand cDNA for RT-PCR was generated using primers TU455 (end of exon 3: 5′-CCAATCACCATCTGCCATCTCCT-3′); TU464 [end of exon 5: 5′-AATCAGACA(G/G/A)ACCAGGACC-3′]; and TU492 (end of exon 4: 5′-GGCGTGCTCCACCACACAGGA-3′).

DNA isolation from agarose slices: Upon their detection in ethidium bromide-stained agarose gels, DNA fragments were isolated from the excised bands by centrifugation at 4° through a glass wool column at 6000 rpm for 15 min. The DNA eluates were further concentrated or purified if necessary and used for cloning, radiolabeling, or PCR amplification.

Cloning of PCR products into plasmid vectors: Fragments ranging in size from 50 to 1500 bp were ligated into plasmid vectors with the aid of the Sureclone (Pharmacia LKB) and TA (Invitrogen, San Diego, CA) cloning kits and used to transform competent *E. coli* XL1 blue competent bacteria (Stratagene, Heidelberg, Germany).

Preparation of plasmid DNA and sequencing: Plasmid DNA was isolated using the Plasmid Mini Kit protocol (Qiagen). Sequencing reactions of double-stranded plasmid DNA were performed by the dideoxy chain-termination method (Sanger *et al.* 1977). Reactions prepared with the Sequenase 2.0 Kit (United States Biochemical, Cleveland) and labeled with ^{32}P were resolved in Bio-Rad (Richmond, CA) electrophoresis systems using 0.5-mm thick Long Ranger gels, after which the gels were dried and transferred to Whatman paper for detection of the signals by autoradiography. Reactions prepared with the Auto Read Sequencing Kit were resolved in automated sequencers (A.L.F.; Pharmacia LKB) using 0.35-mm-thick 6% polyacrylamide gels. DNA sequences were analyzed using the software: MacVector 4.0 (by Sue Olson and Joe Kaufman, Kodak International Biotechnologies, 1990) and the Wisconsin Package (versions 8.0, 9.0, and 9.1, by the Genetics Computer Group, Inc., Madison, WI). All the sequences generated in this study have been deposited in the GenBank database and can be retrieved under the accession numbers AF049939–AF050036.

Southern blotting: Approximately 10 μg of genomic DNA was digested to completion using 6 units of restriction enzyme per μg of DNA. The DNA was precipitated for 2 hr at −80° with 0.1 vol NaOAc and 2.5 vol ethanol, washed with 70% ethanol, and resuspended in 27 μl H_2O or TE buffer, pH 8.0. The samples were loaded onto 0.8% large agarose gels and run overnight at 40 V. After electrophoresis, the genomic digests were transferred to nylon membranes (Amersham, Braunschweig, Germany) either by the capillary method (Maniatis *et al.* 1989) or the vacuum method. In the latter case, vacuum blotting systems from Pharmacia and Bio-Rad were used. The DNA was then fixed to the membrane by UV irradiation. DNA probes were labeled with ^{32}P using the RadPrime DNA Labeling System (Life Technologies, Gaithersburg, MD) and the Ready to Go Labeling Kit (Pharmacia LKB). Residual ^{32}P and dNTP were removed by centrifugation through Sephadex G-50 (Pharmacia) spin-columns at 1200 rpm for 2 min. Blotted membranes were prehybridized for 2 hr at 42° and hybridized to the probe overnight at 42° in a solution containing 50% formamide, 5× SSC, 1× Denhardt's solution, 20 mm NaPi, 0.2% SDS and 10% dextran sulfate. We typically used 20 ng of probe (specific activity $>10^9$ cpm/μg) per ml of hybridization mixture. Nonspecific binding of the probe was removed by washing the membranes once at 60° for 5–15 min with a 2× SSC and 0.1% SDS solution. Blots were exposed to XAR5 film (Kodak, Rochester, NY) and X-OMAT film (Kodak-Pathé, Strasbourg, France) for 72 hr at −80° with an intensifying screen.

Detection of *Mhc* class II *B* transcripts: Fish were anesthetized in 0.02% MS222 (Sigma, Heidenheim, Germany) and dissected. Total RNA was isolated from spleen, liver, intestine or gonad tissues using the RNA Extraction Kit (Pharmacia LKB), RNeasy Total RNA Kit and RNeasy Mini Kit (Qiagen). Messenger RNA was then prepared from total RNA, or directly from the tissues with the aid of the Oligotex Direct mRNA kit (Qiagen), the mRNA Purification and QuickPrep Micro mRNA Purification Kits (Pharmacia LKB). For RT-PCR, first strand cDNA was generated with the First Strand cDNA Synthesis Kit (Pharmacia LKB) using oligo(dT) or random hexamers as primers, and with downstream primers specific for cichlid class II *B* sequences (TU455, end of exon 3, TU464, end of exon 5, TU492, end of exon 4). Tenfold dilutions of the first-strand mixture were used for PCR amplification of specific transcripts. Double-stranded cDNA was prepared with the cDNA Synthesis Kit (Boehringer Mannheim, Mannheim, Germany) and Time Saver cDNA Synthesis Kit (Pharmacia LKB).

RESULTS

Sperm typing: We used several methods to determine the status of the *Mhc* class II *B* homology groups in cichlid fishes, the first of these being sperm typing in *P. zebra*. Preliminary experiments based on whole genome PCR amplification with random primers did not produce satisfactory results, perhaps because of the complexity of the genetic system. We therefore turned to the use of *Mhc*-specific primers applied directly to the single sperm lysates. Initially, we used the generic primer pair TU383-TU377 (Ono *et al.* 1993a,b; D. Klein *et al.* 1993) in the first round of PCR amplification and another generic primer pair, TU383-TU823 (Figure 1), in the second round (= heminested PCR). The first pair amplified intron 1 and exon 2, the second pair intron 1 only. The amplified products (visible by ethidium bromide staining only after the second amplification, Figure 2) were cloned and sequenced. Testing of 55 sperm cells from a single *P. zebra* male yielded four genotypes, according to the homology groups identified in them: (1) groups *2*, *3*, and *11*; (2) groups *4* and *6*; (3) groups *3* and *5*; and (4) group *5* only. This result indicates that groups *2*, *3* and *11* represent distinct loci because otherwise they could not be amplified from a single sperm. Similarly, groups *4* and *6*, as well as *3* and *5*, must be derived from different loci for the same reason. The obtainment of four types from a single male's gametes could mean that the class II *B* loci might be on different chromosomes, but this interpretation is ruled out by the data described in the following section. Most probably, the four types represent an artifact of the single cell PCR: the generic primers used in the first round of PCR may not have been equally efficient in amplifying multiple templates from the limiting amounts of DNA present in the single-cell lysates. This

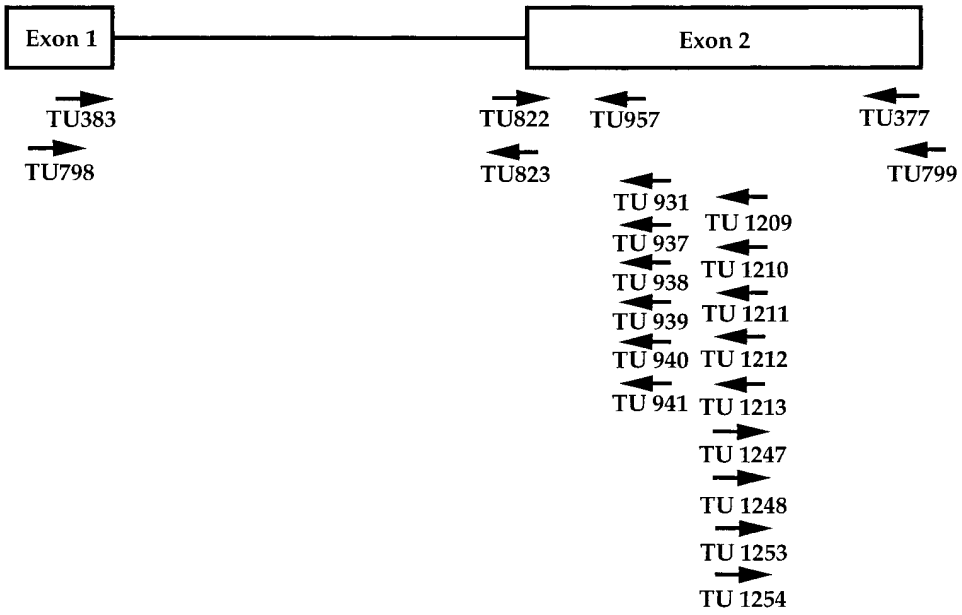

Figure 1.—PCR primers used for the amplification of the intron 1 and exon 2 regions of cichlid *Mhc* class II *B* genes (see materials and methods). Locus-specific primers internal to exon 2 are aligned vertically. Arrows indicate priming sense.

handicap, however, does not change anything in the interpretation of the separate locus status of the detected group.

To obtain more consistent amplification of all groups, we turned to the use of more specific primers. For the first round of amplification, we used the primer pair TU798-TU799 which is generic but based on *P. zebra* sequences exclusively. For the second round, we used primers specific for *P. zebra* groups *1*, *3*, *4*, *11*, *13*, and *14* (TU931-TU937 through TU941; Figure 1). For the other groups, it was not possible to design specific primers. The typing of 70 sperm cells from a second *P. zebra* male yielded two types of results: 33 cells were typed as bearing groups *3* and *4*, while the remaining 37 cells yielded sequences of groups *11* and *13*. We conclude, therefore, that the male was a heterozygote bearing loci *3* and *4* on one chromosome and loci *11* and *13* on a homologous chromosome. In both males, all the loci found by sperm typing could also be demonstrated in somatic cells.

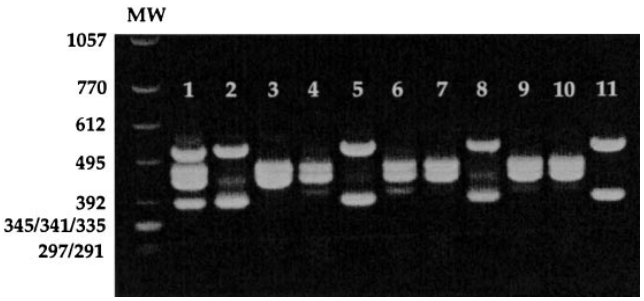

Figure 2.—*Mhc* Class II *B* intron 1 heminested PCR products from 10 *P. zebra* single sperm cells. The fragments were obtained after two rounds of amplification with primer pairs TU383-TU377 and TU383-823, separated in a 2% agarose gel at 50 V for 4 hr, and stained with ethidium bromide. The differences in fragment size are a consequence of the length polymorphism in intron 1. MW: Molecular weight marker (ϕx-174-RF DNA-*Hin*dII digest; Pharmacia LKB); Lane 1: Father's (sperm donor) products; lanes 2–11: single sperm cell's products. Band sizes of the molecular weight marker are expressed in base pairs (bp).

Typing of gynogenetic, androgenetic families and haploid embryos: Because of the laboriousness of the single-cell sorting procedure, and since the sperm-typing provided information about only a limited number of loci, we turned to the analysis of families with offspring produced by gynogenesis or androgenesis from heterozygous parents, as well as families with haploid, maternally-derived, offspring. The generally small size of the haplochromine cichlids makes it difficult to obtain sufficient numbers of eggs for this kind of experiment in which only a small fraction of treated germ cells produces viable progeny. We therefore used *O. niloticus*, a representative of another genus of cichlid fish which, however, shares with haplochromines most of the homology groups at class II *B* loci. The first set of samples, consisting of three gynogenetic and two androgenetic families produced in Stirling was typed by a single round of PCR amplification using generic primers TU383 and TU377, followed by cloning and sequencing. When available, both parents and hybrids derived from them were also typed (Table 1). The segregation pattern of the gynogenetic, androgenetic, and hybrid progeny was used to deduce the haplotypes of the parents (Figure 3). Deviations from the parental types among the offspring were interpreted as interlocus recombinants. A total of six different parental haplotypes were detected in the five families; in addition, three recombinant haplotypes were detected in Family 1 and one in Family 3. The second set of samples included nine families, each consisting of mother, male UV donor and 20 2–3-day-old haploid embryos, and

TABLE 1

***Mhc* class II *B* genotype of five *O. niloticus* families**

Family	Individual	Observed genotype	No. of events
1	Father	*1-15-2-6*	1
	Mother	*4-12*	1
	Mitogynogenetics	*4*	4
		12	3
	Normal F_1	*1-4*	5
		4-15-2-6	5
		12-15-2-6	2
	(Recombinant)	*1-12-2-6*	1
	(Recombinant)	*1-12-15*	1
	(Recombinant)	*1-4-15*	1
2	Father	*1-9-6*	1
	Mother	*1-9-12*	1
	Mitogynogenetics	*1*	3
		9-12	1
	Normal F_1	*1*	5
		1-9-12	4
		9-6-12	1
3	Father	*1-9-12*	1
	Mother	*9-6*	1
	Mitogynogenetics	*9-6*	6
	Normal F_1	*1-9-6*	3
	(Recombinant)	*1-9-6-12*	3
4	Father	*1-12*	1
	Mother	*1-15-2-6*	1
	Androgenetics	*12*	4
		1	3
	Androgenetic, F_2	1	1
	Normal F_1	*15-2-6-12*	2
5	Father	*1-15-2-6*	1
	Mother	*15-2-6*	1
	Androgenetics	*1*	5

Genotypes are designated according to the loci identified through sequencing of the corresponding PCR products.

they were typed using the generic primer pair TU383-TU957. For segregation analysis, the amplified fragments (containing intron 1 and 38 bp of exon 2) were resolved by agarose gel electrophoresis (Figure 4), and identified by cloning and sequencing. A total of 15 different haplotypes could be deduced from this data set: 11 parental and 4 recombinant (Table 2; Figure 5). Taken together, the data from both sets are consistent with the interpretation that the homology groups correspond to separate loci. They indicate also that haplotypes differ in the number and identity of loci they contain.

It is possible, however, that the differences in the number of loci between haplotypes were only apparent, and due to failure of our PCR primers to amplify the same multiple loci from different samples. To exclude this possibility, we typed an additional *O. niloticus* family by Southern hybridization. Since the amount of DNA extracted from 2–3-day-old embryos was not sufficient for Southern hybridization, we extracted DNA from seven gynogenetic embryos ten days postfertilization. DNA samples, along with DNA from the mother and the sperm donor, were digested with the *Hin*dIII restriction enzyme, the digests were separated by electrophoresis, blotted, and the filters were hybridized to a class II *B* exon 3 (highly conserved) probe. The mother's pattern consisted of 18 hybridizing bands, of which six were found in all the embryos, while the remaining 12 segregated into two patterns, A and B. Pattern A, found in four embryos, contained seven bands not found in B, whereas pattern B, present in the other three embryos, contained five bands not found in A (Figure 6). This result is consistent with the haplotype polymorphism detected by the PCR-based method and leads to the conclusion that all the polymorphic class II *B* genes are part of a single cluster; the status of the shared bands presumably corresponding to nonpolymorphic loci remains unclear.

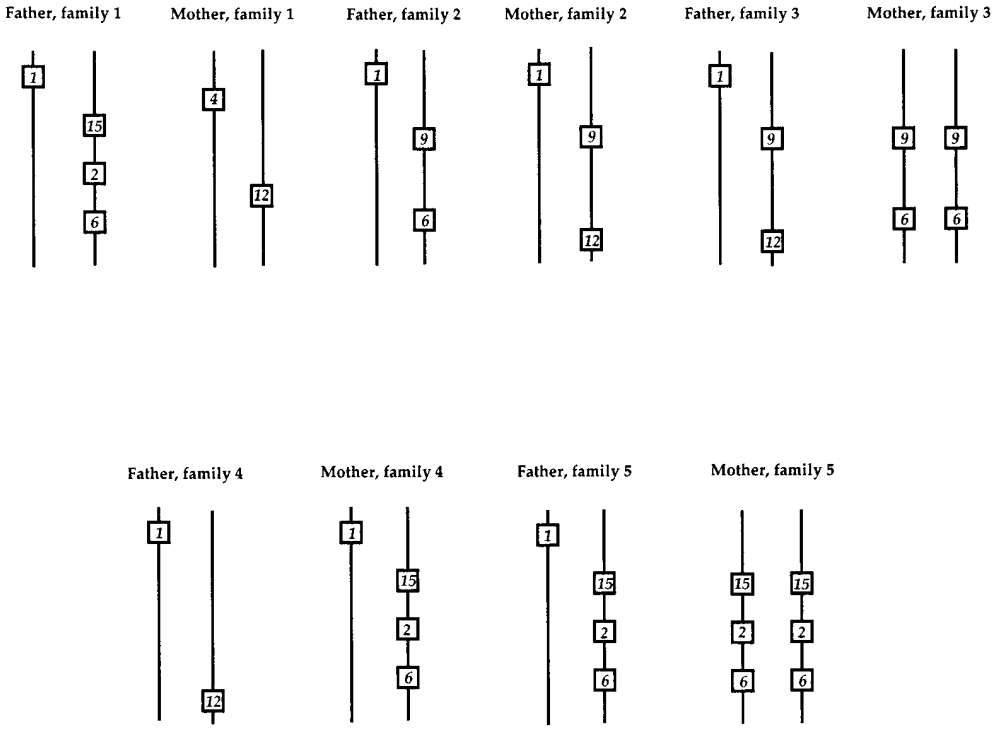

Figure 3.—Parental *Mhc* class II *B* haplotypes deduced from the typing of five *O. niloticus* families. Every parental genotype is represented by the two haplotypes present in their respective homologous chromosomes. Numbered boxes represent each of the loci identified by cloning and sequencing of PCR products. The order of the loci and the distances between them are not known.

Number of class II *B* loci per individual: The total number of class II *B* groups (loci) defined by PCR studies in several cichlid species is greater than what could be found in any of the haplotypes described in the preceding sections. This discrepancy could mean either that not all loci are present in all individuals, or that our PCR primers do not amplify all class II *B* loci. To distinguish between these two possibilities, we carried out two additional experiments. In the first experiment, we digested genomic DNA isolated from two *O. niloticus* individuals (*Orni 1* and *Orni 2*) with the *Hin*dIII restriction enzyme, divided each digest into two aliquots, and separated the digests by electrophoresis in two agarose gels under identical conditions. One gel was then blotted and hybridized to a class II *B* exon 3 probe, whereas the other gel was sliced into 2.5-mm sections along the entire length of the individual lanes. DNA eluted from sections corresponding to positive bands of the Southern blot was PCR-amplified with primers flanking intron 1 (TU383 and TU957), and the PCR products were cloned and sequenced. Of the 14 positive bands detected by Southern hybridization in the *Orni 1* individual (Figure 7A), four bands could not be amplified, presumably because they did not contain intron 1 sequences at all (due to the presence of a *Hin*dIII site between exon 2 and exon 3) or because the class II *B* sequences they contained were not amplifiable with the primers used. Since a survey of complete genomic class II *B* sequences (Ono *et al.* 1993a) from several homology groups did not show the presence of intragenic *Hin*dIII sites, it would be inferred that the four bands contain additional, thus far unidentified loci. The remaining 10 of the 14 bands could be identified as containing intron 1 sequences of seven loci: *1*, *2*, *3*, *4*, *9*, *12*, and *14*. Of the 12 positive bands detected in the *Orni 2* individual (Figure 7A), eight could be assigned to six loci (*1*, *2*, *3*, *4*, *9*, and *14*), while the remaining four could not be amplified at all, presumably for the same reason as the four *Orni* 1 bands. Regardless of what the identity of the additional bands might be, it is clear that neither *Orni* 1 nor *Orni* 2 contains all 17 identified class II *B* loci.

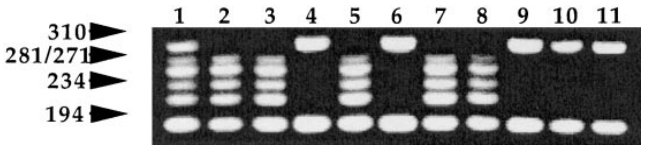

Figure 4.—Segregation analysis of *Mhc* class II *B* loci in *O. niloticus*. Intron 1 generic amplifications were performed from Orni mother 80 and 10 haploid embryos, using primer pair TU383-TU957. The multiple segregating PCR products were resolved in 2% agarose gels and stained with ethidium bromide. The numbers on the left indicate the size (in bp) and position of the molecular size marker bands (ϕx-174-RF DNA-*Hin*dIII digest; Pharmacia LKB). Lane 1: mother's PCR products; lanes 2–11: gynogenetic embryos' PCR products. The two segregation patterns are observed in lanes 2, 3, 5, 7, 8, and in lanes 4, 6, 9, 10, 11, respectively. The genotypes of the mother and embryos were identified by cloning and sequencing of the amplified fragments. Similar experiments were conducted for 8 other *O. niloticus* families (not shown).

In the second experiment, we used genomic DNA from mother 77 in the same way as above, the only difference being the use of an exon 2 instead of an exon 3 probe. The reason for this change was to assure

TABLE 2

Segregation of *Mhc* class II *B* genes in 9 *O. niloticus* haploid embryos

Family	Individual haplotypes	Observed genotype	No. of events	Segregation χ^2
80	Parental A	*14-9-12-10-2*	11	
	Parental B	*14′-1*	6	1.5
	Recombinant A	*14′-1-9-12-10-2*	1	
	Recombinant B	*14*	1	
77	Parental A	*14-9-12-10-2*	13	
	Parental B	*14′-1*	7	1.8
81	Parental A	*14-10-2*	11	
	Parental B	*14′-1-10′-2′*	9	0.2
21	Parental A	*14-10-2*	12	
	Parental B	*14′-1-10′-2′*	8	0.8
79	Parental A	*14-1-6*	19	
	Recombinant A	*14-6*	1	—
9	Parental A	*14-4*	12	
	Parental B	*14′-1*	8	0.8
	Recombinant A	*14-1-6*	1	
20	Parental A	*14-1-12-2-4*	11	
	Parental B	*14′-12′-2′-4′*	9	0.2
69	Parental A	*14-1-9-12-2-4*	20	
34	Parental A	*14-1-9-12-10-2-6*	9	—
	Parental B	*14′-1′-9′*	10	0.1
	Recombinant A	*14′-1′-9′-12-10-2-6*	1	

Families are designated according to the code of the mother used to generate them. Parental haplotypes "A" and "B", in each case, correspond to the two major segregants observed. Loci within a haplotype are designated according to the homology group identified through sequencing of the corresponding PCR products. Second alleles within a locus are designated with the same number and a "′". In all cases, the χ^2 obtained corresponds to $P > 0.05$.

that the primers used would amplify all the bands detected by the hybridization probe, even in the presence of intragenic *Hin*dIII sites. All seven hybridizing bands (Figure 7B) could be PCR-amplified in this case and the seven sequences obtained (*1*, *2*, *9*, *10*, *12*, *14*, and *14′*) corresponded precisely to the six loci found in this mother's progeny in the segregation experiment described in the preceding section. We conclude, therefore, that the differences between haplotypes in the number of loci borne by them are real, that mother 77 does not bear all class II *B* loci and that our primers consistently amplify a significant subset of them.

Genetic map of the class II *B* region: To determine the gene order and genetic distances between some of the class II *B* loci, an additional set of 160 haploid embryos from mother 77 was typed by PCR amplification, cloning, and sequencing (data not shown). This female was chosen because of her two very different haplotypes and the large number of loci per haplotype. The results of the typing reveal that all the tested loci are linked to one another [of the 160 offspring, 130 bore the complete parental haplotypes, which gives a χ^2 value of 62.5 ($P < 0.001$) for parental versus recombinant types], but that recombinations among various pairs of loci occur at appreciable frequencies (Table 3). Among the 30 recombinants, seven different haplotypes could be detected. To explain some of the new haplotypes, eight double recombinations had to be postulated. A linkage map based on the recombination frequencies is shown in Figure 8. The positions of loci *4* and *6* on this map were deduced from the information available for families 9, 20, and 34; the intergenic distances, however, could not be determined.

Tests of gene expression: To determine which of the class II *B* loci are transcribed, we isolated mRNA from three cichlid species, reverse-transcribed it into complementary DNA, prepared λgt10 (Stratagene) cDNA libraries and tested these by PCR, cloning and sequencing using the generic primer pairs TU383-TU377 and TU383-TU957, as well as the *Auha*-specific primers TU 1209-1213, 1247, 1248, 1253, and 1254. The three species were *A. hansbaenschi* from Lake Malawi, *H. nubilus* from Lake Victoria, and *O. niloticus*. The genomic DNA of the same individuals was also PCR tested and the loci borne by them identified. The genomic testing of *A. hansbaenschi* revealed the presence of class II *B* loci

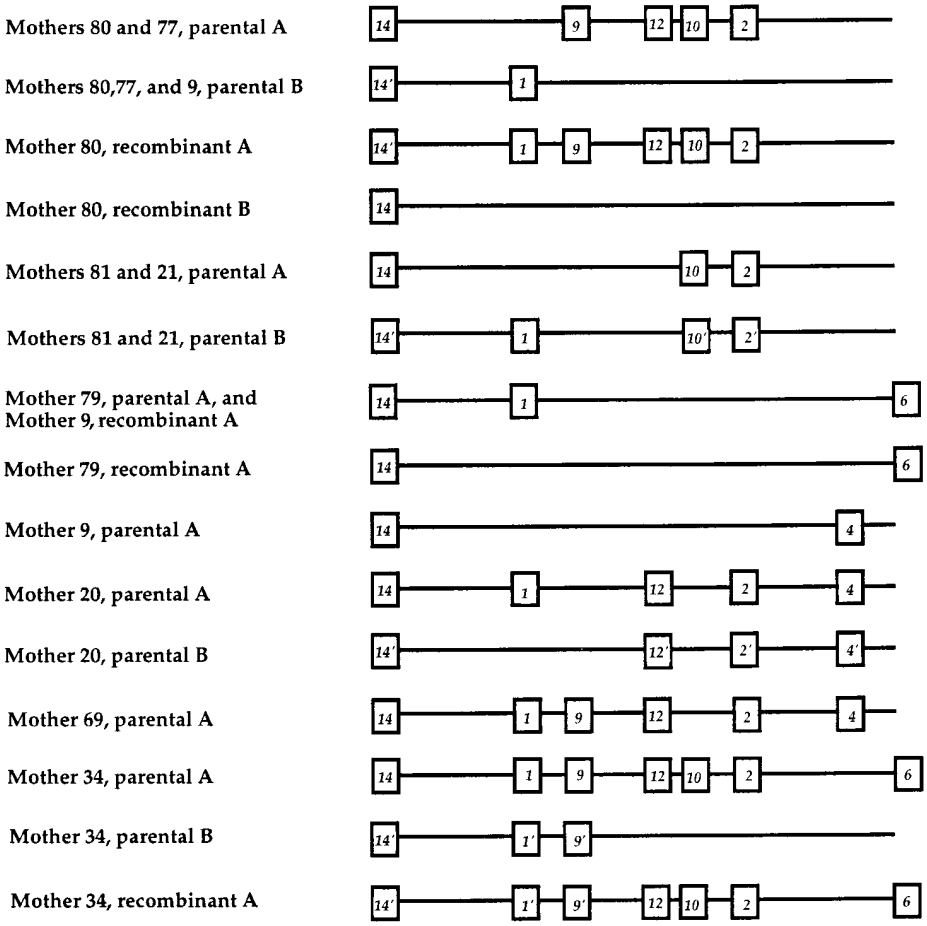

Figure 5.—*Mhc* class II *B* parental and recombinant haplotypes deduced from the segregation analysis of nine *O. niloticus* families. Each individual locus was identified by sequencing of their PCR products. Numbered boxes correspond to distinct loci identified in the offspring. Designations are given by the mother where the haplotype was found. Sharing of haplotypes between different mothers is indicated by multiple designations. Alleles within a locus are indicated by a prime (′) symbol.

1, *2*, *4*, and *6* and the PCR screening of the cDNA library revealed the presence of clones corresponding to these four loci in the mRNA pool. Similarly, testing for *H. nubilus* revealed the presence in both the genome and the mRNA pool of sequences derived from loci *2*, *4*, and *12*. Finally, both the genome and the mRNA pool of two *O. niloticus* individuals yielded sequences derived from loci *1*, *2*, *3*, *9*, *12*, and *14* in one individual, as well as *2*, *3*, *4*, *9*, and *14* in the second individual. Each individual contained one locus whose transcript could not be found in the cDNA libraries (loci *4* and *1* in individuals 1 and 2, respectively). Because transcripts of these loci were found in other species, we assume that the failure to detect them in the two *O. niloticus* individuals might be due to differences in the level of expression. In all cases, the cDNA clones appeared to be derived from correctly spliced transcripts with intact reading frames. From the combined data we conclude that at least loci *1*, *2*, *3*, *4*, *6*, *9*, *12*, and *14* are transcribed in different individuals.

DISCUSSION

Originally, the homology groups of the cichlid *Mhc* class *B* genes were defined on the basis of sequence similarity in intron 1 and exon 2 (the most polymorphic of the six class II *B* exons), clustering on phylogenetic trees, and differences in intron 1 length (D. Klein *et al.* 1993; Ono *et al.* 1993b). In some cases, the intron 1 sequences of different groups were so divergent that they were unalignable. These observations suggested that the groups probably represented distinct loci, but direct evidence for this supposition was not available and the possibility that they were ancient lineages at a few loci could not be ruled out. The data obtained in this study demonstrate clearly that each group corresponds to a separate locus. The sperm typing, the segregation analysis of the haploid embryos, gyno- and androgenetic offspring, and the sequencing of restriction fragments isolated from Southern gels all support this conclusion. In previous studies (D. Klein *et al.* 1993; Ono *et al.* 1993a,b), we identified 10 homology groups; in the present and in a further parallel study (E. Málaga-Trillo H. Tichy, P. Walsh, C. Wood, J. Maina, V. Vincek and J. Klein, unpublished data) we increased the number to 17 groups. Of these, we were able to obtain evidence for a separate locus status of groups *1*, *2*, *3*, *4*, *5*, *6*, *9*, *10*, *11*, *12*, *13*, *14*, and *15*. Because the remaining groups are as different from each other and from those identified as distinct loci, we assume that they, too, represent different loci. The various cichlid species therefore possess together at least

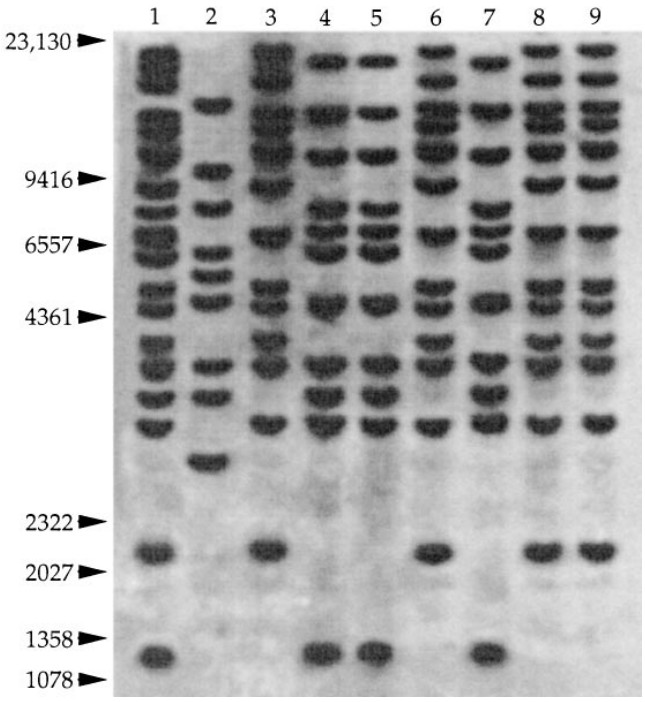

Figure 6.—Southern hybridization of genomic DNA *Hin*dIII restriction fragments from one *O. niloticus* family to an *Mhc* class II *B* (exon 3) probe. Lane 1: mother; lane 2: male UV-donor; lanes 3–9: mitotic gynogenetic offspring. Two types of segregants are observed: pattern A on lanes 3, 6, 8, 9; and pattern B on lanes 4, 5, 7. Arrowheads on the left side indicate the size, in bp, of the molecular weight marker bands.

17 different class II *B* loci, all of which are polymorphic in at least some of the species (loci 7 and 13 are represented by single sequences, but typing of more species might reveal additional alleles). In addition to these, there might also be a few loci which show little variation and could therefore not be studied easily with the present approaches.

No obvious defects were found in the genes at the various loci and hence there was no indication that some of them might be pseudogenes. Furthermore, of the 17 loci, eight could be shown to be occupied by transcribed genes and, as explained above, we assume that, in fact, genes at all 17 loci are transcribed. This assumption makes sense in view of the fact that different haplotypes bear different loci and that each individual must possess at least some functional class II *B* loci. The polymorphism of the 17 loci also supports the assumption that all of them are transcribed in different species because high exon 2 variability is normally associated with functionality (Klein and Figueroa 1986). At the same time, however, our data indicate that probably no individual carries all 17 loci in a single haplotype. The highest number of loci found in one single haplotype found thus far is 13 (Figure 6, pattern A in lanes 3, 6, 8, and 9) and some haplotypes carry as few as one polymorphic locus. The haplotype polymorphism (as reflected in the number and composition of loci) is apparently large: in the small sample tested, we were able to identify 21 distinct haplotypes. Moreover, the "turnover" of the haplotypes is also high: in the segrega-

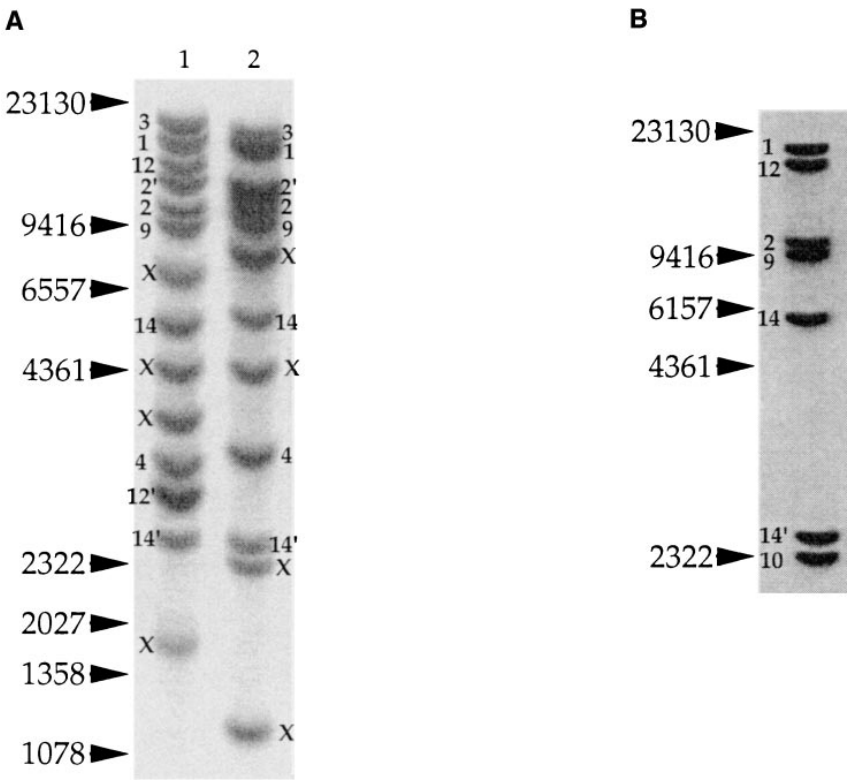

Figure 7.—Southern hybridization of genomic DNA *Hin*dIII restriction fragments from two *O. niloticus* individuals (A) and from *O. niloticus* mother 77 (B) to an *Mhc* class II *B* (exon 3) probe. (A) lane 1: *Orni 1*; lane 2: *Orni 2*. In both A and B, arrowheads on the left side indicate the size, in bp, of the molecular weight marker bands. Small numbers beside each band correspond to the loci identified by PCR amplification of the gel slices. Unidentified bands are designated with an "X".

TABLE 3

Haplotypes found among 160 mitotic gynogenetic embryos from mother 77 and their frequencies

Haplotypes detected	No. of events	Frequency (%)
Parental:		
A: (loci *14*, *9*, *12*, *10*, *2*)	68	42.5
B: (loci *14′*, *1*)	62	38.75
Single recombinants:		
R1: (locus *14*)	12	7.5
R2: (loci *14′*, *1*, *2*)	6	3.75
R3: (loci *14*, *9*)	4	2.5
Double recombinants:		
R4: (loci *14′*, *1*, *9*)	2	1.25
R5: (loci *14′*, *2*, *10*)	2	1.25
R6: (loci *14′*, *2*)	2	1.25
R7: (loci *14′*, *9*, *2*)	2	1.25
Total	160	

Parental haplotypes are designated A and B, and recombinants are shown as R1–R7. Loci contained in each haplotype are shown in parentheses. Segregation χ^2 for parental haplotypes = 0.17 ($P > 0.05$). Linkage χ^2 (parental *vs.* all recombinants together) = 62.5 ($P < 0.001$).

tion studies we were able to witness the generation of 9 new haplotypes by intergenic recombination.

The polymorphic loci are apparently all members of a single class II *B* cluster but, judging from the interlocus recombination frequencies, the physical distances between the loci are probably quite large and hence the physical mapping of the cluster will therefore not be easy. A tentative map of the cluster could be compiled (Figure 8), but since no haplotype contains all the loci, it is an idealized "consensus" map. It is also possible that both the order of loci and the genetic distances between them vary among some haplotypes. Preliminary data (B. Murray, P. Nilsson, E. Málaga-Trillo, H. Sültmann and J. Klein, unpublished data) indicate that a similar situation (in terms of the number of loci in a cluster and haplotype polymorphism) also exists in the case of cichlid *Mhc* class I loci. The data also indicate that, in contrast to the tetrapod *Mhc*, the cichlid class I and class II *B* clusters are not linked (A. Sato, F. Figueroa, E. Málaga-Trillo, B. Murray, H. Sültmann, J. Klein *et al.*, unpublished results).

Comparison of the cichlid class II *B* cluster with the *Mhc*s of other jawed vertebrates reveals some interesting similarities but also important differences. The large number of expressed class II *B* loci is unprecedented. Although class II *B* clusters of other species may contain many members, usually the number of functional loci is low: probably only one in the zebrafish, two in the mouse and in the domestic fowl, and seven in humans of which, however, only three are highly polymorphic (reviewed by Klein 1986 and Trowsdale 1995). It is believed that, as a rule, large numbers of functional, polymorphic *Mhc* loci are not tolerated by natural selection because they reduce the size of the animal's T-cell repertoire and thus its ability to respond to certain parasites (Takahata 1995). Cichlids do not violate this rule because the number of expressed loci per individual is limited by the haplotype polymorphism. At the level of the individual there is therefore presumably no greater *Mhc* influence on the T-cell repertoire in the thymus in the cichlids than in other species.

Haplotype polymorphism is probably a characteristic of many *Mhc*s but in no other species has it been documented to reach the extent observed in cichlid fishes. A case most closely resembling the situation in the cichlids is that of the *HLA-DRB* region (Klein 1991). In the entire human population, however, only five *HLA-DRB* haplotypes have been detected that differ in the number

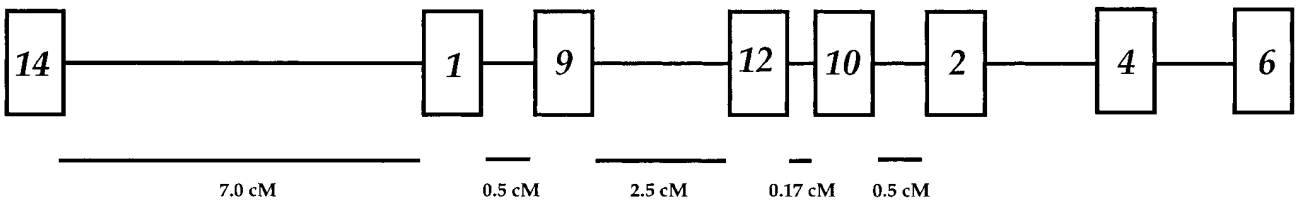

Figure 8.—Genetic map of the *Mhc* class II *B* region in *O. niloticus*. Every numbered box represents a distinct locus. The relative order of the loci and the distances between them were estimated based on the occurrence and frequency of different recombinants among the haploid offspring of all 9 families (Table 3). cM, centimorgan.

of loci they carry. The number of identified class II *B* haplotypes in cichlids, in this early stage of characterization, already exceeds this number.

The aim of this study was to achieve a clear genetic definition of the highly polymorphic *Mhc* loci of cichlids, to be able to use them as markers for phylogenetic reconstruction. Orthologous sequence comparisons can now be made for studying the evolution of the adaptively radiating flocks. Although most of the reported information was obtained for *O. niloticus*, which is only distantly related to the haplochromine species, it should be applicable to other cichlids. Our data indicate that haplochromines, as well as the species and genera of Cichlidae from the other lakes and rivers of Africa, share with *O. niloticus* most if not all of the class II *B* loci. It should, therefore, now be possible to use the *Mhc* class II *B* genes in two ways. First, a specific locus can be selected and its distribution and polymorphism among the populations and species determined. Second, using locus-specific primers, it will be possible to determine the distribution and frequency of haplotypes among the species of all three great lakes of East Africa. Intra- and interspecies comparative analysis of both types of data will provide important information about the size of founding populations, and about the behavior of large ancestral polymorphisms during rapid speciation events. A study using these approaches is underway.

We express our gratitude to Heike Haussmann for the efficient preparation and maintenance of the haploid and gynogenetic samples in Tübingen; Bernd Karstensen and Stefanie Zughorst for excellent technical assistance; Jim Myers and Rafiqul Sarder for their invaluable help and advice with the ploidy manipulation techniques in Stirling; Niamh Ní Bhleithín and Beth S. Coffey for their editorial contribution; and especially Dr. Martin Flajnik in Miami for stimulating discussion of the data.

LITERATURE CITED

Figueroa, F., E. Günther and J. Klein, 1988 *Mhc* polymorphism pre-dating speciation. Nature **335:** 265–267.

Franck, J. P. C., I. Kornfield and J. M. Wright, 1994 The utility of SATA satellite DNAs for inferring phylogenetic relationships among the three major genera of tilapiine cichlid fishes. Mol. Phylogen. Evol. **3:** 10–16.

Fryer, G., and T. D. Iles, 1972 *The Cichlid Fishes of the Great Lakes of Africa: Their Biology and Evolution*. Oliver and Boyd, Edinburgh.

Hussain, M. G., D. J. Penman, B. J. McAndrew and R. Johnstone, 1993 Suppression of first cleavage in the Nile Tilapia, *Oreochromis niloticus* L: a comparison of the relative effectiveness of pressure and heat shocks. Aquaculture **111:** 263–270.

Klein, D., H. Ono, C. O'Huigin, V. Vincek, T. Goldschmidt *et al.*, 1993 Extensive MHC variability in cichlid fishes of Lake Malawi. Nature **364:** 330–334.

Klein, J., 1980 Generation of diversity at *Mhc* loci: implications for T-cell-receptor repertoires, pp. 239–253 in *Immunology 80*, edited by M. Fougereau and J. Dausset. Academic Press, London.

Klein, J., 1986 *Natural History of Major Histocompatibility Complex*. Wiley & Son, New York.

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

Klein, J., 1990 *Immunology*. Blackwell Scientific Publications, Boston.

Klein, J., 1991 Phylogeny of primate major histocompatibility complex *DRB* alleles. Hum. Immunol. **31**(1): 28–33.

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

Klein, J., Y. Satta, N. Takahata and C. O'Huigin, 1993 Trans-specific *Mhc* polymorphism and the origin of species in primates. J. Med. Primatol. **22:** 57–64.

Kocher, T. D., J. A. Conroy, K. R. McKaye and J. R. Stauffer, 1993 Similar morphologies of cichlid fish in Lakes Tanganyika and Malawi are due to convergence. Mol. Phylogenet. Evol. **2:** 158–165.

Maniatis, T., J. Sambrook and E. F. Fritsch, 1989 *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Mayer, W. E., H. Tichy and J. Klein, 1998 Phylogeny of African cichlid fishes as revealed by molecular markers. Heredity **80:** 702–714.

Meyer, A., T. D. Kocher, P. Babasibwaki and A. C. Wilson, 1990 Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. Nature **347:** 550–553.

Moran, P., and I. Kornfield, 1993 Retention of an anchestral polymorphism in the Mbuna species flock (Teleostei: Cichlidae) of Lake Malawi. Mol. Biol. Evol. **10:** 1015–1029.

Myers, J. M., D. J. Penman, Y. Basavaraju, S. F. Powell, P. Baoprasertkul *et al.*, 1995 Induction of diploid androgenetic and mitotic gynogenetic Nile Tilapia (*Oreochromis niloticus*). Theor. Appl. Genet. **90:** 205–210.

Ono, H., C. O'Huigin, V. Vincek and J. Klein, 1993a Exon intron organization of fish major histocompatibility complex class II B genes. Immunogenetics **38:** 223–234.

Ono, H., C. O'Huigin, H. Tichy and J. Klein, 1993b Major histocompatibility complex variation in two species of cichlid fish from Lake Malawi. Mol. Biol. Evol. **10:** 1060–1072.

Sanger, F., S. Nicklen and A. R. Coulson, 1977 DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:** 5463–5467.

Schliewen, U. K., D. Tautz and S. Pääbo, 1994 Sympatric speciation suggested by monophyly of crater lake cichlids. Nature **368:** 629–632.

Srivastava, R., B. P. Ram and P. Tyle, 1991 *Immunogenetics of the Major Histocompatibility Complex*. VCH Publishers, New York.

Sturmbauer, C., and A. Meyer, 1993 Mitochondrial phylogeny of mouthbrooding lineages of cichlid fishes from Lake Tanganyika, Eastern Africa. Mol. Biol. Evol. **10:** 751–768.

Sültmann, H., W. E. Mayer, F. Figueroa, H. Tichy and J. Klein, 1995 Phylogenetic analysis of cichlid fishes using nuclear DNA markers. Mol. Biol. Evol. **12:** 1033–1047.

Takahata, N., 1995 *Mhc* diversity and selection. Immunol. Rev. **143:** 225–247.

Trowsdale, J., 1995 "Both man & bird & beast": comparative organization of *Mhc* genes. Immunogenetics **41(1):** 1–17.

Verheyen, E., L. Ruber, J. Snoeks and A. Meyer, 1996 Mitochondrial phylogeography of rock-dwelling cichlid fishes reveals evolutionary influence of historical lake level fluctuations of Lake Tanganyika, Africa. Philos. Trans. R. Soc. Lond. B Biol. Sci. **351**(1341): 797–805.

Zardoya, R., D. M. Vollmer, C. Craddock, J. T. Streelman, S. Karl *et al.*, 1996 Evolutionary conservation of microsatellite flanking regions and their utility in resolving the phylogeny of cichlid fishes (Pisces: Perciformes). Proc. Soc. Lond. Ser. B Biol. Sci. **263:** 1589–1598.

Communicating editor: N. Takahata