CORINE P. KRUISWIJK

EVOLUTION OF MAJOR HISTOCOMPATIBILITY GENES IN CYPRINID FISH

MOLECULAR ANALYSES AND PHYLOGENIES

EVOLUTION OF MAJOR HISTOCOMPATIBILITY GENES IN CYPRINID FISH

MOLECULAR ANALYSES AND PHYLOGENIES

Promotor:

Prof. Dr Ir Huub F. J. Savelkoul
Hoogleraar in de celbiologie & immunologie

Co-promotor:

Dr René J. M. Stet
Universitair hoofddocent, leerstoelgroep celbiologie & immunologie

Leden promotiecommissie:

Dr Ronald E. Bontrop
Biomedical Primate Research Centre, Rijswijk

Prof. Dr Ir Johan L. van Leeuwen
Wageningen Universiteit

Prof. Dr Rolf F. Hoekstra
Wageningen Universiteit

Prof. Dr Brian Dixon
University of Waterloo, Canada

CORINE P. KRUISWIJK

EVOLUTION OF MAJOR HISTOCOMPATIBILITY GENES IN CYPRINID FISH

MOLECULAR ANALYSES AND PHYLOGENIES

PROEFSCHRIFT TER VERKRIJGING VAN
DE GRAAD VAN DOCTOR OP GEZAG VAN DE
RECTOR MAGNIFICUS VAN WAGENINGEN
UNIVERSITEIT, PROF. DR IR L. SPEELMAN IN
HET OPENBAAR TE VERDEDIGEN OP
VRIJDAG 1 NOVEMBER 2002 DES NAMIDDAGS
TE VIER UUR IN DE AULA

Evolution of Major Histocompatibility Genes in Cyprinid Fish
Molecular Analyses and Phylogenies
PhD thesis, Wageningen University with summary in Dutch

ISBN 90-5808-735-2

C. P. Kruiswijk, 2002
Wageningen Institute of Animal Sciences
PO box 338, 6700 AH Wageningen, The Netherlands

VOORWOORD

Het boekwerkje dat u nu vast houdt is het wetenschappelijk resultaat van vier jaar promotieonderzoek. Echter voor mij betekent het veel meer dan dat. Het waren vier lastige maar ook zeer leerzame jaren. Vanaf het prille begin van het promotieonderzoek ontbrak om verschillende redenen de motivatie. De kennismaking met de duistere kant van mijn geest was niet erg aangenaam. Maar zoals Albert Einstein schreef *'In the middle of difficulty lies the oppertunity'*. Zo ontstond het idee om naast het promotieonderzoek een doctoraal studie Bedrijfskunde te volgen. En Geert, waarschijnlijk was mijn interesse om je drie jaar lang te assisteren bij de workshop anders stukken minder geweest. Het waren geweldig leuke weken die ik niet had willen missen!

U snapt hoe blij ik ben dat het proefschift er, tegen al mijn verwachtingen in, toch is gekomen. Beste René, samen hebben we het dan toch gered en ook nog eens in vier jaar, bedankt! Ook de vele anderen, die een bijdrage hebben geleverd aan mijn promotieonderzoek, wil ik graag bedanken. Een aantal personen wil ik met naam noemen. Beste Huub, pas in een laat stadium raakte je betrokken

bij mijn promotieonderzoek. Je motiverende discussies en enthousiasme waren precies het goede duwtje in de rug. TRUDI, jij stond altijd weer voor mij klaar, dankjewel! Vele DNA fragmenten heb je voor me *gesequenced* en andere experimenten voor me uitgevoerd. Daarnaast maakte je altijd weer tijd om die clonen die ik snel nodig had nog even snel te *sequencen*. BEJA, ook jij bedankt voor de hulp bij het *sequencen*, je enthousiasme en gezelligheid. NICO, bedankt voor de PC en je altijd goede stemming.

Na drie jaar slechts stukjes staartvin of spierweefsel van Afrikaanse grote barbelen te hebben gezien, mocht ik ze in hun natuurlijke omgeving gaan bekijken. NAND en MARTIN, bedankt voor de fantastische tijd in Ethiopië. Ook de mede AIO's, ANK, HEIDI, JEROEN, JOOP, KRYS, MAAIKE, MARC en PATRYCIA, mijn studenten, KIM en JOOST en alle niet bij naam genoemde CBI-ers en EZO-ers, bedankt voor de gezelligheid en prettige samenwerking.

Beste paranimf GEERT, dankjewel voor de vele uren peptalk, ze waren onmisbaar. Paranimf MARK, het waren mooie avonden in de kroeg, met ettelijke glazen bier en altijd weer een discussie over dat 'ene onderwerp' (daar gaan we het vanavond niet over hebben hoor...).

Tenslotte familie en vrienden, door de combinatie studie en promotie bleef er vaak weinig tijd voor julie over, bedankt voor de vriendschap! Een speciaal woordje voor een aantal personen. AURYN meiden en partners, dank jullie wel voor jullie geduld, tolerantie en waardering (echter na wintersport 2002, met zo'n fantastische ontvangst, overweeg ik om gewoon altijd later te komen...). Sommige studiegenoten, jullie weten zelf wel wie ik bedoel, bedankt (zag ik daar een karper...) en een aantal huisgenoten voor de versnaperingen als ik weer eens vergat te ontbijten en te luchen, de altijd werkende server en de mooie plaatjes in dit proefschrift. Tot slot, lieve pa en ma, dank jullie voor de onvoorwaardelijke steun en het vertrouwen, zelfs als ik weer eens aan de zoveelste studie begon.

CORINE KRUISWIJK

CONTENTS

Voor MENNA

CHAPTER 1

GENERAL INTRODUCTION

Corine P. Kruiswijk

Cell Biology and Immunology Group,
Wageningen Institute of Animal Sciences, Wageningen University,
PO box 338, 6700 AH Wageningen, The Netherlands.

INTRODUCTION

All living organisms are exposed to pathogens in all life history stages and must prevent invasion or destruction by them to survive. Various phyla of the animal kingdom exploit different strategies to achieve defence against pathogens. Broadly speaking these strategies fall in two categories: innate (or non-specific) immune responses and adaptive (or specific) immune responses. During evolutionary processes, both categories have not always coexisted. Throughout the animal kingdom, defence strategies involving adaptive immunity appear to have arisen first in members of the vertebrate monophyletic lineage of Gnathostomes (from sharks to man), while adaptive immunity is absent in more primitive vertebrates belonging to the superclass Agnatha (e.g. lamprey, hagfish) and invertebrates. Among Gnathostomes, the essential architecture of the adaptive immune system, as we know it in mammals, seems well preserved with only some minor variations.

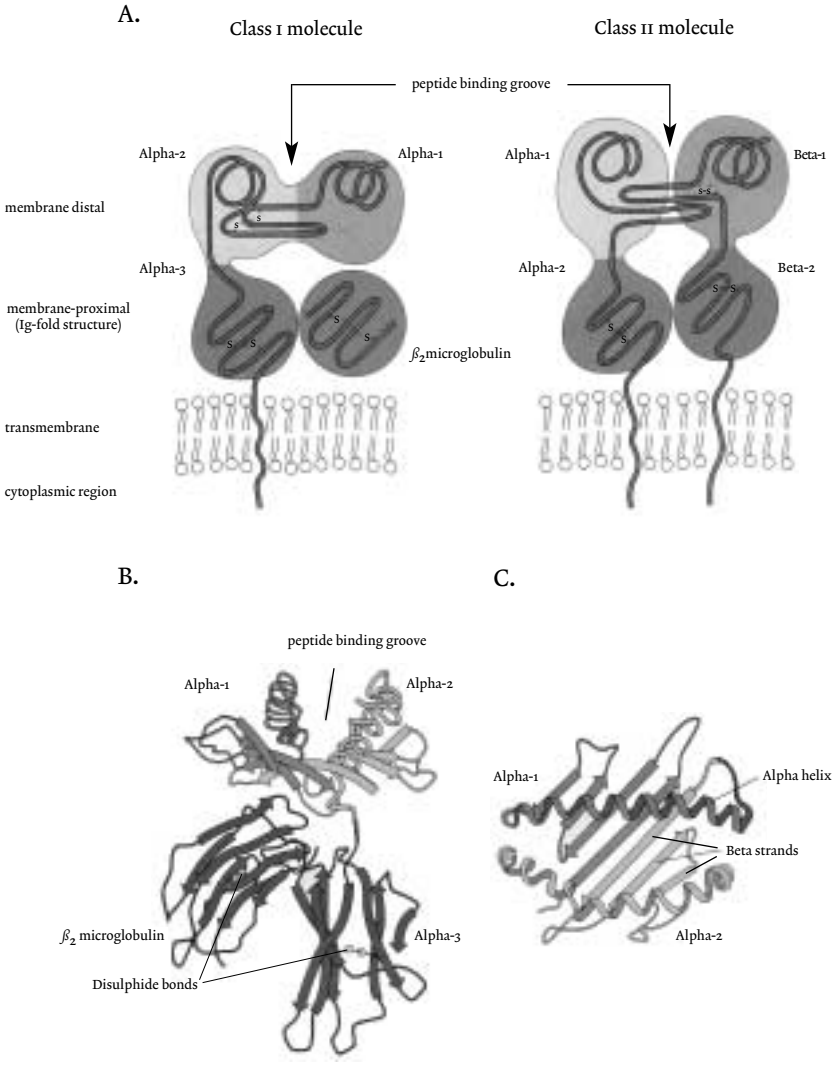


Figure 1: The structure of MHC molecules. (A) Schematic representation of the MHC class I and class II molecules. Disulphide bonds are indicated by s-s. (B) Three dimensional structure of a MHC class I molecule as seen from the side or (C) from the top (Modified[82]).

Two subfamilies of cell surface glycoproteins that play a crucial role in adaptive immunity are the major histocompatibility complex (MHC) class I and class II molecules. The history of major histocompatibility complex molecules starts at the beginning of the previous century with the discovery of their role in transplant rejections[1]. However, their biological function remained unknown until the end of the sixties[2,3]. MHC class I and class II genes encode structurally similar molecules that are equipped to bind degraded self and non-self peptides which they display to particular cells of the immune system. Displaying degraded non-self peptides derived from viral, bacterial or parasitic pathogens will trigger an immune response.

The first MHC gene was isolated from human[4] in 1980, while eight years later the first non-mammalian MHC equivalent was isolated from chicken[5]. The last to follow were class I and class II genes in teleosts[6] (bony fishes) in 1990. Approximately half of all extant vertebrate species belong to the group of bony fishes, with 20,000 representative species with an immense variety, which are found in many different aquatic ecosystems. This large group of bony fishes include cyprinids like the African 'large' barb (*Barbus intermedius*), common carp (*Cyprinus carpio*), and zebrafish (*Danio rerio*).

STRUCTURE AND FUNCTION OF MHC MOLECULES

MHC class I molecules are transmembrane heterodimers of a heavy alpha chain (M_r 45,000) non-covalently associated with a light chain β_2-microglobulin (M_r 12,000). The heavy chain consists of three extra-cellular domains designated alpha-1 to alpha-3, a transmembrane and a cytoplasmic region (*fig. 1*). The alpha-2 and alpha-3 domains both possess intrachain disulphide bridges enclosing approximately 60 to 85 amino acids. The membrane proximal alpha-3 domain of the alpha chain and the β_2-microglobulin are both homologous with the immunoglobulin C1 domains. The membrane distal alpha-1 and alpha-2 domains constitute a platform of eight anti-parallel beta-strands supporting two anti-parallel alpha-helices. The disulphide bridge in the alpha-2 domain connects the N-terminal beta-strand with the alpha-helix of this domain. The two alpha-helices of the alpha-1 and alpha-2 domain are separated by a long groove with occluded ends which accommodates peptides of nine to twelve amino acids in length[7,8].

The extra-cellular domains of MHC molecules are encoded by single exons,

with exception of the class IIB exon-3 of the Acanthopterygii which has acquired an intron[9]. Several exon-intron organisations of MHC genes have been observed among vertebrates. The exon-intron organisations differ in the number of exons encoding regions other than the extra-cellular domains. MHC class I genes usually consist of eight exons and seven introns, while class II genes consist of four to six exons separated by three to five introns. Large differences are known to exist for intron lengths (reviewed[10,11]).

Although MHC class I and class II molecules are similar in overall structure, they differ in class I being a transmembrane heterodimer consisting of a class I heavy chain non-covalently associated with β_2-microglobulin, while class II molecules are transmembane heterodimers composed of an alpha and a beta chain (both approximately M_r 30,000). Each chain consists of two extra-cellular domains designated alpha-1 and alpha-2 or beta-1 and beta-2, a transmembrane and a cytoplasmic region (*fig. 1*). The membrane proximal alpha-2 and beta-2 are both homologous with the immunoglobulin C1 domains. The membrane distal alpha-1 and beta-1 domains together constitute a platform of eight anti-parallel beta-strands that supports two anti-parallel alpha-helices and form a peptide binding groove, similar to that of MHC class I molecules. However, the peptide binding groove of MHC class II molecules is more open at both ends which allows peptides to extend beyond the groove. Due to this characteristic it can accommodate longer peptides, varying from 12 to 25 amino acids[12]. Although both MHC class I and class II molecules bind small peptides, they present these peptides to different cells of the immune system and they differ in their tissue distribution.

MHC class I molecules are expressed on all nucleated cells and present endogenous (intra-cellular) derived self and viral peptides to CD8-positive T lymphocytes. The presentation of viral peptides elicits a cytotoxic immune response. The CD8 receptor interacts with a conserved region in the alpha-3 domain of class I molecules. In addition to their role in antigen presentation, class I molecules are also known to play a role in natural killer (NK) cell mediated responses as part of innate immunity. Specific membrane receptors present on NK cells are capable of detecting down regulation of MHC class I molecules on cell surfaces. Two categories of these receptors have been identified in humans; the CD94/NKG2A heterodimer belonging to the C-type lectin-like superfamily and the killer-cell immunoglobulin-like receptor (KIR) family[13].

Expression of MHC class II molecules is restricted to epithelial cells, B-cells, activated T-cells and most importantly professional antigen presenting cells of the immune system. Although in diseased states it is known that class II molecules are also expressed on other cells which can result in auto-immune responses. The class II molecules present exogenous (extra-cellular) derived, internalised and degraded peptides from pathogens to CD4-positive helper T lymphocytes that, when activated, provide help and guidance to B-cells through secretion of cytokines resulting in a humoral immune response. The CD4 receptor interacts with conserved regions the alpha-2 and beta-2 domains of the class II molecule.

HUMAN MAJOR HISTOCOMPATIBILITY COMPLEX GENES

The major histocompatibility complex genes derived their name from clustering in a single genomic region and provoking strong allograft rejection. Characteristics of MHC genes are essentially the same for all mammalian species investigated with human being the most extensively studied to date. The human MHC (or HLA complex) is located on chromosome 6, spanning a region of approximately 4 million basepairs (MB). This region is divided into three gene clusters[14], in order of physical location from centromer to telomer: class II, class III and class I (*fig. 2*).

The class I cluster includes genes encoding the heavy alpha chain, *HLA-A, -B, -C, -E, -F, -G, -H, -J, -K, -L,* and *-X*, of which the first three encode the classical class I (Ia), followed by three non-classical class I (Ib) genes and five pseudogenes. Other class I-like encoding genes, *MICA, -B, -C, -D,* and *-E*, are found dispersed through the class I region[15]. Recently, another class I-like molecule has been isolated, the hereditary hemochromatosis gene (*HFE*), which maps approximately 4 MB telomeric of *HLA-F*, extending the traditional MHC region to a total of 8 MB. As noted in the literature, this *HFE* gene should not be confused with the *HLA-H* gene[16]. While the gene products of classical and non-classical class I genes are similar in structure, they differ in function and tissue distribution. The class I light chain encoding gene, β_2-microglobulin, is not located in the MHC, but on chromosome 15. The class II cluster comprises several loci, *HLA-DM, -DN, -DP, -DO, -DQ,* and *-DR* each comprising genes encoding alpha and beta chains. The *HLA-DP* and *-DM* regions are most proximal to the centromer and separated from the other class II genes by the low-molecular-weight protein

centromeric

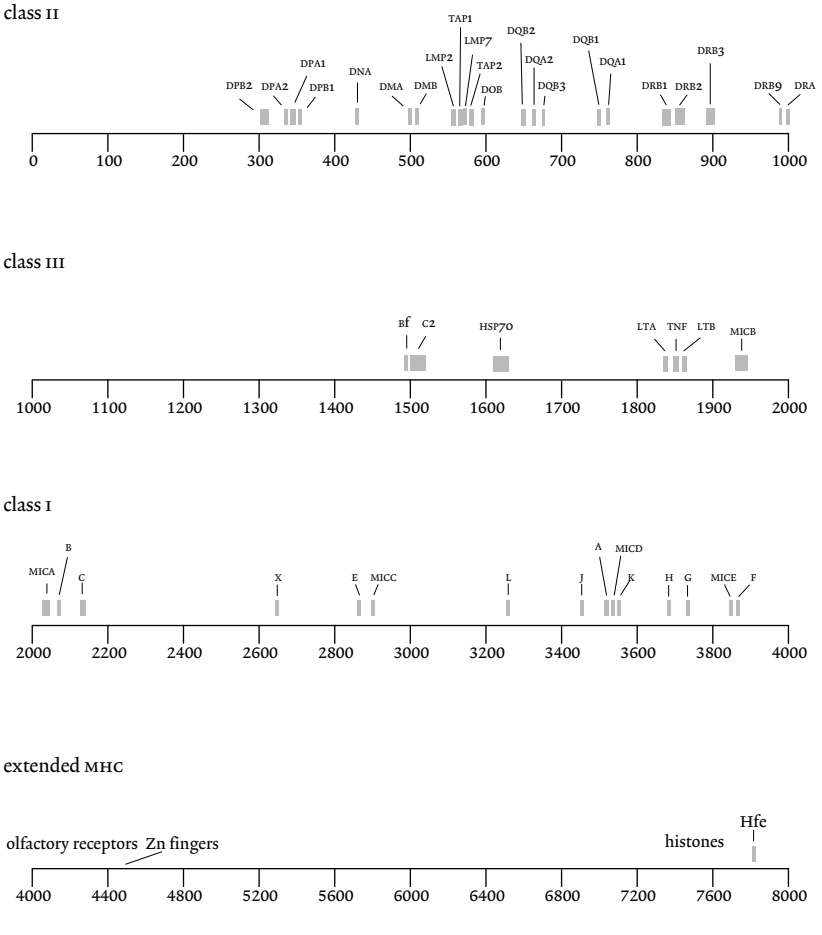


Figure 2: Organisation of the extended major histocompatibility complex in humans. (Modified after RHODES & TROWSDALE, http://www.path.cam.ac.uk/~mhc/mhc.html).

(*LMP*) encoding genes and transporter associated with antigen processing (*TAP*) encoding genes. The LMP and TAP molecules are involved in generating peptides in the cytosol and peptide transport into the endoplasmatic reticulum (ER)[17,18]. The *HLA-DM* and *-DO* genes encode non-classical class II molecules with a similar structure but different function. These molecules have a chaperone function involved in loading peptide onto class II molecules[19].

A region of 1 MB of DNA between the class II and class I region contains a variety of densely packed genes that do not possess an antigen presentation function proper. This region was assigned class III. Some products encoded by this region play a role in the immune system for example members of the complement cascade (C2, C4 and BF) and tumour necrosis factor family members (TNF, LTA and LTB). However, many gene products encoded in this region possess functions that so far seem to bear no relation to the immune system.

HUMAN NON-CLASSICAL CLASS I AND CLASS II GENES

Originally, non-classical MHC class I molecules were distinguished from classical MHC class I molecules by their limited polymorphism and tissue specific expression, although both types of molecules are encoded by genes located in the MHC. Initially, several characteristics of the non-classical class I molecules led to the hypothesis that these genes are essentially non-functional relics of classical class I genes whose ultimate extinction is inevitable[20,21]. Now it has become clear that these non-classical class I molecules are involved in a variety of biological processes, some having little to do with the immune system, and interact with a range of different ligands[22,23]. Although they possess the characteristic structure of classical class I proteins, it remains to be seen whether the molecules are evolutionary relics of once classical class I molecules.

To date several other class I-like genes have been discovered that reside on chromosomes other than the one carrying the MHC proper. All non-classical class I molecules have in common that they exhibit low polymorphism, which contrasts with the high degree of polymorphism of classical class I molecules. The different MHC class I-like molecules also show some other variations to the classical paradigm. For example, the only known soluble class I-like protein is the Zn-alpha-2-glycoprotein (ZAG), which is present in serum and other body fluids and does not associate with the soluble β_2-microglobulin. The groove formed by the alpha-1 and alpha-2 domains of the ZAG molecules binds a polyunsaturated

fatty acid, which is relevant for its function in lipid catabolism[24,25]. Another class I-like molecule that does not associate with β_2-microglobulin is MICA. However, it is highly unlikely that the peptide binding groove of MICA is occupied with a molecule of some sort because the groove formed by alpha-1 and alpha-2 domains is flipped over and probably pointing downwards to the cell membrane[15,26]. MIC molecules act as ligands for the activatory natural killer-cell receptor (NKG2D) expressed on certain cells of the immune system.

The hereditary hemochromatosis protein (HFE[27]) and the neonatal Fc receptor[28] (FcRn) are both class I-like molecules that do associate with the β_2-microglobulin, but the peptide binding groove of these molecules is either narrowed or closed. Thus, these molecules are unable to bind small molecular weight ligands. Yet another group of class I-like molecules are the CD1 molecules that do associate with β_2-microglobulin. They present self and microbial lipids to a specialised subset of T lymphocytes.

The MHC encoded HLA-E and -G non-classical molecules that are phylogenetically most closely related to classical class I molecules and associate with β_2-microglobulin. They bind small peptides, but are oligomorph in nature instead of highly polymorphic like the classical class I molecules. HLA-E molecules in complex with peptides, which are derived from the leader peptides of classical class I molecules, serve as ligands for NK receptors like the heterodimer CD94/NKG2D. The peptides bound are thus not presented to T lymphocytes, but are most likely important in the stabilisation of the HLA-E molecules at the cell surface.

Due to the variety of different class I-like molecules now known, a new nomenclature for these molecules in mammals was proposed based on phylogenetic analyses and chromosome location[29]. Although class I-like molecules have been identified outside the MHC, class II-like molecules have only been identified within the MHC.

EVOLUTION AND POLYMORPHISM

Mutational change in DNA sequences is the primary driving force of the evolution of genes. Such changes in DNA sequence will either be lost in time or spread through a population and eventually be fixed in a species by genetic drift and/ or natural selection. Four basic types of changes may occur in DNA; 1) point mutations, replacement of one nucleotide by another, 2) insertions or deletions, the addition or removal of nucleotides, 3) inversions, the rotation of double-stranded DNA

segments, 4) recombination which includes crossing-over and gene conversion. Another process that leads to rapid mutational change of genes is exon shuffling.

Two processes, genetic drift and natural selection, influence whether such mutational changes will ultimately be lost or fixed in a species. Random genetic drift is a process by which changes in allele frequencies occur solely as chance effects, while natural selection is a non-random process. Natural selection is defined by the non-random differential reproduction and survival of individuals carrying alternative inherited traits. It involves differences in the relative contributions of various genotypes to the next generation.

Most mutational changes arising in a population reduce the fitness of the individuals carrying them and will be selected against (negative selection). Occasionally, a mutational change is as fit as the best allele in the population, thus being selectively neutral. Its fate will be determined by chance events. In rare occasions the fitness of a mutational change is advantageous and will be subject to positive selection. Sometimes mutational changes are only advantageous in heterozygotes and not in homozygotes, resulting in overdominant selection.

Loci known to encode multiple alleles within a population are MHC class I and II. The high variability between alleles of a MHC locus is restricted to positions known as the peptide binding residues of the peptide binding domains. Analyses of polymorphic MHC alleles have shown that point mutations, recombination, and exon shuffling are the main mechanisms creating mutational changes[30-33]. However, only nonsynonymous substitution in the peptide binding region seemed to be advantageous and thus alleles acquiring mutations in this region may be subject to positive selection. Evidence that this region is subjected to positive selection came from examination of nucleotide substitutions in different regions of MHC molecules. Comparison of synonymous nucleotide substitution rates (d_s) and nonsynonymous nucleotide substitution rates (d_n) of peptide binding residues and non-peptide binding residues in the coding regions of MHC genes revealed that d_n rates substantially exceed d_s rates of peptide binding residues[34-37]. In addition, polymorphic alleles are maintained selectively by overdominant selection[34] (heterozygotic advantage).

The basis of selective maintenance of MHC polymorphism is peptide binding and thus pathogen driven. Overdominant selection and other forms of balancing selection are known to be capable of maintaining polymorphism for longer periods than neutral selection[38]. Polymorphic lineages of MHC class I and class II

are known to be maintained for millions of years, predating speciation events, a process known as *trans*-species evolution[39-42]. In general, mammalian MHC class I genes seem to be more rapidly evolving than class II genes. It has been shown that allelic HLA class I lineages were maintained for up to six million years, while certain HLA class II lineages were maintained for up to 35 million years[43-45].

Phylogenetic analysis of DNA or protein sequences is an important tool for studying the evolutionary history of genes or organisms. All life forms are related by descent to one another with closely related genes or organisms descending from more recent common ancestors than distantly related ones. Phylogenetic studies attempt to reconstruct the correct genealogical ties between genes or organisms and estimate the time of divergence between genes or organisms since they last shared a common ancestor. A phylogenetic tree is composed of nodes and branches, with only one branch connecting two adjacent nodes. A node represent the ancestral gene or organism. The branch length usually represents the number of changes in the DNA or protein sequence that have occurred in that branch. Statistical methods are used to estimate these branch lengths and to reconstruct the phylogeny of DNA or protein sequences of interest[180,183].

ORIGIN AND EVOLUTION OF THE MAJOR HISTOCOMPATIBILITY COMPLEX GENES

One of the intriguing questions concerning the MHC involves its origin and evolution. To date there has been no hints of the presence of MHC molecules, as well as other components of the adaptive immune system, in jawless fish or invertebrates. This suggests that the MHC molecules arose rather abruptly in an ancestor of the jawed vertebrates.

All tetrapod species studied to date, including several mammalian species, birds[46] and *Xenopus*[47] also possess tightly linked class I and class II loci. On the other hand all bony fish species investigated including zebrafish[48], medaka[47,49], common carp[50], Atlantic salmon[51,52], and rainbow trout[53] classical class I loci were shown not to be linked to the class II loci, but were found on different linkage groups. Recently, segregation analyses reported tightly linked class I and class II loci in the oldest vertebrate taxon of cartilaginous fishes[54] (Chondrichthyes: sharks, skates and rays). This suggests that tight linkage of class I and class II loci appears to be the ancestral organisation since the bony fish arose after the cartilaginous fish but before the tetrapods. However, it should be kept in mind

that, although the Chondrichthyes are commonly accepted as being sister group to the other extant Gnathostomata (jawed vertebrates), the basal position of the Chondrichthyes relative to the other Gnathostomes is still under debate[55,56].

Several possibilities have been suggested for the absence of linkage of class I and class II loci in teleostean fishes. First, two rounds of genome-wide duplications have been suggested to result in two or four duplicates of the ancestral MHC-bearing region[57-60] that in the teleostean lineage over time has lost either class I loci or class II loci. Second, block duplication of parts of a chromosome bearing an MHC region followed by translocation and subsequent loss of loci. Third, translocation of class II loci from a prototype MHC to other chromosomes in the ancestor of the teleosts[61]. A definitive answer will have to come from genome analysis of teleost fish allowing identification of ancient synteny. As indicated by the structural and functional similarity of class I and class II molecules, they are clearly evolutionarily related. At some point during evolution one of the molecules must have preceded the other. However, whether class I or class II molecules arose first in evolution is still under debate. Some argue that class II molecules emerged first based on phylogeny[62,63] and on thermodynamic grounds[64]. However, others argue that class I molecules arose first by exon shuffling of a proto-peptide binding region (PBR) onto an exon encoding an Ig-like C1 domain[65]. Whether such ancestral PBR existed and whether it was used to bind antigen, peptide or otherwise, remains to be proven. Likewise, it is not known whether classical or non-classical class I molecules arose first. Such problems can only be addressed by examining the molecular evolution of antigen receptor molecules, in particular those of the immunoglobulin superfamily.

CYPRINID FISHES

Cyprinid fishes play an important role in aquaculture, making up some 80% of the 9.5 million tons of fish produced world-wide. They belong to the largest group of vertebrates, the bony fish, that comprises almost half of all 43,000 extant vertebrate species. In cyprinids diploid, tetraploid, and hexaploid species have been identified. Zebrafish (*Danio rerio*[66], 2n=50), common carp (*Cyprinus carpio*[67], 2n=100), and the African 'large' barb (*Barbus intermedius*[68], 2n=150), including the Lake Tana African 'large' barb species flock, represent such different ploidy status. These cyprinid species, representing highly divergent cyprinid genera, *Danio, Cyprinus,* and *Barbus*, are estimated to have diverged 50 and 30 million

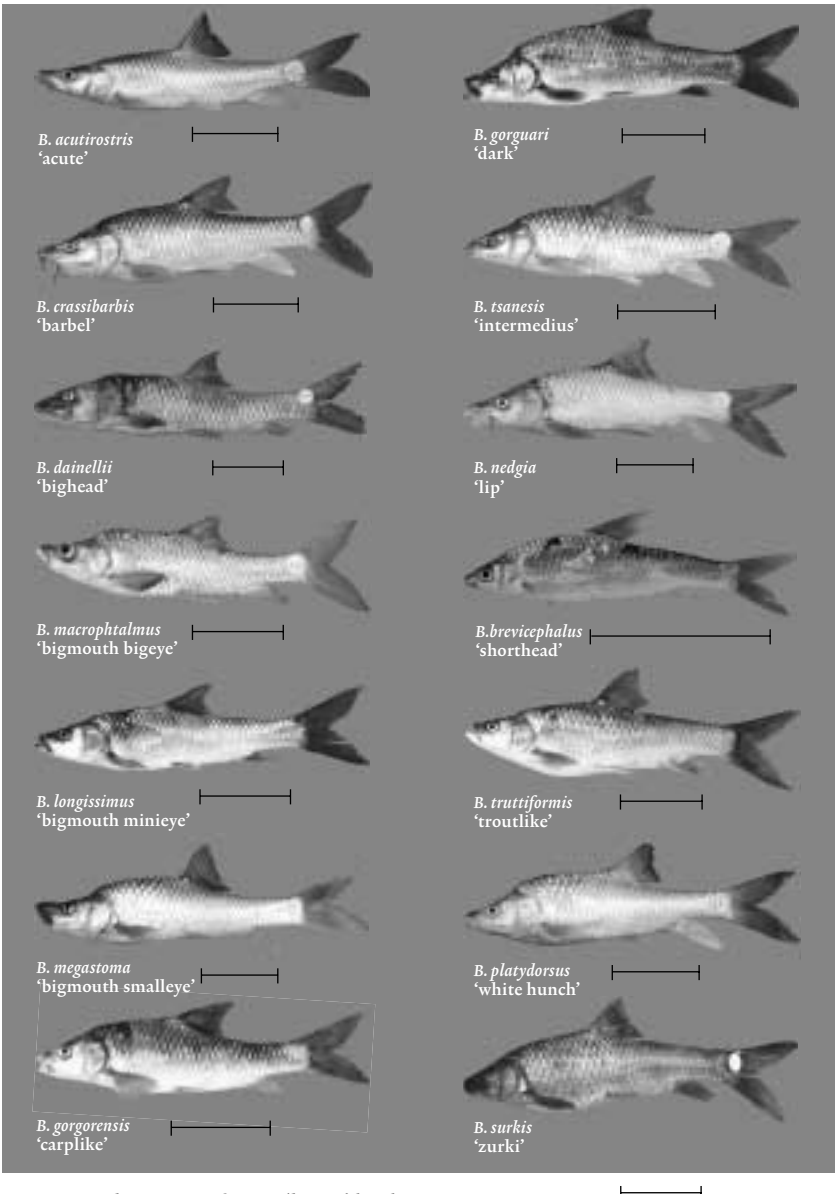


Figure 3: Lake Tana African 'large' barb species. (*B. intermedius* and *B. osseensis* are not shown)

years ago, respectively[69-71]. The different ploidy status, which may have implications for evolution, inheritance and usage of MHC genes, and their evolutionary distant relationship, makes these species interesting to study.

Zebrafishes are relatively small making them somewhat less suitable for the study of the immune system. Carps are much larger fishes and thus more suitable. Their main advantage lies in the availability of inbred or clonal carp, which are imperative experimental animals to be used in immunological studies[72-74]. The Lake Tana African 'large' barb species are of great interest since they have undergone adaptive radiation during 5 million years of isolation of most likely a common ancestral *Barbus intermedius* population, resulting in 15 novel species[75] (*fig. 3*). The isolated aquatic ecosystem of Lake Tana provides a unique natural system to study evolution of the MHC. The lake is situated in the north-western highlands of Ethiopia and was formed in late Pleistocene times by a tectonic event, creating a 40 meter high waterfall, that blocked the outlet to the Blue Nile[76]. This effectively prevented gene flow between the lake and the Blue Nile and its tributaries.

AIM OF THE THESIS

The abundance of modern bony fishes among vertebrates implies that immune function in general is not impaired, although they possess unlinked class I and II genes which may have influenced MHC function. A thorough knowledge of MHC genes in this group may provide important insights into the evolution of these genes in bony fish. In addition, comparative genomics with other vertebrate lineages may help to understand the mechanism that led to the evolution of a complex of class I and II genes in other vertebrates.

In the past, the MHC of two cyprinid species, common carp, and the African 'large' barb was thoroughly studied[77,78]. These studies are reviewed in chapter 2. In addition, this chapter includes several pilot experiments that provided the basis of chapter 3 to 5. A study performed on the hexaploid Lake Tana African 'large' barbs[68] (2n=150) indicated that fewer than six class IIB genes in a single individual (BOB) were found to be expressed, suggesting some sort of silencing mechanism for MHC genes[70,74]. Mammalian MHC class I and class II genes are known to be co-dominantly expressed. However, it has been suggested that expression of too many alleles in a single individual might be deleterious. Therefore, in polyploid fishes a silencing mechanism might be present to overcome expression of too many MHC class I and II molecules, similar to that observed in several

Xenopus species. In these species the number of class I and class II genes does not increase with higher ploidy of the genomes[79,80]. This prompted the study of the expressed and genomic MHC class I and class II sequences in a single Lake Tana African 'large' barb shore-complex individual as described in chapter 3.

The study of class IIB genes in the African 'large' barb species flock revealed that the four species investigated seem to have a unique pool of class IIB sequences which probably evolved from a common ancestral suit within a time span of 5 million years[70]. However, a pilot study on evolution of class I sequences in this species flock indicated that they were shared between the different species of the species flock. In chapter 4, these differences in modes of evolution of class Ia and IIB in the African 'large' barb species flock were investigated. We extended the study on MHC class IIB encoding sequences performed by DIXON and coworkers[70] with six Lake Tana African 'large' barb species (*B. intermedius, B. brevicephalus, B. macrophtalmus, B. megastoma, B. platydorsus,* and *B. surkis*). In addition, African 'large' barbs (*B. intermedius*) from the Blue Nile and its tributaries were included. To test whether MHC class I sequences evolved in a similar species specific manner, class I encoding sequences from seven different Lake Tana African 'large' barb species and African 'large' barbs from the Blue Nile, comprising in total 35 individuals, were analysed.

Previously, Southern blot hybridisation performed at extremely low stringency on restriction-enzyme digested high molecular weight DNA of common carp of different geographical origins, using a class I Z exon-4 probe, detected 9-12 hybridising fragments[81]. These data suggested the existence of additional class I Z sequences in common carp. A pilot study, described in chapter 2, revealed the partial coding sequence of the extra-cellular domains of two novel class I Z lineage sequences[74], *Cyca-Zr2* and *Cyca-Zr3* . In a phylogenetic analysis, these sequences clustered with other cyprinid class I Z lineage sequences, but formed a separate cluster and were therefore renamed to *Cyca-ZE*0101* and *Cyca-ZE*0201* in chapter 5. In this chapter, we identified the complete coding sequence of *Cyca-ZE*0101*. Further, we analysed the presence of these class I ZE genes in zebrafish (*Danio rerio*[66], 2n=50) and African 'large' barb (*Barbus intermedius*[68], 2n=150), investigated their expression, exon-intron organisation and the characteristics of polymorphic residues involved in peptide binding. Finally, in the last chapter the observations described in each chapter are combined, discussed, and compared with MHC characteristics described for warm-blooded vertebrates.

CHAPTER 2

MAJOR HISTOCOMPATIBILITY GENES IN CYPRINID FISHES: THEORY AND PRACTICE

René J. M. Stet, Corine P. Kruiswijk,
Jeroen P. J. Saeij, Geert F. Wiegertjes

Cell Biology and Immunology Group,
Wageningen Institute of Animal Sciences, Wageningen University,
PO box 338, 6700 AH Wageningen, The Netherlands.

Published in: Immunological Reviews (1998) Vol. 166: 301-316
Reprinted with kind permission of Munksgaard
International Publishers Ltd.

ABSTRACT

The first teleostean MHC sequences were described for carp. Subsequent studies in a number of cyprinid fishes showed that the class I sequences of these fishes are of particular interest. Two distinct lineages (*Cyca-Z* and *Cyca-U*) are found in common and ginbuna crucian carp, but only the *U* lineage is present in zebrafish, and other non-cyprinid species. The presence of the *Z* lineage is hypothesised to be the result of an allotetraploidisation event. Both phylogenetic analyses and amino acid sequence comparisons suggests that *Cyca-Z* sequences are non-classical class I sequences, probably not unlike CD1. The comprehensive phylogenetic analyses of these sequences revealed different phylogenetic histories of the exons encoding the extra-cellular domains. The MHC genes are studied in laboratory and natural models. The natural model addressed the evolution of MHC genes in a *Barbus* species flock. Sequence analysis of class I and class II supported the species designation of the morphotypes present in the lake, and as a consequence the *trans*-species hypothesis of MHC polymorphism. The laboratory model involves the generation of gynogenetic clones,

which can be divergently selected for traits such as high and low antibody response. The role of MHC molecules can further be investigated by producing a panel of isogenic lines.

INTRODUCTION

Cyprinid fishes provide useful models for fundamental and applied research. The zebrafish is used extensively as a model in developmental biology, making it the vertebrate '*Drosophila*'[83]. Several inbred strains and an extensive genetic linkage map are available[84,85] and have been used to characterise major histocompatibility complex (MHC) genes[86-88], *trans*-species polymorphism[89], and to map the class I and class IIA and B genes to different linkage groups[48]. However, due to its small size the zebrafish is less suitable for the study of the immune system of cyprinid fishes beyond the molecular level. To this end inbred or clonal carp are imperative experimental animals to be used in immunological studies. The common carp and other cyprinid fishes constitute the larger part of fish protein produced in aquacultural systems, making up some 80% of the 9.5 million tons of fish produced world-wide. Aquaculture systems invariably involve high stocking densities with the increased risk of disease outbreaks. To reduce this risk, better control of disease, through for example vaccination or selective breeding, is essential and requires a better understanding of the immune system of these fish[90]. MHC molecules play a pivotal role in the initiation of a specific immune response. Therefore, a thorough knowledge of the structure and genetics of the carp MHC genes may be instrumental in increasing disease resistance of cultured stocks.

Studies on the population biology and evolution of mammalian MHC genes have attracted much attention although sometimes different explanations are put forward for the same observations[91,92]. Fish species flocks are particularly interesting for evolutionary studies, such as on the cichlid species flocks of the great African Lakes[93]. Two species flocks of cyprinid fishes are known: the almost extinct cyprinid species flock of Lake Lanao on the Philippines[94] and the *Barbus* species flock of Lake Tana[95].

The Cypriniformes is one of the largest orders of extant fishes, containing some 3,000 species, and only surpassed by the Perciformes[96]. All of the species described live in freshwater. The species for which the MHC genes have been reported belong to the genera *Danio, Barbus, Carassius,* and *Cyprinus* and these are part of the family Cyprinidae[10, 70]. Karyotyping has demonstrated that the average number of chromosomes in a diploid arrangement is fifty[66]. The species from the

genera mentioned above have diploid (2n=50; *Danio rerio*[66]), tetraploid (2n=100; *Cyprinus carpio* and *Carassius auratus*[67]) and hexaploid (2n=150; *Barbus intermedius*[68]) number of chromosomes. The different ploidy status may have implications for the evolution, inheritance and usage of MHC genes within the cyprinid fishes.

It is suggested that tetraploidy in the common carp (*Cyprinus carpio*) is the result of an allotetraploidisation event, indicated by the fact that only bivalents are seen during meiosis[67]. The timing of cyprinid polyploidisation event has been set at 16 MYR ago as deduced from comparisons of duplicated genes[97]. However, other studies based on the tetraploid genome of catostomid fishes have estimated the event to have occurred 50 MYR ago[98]. Such alloploidisation events are not uncommon in cyprinid fish, since diploid cyprinids have been observed to form hybrids under natural conditions[99]. Many of the duplicated genes in tetraploid fish may persist as functional genes for millions of years, although they may accumulate non-deleterious amino acid replacements. Alternatively, some of the genes become pseudogenes, as has been observed in alloenzyme studies.

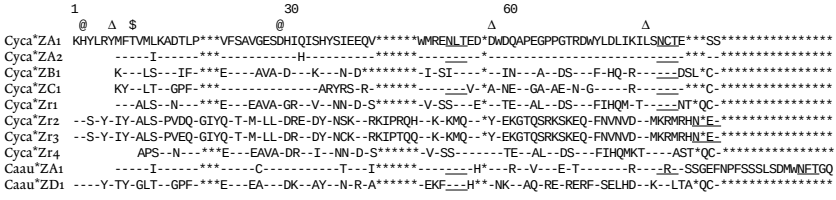
Comparable scenarios can also be postulated for the different *Barbus* species, which have a different ploidy status (6n). Calculations based on conserved Ig-like domains of different genes, such as class II and $ß_2$-microglobulin suggests that the hexaploid *B. intermedius* and carp must have diverged some 30 MYR ago[70]. This implies that the hexaploid status of this species has arisen independently of the tetraploid character of the carp, under the assumption that the estimation for the polyploidisation event of the latter species (16 MYR) is accurate.

In this review we will focus on three major aspects of cyprinid MHC research: 1) analysis and comparison of classical and non-classical class I sequences between three cyprinid species, 2) a natural model of the *Barbus* species flock for evolutionary studies, and 3) a laboratory model of the common carp for immunological studies. We have chosen not to go into a detailed comparison between the class II alpha- and beta-chain, and $ß_2$-microglobulin of the different cyprinid species, although these molecules obviously are part of MHC molecules.

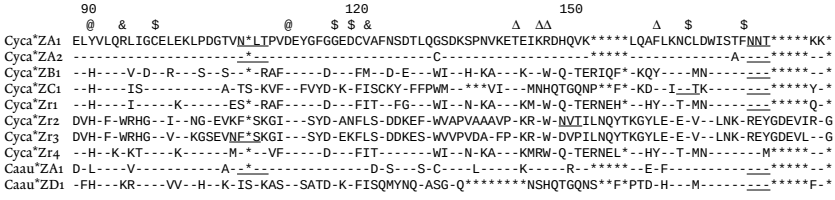
MHC CLASS I GENES IN CYPRINID FISHES

MHC genes in teleostean fish have been rather elusive for a long time[100,101]. Although indirect evidence for the existence of MHC genes has been accumulating since the pioneering work of HILDEMANN[102] and KALLMAN[103], it was not until

Alpha-1



Alpha-2



Alpha-3

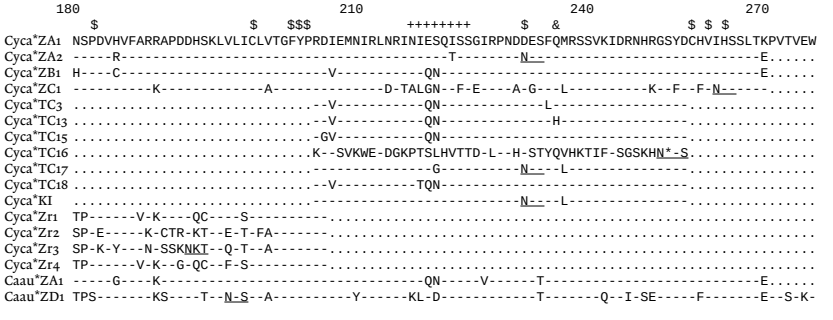


Figure 1: Multiple alignment of amino acid sequences from cyprinid non-classical class I alpha-1, alpha-2 and alpha-3 domains. Cyca-ZA1, -ZA2 -ZB, -ZC, common carp[106]; Cyca-ZR1, -ZR2, -ZR3, -ZR4, common carp (C. P. KRUISWIJK, R. J. M. STET, unpublished observations); Caau-ZA1, and -ZD1 ginbuna crucian carp[106]. *Dashes* indicate identity to Cyca-ZA1 sequence, *asterisks* gaps, and *dots* denote absence of sequence information. Residues binding peptide termini (*triangles*), conserved β_2-microglobulin contact residues (&), residues involved in salt-bridges (@), highly conserved residues ($), putative CD8 binding loop (+), and potential N-linked glycosylation sites (*underlined*) are indicated.

1990 that the first MHC genes sequences became available for carp[6] (*Cyprinus carpio*). Initially, only genomic sequences of class I (*Cyca-Z*) and class II beta-chain (*Cyca-YB*) exons encoding extra-cellular domains were described. The latter sequence, although incomplete, seems to be a regular *Cyca-DAB* sequence as shown by subsequent isolations of similar sequences in European carp[104,105]. The class I (*Cyca-Z*) sequence turned out to be the most enigmatic one in many respects.

Despite the relative short intron lengths between the exons encoding the extracellular domains no exons encoding the leader peptide or exons for the transmembrane and the cytoplasmic region were found in the genomic clone of which 6.5 kb was sequenced. This suggests that the introns separating exon-1 and -2, and exon-4 and -5 must be of considerable length. Subsequent studies describing an expansion of the number of *Cyca-Z* sequences made use of PCR on cDNA with primers designed on the N-terminal region of the alpha-1 domain and the C-terminal region of the alpha-3 domain[106]. Thus, no information is available on the transmembrane nature of the molecule encoded by the *Cyca-Z* sequences.

Southern blot analyses of carp genomic DNA, using probes, which included the alpha-1 and alpha-2 encoding exons of different cDNAs (*Cyca-ZB; Cyca-ZC*) and the alpha-3 encoding exon of *Cyca-ZA1*, revealed the presence of a large number[90-92] of *Cyca-Z* genes per individual carp[81,106]. In most cases the intensity of the hybridising bands, within an individual carp using an exon-4 probe, was of the same magnitude, suggesting low divergence in this part between the different genes. To date, the only other species in which these class I Z sequences have been identified is the ginbuna crucian carp (*Carassius auratus langsdorfii*), a close relative of the common carp[106]. The similarity between the *Caau-Z* and *Cyca-Z* sequences is highlighted by the extensive cross-hybridisation on Southern blots. In zebrafish (*Danio rerio*), a diploid cyprinid species, despite extensive searches of genomic DNA, genomic libraries and cDNA libraries no evidence has been obtained for the existence of class I Z sequences[48]. Apart from the three species mentioned no other cyprinid species, including *Barbus* species, have been probed for the presence of these genes. Neither are these genes reported for teleostean families in which MHC genes have been characterised[10].

The nature of non-classical class I genes in cyprinid fish

Most of the class I Z cDNA sequences, including the recently characterised *Cyca-Z1-4* (Acc. No. AJ007848-AJ007851), are truncated sequences, some at the 5'- and

some at the 3'-end (*fig. 1*). The sequences that are complete at the 5'-end are *Cyca-zr1* and *Cyca-zr2*, which both encode a hydrophobic leader sequence of 26 residues. These sequences support the notion that the intron between exon-1 and -2 must be of considerable size. A long intron has been reported for class I sequences in carp (*Cyca-UA*) between exon-2 and -3, and was estimated to be 14 kb long[107]. None of the Z sequences are complete at the 3'-end, lacking sequence encoding the connecting peptide, transmembrane and cytoplasmic regions, and it is therefore not possible to predict whether these sequences encode functional class I heavy chain membrane molecules. The alignment shows a large number of indels, especially in the alpha-1 and less so in the alpha-2 domain. The alpha-3 domains show a good alignment. Most notably are the insertions in the alpha-1 and alpha-2 domains of Caau-ZA1, Cyca-ZR2, and Cyca-ZR3. Assuming a high degree of homology between the structure of HLA-A2 molecule[7] and the carp class I molecule encoded by these genes, the indels are located in the alpha-helices at either ends of the peptide binding groove. It remains to be seen whether this may affect peptide binding characteristics of these carp class I molecules, as it has been shown that the ends of the peptide binding groove are occluded allowing binding only of peptides 8-12 residues long[108].

The nature of the functionality can also be predicted from the presence of a number of key residues, which have been shown to be of crucial importance for folding, glycosylation signals, interaction with β_2-microglobulin, CD8 T-cell co-receptor, and peptide binding[64]. The cysteines in the alpha-2 and -3 domains are well conserved and are at a distinct amino acid sequence length apart to be able to form disulphide bridges, stabilising the folding of this part of the molecule. There are a number of cysteines in the alpha-1 domain in almost all of the Z sequences, except Cyca-ZR3. However, it is unlikely that these are involved in forming disulphide bridges. In all class I Z sequences there are several N-linked glycosylation sites (*fig. 1*). These sites are present in all three extra-cellular domains. This means that the Cyca-Z encoded class I molecules are probably more heavily glycosylated when compared to the Cyca-UA class I heavy chain, which has only one N-linked glycosylation site[107], situated between the alpha-1 and alpha-2 domains and found in all class I molecules, except CD1 and MIC molecules[64]. The CD1 molecule also contains multiple N-linked glycosylation sites, and in this respect the Z sequences are similar in nature to the CD1 molecule.

A number of residues that are involved in binding of the termini of the peptides are more or less conserved in all of the Z sequences. Whereas the Y6 and T142 are conserved, however, others such as K145 and W146 are present in some, but not all sequences. This observation applies to most of the conserved peptide binding residues. The highly conserved contact pairs involved in the interaction with the β_2-microglobulin (Q93, D118 and Q236) are present, except in Cyca-Zr2 (N118), Cyca-Zr3 (K118) and Cyca-Zr4 (K93), suggesting the possibility of interaction with the carp β_2-microglobulin, based on the presence of the conserved contact counterparts in the latter molecule[109].

Thus, although previously it has been suggested that Cyca-Z sequences are the classical (class Ia) type of sequences[106], they seem to be more of an intermediate form between the class Ia and the non-classical class I sequences, based on differences in the conserved peptide binding residues. This is also corroborated by the presence of a phenylalanine residue at position 166 (*fig. 1*), which is thought to be implicated in the binding of the N-formylated form of peptides to non-classical class I molecules[110]. In addition, the CD8 binding region in the alpha-3 domain[111], which should have many residues bearing surface negatively-charged side chains, such as E and D, is not very well conserved in the Cyca-Z sequences in terms of acidic nature of the residues present. The only conserved amino acids are perhaps the two glutamine residues (Q218 and Q220). Apart from these, the region is highly variable between the different Z sequences. As in the chicken it might be that the CD8 co-receptor has co-evolved with the diverged class I molecule, thereby obscuring the identification of the contact residues[64]. However, the classical carp class I sequence Cyca-UA does show the preponderance of acidic and conserved residues in the position implicated in binding to the CD8 co-receptor[107]. The CD8 contact region between mammalian classical and non-classical class I molecules is well conserved, allowing the co-receptor to interact with both type of molecules. This leaves open the question of how the Cyca-Z encoded molecules interact with the T-cell. Unfortunately, no information is as yet available on the teleostean CD8 homologue. Hydrophobicity plots of Cyca-Z amino acid sequences are comparable to the plot for CD1. Thus, the Z class I lineage molecules, being analogous to CD1 in terms of hydrophobicity and glycosylation patterns, may be involved in presenting antigens of a different chemical nature compared to the peptides presented by classical class I molecules[112].

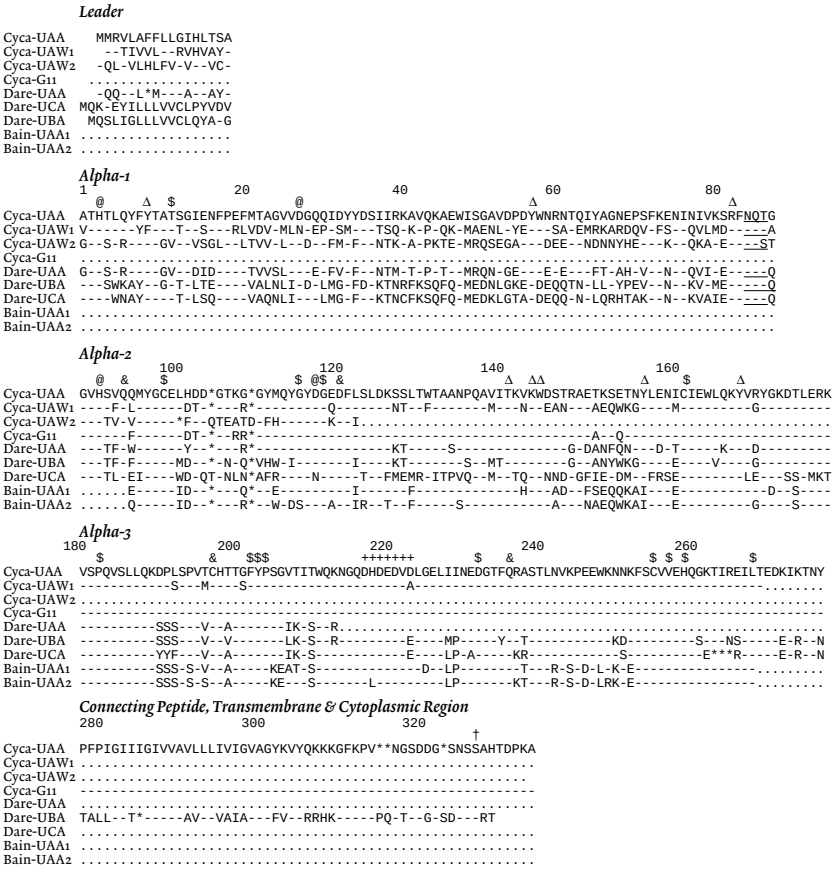


Figure 2: Multiple alignment of amino acid sequences from cyprinid classical class I alpha-1, alpha-2 and alpha-3 domains. Cyca-UAA, -UAW1, -G11 common carp[107], Cyca-UAW2, common carp (C. P. KRUISWIJK, R. J. M. STET, unpublished observations), Dare-UAA, -UBA, -UCA, zebrafish[86], Bain-UAA1, -UAA2, African 'large' barb (C. P. KRUISWIJK, R. J. M. STET, unpublished observations). *Dashes* indicate identity to Cyca-UAA sequence, *asterisks* gaps, and *dots* denote absence of sequence information. Residues binding peptide termini (*triangles*), conserved β_2-microglobulin contact residues (&), residues involved in salt-bridges (@), highly conserved residues ($), putative CD8 binding loop (+), potential serine kinase site (†), and potential N-linked glycosylation sites are *underlined*.

In the course of a study into the expression of *Cyca-Z* sequences in the thymus, cDNA amplification was performed using degenerate primers based on the conserved regions flanking the cysteines in exon-4 of *Cyca-ZA1*. Several *Z* sequences were obtained which were designated *Cyca-TC3, -TC13, -TC15, -TC17,* and *-TC18* (*fig. 1*). In addition to these, a sequence (*Cyca-TC16*) was obtained which showed the highest similarity (42-46%) to the class I sequences of the coelacanth (*Lach-UA, -UB*), and the lowest to the *Cyca-Z* and *Cyca-UA* [113] (both 20%). As a result, the *Cyca-TC16* clusters together with the *Lach* class I sequences in a neighbour-joining tree, at a considerable genetic distance from both the *Cyca-Z* and *Cyca-UA* clusters. Unfortunately, despite intensive efforts screening genomic and cDNA libraries and performing anchored PCR, the remainder of the *Cyca-TC16* gene was not found. The TC16 sequence shows a number of conserved class I features, such as the β_2-microglobulin contact residues. In addition, there is a high degree of congruity between the predicted positions of the beta-strands with the HLA-A2 structure[114]. However, the conservation of the CD8 binding residues is completely absent compared to the Cyca-Z sequences, which at least have the two conserved glutamine residues. Thus, it remains to be seen whether the *Cyca-TC16* sequence is part of a functional class I gene, or just a evolutionary remnant of an ancestral class I gene, which gave rise to the coelacanth class I lineage.

The expression of the *Cyca-Z* sequences is not well studied. The sequences reported were obtained from either kidney[106] or thymus[113], or from activated macrophages (*Cyca-Zr1-Zr4*). It is not clear whether these sequences have the same distribution of expression compared to the classical class I sequences, such as the *Onmy-C32*, which is expressed at high levels in brain, heart, intestine, kidney, thymus, liver and spleen[115]. Aberrant and low expression levels of the *Z* sequences might underpin their non-classical nature, similar to the situation found in rainbow trout non-classical gene *Onmy-UAA* [116]. Another salient gap in our understanding of these genes is the lack of alleles. The picture that is emerging is that individual carp have a set of occasionally duplicated *Cyca-Z* genes, but no alleles. The only alternative explanation is that the different loci (*ZA, ZB, ZC*) are in fact old alleles, which due to their long divergence time have come to look like loci. The answer to the nature of the sequence will have to come from segregation studies. Currently, we are studying F1 hybrids with respect to segregation of *Cyca-Z* sequences. The partial *Cyca-TC* and *Cyca-Zr1-4* sequences originate from this F1 hybrid group of carp and can be used to elucidate the

allele/locus enigma. Alternatively, the most accurate answer may come from *in situ* hybridisation experiments on chromosome spreads, known as FISH.

The nature of classical class I genes in cyprinid fish

In contrast to the alignment of the Z class I sequences, the U sequences can be aligned perfectly between the three cyprinid species (common carp, zebrafish and the African 'large' barb), with only minor indels in the alpha-2 domain, transmembrane and cytoplasmic region (*fig. 2*). The alignment shows conservation of the structural and functional features of the class I heavy chain. The ability to bind peptides critically depends on eight residues located in the A and F pockets, which interact with the amino and carboxyl termini of the peptide[64]. These residues are Y7, Y57, Y82, Y156, Y168, T139, K142, W143, which are found in all cyprinid sequences with a few exceptions. Most notably is the consistent replacement of the Y82 with arginine (R82) in all cyprinid sequences. This replacement is also seen in salmonids[115,117,118] and the guppy[119] (*Poecilla reticulata*), but not in the coelacanth[120]. The conserved peptide binding residue Y7 is not present in the Dare-UBA and Dare-UCA molecules, but a tyrosine present at position 8 in both sequences may serve a similar role in the binding pocket. Three contact residues (Q93, D120, Q236) between the class I domains and the β_2-microglobulin are highly conserved, and are similar to those in mammals[121]. In addition to these residues many other key amino acids such as the cysteines (C98, C161, C197 and C255) forming disulphide bridges, those forming salt-bridges[7] (D27/H3; H90/D117) , and the serine at position 327 involved in phosphorylation are all conserved with a few exceptions. Although most of these structural features are also present in class I Z sequences, this is not the case for the conserved peptide binding residues. A more striking difference is the presence of a single N-linked glycosylation site (N84) at the end of the alpha-1 domain, compared to the multiple glycosylation sites in the non-classical Z sequences. The acidic nature of the CD8 binding region (residues 217-223) is well preserved in all of the cyprinid species, opening the possibility for interaction with positively charged residues in a putative CD8 homologue.

Expression of these genes was studied in some detail only in the common carp[107]. No Northern blots were performed, unlike in the case of the rainbow trout[115], where expression of *UA* sequences were studied on RNA extracted

from whole organs, but alternatively a polyclonal antiserum was used raised against a recombinant Cyca-UA protein. This revealed that Cyca-UA is expressed on leukocytes from blood, spleen and thymus, but not on erythrocytes and thrombocytes. Lack of expression on the latter two cell types is surprising, as they are nucleated cells. In another well studied ectothermic vertebrate *Xenopus laevis*, nucleated erythrocytes do express class I molecules[122]. In mammals, classical class I molecules can be found on all nucleated cells, although the levels may vary[123]. Subsequent studies on the ontogeny of class I molecules using the anti-Cyca-UA serum, in conjunction with an antiserum raised against carp β_2-microglobulin[124] (Cyca-β_2m), revealed that in the thymus an alternative class I heavy chain is expressed[125]. A higher percentage of Cyca-β_2m positive cells was noted compared to the number of Cyca-UA positive cells. The most likely candidate seems to be the *Cyca-Z* encoded heavy chain, or a CD1 analogue as has been suggested for *Xenopus*[126]. The *Cyca-UA* gene is expressed constitutively starting 1 day after fertilisation. In contrast, the expression of the Cyca-β_2m genes is detected 7 days after fertilisation[125]. This suggests that expression of class I molecules, which is critically dependent on the presence of β_2-microglobulin, is delayed. A down-regulation of expression of class I molecules, due to the absence of transcription of Cyca-β_2m, was also observed in experiments where carp were transferred to lower permissible temperatures[124]. These experiments showed an important role of ambient temperature on the expression of class I molecules, due to a temperature-sensitive transcriptional regulation of the *Cyca-β_2m* gene.

Similar to the situation with the cyprinid Z sequences there is scant support for the existence of different alleles of the same locus in a single cyprinid fish. In zebrafish three loci have been postulated[86] designated *Dare-UAA, -UBA*, and *-UCA*. Two class I sequences were obtained from a cDNA library made from 20 individuals from a non-inbred KOC strain. Based on sequence comparisons these were assigned to different loci (*UAA* and *UBA*). The third locus *Dare-UCA* was obtained from a commercial genomic library made from several male and female zebrafish[127]. In a later study based on segregation studies the *UAA* and *UBA* loci were renamed as alleles of the same *UAA* locus[48]. Similar difficulties have been encountered in our studies on carp, where it seemed impossible to assign locus or allele nature of the sequences obtained, with the possible exception of the *Cyca-UAA* and the *Cyca-G11* sequences. These two sequences are completely

identical from the 3'-end of the exon-3 to exon-8, which encodes the 3'-untranslated region[107]. This identity across 1,400 bp of cDNA suggests a recent recombination event, resulting in homogenisation of this part of the sequence. In the *Barbus* species we found evidence for the expression of only two closely related sequences (*Bain-UAA1* and *Bain-UAA2* (Acc. Nos. AJ007852-53)) in a wild specimen of *B. acutirostris*. The Dare-UAA and -UBA differ at 22 positions in the alpha-2 domain, whereas the Bain-UAA1 and -UAA2 differ by only 16 residues. Thus, these *Barbus* class I sequences may be alleles at a single locus.

How many cyprinid class I loci are there?

This subject cannot be discussed without considering the ploidy status of a given cyprinid fish. The most simple answer must come from zebrafish since these are diploid (n=50). Bingulac-Popovic and co-workers[48] showed that the two class I loci (*Dare-UAA* and *-UCA*) map to the same linkage group. The diploid configuration seems therefore a single class I region, underpinned by the fact that an extensive search for additional class I loci was unsuccessful. The common carp, which is the tetraploid representative, might possess two regions containing *UA*-like loci. Although it can not be excluded with absolute certainty, this seems unlikely as extensive screening of a genomic library of a single individual and cDNA libraries prepared from clonal carp invariably produced only a single locus[107]. Divergent loci, such as *Cyca-UAW1* and *Cyca-UAW2* (Acc. No AJ007901) have been found, but they originated from different groups of carp. The second set of class I loci might be represented by the *Cyca-Z* sequences. However, assuming an orthologous relationship between different *U* lineage sequences, the divergence time between carp and zebrafish, assumed to be in the order of 50 MYR, cannot account for the differences between *U* and *Z* sequences if the similarities between the *UA* sequences from the two species are taken into account. The presence of *Cyca-Z* lineage sequences in carp and ginbuna crucian carp might be explained by the fact that their tetraploid nature arose through an alloploidisation event, as suggested by Larhammar & Risinger[97]. This hypothesis implies that the *Z* lineage should be present in other cyprinid species closely related to the carp, such as the *Barbus* species. No evidence for this has been reported, due to the fact that these species have not been extensively studied with respect to the presence of the *Z* lineage sequences.

Currently we are studying the *Barbus* species flock of Lake Tana for the presence of class I sequences. Initial studies focused on *UA* lineage sequences, as these seem to represent the classical class I sequences. Multiple *Bain-UA* sequences are found in single individuals (*table 1*), which suggests that the hexaploid nature of the *Barbus* species is more likely to be the result of autoploidisation, rather than alloploidisation. However, proof of absence or presence of *Z* lineage sequences needs further research. An alternative explanation of the absence of class I *Z* sequences in the zebrafish could be the expansion and contraction of the class I loci exemplified by the accordion model of MHC evolution proposed by KLEIN and co-workers[128]. However, it is likely that remnants of the sequences could still be present in the genome.

Table 1. MHC class I exon-3 sequences obtained from Lake Tana *Barbus* species

Species[a]	Individual	*UA* sequences[b]													
B. nedgia (lip)	55	L01	L02	L03	L04	L05									
	87				L04		L06	L07	L08	L09	L10	L11	L12	L13	
B. acutirostris (acute)	Bob	**A01**	**A02[c]**					A03	A04	A05	A06				
B. intermedius (shore-complex)	Flip		**L02**				L06			**A04**		F01	**F02**	F03	

[a] Species names are from [132]. The morphotype designation used formerly is given in parenthesis

[b] Sequences obtained per individual. Identical sequences shared between individuals of the same species are given the same sequences designation.

[c] Sequences in bold have been identified both as genomic and as cDNA sequences. No sharing is found between the *B. nedgia* and *B. acutirostris* sequences, but *B. intermedius* shares sequences with the other *Barbus* species. Sequences were obtained from database entries.

Phylogenetic analysis of class I sequences

From the structural analysis it is clear that the Cyca-Z represents non-classical and the Cyca-UA the classical class I heavy chain. The question remains whether these sequences are orthologous or paralogous within and between the group of sequences. To answer these questions we analysed the sequences by constructing separate neighbour-joining trees for exon-2 to -4, encoding the extracellular domains (*fig. 3*).

The neighbour-joining tree of exon-4 seems to be the most stable tree with almost all of the clusters supported by high bootstrap values, the most notable exception being the cluster containing the *Dare-UBA* and the *Cyca-UA* sequences. The *Z* and *Cyca-TC* sequences form a large cluster of closely related sequences with the *Cyca-TC16* as an outgroup. The *Cyca-TC3*, *-TC13*, *-TC15*, and *-TC18* belong to

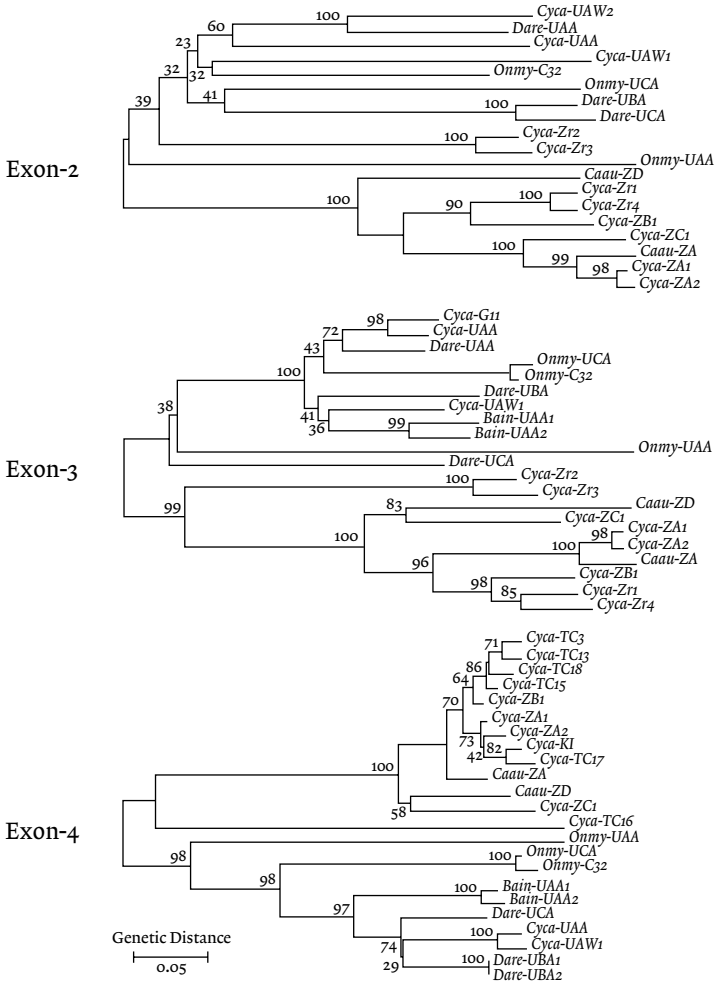


Figure 3: Neighbour-joining tree of exons-2, -3, and -4 of cyprinid MHC class I nucleotide sequences. Trees were constructed using the method of SAITOU and NEI[161], based on uncorrected p-distances in MEGA software. Numbers at branch nodes represent bootstrap confidence levels of 1000 bootstrap replications. Sequences were compiled from references[86,106,107,113,115,118] and from sequences submitted to the EMBL database.

the *Cyca-ZB* lineage, while *Cyca-TC17* and *-K1* belong to the *Cyca-ZA* lineage. We found no representatives of the *Cyca-ZC* lineage in our study on the expression of *Cyca-Z* sequences[113]. This raises the possibility that not all carp have the same set of class I loci. OKAMURA and co-workers[106] failed to isolate the carp gene equivalent of the *Caau-ZD1*, but stated that the carp *ZD* sequence must be present, on the grounds that they had identified another alpha-3 domain sequence divergent from the *Cyca-ZA1*. In both exon-4 and exon-3 trees the *Cyca-ZC1* clusters with the *Caau-ZD1*, which seems to implicate that *Cyca-ZC1* is the equivalent of the ginbuna crucian carp *Caau-ZD*. However, this clustering dissolves when analysing exon-2 (*fig. 3*).

We included the rainbow trout sequences in the phylogenetic analysis of the *UA* lineage sequences. The main reason for this inclusion is the characterisation in this species of the two types of class I sequences, namely classical sequences[115], *Onmy-C32* and *Onmy-UCA* (Acc. No. AJ007847) and a non-classical (*Onmy-UAA*) sequence[116], which can now be compared to the carp equivalents. The non-classical rainbow trout class I sequences cluster with the *UA* lineage sequences and not with the non-classical *Z* suggesting a paralogous relationship between the two. The *Onmy-UAA* sequence does, however, form an outgroup in the *UA* cluster. The other *UA* sequences cluster reasonably well according to species, with the exception of the *Dare* sequences. However, this part of the tree topology is not supported by high bootstrap values.

In the analysis of exon-3 a number of new sequences (*Cyca-Zr1-r4*) are now included, whereas no information is available for the *Cyca-TC* sequences. The classical (*UA*) and non-classical (*Z*) sequences remain as two main clusters (*fig. 3*). The *Z* lineages cluster in a similar topology as seen in the exon-4. The *Cyca-Zr1* and *-Zr4* can be considered *Cyca-ZB* lineage sequences, whereas the *-Zr2* and *-Zr3* form a completely separate group, from all of the other *Z* sequences. This implies that they represent a new locus, but most likely not the missing *Caau-ZD* sequence in the carp, as the *Caau-ZD* clusters away from the new sequences with the *Cyca-ZC* sequence. The *Z* clustering is supported by high bootstrap values, contrasting with the values found for the *UA* clusters. The position of some of the sequences has changed dramatically. The *Dare-UCA* is now found outside the main clusters, similar to the non-classical *Onmy-UAA* sequence. The *Cyca-UA* sequences now intermingle with class I exon-3 sequences of other cyprinid and even salmonid species. The close relationship of salmonid, especially exon-3

sequences, has been noted before, based on a high identity between the different salmonid and cyprinid sequences[53]. This close relationship between the two might be explained by a recombination event involving exon-3, which occurred prior to the divergence of the salmonid and cyprinid fishes. The tight clustering of the *Onmy-UCA* sequence with the *Onmy-C32* seems to support this hypothesis. The branch lengths between the different species are more or less comparable, whereas the presumed divergence times between the different genera such as *Danio*, *Cyprinus* and *Barbus* are quit different. This also suggests that ancient recombination events have an effect on the current divergence estimates of class I sequences of extant species. In addition, consistent clustering of a particular sequence with others in all exon trees is not observed. For example the *Dare-UCA* clusters in three different positions in relation to the other sequences.

Although the *Z* clusters showed a consistent topology in the exon-4 and -3 trees, this is breaking in the exon-2 tree (*fig. 3*). The position of the *Cyca-ZC1* and *Caau-ZD* have changed and are not found clustered together. More importantly, the *Cyca-Zr2* and *-Zr3* are now found to cluster with *UA* sequences. The remainder of the *Z* sequences showed consistent positions compared to the exon-3 tree. The branch lengths of the *UA* cluster in the exon-2 tree are much longer and the bootstrap values are much lower compared to the other trees. The *Onmy-UAA* exon-2 sequence shows a consistent position, whereas the *Dare-UCA* is now found clustered together with *Dare-UBA*.

In conclusion, this analysis suggests that the different exons, especially *UA* class I sequences, have different phylogenetic histories, shown by the different clusterings both within and between species. This could probably be related to interlocus recombinations predating species divergence. An analogous mechanism in mammals of inter- and intralocus recombination is hypothesised to be the driving force for the generation of new class I allotypes[129]. The mammalian classical and non-classical class I genes seem to have diverged within the different orders (e.g. rodents and primates), resulting in a species-specific clustering of the class I sequences, except in rare occasions when the recombination level is uniquely high[130]. This is in contrast to the finding of homologues of the different class II loci in different mammalian orders which intermingle[30]. In cyprinid and salmonid fish this situation is now also found for class I sequences. The driving force which may be responsible for the conserva-

tion of the close relationship between some of the class I exons is selection. The exon-2 and -3 contain the peptide binding pockets, and these are selectively favoured for diversity.

LAKE TANA BARBUS SPECIES FLOCK

Lake Tana is the source of the Blue Nile. It was formed in the late pleistocene as a result of a volcanic eruption during which the flow of lava blocked the river to form a waterfall about 40 m high. At the time of the formation of the lake, the Blue Nile contained riverine species among which was the ancestral *Barbus* stock. This stock was probably similar to the present riverine *Barbus intermedius*. Geological evidence suggests that the species flock found nowadays evolved sympatrically, because the lake is shallow and regularly saucer-shaped, without separate basins. There is no information as to whether a situation occurred similar to that in the large African Lakes of the rift valley, which are thought to have been completely desiccated some 12,000 years ago[131]. These data suggest an extremely fast evolution of the resident cichlid species flock.

The Lake Tana 'large' barbs especially those forming the 'shore complex' population, were highly variable in their morphology as were their riverine counterparts; thus, they had the ability to radiate out into the newly formed lacustrine habitats. These incipient morphotypes could occupy new niches, especially those in deep, open water. Genetically determined differences between the morphotypes could become fixed and, following assortative mating, lead to speciation. Initially, the different barbs were recognised as different morphotypes[95] but, recently, a revision has taken place resulting in the description of 14 closely related species[132]. This virgin species flock presents a simple model for the study of the natural selection of MHC genes, as it has been postulated that selection is most evident during adaptive radiation[39]. Similar studies were also undertaken involving the cichlid species flock of Lake Malawi[133]. In addition to being an interesting model for evolutionary studies, the African 'large' barb species are hexaploid[68]. The ploidy status of the *Barbus* species complements the diploid zebrafish and the tetraploid common carp, and allows comparisons of different aspects of MHC characteristics. Polyploidisation events must have had major impacts on the evolution of the MHC, as it is suggested that expression of too many alleles in a single individual might be deleterious.

MHC genes in the lake tana barbus species

A total of 45 genomic class IIB sequences were obtained from 4 *Barbus* species (*B. acutirostris* (acute), *B. tsanensis* (intermedius), *B. truttiformis* (trout-like) and *B. nedgia* (lip))[70]. The genomic sequences analysed include intron-1 and exon-2, and were shown to be *bona fide* class IIB sequences, with one exception, which contained an in frame stop codon. Each of the sixteen individuals studied contained a variable number of class IIB sequences, ranging from two to seven. The *Barbus* class II genes are orthologous to *Cyca-DAB* and *Dare-DAB*. The latter has been implicated as being the only class IIB gene that is expressed and is represented by four allelic lineages[89,127].

Table 2. MHC class II B exon-2 sequences obtained from Lake Tana *Barbus* species

Species[a]	Individual	*UA* sequences[b]
B. acutirostris	2	A02 A05
(acute)	4	A01 A02 A03 A05
	9	A02 A04 A06
	Bob[c]	A01 A04 A05 A06 A07 **A11 A12**
B. tsanensis	67	I01 I03 I09 I10 I11
(intermedius)	76	I01 I02
	78	I03 I04 I05 I06 I07 I08
	58	I12 I13 I19
	438	I13 I14
	481	I01 I11 I12 I15 I17 I18
	60	I03 I15 A16
B. nedgia	55	L03 L04 L05 L06 L07
(lip)	87	L01 L02
B. truttiformis	457	T01 T02 T04
(troutlike)	458	T01 T02 T03 T04
	471	T05 T06 T07 T08

[a] Species names are from [132]. The morphotype designation used formerly is given in parenthesis
[b] Sequences obtained per individual. Identical sequences shared between individuals of the same morphotype are given the same sequences designation. No sharing of sequences is observed between the different species.
[c] cDNA from this individual was prepared and analysed for the presence of *DAB* sequences. These are given in bold. (Modified from [70])

The analysis of the *Barbus DAB* sequences revealed a number of interesting features. The peptide binding positions are much more variable compared, for example, to the variability observed in Atlantic salmon (*Salmo salar*) class IIB sequences[134]. At similar positions *Bain-DAB* have up to eight different residues compared to only three different amino acids in the position in *Sasa-DAB*. More

importantly, the 45 *Bain-DAB* alleles obtained showed a species-specific distribution i.e. no sharing of alleles was found between the four closely related species studied (*table 2*).

This observation could be regarded as a validation of the species designation[132], which has replaced the notion that the different morphotypes belong to the same species. The neighbour-joining tree of the exon-2 revealed a clear intermingling of sequences from different species within clusters, supported by high bootstrap values. In contrast to the situation in the *Barbus* species, no sharing of allelic lineages have been observed between carp and the *Barbus* species. This may be due to the fact that the number of class II beta sequences for carp, especially alleles of the same locus, are rather scarce, and therefore have not yet been identified. Alternatively, the long divergence time between these species may have obscured the identification of alleles from the same locus, which may now appear as different loci[135]. Also little evidence has been obtained for similar clustering of class II sequences from different salmonid species[136]. The main reason for this observation is suggested to be the recent bottleneck undergone by these pacific salmonids during the last glacial period. This observation may also impinge on the differences in variability seen at the peptide binding residues between the African 'large' barb species and Atlantic salmon. Collectively, these observations support the proposition that the *trans*-species mode of evolution of class II genes is not limited to mammals, but now also applies to teleostean fishes[89].

The *Barbus* species flock, resulting from a rapid adaptive radiation event, also provides a unique model for the study of the effects of selection on the MHC genes. The magnitude and nature of the selection can be measured by comparing the ratio of synonymous to nonsynonymous substitution per site. The ratio, expressed as a gamma value, is 1 under neutral conditions. However, in the case of positive Darwinian selection the gamma value should exceed 3, whereas under purifying selection the value is below 1. The analysis of the *Bain-DAB* sequences revealed that invariably the gamma values of the peptide binding residues are well above 3, in contrast to the non-peptide-binding residues, which are below 1. These data suggests that similar to the situation in mammals the polymorphism observed is driven by overdominant selection[36] and that the high gamma values for the peptide-binding residues are consistent with values obtained from orthologous genes of different mammalian species[137].

In each of the individuals studied a variable number of *DAB1* sequences (n=2-7) were found (*table 2*). In one individual (BOB; *B. acutirostris*) an analysis was performed with the aim to establishing if all of the class IIB genes present are actively used. To this end cDNA was prepared from a number of tissues/cells, and it was apparent that not all class IIB genes are expressed. Only two of the seven *DAB1* sequences were found to be expressed, which raises the question whether the other genes are silenced by inactivation of parts of the complete gene or whether there is transcriptional regulation. In polyploid *Xenopus* species the number of class I and class II genes does not increase with higher ploidy genomes[79,80]. This led to the suggestion that MHC genes in polyploids are silenced by deletion, probably leaving remnants of the inactivated genes. Whether a similar mechanism operates in the hexaploid *Barbus* species remains to be investigated. However, the presence of variable numbers of *DAB1* sequences seems to lend support to a similar mechanism of inactivation by physical deletion of the genes. In zebrafish six class IIB genes have been reported, only two of these genes (*Dare-DAB, -DEB*) are complete, whereas the other four are truncated and probably pseudogenes[127]. However, it is unclear how many class IIB genes are present in a single zebrafish individual, because the genomic libraries used were prepared from multiple individuals. Carp individuals have shown to carry different number of class II loci, reflecting their tetraploid status but also suggesting haplotypic polymorphism for the class II region[105]. Currently the studies on the African 'large' barb have now been extended to include additional species from the species flock and more importantly the analysis of class I genes.

Similarly to the MHC class IIB study, primers based on *Cyca-UA* were used to amplify class I exon-3 sequences from different *Barbus* species (*table 1*). The sequences obtained (Acc. Nos. AJ007879-7900), showed a high degree of similarity to the equivalent *Cyca-UA* sequences (80.5%). No sequences were shared between the *B. nedgia* (lip) and *B. acutirostris* (acute), although sequences were shared between the two *B. nedgia* individuals. This observation is in compliance with the class IIB study. However, the individual FLIP, identified as a shore-complex specimen, shared sequences with both *B. acutirostris* and *B. nedgia* individuals included in this study. This observation is most likely not a peculiar characteristic of class I sequences, but can be more likely attributed to the fact that the individual (FLIP) belongs to the shore complex. This is a group of African

'large' barbs that inhabit the shore area of Lake Tana[138], and it is suggested that these fish belong to the highly variable *B. intermedius*. This species is probably reminiscent of the riverine African 'large' barbs, considered to be the ancestral African 'large' barb restricted to lake by the waterfall. This is substantiated by the fact that the shore habitat resembles the riverine circumstances. The shore complex is therefore thought to consist of the descendants of the ancestral recruitment population from which individuals radiated out, especially into the open water, leading to the distinct species observed nowadays.

INBRED STRAINS OF COMMON CARP

Inbred strains of carp are indispensable experimental animals in immunological studies. The common carp is easy to breed under laboratory conditions. Injections with carp pituitary suspension stimulate ovulation or spermiation, allowing the exact timing of in vitro fertilisation. The fecundity of the mature female is enormous ($> 1 \times 10^6$ eggs/female). In theory, one single breeding would suffice for one full year of large-scale fundamental and applied research on carp in the whole of the european union. Although potentially a challenging idea, the logistics would be unfeasible and the solution should be sought in production and distribution of inbred fish strains.

Inbred strains of laboratory rodents have typically been established via successive mating of related animals, thereby reducing the genetic variation within the mating population. Successive crosses within single litters slowly but inevitably results in a population with an increasing probability of having two identical genes at each locus by descent. By definition, at least 20 successive generations are needed to define a strain as inbred, as decided in 1952 by the Committee on Standardised Genetic Nomenclature. Clearly, short generation intervals of a few months are a necessity, and the fish species to which this classical inbreeding strategy have been applied are without exception small aquarium species showing early maturation (*Xiphophorus maculatus*[103]; *Danio rerio*[139]; *Oryzias latipes*[140]). These strains offer great benefits for studies on fish genetics because the inheritance of genes in consecutive generations can be studied with some measure of expedience.

Although, by default, fish are themselves the target species for fish-based studies, there are cases where they can serve as a model for other species: for

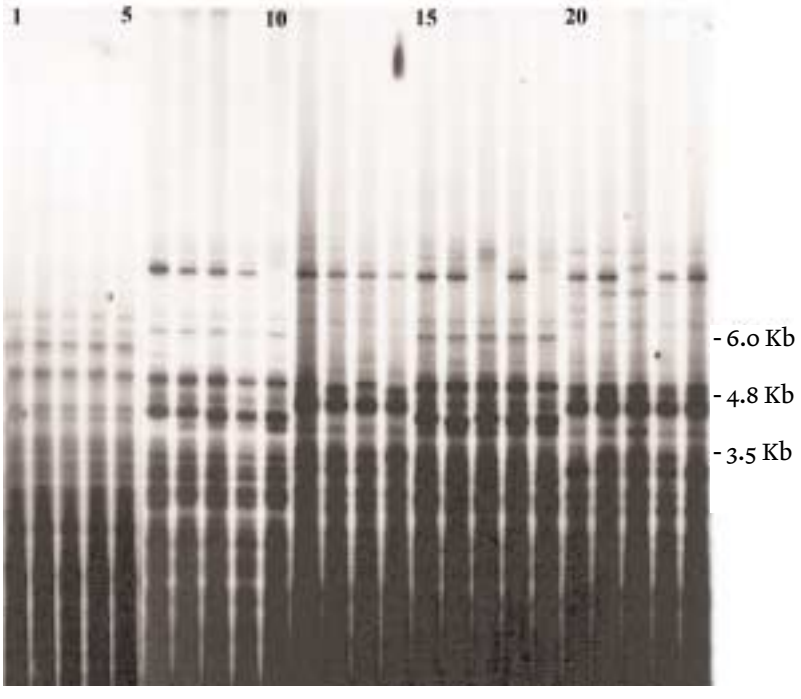


Figure 4: DNA fingerprints of gynogenetic inbred strains of common carp. Genomic DNA was digested with *Hinf*1 restriction enzyme, separated on 0.8% agarose gel and transferred to a nylon filter. The filter was probed with a (GGAT)4 microsatellite sequence. Lanes 1-5: 69E/E13 (mitotic gynogens); lanes 6-10: 69E/E26 (meiotic gynogens); lanes 11-14: 85E/E76 (meiotic gynogens); lanes 15-19: 69E/E26 (spontaneous diploids); lanes 20-24: 85E/E76 (spontaneous diploids).

example, the Amazon molly (*Poecilia formosa*) has been used as a model for human cancer research[141]. Amazon mollies carrying a microchromosome containing a locus for macromelanophore pigmentation can spontaneously develop papillomatous pigment cell tumours. The unisexual Amazon molly, which is of interspecific hybrid origin, reproduces clonally by natural gynogenesis. Sperm cells of a related species trigger ameiotic diploid eggs to start embryonic development without karyogamy. In most cases the offspring is genetically identical to the mother, as shown with DNA fingerprinting[142]. In rare cases paternal DNA, in the form of microchromosomes, can be included into the maternal genome[143].

Likewise, the Japanese ginbuna crucian carp (*Carassius auratus langsdorfii*) reproduces clonally by gynogenesis; the triploid females reproduce without a reduction in ploidy number. Several clones which are allogeneic to one another have been established. Its larger size allowed for in vivo experiments to be carried out on the transfer of immunity by intravascular introduction of splenic lymphoid cells into naive recipients. Transfer was only successful in histocompatible donor-recipient systems, as established by scale transplantations[144]. In these earlier studies molecular typing of MHC genes was not yet available.

Induced inbreeding

Manipulations of fish reproduction are mainly driven by commercial rather than by scientific needs. For example, triploidy control and induced gynogenesis or androgenesis are manipulations expected to lead to sterility, single-sex populations or faster growth[145]. Artificial meiotic gynogenesis is based on an induced retention of the second polar body by temperature or pressure shock treatment after activation of the oocyte with sperm which has been inactivated genetically, preferably by treatment with U.V. light[146]. A dominant external marker (scalation or colour) is used as an early check for paternal inheritance. This reproduction technique has been developed for a large number of commercial fish species[145]. Although it leads to a homogeneous all-female progeny nearly identical to the parental fish, crossovers during meiosis I determine the degree of heterozygosity, which may be different for each individual within the offspring. Therefore, induced meiotic gynogenesis does not result in completely homozygous offspring. The proportion of heterozygous animals in

the offspring can be high[145] and is directly related to cross-over frequencies and therefore to the distance between a given locus and the centromere. Application of this method to carp gynogenetic offspring for the linked *Cyca-DAB3* and *Cyca-DAB4* loci, in two independent experiments, estimated 8-25% crossing-over.

Occasionally, in experiments on induced gynogenesis, we observed high numbers of diploid female offspring without our having induced diploidisation (G. F. WIEGERTJES, R. J. M. STET, unpublished results). Spontaneous diploidisation is a heritable trait in carp[147]. These spontaneous diploids most probably arise from the same cytological mechanism that underlies induced gynogenesis, i.e. retention of the second meiotic division, since induced and spontaneous diploids derived from a single mother show the same DNA restriction pattern (*fig. 4*). These spontaneous diploids could be artificial effects of the forced ovulation, or could well be remnants of an evolutionary strategy similar to the one used in unisexual species.

Table 3. Antibody response of inbred carp (*Cyprinus carpio* L.) strains selected for high and low antibody respons to DNP-KLH and to various antigens measured with ELISA.

	Antibody response (ELISA)			
	DNP-KLH	DNP-HSA	TNP-LPS	Ovalbumin
Strain				
69e/E13	High			High
69e/E45	High	Low	Low	Low
85e/E29	Low			Low
85e/E61	Low			Low
85e/E76	Low	High	Low	

Induced mitotic gynogenesis does result in a population of homozygous individuals. After activation of the oocyte with irradiated sperm, completion of meiosis is allowed. Shock treatment at the moment of first mitotic division induces fusing of the two daughter cells into one diploid homozygous zygote. Subsequently, each homozygous individual that becomes sexually mature can be reproduced again by gynogenesis, preferably by the meiotic method because of its higher yield, to establish a clonal strain. Although this technique was already successfully applied to zebrafish[139] as early as in 1981, apparent difficulties, such as severe inbreeding depression acting on sexual maturation,

resulted in only few other fish species being artificially cloned by this technique: medaka (*Oryzias latipes*[148]), ayu (*Plecoglossus altivelis*[149]), the common carp (*Cyprinus carpio*[150]), and most recently, hirame (*Paralichthys olivaceus*[151]). Once cloned, part of the offspring can be sex-reversed with hormonal treatment to facilitate further reproduction[152].

As indicated above, production of homozygous fish strains is associated with a number of technical difficulties. Inactivation of sperm DNA by U.V. light to minimise the risk of paternal inheritance requires optimisation in order to retain fertilisation capacity. In addition, the window for disruption of the spindle formation during first mitosis is narrow and depends on the species, age and temperature[146]. Yields are typically low (5-10% hatching), largely because the fertilised egg is physically damaged by the chromosome manipulation treatment[153]. Although we assume random fertilisation and sampling, the low yields might skew the data set obtained. Also, the introduction of single copies of multiple genes into the male (androgenesis) or female (gynogenesis) genome is not essentially different from introducing multiple copies of a single gene (transgenesis). The abundance of repeats generally observed in the genome of fish[154] may act as potential recombination hot spots, facilitating the integration of introduced genes. Indeed, we have observed the occasional integration of paternal MHC loci in female gynogenetic offspring as well as integration of maternal MHC genes in androgenetic offspring. Therefore, application of genetic markers such as DNA fingerprints (*fig. 4*) is necessary to examine homozygous offspring for unexpected fragments that may indicate gene integration or spontaneous diploidisation. On the basis of DNA fingerprints YOUNG and co-workers[155] concluded that their only reproductive gynogenetic female was a in fact a spontaneous diploid by origin, but that their androgenetic rainbow trout were true homozygous clones. Androgenesis, the duplication of the male genome by suppression of the first cleavage, results in female XX and male YY homozygous offspring. In carp and rainbow trout, irradiation of eggs with U.V. light or gamma irradiation, respectively, inactivates the female genome after which temperature shock treatment leads to viable homozygous androgenetic diploids[155,156].

After the successful establishment of the techniques needed to produce mitotic gynogenetic carp[150], and the confirmation of their clonal nature by reciprocal exchange of skin allografts[157], we initiated a two-way selection using artificial

gynogenesis to selectively breed for level of antibody production. As a model, we chose a defined antigenic determinant: dinitrophenyl coupled to a carrier protein, keyhole limpet haemocyanin (DNP-KLH) to measure the primary antibody response with a reliable read-out system (enzyme-linked immunosorbent assay (ELISA))[72]. The divergent selection procedure resulted in a number of isogenic carp strains with an intrinsic difference in their primary antibody response[73] and their secondary antibody response[158] to DNP-KLH. In addition, these high and low responder strains can sometimes respond adversatively to immunisation with antigens other than DNP-KLH (*table 3*). These differences could be linked to Ir gene control.

Mapping of the genes controlling quantitative antibody production in divergently selected Biozzi mice not only detected three previously unidentified immunomodulatory regions, but confirmed the previously inferred contribution of MHC and IgH genes to the two distinct phenotypes[159,160]. We have started an extensive characterisation of the MHC genes present in our inbred carp strains. Although the high- and low-antibody responder carp strains indeed have different MHC class II beta[73] and class I genes (G. F. WIEGERTJES, C. P. KRUISWIJK, J. P. J. SAEIJ, R. J. M. STET, unpublished data), segregation studies in clonal strains derived from hybrids between high and low responders will have to be performed in order to prove the co-segregation of antibody responsiveness and the MHC genes present in carp. In addition, correlations between presence of expressed MHC genes and disease resistance will be tested on a large scale using a large number of different families or strains and standardised challenge protocols. The presence of unlinked class I and class II MHC genes in teleostean fishes might give rise to some interesting results.

ACKNOWLEDGEMENTS

J. P. J. SAEIJ is supported by a grant from SLW (46.033). We thank DR F. A. SIBBING for providing the material and background information on the Lake Tana African 'large' barb species flock.

CHAPTER 3

ANALYSIS OF GENOMIC AND EXPRESSED MAJOR HISTOCOMPATIBILITY CLASS IA AND CLASS II GENES IN A HEXAPLOID LAKE TANA AFRICAN 'LARGE' BARB INDIVIDUAL (*BARBUS INTERMEDIUS*)

Corine P. Kruiswijk*, Trudi Hermsen*, Kazuhiro Fujiki**, Brian Dixon**, Huub F. J. Savelkoul*, René J. M. Stet*

*Cell Biology and Immunology Group,
Department of Animal Sciences, Wageningen University,
PO box 338, 6700 AH Wageningen, The Netherlands

**Department of Biology, University of Waterloo,
200 University Ave West, Ontario, N2L 3G1, Canada

ABSTRACT

Mammalian major histocompatibility complex (MHC) genes are co-dominantly expressed. However, expression of too many alleles is thought to be detrimental to the proper functioning of the immune system. Polyploidy of the genome, as seen in several species, will increase the number of expressed MHC genes unless these genes are prone to a silencing mechanism. Physical silencing was observed in polyploid *Xenopus* species in which the number of class I and class II genes does not increase with higher ploidy genomes. Another mechanism to reduce the number of expressed MHC molecules is functional silencing. In the zebrafish genome only two genomically *bona fide* loci, *DAA/DAB* and *DEA/DEB*, are found while several other class IIB loci are present. Earlier studies indicated that such silencing mechanisms might operate on MHC genes of the hexaploid African 'large' barb. The hexaploid nature of African 'large' barb species implies an increased number of expressed MHC genes unless some MHC genes have been silenced. We studied the number of MHC genes present in the genome of an African 'large' barb individual (*Barbus intermedius*) in relation to

those expressed. We adopted the following strategy. First, full-length cDNA sequences were generated from mRNA. Secondly, partial genomic class Ia and class II sequences were obtained and compared to those from cDNA using the same primer set. In addition, we performed Southern hybridisations to obtain a verification of the number of class I and class IIB genes sequences as identified by PCR analyses and DNA sequencing. Our study revealed the presence of five MHC class Ia, four class IIA, and four class IIB genes at the genomic level, which were shown to be expressed in this hexaploid African 'large' barb individual. These data indicate that the ploidy status does not correlate with the presence and expression of MHC genes. This suggests that gene silencing is most likely a result of physical exclusion. However, functional silencing also played a role in reducing the number of expressed class II genes, as class IIB pseudogenes were identified in the genome of this African 'large' barb individual.

INTRODUCTION

The species of the Lake Tana African 'large' barb flock belong to the genus *Barbus*, which is part of the teleost family Cyprinidae[162]. Lake Tana, the source of the Blue Nile, was formed in the late Pleistocene as a result of a volcanic eruption blocking the outlet to the river with a 40 m high waterfall. This tectonic event effectively isolated the lake from the Blue Nile and its tributaries. Adaptive radiation of the ancestral *Barbus intermedius* population has led to an African 'large' barb species flock comprising fifteen different species[75]. In addition to these novel species, a shore complex population, occupying the shores of the lake is present. This species resembles the riverine *Barbus intermedius*. The fifteen novel African 'large' barb species occupy other ecological niches in the lake, have different spawning behaviour, differ in maximal size, and were shown not to share MHC class IIB alleles[70].

Since natural selection is strong during speciation events, this African 'large' barb species complex provides an excellent model for studying natural selection of MHC class I and class II genes[39]. MHC class I and class II genes encode structurally similar proteins that present peptides to T lymphocytes. The MHC class I genes are considered to be highly polymorphic and co-dominantly expressed on all somatic cells. Class I molecules are composed of a large alpha chain, encoded in the MHC, non-covalently associated with a much smaller β_2-

microglobulin (β_2m) molecule, encoded outside the MHC. The class I alpha chain consists of three extra-cellular domains with two membrane-distal domains that form the peptide binding region.

MHC class II genes are polymorphic and co-dominantly expressed, but expression is restricted to epithelial cells, B-cells, activated T-cells and antigen presenting cells of the immune system. The class II molecules are heterodimers of non-covalently associated alpha and beta chains that both consist of two extra-cellular domains, a transmembrane and a cytoplasmic region. The membrane distal domains of both chains combine to form a peptide binding region.

Polymorphic residues within the peptide binding region of class I and class II molecules interact with self and non-self peptides and are under positive Darwinian selection[35]. MHC polymorphism evolves in a *trans*-species fashion[39] and, in general, MHC class I genes seem to be more divergent and evolve more rapidly than class II genes. Allelic HLA class I lineages were shown to be maintained for up to 6 million years while certain HLA class II lineages were shown to be maintained for up to 35 million years[43-45].

In addition to being an excellent model for studying natural selection, the Lake Tana African 'large' barb species are hexaploid[68] (2n=150) in contrast to the distantly related species common carp (2n=100), which is tetraploid[67], and the diploid zebrafish[66] (2n=50). The polyploidisation events most likely influenced the evolution of the MHC genes since mammalian MHC class I and class II genes are known to be co-dominantly expressed and expression of too many MHC alleles in a single individual is thought to be deleterious.

In polyploid *Xenopus* species the number of class I and class II genes does not increase with higher ploidy genomes due to physical silencing[79,80]. Another mechanism to reduce the number of expressed MHC molecules is functional silencing. For example, in the zebrafish genome only two genomically *bona fide* loci, *DAA/DAB* and *DEA/DEB*, are found while several class IIB loci are present. The *DEA/DEB* genes are only present in some strains and expression of the *DEA/DEB* locus has never been demonstrated, suggesting functional silencing perhaps due to promoter disruption. The remaining loci were truncated or corrupted in their coding regions and thus functionally silenced pseudogenes[127]. A large number of expressed sequences MHC sequences are expected in the hexaploid Lake Tana African 'large' barb species since these genes are believed to be co-dominantly expressed. However, the MHC sequence in the hexaploid

African 'large' barb could be prone to silencing mechanisms. Previous studies indicated that such silencing mechanisms might operate on MHC genes[70,74]. To further investigate the presence of these mechanisms, we studied MHC class I and class II genes by sequence analyses of cDNA and genomic DNA from a single Lake Tana African 'large' barb shore-complex individual (*Barbus intermedius*) and performed Southern blot analyses.

MATERIALS AND METHODS

Fish

A liver sample from a Lake Tana *Barbus intermedius* individual was collected to extract genomic DNA and total RNA for identification of MHC class I light (β_2m) and heavy chains (IA) and class IIA and IIB sequences.

Genomic DNA and total RNA extraction

Genomic DNA was isolated using a Wizard® genomic DNA purification kit (Promega, Madison, WI, USA) according to the protocol provided. Total RNA extraction from tissue samples was performed according to the protocol described by DIXON *et al.*[70]. DNA and RNA concentrations were determined using the GeneQuant system (Amersham Pharmacia Biotech, Little Chalfont, UK).

PCR and Expand™ long template PCR conditions

Standard PCR reaction conditions were 1x reaction buffer, 1.5 mM $MgCl_2$, 1 unit of Taq polymerase (Goldstar; Eurogentec, Seraing, Belgium), 0.2 mM dNTPs, 0.2µm of each primer and 100 ng genomic DNA or cDNA. The cycling profile was 1 cycle at 94°C for 5 min followed by 30 cycles consisting of denaturing at 94°C for 30 sec, annealing at 50°C or 55°C for 30 sec, polymerisation at 72°C for 1 min, and a final cycle of 10 min at 72°C. Expand™ long template PCR was performed according to the protocol described for amplification of cDNA (Boehringer Mannheim, Ingelheim, Germany). The standard and Expand™ long template PCR were performed using a GeneAmp PCR system 9700 (PE Applied Biosystems, Foster City, CA, USA).

Amplification of expressed MHC class I and class II genes

Full-length African 'large' barb liver cDNA was synthesised using the Generacer™ kit for full-length RNA ligase-mediated rapid amplification of 5'- and 3'-ends

(Invitrogen, Carlsbad, CA, USA) according to the protocol described. Expressed African 'large' barb MHC class I alpha, β_2-microglobulin, class IIA and IIB sequences were amplified by PCR using conserved anti-sense primers, primer D, E, H, and J, respectively (*table 1*), in combination with the Generacer™ 5' primer. The anti-sense primers matched the nucleotide sequence of cytoplasmic or 3'-untranslated regions of known cyprinid MHC genes. Partial class I sequences were amplified using primers matching the conserved nucleotide sequence encoding part of the leader peptide or the start of alpha-2 domain in combination with a primer matching the end of exon-3 (alpha-2 domain) of known cyprinid class I sequences (*table 1*: A2, B, C).

Amplification of genomic MHC class I alpha chains and class IIA and IIB chains

Genomic class I sequences were amplified using a sense primer matching the nucleotide sequence of the leader peptide of *Bain-UA*0101* (*table 1*: A1) in combination with the conserved anti-sense primer matching the nucleotide sequence of the cytoplasmic region of known cyprinid class I alpha chains (*table 1*: D). Alpha-2 domains of class I alpha chains were amplified using conserved primers matching the nucleotide sequence of start and the end of known cyprinid class I exon-3 sequences. (*table 1*: B, C)

Genomic class IIA sequences were amplified using two sense primers (*table 1*: F, G) matching the nucleotide sequence of the leader peptides of *Bain-DAA*0101/-0201* or *Bain-DAA*0301/-0401* in combination with the conserved anti-sense primer (*table 1*: H). The anti-sense primer matched the nucleotide sequence of the cytoplasmic region of known cyprinid class IIA chains.

Genomic class IIB sequences were amplified using a sense primer matching the nucleotide sequence of the leader peptides of *Bain-DAB*0101-0401* in combination with the conserved anti-sense primer matching the nucleotide sequence of the cytoplasmic region of known cyprinid class IIB chains (*table 1*: I, J).

Cloning and DNA sequencing

PCR products were ligated and cloned using the PGEM® T-easy kit (Promega, Madison, WI, USA) following the manufacturer's description. Plasmid DNA was isolated from bacterial cultures using the QIAprep spin miniprep kit (QIAGEN, Valencia, CA, USA) according to the protocol provided. Subsequently, plasmid DNA was sequenced using the ABI prism bigdye™ terminator cycle sequencing

Table 1: Primers for PCR amplification of genomic and cDNA

Genes	*No*	*Nucleotide sequence (5' - 3')*	*Location*
Class I alpha chain	A1	ATGATGCGGCATGTAGTGCTTTTGCTCCTT	Leader
	A2	ATGTGGTCAGCAGCCCTTCTGATATTTG	Leader
	B	GGTGTTCACTCAGCTCAG	Alpha-2
	C	CTTTTCTCTCCAGAGAGTCCTT	Alpha-2
	D	GGAGTTGTTTGAACCATCATCAG	CT[1]
ß2-microglobulin	E	GTTTGTAAAATTGGAACAGCAG	3' ut
Class IIA chain	F	ATGGAGCTGTATGGTGTCCTGCTT	Leader
	G	TGTTTTCCTGTGAGCGAAGACTGCTTTCTG	Leader
	H	TCAGTTGCAGTTGTTTCCTTTAATGAG	CT
Class IIB chain	I	CTGATGCTGTCTGCTTTCACTGGAGCAG	Leader
	J	TTCTTCTTGTAGTAAATGAGTCCAGCA	CT
Class I probe (265 bp):			
*Baic-UA*0102*	K	TCTCTCCTCAGGTGTCTCTGCTGC	Alpha-3
	L	AGTATCTCTCTGATGGTTTTG	Alpha-3
Class II exon 3 probe (275 bp):			
*Bain-DAB*0101*	M	GCACCAAAGGTTAAGCTCAGTTC	Beta-2
	N	TCTACAATGATGGGTTTAGTTAAG	Beta-2
Class II exon 3 probe (275 bp):			
*Bain-DAB*0401*	O	TACAACCGAAGATTACACTCAAGTC	Beta-2
	P	TGATGATGGGTTTATTGAAGCTGGC	Beta-2

[1]Cytoplasmic tail

ready reaction kit (Perkin-Elmer, Branchbury, NJ, USA) and analysed using an ABI 377 sequencer (Applied Biosystems, Foster City, CA, USA).

Southern blotting

Southern blot analyses was performed according to the protocol described by FUJIKI and co-workers[163]. DNA probes were prepared by PCR amplification. Primers used matched the start and end of African 'large' barb class I *Baic-UA*0102* exon-4, *Bain-DAB*0101* exon-3, and *Bain-DAB*0104* exon-3 (*table 1*, K to P).

Accession numbers and nomenclature

The new sequences reported here were deposited in the EMBL database under the following accession numbers: AJ506998 to AJ506011. The nomenclature used to assign full-length sequences was similar to the suggested nomenclature used to assign class I *UA* exon-3 and class II *DAB*01* exon-2 sequences described by KRUISWIJK and co-workers[164] and adheres, in part, to the recommendations described in the HLA facts book[165]. Full-length sequences possessing exon sequences similar to those previously described by KRUISWIJK and co-workers [164] were given similar gene designations.

Nucleotide sequence, amino acid sequence and phylogenetic analyses

Genomic and expressed MHC nucleotide sequences were represented by at least two identical clones amplified in two separate PCRs. Sequence data obtained using a ABI sequencer were analysed with SEQUENCER 4.1 software (Gene Codes, Ann Arbor, Michigan, USA). Multiple alignments were performed using the program[166] CLUSTAL-W version 1.8 . Signal peptides were predicted[167] using PSORT II (http://psort.nibb.ac.jp). Phylogenetic trees were constructed using MEGA 2.1 software[168] with p-distances for amino acid sequences using the neighbour-joining algorithm[161].

RESULTS AND DISCUSSION

To investigate the number of MHC genes present in the genome of this African 'large' barb individual in relation to those expressed, we adopted the following strategy. First, 'full-length' CDNA sequences were generated from MRNA using a single gene specific primer in combination with a 5' RACE primer. It should be noted that the 'full-length' CDNA sequences lack part of the cytoplasmic region (see *fig. 1, fig. 5, fig. 6*). Secondly, partial genomic class Ia and class II sequences were obtained and compared to those from CDNA using the same primer set. In addition, we performed Southern hybridisations to obtain a verification of the number of class I and class IIB genes as identified by PCR analyses.

African 'large' barb class I alpha genes

Anchored PCR was performed on full-length African 'large' barb liver CDNA, synthesised with the Generacer™ kit, to amplify MHC class I heavy chains. A reverse

Alignment of class I alpha chains

Numbering is based on the Baic-UA*0101 sequences including gaps. *Dashes* indicate identity to the Baic-UA*0101 sequence, *asterisks* indicate gaps and *dots* absence of sequence information. Cysteine residues, residues involved in salt-bridges and the FYP motif are marked by *light grey boxes*. Nine putative evolutionary conserved residues important for anchoring peptide termini are marked with *deep grey boxes*. The putative CD8 binding loop is indicated by (+), putative N-linked glycosylation sites are *underlined* and conserved putative β_2m contact residues are marked in **bold** and *italic*. Relevant accession numbers are shown in *Figure 2*.

Figure 1: Comparison of the deduced African 'large' barb class I amino acid sequences with common carp (Cyca), zebrafish (Dare) and human (HLA) classical class I sequences.

primer matching the cytoplasmic region of known cyprinid class I sequences in combination with the 5' Generacer™ primer revealed a PCR product of approximately 1100 bp. Cloning and sequence analyses revealed three unique nucleotide sequences. An additional partial nucleotides sequence was obtained by PCR using a reverse primer matching exon-4 of known cyprinid class I sequences in combination with a specific forward primer matching the first 28 nucleotides encoding part of the leader peptide of a known African 'large' barb class I sequence. Alignment of the four deduced amino acid sequences of African 'large' barb (Baic-UA*0101, Baic-UA*0102, Bain-UA*0401, Bain-UA*0701) with common carp, zebrafish and human class I sequences underpinned their classical nature (*fig. 1*).

Most features known to be conserved in classical class I molecules were present in the deduced African 'large' barb amino acid sequences. They possessed the conserved cysteine residues in the alpha-2 and alpha-3 domain to form disulfide bonds within these domains (*fig. 1*: C102, C167, C206, C264), conserved residues in alpha-1 and alpha-2 domains to form two salt bridges within these domains (*fig. 1*: H3, D29, H94, D120) and the FYP motif in alpha-3 domain (*fig. 1*: No 211-213). An exposed loop in the alpha-3 domain, that contains three acidic residues, forms a major CD8 binding site in HLA-A*02011 (*fig. 1*: D226, D230, E232). The African 'large' barb sequences all possess five acidic residues in this region of the alpha-3 domain. The preponderance of acidic residues in this region and the conservation of residues between cyprinid class I UA sequences could implicate binding of a CD8 co-receptor similar to that seen mammals[111]. Cyprinid class I ZE sequences also possess acidic residues in this region, though to a lesser extent. This region is less conserved among cyprinid ZE sequences[169], while it is conserved among mammalian class I and among cyprinid classical class I sequences. The acidic nature of this region is almost absent in the non-classical common carp Cyca-ZA, -ZB and -ZC and goldfish Caau-ZA and -ZD lineage sequences[74]. This suggests that two or more co-receptors may exist in these species; one that co-evolved with non-classical class I sequences, while the other remained conserved. Co-evolution of the CD8 co-receptor has been observed in chicken[64]. The tetraploid status of the species[67] (2n=100; *Cyprinus carpio* and *Carassius auratus*) may have facilitated the emergence of multiple CD8 co-receptors. Alternatively, sequences lacking the acidic nature of this region might not interact with a CD8 co-receptor. Class I molecules that serve as ligands for NK receptors do not require such region. The likelihood of the presence

of multiple CD8 co-receptors is conjectural and in contrast to the situation seen in mammals. The CD8 co-receptor region is well conserved between mammalian classical and non-classical class I sequences, allowing the co-receptor to interact with both types of molecules.

Four residues (*fig. 1*: T10, Q97, D120, Q245) known to be involved in β_2m binding of human class I molecules are conserved in the African 'large' barb sequences indicating similar contacts between the African 'large' barb β_2m molecule and class I molecules as seen in mammals. A single putative N-linked glycosylation site is present at the end of alpha-1 domain as found in most class I molecules. On the other hand, CD1, MIC[64] and the class I Z lineage molecules[74,169], with exception of the ZE sequences[74,169], possess multiple putative N-linked glycosylation sites. The alpha-1, alpha-2 and alpha-3 domains of the African 'large' barb sequences, like other cyprinid class I sequences, are comparable in length to human class I sequences with only small insertions and deletions up to three amino acids. Remarkably, the African 'large' barb sequences do not possess the insertion at the end of the alpha-3 domain as observed for common carp and zebrafish class I UA sequences (*fig. 1*), when compared to HLA-A*02011.

The presence of conserved class I features in the identified African 'large' barb sequences indicate that these cDNA sequences encode *bona fide* class I molecules. The nine evolutionarily conserved putative peptide-anchoring residues present in the amino acid sequence of classical class I genes have shown to be a useful criterion in discriminating classical class I genes and non-classical class I genes in many vertebrates[79,170-173]. This motif is present in most African 'large' barb sequences (*fig. 1*: Y7, Y60, R85, F124, T144, K147, W148, Y162, Y184) with the exception of the tyrosine residue (Y7) which is replaced by a serine residue in Baic-UA*0101 and the tyrosine residue (Y60) which is absent due to a deletion of three amino acids in the Bain-UA*0401 sequences. This observation indicates that the isolated cDNA sequences indeed encode classical class I molecules.

To investigate the relationship between genomic and expressed genes we tried to establish the genomic sequences from which the *UA* cDNAs were derived. Large introns separating the exon sequences hampered the analyses of full-length genomic class I sequences. Expand™ long template PCR fragments were amplified using the reverse primer matching the cytoplasmic region of known cyprinid class I sequences in combination with a primer matching the nucleotide

sequence encoding the leader peptides from *Baic-UA*0101*, as well as *Baic-UA*0102* and *Bain-UA*0701*. Agarose gel electrophoreses revealed fragments of 15 to 20 kb in length, indicating large introns. Such large introns separating class I exons, were previously also observed in common carp[107], coelacanth[120] and rainbow trout[11]. In addition, as a consequence of the nucleotide variability of exon-2, designing conserved primers at the start and end of this exon was impossible. Therefore, to be able to further investigate the presence of additional class I sequences at the genomic level, primers were designed matching the start and end of exon-3 of African 'large' barb *UA* cDNA sequences. PCR was performed on cDNA as well as genomic DNA and revealed four unique African 'large' barb *UA* exon-3 nucleotide sequences isolated both from genomic DNA and cDNA. Three sequences were identical to exon-3 sequences of *Baic-UA*0101, Baic-UA*0102*, and *Bain-UA*0401* identified by 5' RACE. The fourth sequence encoded a novel UA alpha-2 domain[164] and was assigned *Bain-UA*0301*.

In conclusion, the analyses revealed four unique exon-3 cDNA nucleotide sequences that were identical to those amplified from genomic DNA. We were, however, unable to isolate the *Bain-UA*0701* cDNA sequence from genomic DNA. Thus, the total number of five expressed genes (*Baic-UA*0101, Baic-UA*0102, Bain-UA*0301, Bain-UA*0401*, and *Bain-UA*0701*) does not correspond to the actual number of genomic sequences. It should be taken into account that isolation of genomic and/or expressed sequences might be biased by the primer set used since it is known that exon-3 sequences do show some nucleotide variability. Therefore, we cannot exclude with absolute certainty that we might have missed additional distantly related sequences to those identified in this study. To exclude this possibility we extended our analysis using DNA hybridisation.

Southern blot analyses were performed on *Hind*III or *Pst*I digested DNA using a *Baic-UA*0102* exon-4 probe (*fig. 8A*). Southern blot hybridisation patterns revealed either one or four hybridising fragments. The isolated exon-4 sequences of *Baic-UA*0101, Baic-UA*0102*, and *Bain-UA*0701* are 88% to 92% identical, which would allow cross-hybridisation even under stringent conditions. Therefore, observing a single intense hybridising *Pst*I fragment is most likely due to restriction enzyme digestion, which generated fragments similar in length.

The observation of four hybridising *Hind*III fragments is inconsistent with the presence of at least five class I *UA* cDNA sequences (*Baic-UA*0101, Baic-UA*0102, Bain-UA*0301, Bain-UA*0401* and *Bain-UA*0701*). Detection of four hybridising

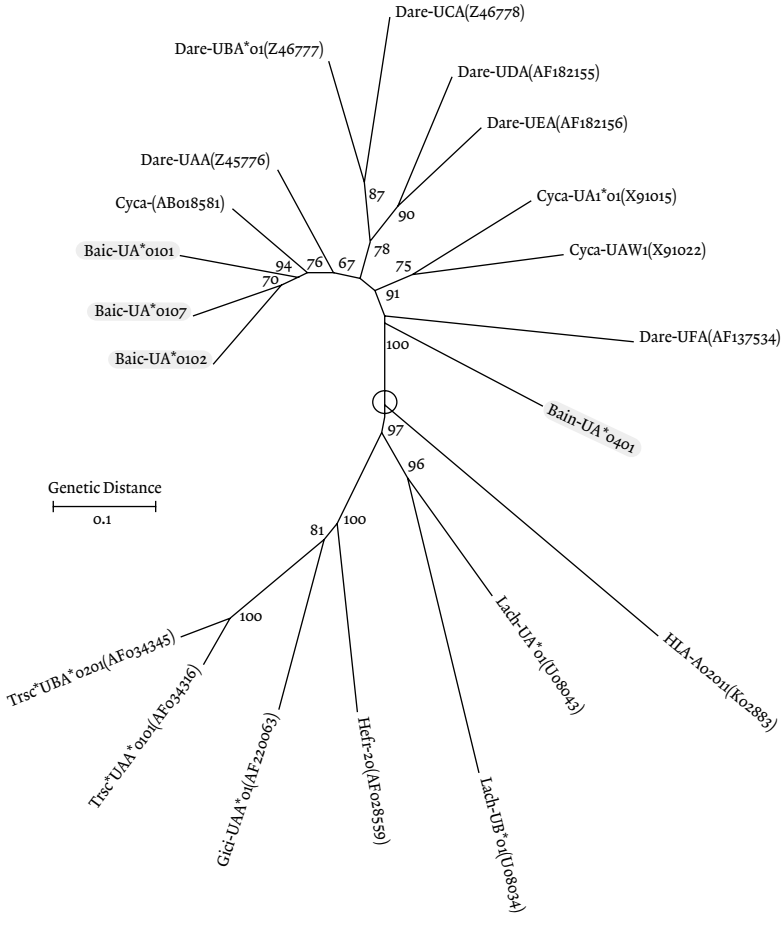


Figure 2: Neighbour-joining tree of classical cyprinid, shark and coelacanth class I amino acid sequences. Trees were constructed using the method of SAITOU and NEI[161] using p-distances in MEGA software. *Numbers* at branch nodes indicate bootstrap confidence levels of 1000 bootstrap replications. Accession numbers for African 'large' barb (Bain), common carp (Cyca), zebrafish (Dare), coelacanth (Lach), nurse shark (Gici), horned shark (Hefr), banded hound shark (Trsc) and human (HLA) are shown between brackets. The starting point of cluster formation is indicated by an *open circle*.

fragments suggests that the exon-4 sequences of either *Bain-UA*0301* or *Bain-UA*0401* are divergent and may not hybridise with the *Baic-UA*0102* exon-4 probe. Alternatively, restriction enzyme digestion could have generated fragments identical in length. However, this usually increases the intensity of the hybridisation signal.

Phylogenetic analyses of cyprinid classical and non-classical class I sequences have been extensively reviewed by STET and co-workers[74]. In the analyses presented in this study only classical class I sequences of several shark species, several cyprinids, and coelacanth were included (*fig. 2*). Two major clusters are formed that were supported by high bootstrap values; one comprising all cyprinid sequences and one including the coelacanth and shark sequences. Such clustering seems to suggest that the coelacanths are evolutionary more related to sharks. This is in contrast with the current views on Gnathostome (jawed vertebrates) evolution. Chondrichthyes (cartilaginous fishes) are thought to be ancestral to Osteichthyes, which includes the lungfishes (class Sarcopterygii, lobe-finned fishes) and the ray-finned fishes (class Actinopterygii). Lungfishes, together with the coelacanths, are considered to be the closest relatives to the tetrapods[174.175]. However, recent molecular studies using mtDNA sequences suggested that the Osteichthyes are ancestral to the Chondrichthyes. As a consequence the latter are therefore evolutionary more related to the coelacanths[56]. This notion seems to be supported by the phylogenetic analysis presented in this study.

Within the cyprinid cluster, the 'full-length' African 'large' barb sequences (Baic-UA*0101, Baic-UA*0102 and Bain-UA*0701) cluster together with a common carp sequence (Acc. No AB018581) and the zebrafish UAA locus, away from the other cyprinid sequences. These sequences may represent alleles from a single locus. The Bain-UA*0401 sequence clusters with all other cyprinid sequences and seems more closely related to *Dare-UFA* and may represent a second locus. However, phylogenetic analyses of exon-3 sequences only, which were isolated from individuals of the Lake Tana species flock, revealed a different topology[164]. In that analysis *Baic-UA*0102* seemed to have diverged from an ancestral gene that gave rise to the *Dare-UFA* sequence. In addition, the *Baic-UA*0101* exon-3 sequence was found together with *Dare-UBA* in the cluster comprising *Dare-UAA, Cyca-UA1*01* and *Cyca* (Acc. No AB018581) away from other clusters. The *Bain-UA*0401* and *Bain-UA*0701* sequences were located within a cluster comprising most of the other African 'large' barb class I exon-3 sequences away from other cyprinid

Numbering is based on the Bain-β_2m*01 sequence including gaps. *Dashes* indicate identity to the Bain-β_2m*01 sequence and *asterisks* indicate gaps. Conserved cysteine residues involved in the formation of the immunoglobulin domain are marked by *light grey boxes*. The *grey box* indicates the Ig-motif. The contact residues with the alpha-1 domain (#), the alpha-2 domain ($) and the alpha-3 domain (+) of the class I molecules are based on the human HLA-A2 molecule[114]. Accession numbers are shown in *Figure 4*.

Figure 3: Comparison of the deduced African 'large' barb β_2m amino acid sequences with common carp (Cyca), zebrafish (Dare) and human (HLA) β_2m sequences.

sequences. The phylogenies and the extensive sequence variation of class I exon-3 domains observed in Lake Tana African 'large' barb individuals seems to suggest haplotypic variation as seen in other bony fishes[176-178]. However, the extend of haplotypic variation and the total number of loci present in Lake Tana barbs still remain to be investigated.

African 'large' barb ß2-microglobulin genes

A 550 bp fragment was amplified by anchored PCR on full-length African 'large' barb liver cDNA synthesised with the Generacer™ kit using a reverse primer matching the 3'-untranslated region of putative partial *Bain-ß2m* sequences. Analyses indicated the presence of three unique protein-encoding nucleotide sequences (*Bain-ß2m*01-03*). Alignment of the deduced amino acid African 'large' barb sequences with known cyprinid and human ß2m sequences revealed the *bona fide* nature of these sequences (*fig. 3*).

The complete African 'large' barb ß2m proteins, 114 amino acids in length, possessed a leader peptide of 19 amino acids as predicted by signal peptide analyses[167] and mature peptides of 97 amino acids. These sequences are two amino acids shorter in length compared to the human ß2m molecule. The cysteine residues important for Ig-folding of the molecule are conserved (*fig. 3*, C25, C80) as is the YXCXVXH Ig-motif (*fig. 3*, 78-84). The African 'large' barb ß2m sequences possess ten identical contact residues for interaction with the class I heavy chain when compared to the human sequence. The other twelve residues known to be involved in the interaction with the class I heavy chain were shown to be conserved in all cyprinid ß2m sequences. Taken together this indicated that the sequences isolated encode *bona fide* ß2m molecules, that most likely are non-covalently associated with African 'large' barb MHC class I UA molecules. The ß2m molecule is usually encoded by a single copy gene in warm-blooded vertebrates[20,179]. Some fish species possess multiple ß2m genes. Common carp has two copies, which reflects the tetraploid nature of this species[109].

On the other hand rainbow trout possesses multiple ß2m genes, far exceeding the ploidy status of this species[116]. The presence of three expressed sequences reflects the hexaploid nature of African 'large' barb individual with ß2m loci present on three homologous chromosome pairs.

Phylogenetic analyses of the African 'large' barb ß2m amino acid sequences together with other fish and human ß2m genes showed clustering of all

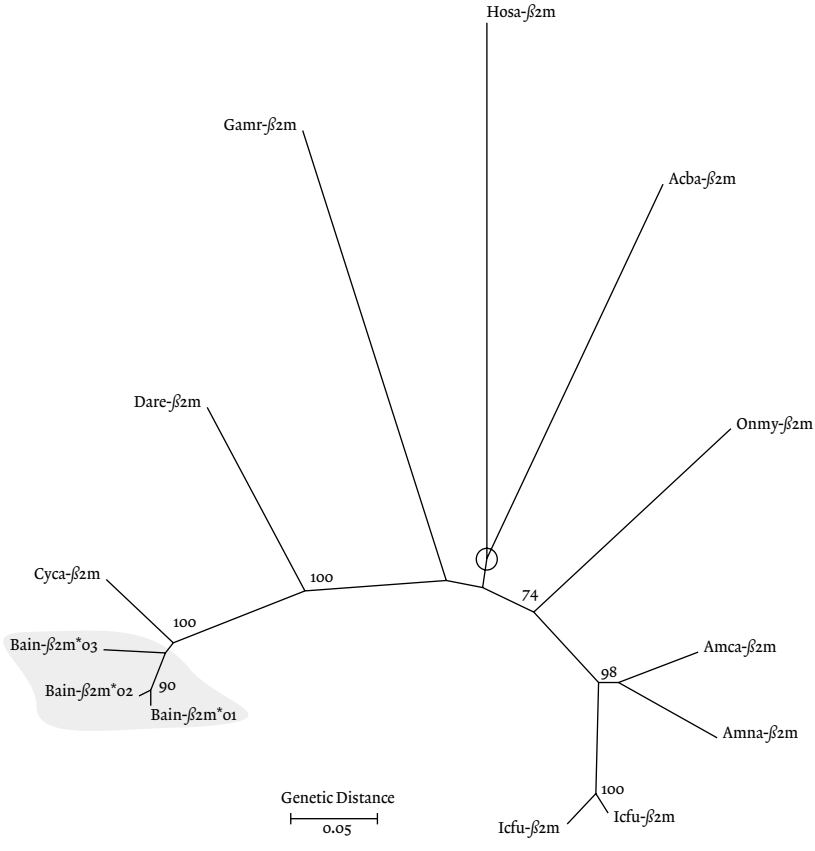


Figure 4: Neighbour-joining tree of $ß_2$m amino acid sequences. Trees were constructed using the method of Saitou and Nei[161] using p-distances in MEGA software. *Numbers* at branch nodes indicate bootstrap confidene levels of 1000 bootstrap replications. References for $ß_2$m sequences included are; common carp[109] (Cyca), zebrafish[185] (Dare), rainbouw trout[116] (Onmy,), Atlantic cod[186] (Gamr,), Siberian sturgeon[187] (Acba), catfishes[188] (Icpu, Icfu, Amna, Amca) and human[189] (Hosa). Starting point of cluster formation is indicated by an *open circle*.

cyprinid sequences (*fig. 4*). Divergence time calculations using a fourfold degenerated nucleotide substitutions rate[180] of 3.5 * 10^{-9} estimated that common carp and zebrafish β_2m have been separated for approximately 50 million years and common carp and African 'large' barb β_2m genes for approximately 24 million years. This is in accordance with the other studies that estimated that the genera *Danio*, *Cyprinus* and *Barbus* diverged 50 and 30 million years ago, respectively[69-71].

African 'large' barb class II alpha genes

Anchored PCR on full-length African 'large' barb liver cDNA using the 5' Generacer™ primer in combination with reverse primer designed to match the conserved part of the cytoplasmic region of known common carp and zebrafish sequences resulted in a PCR fragment of 700 bp in length. Nucleotide sequences analyses revealed four unique sequences and comparison of the deduced amino acid sequences with other cyprinid and human class IIA chains showed that they encoded African 'large' barb class IIA molecules (*fig. 5A*). The coding sequences, 217 amino acids in length, possessed a leader peptide, an alpha-1 domain, an alpha-2 domain, a connecting peptide, a transmembrane, and a cytoplasmic region comparable in length to human and other cyprinid class IIA sequences. Overall, only minor insertions or deletions were observed with exception of a consistent deletion of five amino acids in the alpha-1 domains in all fish sequences compared to the human sequence (*fig. 5A*). Possible cleavage sites were predicted to result in N-terminal leader peptides of 18, 18, 16 and 16 amino acids in length for Bain-DAA*0101 to -0401, respectively.

The alpha-1 and alpha-2 domains each possessed two cysteine residues to form disulfide bonds within these domains (*fig. 5A*, C19, C72, C113, C169). Potential N-linked glycosylation sites were present in some of the African 'large' barb class IIA sequences (*fig. 5A*). The typical regularly spaced glycine residues in the transmembrane region, which play a role in the interaction with the class IIB chain were present in all African 'large' barb class IIA sequences (*fig. 5A*, G203, G205, G209, G212, G216). Human class II molecules possess a conserved hydrophobic pocket formed by the MHC class II alpha-2 and beta-2 domain residues into which the phenylalanine residue of the CD4 molecule inserts[181].This region comprises the motif PphTVhsppPV-L-pP[xx]W, corresponding to aa 93-108 and 185 (*fig. 5A, 5C*, HLA-DRA*0101 & consensus respectively). Only four (P93, V103,

A. Alignment of class II alpha chains

A) *Numbering* is based on the Bain-DAA*0101 sequences including gaps. *Dashes* indicate identity to the Bain-DAA*0101 sequence, *asterisks* indicate gaps and *dots* absence of sequence information. Cysteine residues are marked with *light grey boxes*, conserved residues involved in peptide binding with *deep grey boxes*, residues contacting CD4 with *grey boxes*, (+) indicated regularly spaced glycine residues in the transmembrane region and putative N-linked glycosylation sites are *underlined*. Relevant accession numbers are shown in *Figure 7*. B) connecting peptide (cp), transmembrane (tm) and cytoplasmic tail (ct). c) Residues conserved in the human consensus are indicated by their amino acid code. *Dots* indicated absence of conservation, *p* indicates conservation of a polar residue, *h* indicates conservation of a hydrophobic residue, *s* indicates conservation of a small side-chain-containing residue. *Dashes* indicate identity to the Dare-DAB2*01 sequence.

Figure 5: Comparison of the deduced African 'large' barb class IIA amino acid sequences with common carp (Cyca), zebrafish (Dare) and human (HLA) classical class IIA sequences (A), the genomic organisation of African 'large' barb class IIA chains (B) and comparison of CD4 contact residues (C)

B. Genomic organisation of class *II* alpha chains

Bain-DAA*0101	58	322	246	541	285	405	123
Bain-DAA*0201	-	-	-	213	-	-	-
Bain-DAA*0301	-	370	-	128	279	241	-
Bain-DAA*0401	40	398	-	177	279	244	-

C. Alpha-2 domain residues contacting CD4

Consenses Human [181]	PphTVhsppPV.L.pP[75]W
Bain-DAA*0101	PDASMYSKEDVVLGAE[75]W
Bain-DAA*0201	--V-I---D-------[75]-
Bain-DAA*0301	-HT-I-A-D--Q-NV-[75]-
Bain-DAA*0401	-QT-I-AED--Q-NV-[75]-
Cyca-DXA1*01(X95432)	-------EG-----VQ[75]-
Cyca-DXA2*01(X95433)	--------D-----VQ[75]-
Dare-A4(L19451)	-QT-I--RD--Q-DI-[75]-
Dare-A1(L19445)	-VT-I--EDE---DEK[75]-
Dare-A3(L19450	-VT-I--EDE---DER[75]-
Gici-(M89950	-EV----EDL-EW-QL[75]-
Gici-(M89951)	-EV----EDL-EW-QL[75]-

L105, W185) out of eight residues (P93, T96, V97, P102, V103, L105, P108, W185), which were shown to be identical in human and murine sequences, are present in the cyprinid sequences included in the analyses. The shark sequences possessed similar residues to those present in cyprinid sequences with the exception of the leucine residue (L105). In shark this residue is replaced by a tryptophan (W105). A conserved substitution of the tyrosine residue (Y96) to a serine residue was observed in cyprinid and shark sequences. In addition, all these sequences possessed a tyrosine residue at position 98. Although this region seems to be only partly conserved between fish and mammalian sequences, it is likely to be involved in CD4 interaction. The data seem to indicate that interaction of class IIA with CD4 has changed, suggesting co-evolution of CD4. However, it is also possible that the five conserved residues (P93, T/S96, V103, L105, W185) are of more importance for proper interaction with CD4, while other residues provide only minor contributions. It must be taken into account that

interaction with CD4 depends both on class IIA and IIB chains, which together form a hydrophobic pocket for the phenylalanine residue (F43) of CD4. Therefore, the class IIB residues involved should be taken into account as well (see below).

Some features present in most mammalian class IIA were not conserved in the African 'large' barb sequences. They did not possess the glutamic acid residue directly preceding the transmembrane region commonly found in mammalian sequences. However, as observed for most cyprinid sequences, the African 'large' barb sequences possessed a proline residue at this position (*fig. 5A*, P198). Conserved residues involved in peptide binding in mammalian molecules are also less conserved in African 'large' barb sequences. Similar to the situation in other teleost class IIA sequences, only one (*fig. 5A*, N75) of the three conserved peptide binding residues in human class IIA molecules (*fig. 5A*, N68, N75, R82) is conserved in African 'large' barb class IIA sequences. Taken together, the isolated African 'large' barb sequences could be considered as *bona fide* class IIA molecules and most likely form a MHC class II molecule by non-covalent association with a class IIB chain.

To study the presence of additional class IIA sequences, sequences were amplified by PCR on cDNA and genomic DNA using the same primerset. The reverse primer was designed to match the conserved part of the cytoplasmic region of known common carp and zebrafish sequences. Two different forward primers were designed matching nucleotide sequence of leader peptides from either *Bain-DAA*0101* and *Bain-DAA*0201* or *Bain-DAA-0301* and *Bain-DAA*0401*. PCR on while genomic as well as cDNA resulted in four African 'large' barb class IIA sequences identical to those isolated by anchored PCR on cDNA.

Analyses of the genomic organisation of class IIA revealed 4 exons separated by three phase I introns[10]. Exon-1 to exon-3 encodes the leader peptide, the alpha-1 domain and the alpha-2 domain, respectively (*fig. 5B*). Exon-4 encodes the connecting peptide, the transmembrane and the cytoplasmic region (*fig. 5B*). Among mammals and other vertebrates distinct class IIA exon-intron organisations have been observed, including that of the African 'large' barb class IIA genes presented in this study[10]. Intron-1 is in usually greater than 1 kb in size, the remaining introns are generally smaller than 1 kb[10]. All African 'large' barb class IIA introns were relatively short and varied between 128 and 541 bp.

Similar introns of the different African 'large' barb class IIA sequences were shown to vary in length. Comparing the *Bain-DAA*0101* and *Bain-DAA*0201*

introns showed highly similar intron-1, intron-2, and intron-3 sequences with only a few nucleotide substitutions. However, the intron-2 sequence of *Bain-DAA*0201* showed a deletion of 328 bp compared to intron-2 of *Bain-DAA*0101* (data not shown). The other two sequences (*Bain-DAA*0301* and *Bain-DAA*0401*) possessed highly similar intron-3 sequences with only a few nucleotide substitutions and a deletion of three nucleotides in the intron-3 sequences of *Bain-DAA*0301*. Both intron-1 and intron-2 sequences are highly divergent between *Bain-DAA*0301* and *Bain-DAA*0401* (data not shown). Differences in intron sequences divide the four sequences in two groups (*Bain-DAA*0101/Bain-DAA*0201* and *Bain-DAA*0301/Bain-DAA*0401*), suggesting that these sequences represent two loci. However, the high similarity between coding sequences and the intron sizes between the two representatives of each group does not necessary imply that these are alleles of a distinct locus[182]. In common carp the two class IIA genes characterised co-segregate, suggesting that these sequences represent closely linked genes[105,183].

Phylogeny of cyprinid class IIA and IIB chains was constructed from amino acid sequences (*fig. 6*). Class IIA and IIB sequences of shark and human were included in the analyses. As expected, the class IIA and IIB sequences form two clusters supported by a high bootstrap value. The class IIA sequences of the cyprinid species intermingle forming two separate clusters. The cluster comprising, Bain-DAA*0301, Bain-DAA*0401, and Dare-A4 clustered away from the cluster with Bain-DAA*0101, Bain-DAA*0201, Dare-A1, Dare-A3, Cyca-DXA1*01 and Cyca-DXA2*01. Both clusters were supported by high bootstrap values. This topology suggests the presence of two loci present in the zebrafish and African 'large' barb of which only one is present in common carp. Although, closely related sequences from an individual may in fact not be alleles but reflect haplotypic variation with closely linked duplicated genes as observed for common carp class II beta[50,104,105].

Despite its hexaploid status, the African 'large' barb individual seems to possess only four expressed class IIA sequences with no additional sequences present at the genomic level. However, it should be taken into account that the isolation of the class IIA sequences might be primer biased due to nucleotide variability in the target sequences. Therefore, we can not exclude with absolute certainty the possibility that we may have failed to amplify sequences distantly related to those identified. Phylogenetic analyses and sequence identities suggest that

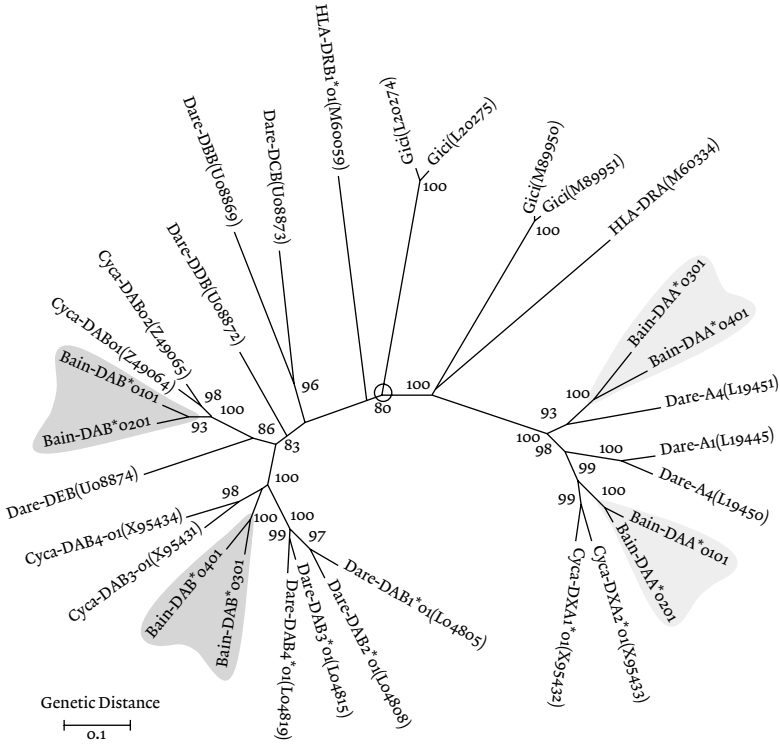


Figure 6: Neighbour-joining tree of cyprinid, shark and coelacanth class IIA and IIB amino acid sequences. Trees were constructed using the method of SAITOU and NEI[161] based on p-distances in mega software. *Numbers* at branch nodes indicate bootstrap confidence levels of 1000 bootstrap replications. Accession numbers for African 'large' barb (Bain), common carp (Cyca), zebrafish (Dare), nurse shark (Gici), and human (HLA) are shown between brackets. Starting point of cluster formation is indicated by an *open circle*.

the four sequences represent two loci. Assuming the presence of three pairs of chromosomes each harbouring a class IIA locus and no haplotypic variation, a maximum of six sequences is to be expected. Finding fewer than six may simply reflect homozygosity or haplotypic variation. Haplotypic variation of class IIA loci has been observed in other bony fish[48,177], and complicates the unambiguous assignment of loci/alleles present in this individual and in general among the Lake Tana African 'large' barbs.

African 'large' barb class II beta genes

Anchored PCR was performed on full-length African 'large' barb liver cDNA synthesised with the generacer™ kit to amplify MHC class IIB sequences. A reverse primer matching the cytoplasmic region of known cyprinid class IIB sequences in combination with the 5' generacer™ primer revealed a PCR product of 750 bp. Cloning and sequences analyses revealed four unique nucleotide sequences. Alignment of the deduced amino acid sequences with zebrafish, common carp and human class IIB sequences revealed the *bona fide* nature of the four African 'large' barb deduced amino acid sequences as class IIB chains (*fig. 7*A). They possess many protein features present in other fish and mammalian class IIB sequences. The cysteine residues involved in forming disulfide bridges are present in the beta-1 as well as the beta-2 domains of all African 'large' barb sequences (*fig. 6*A, C15, C80, C116, C173). An N-linked glycosylation site is present at amino acid positions 42-44 (*fig. 7*A). Conserved residues involved in peptide binding in mammalian class II molecules are well conserved (*fig. 7*A, W62, H82, N83) with the exception of the histidine residue which is replaced by a proline residue in the Bain-DAB*0101 sequence. The transmembrane region of all African 'large' barb sequences possess the regularly spaced glycines that play a role in interaction with class IIA chains. At amino acid positions 144-160 a putative stretch of amino acids is found corresponding to the region in mammalian class IIB sequences known to be involved in CD4 binding. The CD4 binding region is only partly conserved between fish and mammalian sequences. Only five conserved amino acids present in human sequences (S145, T146, G152, W154, Q157) are observed in cyprinid sequences (*fig. 7*B). This suggests that interaction of class IIB with CD4 has changed, implying co-evolution of CD4. However, it is also possible that the five conserved residues are of more importance

A. Alignment of class II beta chains

A) *Numbering* is based on the Bain-DAB*0101 sequences including gaps. *Dashes* indicate identity to the Bain-DAB*0101 sequence, *asterisks* indicate gaps and *dots* absence of sequence information. Cysteine residues are marked with *light grey boxes*, conserved residues involved in peptide binding with *deep grey boxes*, residues contacting CD4 with *grey boxes*, (+) indicated regularly spaced glycine residues in the transmembrane region and putative N-linked glycosylation sites are *underlined*. Relevant accession numbers are shown in *Figure 7*. B) Residues conserved in the human consensus are indicated by their *amino acid code*. *Dots* indicated absence of conservation, *p* indicates conservation of a polar residue and *h* indicates conservation of a hydrophobic residue. *Dashes* indicate identity to the Dare-DAB2*01 sequence.

Figure 7: Comparison of the deduced African 'large' barb class IIB amino acid sequences with common carp (Cyca), zebrafish (Dare) and human (HLA) classical class IIB sequences (A) and comparison of cd4 contact residues (B).

B. Beta-2 domain residues contacting CD4

Consenses Human [181]	.ST.LhpNGDWTFQ.LV
*Dare-DAB2*01(L04805)*	TSTMEMANGNWYYQIHS
*Dare-DAB1*01(L04808)*	---I------------
*Dare-DAB3*01(L04815)*	-------D---------
*Dare-DAB4*01(L04819)*	-------D---------
Dare-DCB(U08873)	--IE---D---------
Dare-DDB(U08872)	-------D---------
Dare-DEB(U08874)	---E-L-D---------
*Bain-DAB*0101*	---E--P-----K----
*Bain-DAB*0201*	---E--P-------V--
*Bain-DAB*0301*	-----------------
*Bain-DAB*0401*	-----L-D---------
*Cyca-DAB*01(Z49064)*	---E--P----------
*Cyca-DAB*02(Z49065)*	---E--P----------
*Cyca-DAB3*01(X95431)*	-------D---F-----
*Cyca-DAB4*01(X95434)*	----------F-----
Gici-(L20274)	-TVELLS----T--VRQ
Gici-(L20275)	-TVELLS----T--VRQ

for proper interaction with CD4, while other residues provide only minor contributions.

To investigate the presence of additional class IIB sequences at the genomic level, PCR was performed on genomic DNA, and the resulting sequences were compared to those amplified from cDNA. A reverse primer matching the cytoplasmic region of known cyprinid class IIB sequences in combination with a forward primer matching the nucleotide sequences encoding the leader peptides of common carp and African 'large' barb class IIB sequences. Using these primers, four unique African 'large' barb sequences, corresponding to those identified by anchored PCR, were amplified using cDNA as template. Agarose gel electrophoreses of a PCR product amplified using genomic DNA revealed fragments of approximately 1350 bp, 1450 bp and 1600 bp. Subsequent cloning and sequencing of the PCR product corresponding to the smallest fragment observed in agarose gel analyses, resulted in two unique sequences. Analyses of these nucleotide sequences revealed multiple small and large deletions and insertions in the exon sequences and absence of several splice sites. These genomic sequences did not seem to possess an exon-2 sequence. Alternatively, the exon-2 sequence might be unrecognisable due to extreme diversification of this region. The exon-3 sequences were 83% identical to each other. One of the two exon-3 sequences was 71% identical to the exon-3 sequences of *DAB*0301* and *DAB*0401*

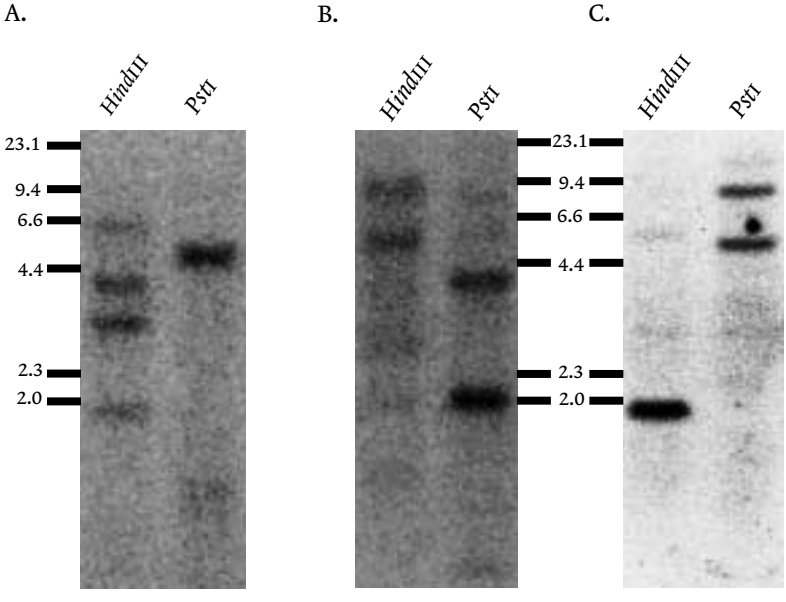


Figure 8: Southern blot hybridisation pattern of *Hind*III or *Pst*I-digested DNA, hybridised with a *Baic-UA*0102* exon-4 probe (A), a *Bain-DAB*0101* exon-3 probe (B), or a *Bain-DAB*0401* exon-3 probe (C).

and the other was 75% identical to the exon-3 sequences of *DAB*0101* and *DAB*0201*. Amplification of *DAB* intron-1 and exon-2 genomic fragments did reveal the presence of *Bain-DAB*0101, Bain-DAB*0201, Bain-DAB*0301*, and *Bain-DAB*0401* in this individual[164]. Most likely these sequences are present in the two larger fragments (1450 and 1600 bp), that resisted cloning. To verify the number of genomic sequences obtained we performed DNA hybridisation to establish the number of class IIB genes.

Southern blot analyses were performed on restriction enzyme digested DNA (*Hind*III or *Pst*I) with two different probes, *Bain-DAB*0301* exon-3 and *Bain-DAB*0401* exon-3. Southern blot hybridisation using the *DAB*0101* or the *DAB*0401* exon-3 probe revealed one or two strongly hybridising fragments and in some cases additional weaker bands (*fig. 8B, 8C*). The stronger signals correspond, most likely, to *DAB*0101* and *DAB*0201* or to *DAB*0301* and *DAB*0401* since their exon-3 coding sequence showed the highest similarity. The weaker bands are most likely due to cross-hybridisation of the *DAB*0101* exon-3 probe with *DAB*0301* and *DAB*0401* genes or the *DAB*0401* exon-3 probe with *DAB*0101* and *DAB*0201* genes. Alternatively, the weaker bands are the result of cross-hybridisation with the identified pseudogenes.

The presence of six class IIB sequences in the genome of the African 'large' barb is consistent with the hexaploid status of this individual possessing three class IIB loci. However, the six class IIB sequences do not necessarily represent alleles of three different loci. Haplotypic variation combined with the presence of closely linked duplicated genes, as seen in other vertebrates[48,50,104,105,184], could also account for the presence of six class IIB sequences. Such a situation would severely complicate the assignment of loci similar to the situation for class IIA genes.

In the phylogenetic analyses of cyprinid, shark and human class IIA and IIB sequences (*fig. 6*), each formed a separate cluster. The class IIB sequences of different species intermingle in a *trans*-species manner. Two sub-clusters are formed comprising sequences of African 'large' barb, common carp and zebrafish. The zebrafish DAB sequences cluster together with Bain-DAB*0301, Bain-DAB*0401, Cyca-DAB*03 and Cyca-DAB*04, suggesting an orthologous relationship. On the other hand *Bain-DAB*0101, Bain-DAB*0201, Cyca-DAB*01* and *Cyca-DAB*02* genes seem to be more related to the *Dare-DAB* gene. The *Dare-DEB* is a functional gene, although its expression has not been demonstrated[48]. The

Dare-DBB, -DCB and *-DDB* pseudogenes are found at considerable distance from the two cyprinid sub-clusters. No African 'large' barb or common carp orthologues of these sequences have yet been found. The phylogeny and sequence identities seem to suggest the presence of two loci present in the common carp and African 'large' barb of which only one is present in present in zebrafish.

The African 'large' barb individual investigated in this study seems to express up to four class IIA and class IIB sequences, similar to the situation seen in common carp and zebrafish. This indicates that an increased ploidy status does not necessarily result in an increased number of expressed genes. These data suggest that a silencing mechanism plays a role in reducing the number of expressed genes. However, it can not be excluded that other African 'large' barb individuals might express more genes since up to six genomic *DAB*01/DAB*02* genomic exon-2 sequences were identified in representatives of other species belonging to the African 'large' barb species flock[70,164].

In conclusion, our data from this hexaploid African 'large' barb individual seems to indicate that the ploidy status does not correlate with the presence and expression of MHC genes. Only five class I and four class IIA genes present at the genomic level were shown to be expressed. This seems to suggest that gene silencing is most likely a result of physical exclusion. However, functional silencing also seems to have played a role in reducing the number of expressed class II genes, as class IIB pseudogenes were identified in the genome of this African 'large' barb individual. We realise that our conclusions are based on a single individual, which could be the odd-one out. However, our interpretation that both mechanisms are operative is corroborated by another study in which we have shown that a *Barbus acutirostris* individual expressed only three out of the six identified class IIB sequences present in its genome[70].

ACKNOWLEDGMENTS

The authors would like to thank MARTIN DE GRAAF, LEO NAGELKERKE and NAND SIBBING for providing tissue samples of a Lake Tana *Barbus intermedius* individual.

CHAPTER 4

DIFFERENT MODES OF MAJOR HISTOCOMPATIBILITY CLASS IA AND CLASS II*B* EVOLUTION IN THE LAKE TANA AFRICAN 'LARGE' BARB SPECIES FLOCK.

Corine P. Kruiswijk*, Trudi Hermsen*, Joost van Heerwaarden*, Brian Dixon**, Huub F. J. Savelkoul*, René J. M. Stet*

*Cell Biology and Immunology Group,
Department of Animal Sciences, Wageningen University,
PO box 338, 6700 AH Wageningen, The Netherlands

**Department of Biology, University of Waterloo,
200 University Ave West, Ontario, N2L 3G1, Canada

ABSTRACT

The sixteen Lake Tana African 'large' barb species inhabit different ecological niches, exploit different food webs and have different temporal and spatial spawning patterns within the lake. This Lake Tana African 'large' barb species flock is the result of adaptive radiation within the last 5 million years. During speciation, selection pressure on peptide binding residues of major histocompatibility molecules is hypothesised to be extreme. Individuals move to new environments and as a consequence their major histocompatibility molecules encounter new pathogens. The variability of class IIB exon-2 encoding gene in individuals that belonged to four Lake Tana African 'large' barb species (*B. acutirostris, B. nedgia, B. tsanensis, and B. truttiformis*) was studied using polymerase chain reaction amplification and DNA sequencing. This revealed that these genes are indeed under strong selection pressure. In addition, no sharing of class IIB alleles between species was observed for the four Lake Tana African 'large' barb species.

We extended the study on class IIB exon-2 encoding gene with six Lake Tana African 'large' barb species (*B. intermedius, B. brevicephalus, B. macrophtalmus, B.*

megastoma, B. platydorsus and B. surkis) and African 'large' barbs (*B. intermedius*) from the Blue Nile and its tributaries. In addition, we studied the presence and variability of class I *UA* exon-3 sequences in seven Lake Tana African 'large' barb species (*B. intermedius, B. megastoma, B. platydorsus, B. acutirostris, B. nedgia, B. tsanensis, B. truttiformis*) and African 'large' barbs (*B. intermedius*) from the Blue Nile. Comparisons of nonsynonymous and synonymous substitutions in peptide binding and non-peptide binding regions revealed that variability of Lake Tana African 'large' barb class Ia and class IIB genes is generated by positive Darwinian selection acting upon peptide binding regions. In general, phylogenetic lineages are maintained by purifying or neutral selection on non-peptide binding regions. Positive selection on peptide binding regions between phylogenetic lineages of class Ia and class IIB genes and recombination events seem to have contributed to diversification of these genes. Remarkably, class IIB intron-1 and exon-2 sequences were not shared between African 'large' barb species. In contrast, class I *UA* exon-3 sequences were shared. These observations suggest that the turn-over of class IIB alleles in the Lake Tana African 'large' barb species is much higher than that of class Ia alleles.

INTRODUCTION

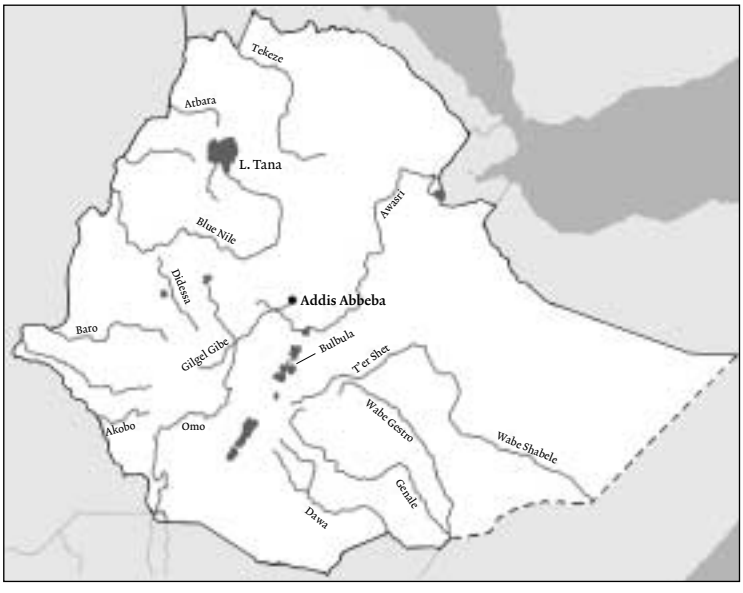
Lake Tana situated in the north-western highlands of Ethiopia harbours a unique aquatic ecosystem system formed during 5 million years of isolation. In late pleistocene times a volcanic eruption blocked the outlet to the Blue Nile by creating a 40 m high waterfall[76]. This event effectively prevented gene flow between the lake and the Blue Nile and all other fresh-water systems that are connected to this river. During this period of isolation, an ancestral *Barbus intermedius* population underwent speciation as a result of adaptive radiation to different ecological niches within the lake resulting in fifteen novel African 'large' barb species[75]. In contrast, only a single extant species, *Barbus intermedius*, is found in the Blue Nile and its tributaries. Beside these fifteen species, the lake harbours a population of African 'large' barbs occupying the shores of Lake Tana resembling the *Barbus intermedius* of the Blue Nile system. In general, during such speciation events, selection pressure is expected to be strong as predicted by the punctuated equilibrium model[190]. As such, it is expected that selection pressure on peptide binding residues of major histocompatibility complex (MHC) class I and class II molecules would be extreme[39], since individuals

move to new environments and as a consequence their MHC molecules encounter new pathogens.

The MHC class I and class II genes encode structurally similar cell surface glycoproteins that present peptides to cytotoxic T or helper T lymphocytes, respectively. Unlike all other jawed vertebrates, class I and class II genes of bony fish are located on different linkage groups[48,53,191,192]. MHC class I and class II loci include genes known to be highly polymorphic. Differences between nucleotide sequences, encoding allotypic MHC molecules, can be created by point mutations[30]. However, other mechanisms such as recombination (reciprocal and non-reciprocal) and exon shuffling have been demonstrated to play a major role[30-33]. Polymorphism is maintained selectively by natural selection known as a form of balancing selection such as overdominant selection or heterozygotic advantage[34]. Evidence that polymorphism is maintained by a form of balancing selection came from examination of nucleotide substitutions in different regions of MHC molecules. Comparison of synonymous nucleotide substitution rates (d_S) and nonsynonymous substitutions rate (d_N) of peptide binding residues and non-peptide binding residues in the coding regions of MHC genes revealed that d_N rates substantially exceed d_S rates of peptide binding residues[34-37]. The basis of selective maintenance of MHC polymorphism is peptide binding and thus pathogen driven. Overdominant selection and other forms of balancing selection are known to be capable of maintaining polymorphism for longer periods than neutral selection[38]. Polymorphic lineages of MHC class I and class II alleles are known to be maintained for millions of years, predating speciation events, a process known as trans-species evolution[39-42]. In general MHC class I genes seem to be evolving more rapidly than class II genes. It has been shown that allelic mammalian class I lineages were maintained for up to 6 million years, while certain class II lineages were maintained for up to 35 million years[43-45].

DIXON and co-workers[70] studied MHC class II polymorphism in four (*B. acutirostris, B. nedgia, B. tsanensis, B. truttiformis*) out of fifteen species belonging to the Lake Tana African 'large' barb species flock. Comparison of nonsynonymous and synonymous nucleotide substitutions rates of PBR and non-PBR of the isolated class IIB sequences revealed indeed a strong selection pressure acting upon peptide binding residues. The four species under study did not possess identical sequences. This suggests that in each species a unique set of class IIB

A.



B.

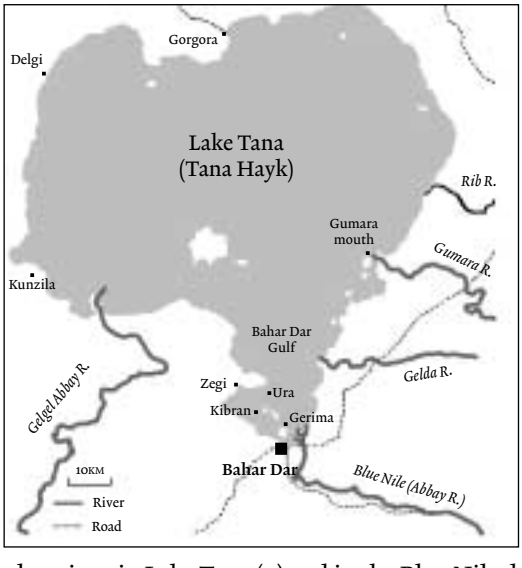


Figure 1: Sample locations in Lake Tana (A) and in the Blue Nile drainage system (B) of the specimens used in this study.

sequences evolved from a common ancestral suit within a time span of 5 million years due to an expansion into a new environment. As a likely consequence the class II molecules were faced with new pathogens.

To test this hypothesis we extended the study on MHC class IIB encoding sequences performed by DIXON and co-workers[70] with six Lake Tana African 'large' barb species (*B. intermedius, B. brevicephalus, B. macrophtalmus, B. megastoma, B. platydorsus, B. surkis*) and African 'large' barbs (*B. intermedius*) from the Blue Nile and its tributaries comprising in total eighteen individuals. In addition, to validate whether MHC class I sequences evolved in a similar species specific manner during a time span of 5 million year, class I encoding sequences from seven different Lake Tana African 'large' barb species *(B. intermedius, B. megastoma, B. platydorsus, B. acutirostris, B. nedgia, B. tsanensis, B. truttiformis*) and from Blue Nile African 'large' barbs *(B. intermedius*), comprising in total 35 individuals, were analysed.

MATERIALS AND METHODS

Sampling of fish and DNA extraction

Muscle or fin clip samples from 30 Lake Tana *Barbus intermedius* species flock individuals and seven *Barbus intermedius* individuals of the Blue Nile system were collected from various locations in the lake and the Blue Nile system (*fig.1*). Genomic DNA was isolated using a Wizard® genomic DNA purification kit (Promega, Madison, WI, USA) according to the protocol provided. DNA concentration was determined using the GeneQuant system (Amersham Pharmacia Biotech, Roosendaal, Netherlands).

Polymerase chain reaction conditions

Standard PCR reaction conditions were 1 x reaction buffer, 1.5 mM $MgCl_2$, 1 unit of Taq polymerase (Goldstar, Eurogentec, Seraing, Belgium), 0.2 mM dNTPs, 0.2µm of each primer and 100 ng genomic DNA. The cycling profile was 1 cycle at 94°C for 5 min followed by 30 cycles consisting of denaturing at 94°C for 30 sec, annealing at 55°C for 30 sec, polymerisation at 72°C for 1 min, and a final cycle of 10 min at 72°C.

Amplification of genomic MHC class I exon-3 and class II intron-1 and exon-2 sequences

Genomic class II intron-1 and exon-2 sequences were amplified using similar primers as used by DIXON and co-workers[70]. Exon-2 of class IIB genes encodes

the beta-1 domain which is known to be involved in peptide binding[12]. Alpha-2 domains of class I alpha chains were amplified using primers matching the nucleotide sequence of the start (5' GGTGTTCACTCAGTCCAG 3') and the end (5' CTTTTCTCTCCAGAGAGTCCTT 3') of known cyprinid class I alpha-2 encoding exons. Together with exon-2, exon-3 of class I genes encodes the peptide binding groove[114]. In general exon-2 sequences possess the highest variability making it more suitable for the analyses performed in this study. However, amplification of exon-2 sequences was hampered by the inability to design a conserved primer matching the start of exon-2. In addition, introns separating exon-1 and exon-2 and exon-2 and exon-3 have shown to be large[193]. Therefore, only exon-3 sequences were amplified and analysed.

Cloning and DNA sequencing

PCR products were ligated and cloned using the PGEM® T-easy kit (Promega, Madison, WI, USA) following the manufacturer's description. Plasmid DNA was isolated from bacterial cells using the QIAprep spin miniprep kit (QIAGEN, Valencia, CA, USA), according to the protocol provided. Subsequently, plasmid DNA was sequenced using the ABI prism bigdye™ terminator cycle sequencing ready reaction kit (PE Applied Biosystems, Foster City, CA, USA) and analysed using an ABI 377 sequencer (PE Applied Biosystems, Foster City, CA, USA).

Accession numbers and nomenclature

The new sequences reported here were deposited in the EMBL database under the following accession numbers: AJ506649 to AJ506736. The nomenclature used to assign new sequences or rename existing sequences adheres, in part, to the recommendations described in the HLA facts book[165]. The first four letters refer to the species names and are followed by a dash for the locus designation (*UA* for class I sequences and *DAB* for class IIB), an asterisk and four digits. According to the HLA facts book[165], the first two digits following the asterisk describe the lineage and the third and fourth digits that follow assign alleles. However, it is not yet clear whether the class Ia or class IIB sequences isolated belong to a single locus or multiple loci and which lineage they represent. Therefore, the nomenclature given assumes that all class I and all class IIB sequences belong to a single locus *UA* for class I and *DAB* for class IIB and represent a single lineage, 01. The following two digits, in general describing the alleles, indicate in

this study the order in which a novel sequence was isolated from a species. Sequences that were shared between African 'large' barb species started with the abbreviation *Baic*, referring to the *Barbus intermedius* species complex. Fourteen out of twenty-two previously described class I exon-3 sequences characterised in a earlier study[74] were renamed and included in this study. The fourteen sequences were isolated from two *Barbus nedgia* (lip) individuals, a *Barbus acutirostris* (acute, BOB) and a *Barbus intermedius* (shore-complex, FLIP) individual. The sequences were previously designated *L01 to L04, L06 to L09, A01* to *A05* and *F02* and renamed to *Bane-UA*0101(li), Baic-UA*0101, -0112, -0111* and *-0114, Bane-UA*0107(li), -0108(li)* and *-0109(li), Baac-UA*0101(ac), -0102(ac), -0103(ac), -0104(ac)* and *-0105(ac), Bain-UA*0104* (Acc. Nos. AJ007885 to 88, -90 to -93, AJ007879 to -83, and -99), respectively. Some sequences were only found as single clones in an individual; *L05, L10, L11* and *L13* isolated from a lip, *L06, A04* and *F03* isolated from a shore-complex and *A06* isolated from an acute individual. Thorough analyses by multiple PCRs and subsequent sequences analyses did not result in isolation of these sequences and were therefore not included in the current analyses. The analyses performed only included class I *UA* exon-3 sequences found at least two times in two independent PCRs. Two other sequences, designated *L12* and *F01*, that seemed pseudogenes in an earlier study[74], were shown to be sequence artefacts. Thorough analyses revealed that these sequences were similar. Subsequently, this sequence was designated *Baic-UA*0102*.

Nucleotide sequence and amino acid sequence analyses

The isolated class I and class II nucleotide sequences were represented by at least two identical clones from two independent PCR amplifications. Sequence data obtained using the ABI sequencer were analysed with sequencer 4.1 software (Gene Codes, Ann Arbor, Michigan, USA). Multiple alignments were performed using the program[166] CLUSTAL-W version 1.8. Phylogenetic analyses were constructed in MEGA 2.1 software[168] using the neighbour-joining algorithm[161] with the JUKES-KANTOR method for nucleotide sequences. Synonymous (d_S), nonsynonymous (d_N) substitution, nucleotide substitution rates (d), and nonsynonymous minus synonymous substitution (d_N-d_S) were calculated with the MEGA 2.1 software package using the JUKES-KANTOR method. Standard errors were estimated by thousand bootstrap replications. We realise that bootstrap tests in some cases may lead to erroneous values when the number of

Table 1: Presence of class II *DAB*01* sequences in African 'large barb' individuals

Individuals	Location	Sequence number																
		01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19
B. intermedius	Blue Nile system																			
River barbs																				
No 10971	Didessa river	rb	rb																	
No 30971	Bulbula river			rb																
No 08973	Gibe river						rb													
No 09973	Loko river					rb														
No 30973	Bulbula river				rb															
B. intermedius	Lake Tana																			
Shore-complex																				
Flip		sc	sc																	
B. acutirostris	Lake Tana																			
Acute																				
Bob		ac			ac	ac	ac	ac				ac								
No 2	Bahar Dar gulf		ac			ac														
No 4	Bahar Dar gulf	ac	ac	ac		ac														
No 9	Bahar Dar gulf		ac		ac		ac													
B. brevicephalus	Lake Tana																			
Shorthead																				
No 43	Ura	sh	sh	sh																
B. macrophtalmus	Lake Tana																			
Bigmouth big-eye																				
No 29	Kibran	be	be	be	be															
B. megastoma	Lake Tana																			
Bigmouth small-eye																				
No 16	Gerima				se				se											
No 17	Gerima				se															
No 75	Gumara river	se	se	se				se												
No 84	Gumara mouth		se	se		se	se													
No 85	Gumara mouth		se	se			se													
B. nedgia	Lake Tana																			
Lip																				
No 55	Bahar Dar gulf			li	li	li	li	li												
No 87	Bahar Dar gulf	li	li																	
B. platydorsus	Lake Tana																			
White hunch																				
No 22	Kibran	wh	wh																	
No 24	Kibran	wh	wh		wh															
No 37	Kibran				wh															
No 65	Ura			wh	wh															
B. surkis	Lake Tana																			
Zurki																				
No 62	Ura	zu	zu																	
B. tsanensis	Lake Tana																			
Intermedius																				
No 58	Bahar Dar gulf												in	in						in
No 60	Bahar Dar gulf			in												in	in			
No 67	Bahar Dar gulf	in		in						in	in	in								
No 76	Bahar Dar gulf	in	in																	
No 78	Bahar Dar gulf			in	in	in	in	in	in											
No 438	Bahar Dar gulf													in	in					
No 481	Bahar Dar gulf	in										in	in			in		in	in	
B. truttiformis	Lake Tana																			
Troutlike																				
No 457	Ura	tr	tr		tr															
No 458	Ura	tr	tr	tr	tr															
No 471	Gumara river					tr	tr	tr	tr											

synonymous and nonsynonymous substitutions are small. NEI and JIN[194] provided a better method for computing variances and covariances. However, this method is difficult to perform with large data sets.

The phylip 3.6 DNAml program[195] and PLATO[196] were used to identify recombination tracts.

RESULTS

*Analyses of African 'large' barb class II DAB*01 genes*

DIXON and co-workers[70] amplified class II intron-1 and exon-2 coding sequences by PCR on genomic DNA of 16 Lake Tana African 'large' barb species flock individuals using a primer set designed to match the leader sequence and the end of exon-2 of the *Cyca-DAB2*01* gene. The sixteen individuals represented four different African 'large' barb species; *B. acutirostris* (acute), *B. nedgia* (lip), *B. tsanensis* (intermedius) and *B. truttiformis* (troutlike).

In this study we extended the analyses of class II intron-1 and exon-2 encoding sequences in the Lake Tana African 'large' barb species flock with 18 individuals using a similar approach as DIXON and co-workers[70]. Genomic DNA was obtained from one *B. brevicephalus*, one *B. macrophtalmus*, five *B. megastoma*, four *B. platydorsus*, one *B. surkis* and six *B. intermedius* individuals. The latter represented five individuals from the Blue Nile system (river barbs) and one shore complex individual from Lake Tana (*fig. 1, table 1*).

All PCRs yielded products of approximately 480 bp for each African 'large' barb individual, the expected product size of African 'large' barb *DAB*01* equivalents[70]. In PCRs of some individuals an additional product of approximately 800 bp was obtained, which is the expected product size of African 'large' barb *DAB*03* equivalents. Sequence analyses of the 480 bp fragments revealed 29 novel *DAB*01* sequences, 23 identified in individuals of the Lake Tana African 'large' barb species flock and six identified in African 'large' barb individuals from the Blue Nile system. Together with the 40 *DAB*01* sequences identified by DIXON and co-workers[70], a total of 69 different *DAB*01* sequences were obtained (*fig. 2, table 1*).

No sharing of these *DAB*01* sequences was observed for individuals from different African 'large' barb species. However, *DAB*01* sequences were shared by individuals of the same species (*table 1*). The *B. intermedius* individuals from different locations in the Blue Nile system did not share any *DAB*01* sequences. The

African 'large' barb DAB*01 exon-2 amino acid sequences

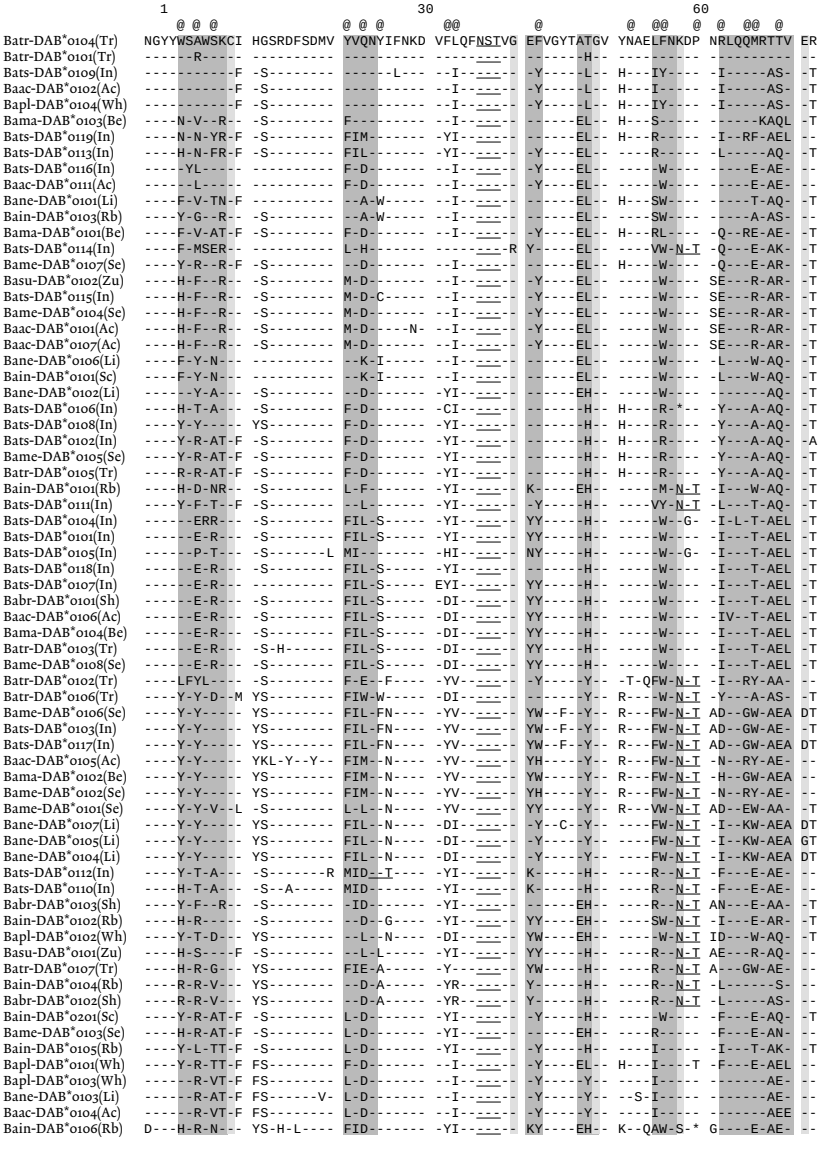


Figure 2.

number of different *DAB*01* sequences observed in a single African 'large' barb individual varied from one to six.

Alignment of the deduced amino acid sequence of all DAB*01 partial exon-2 sequences confirmed that these sequences were equivalent to the previously identified African 'large' barb DAB*01 sequences and demonstrated the presence of conserved amino acid residues in the deduced protein sequences of the African 'large' barb *DAB*01* genes (*fig. 2*). These conserved amino acid residues are thought to be important for MHC function, as determined for human MHC class IIB genes[12]. These included the cysteine residue at position 9, the glycine residues at position 40 and 49 that produce turns important for the secondary structure of the protein, and the arginine or lysine residue at position 58 which forms a salt bridge with a aspartic acid or a glutamic acid residue at position 71 (*fig. 2*). A putative N-linked glycosylation site conserved between all African 'large' barb DAB*01 sequences was present at position 36 to 38. Several sequences possess an additional glycosylation site at positions 58 to 60. Based on motifs at variable positions, the 69 African 'large' barb DAB*01 sequences could be grouped (*fig. 2*).

‹ *Figure 2*: The deduced African 'large' barb class II DAB*01 exon-2 amino acid sequences. The *numbers* above denote the amino acid position relative to the mature protein. *Dashes* indicate identity to the Bats-DAB*0118(in) sequence and *asterisks* indicate gaps. Highly conserved cysteine residues, residues involved in salt-bridges and residues important for the secondary structure are marked with *grey boxes*. Residues known to be involved in peptide binding in mammalian class IIB molecules[12] are indicated by @ below numbering. Putative recombination tracts (*table 8*) are marked with *deep grey boxes*. Putative N-linked glycosylation sites are underlined. African 'large' barb sequences starting with Bats-, Baac-, Bane-, Batr- are renamed sequences described by DIXON and coworkers[70]. Previously, the Lake Tana African 'large' barb species were indicated by the morphotype nomenclature. To facilitate identification, morphotype abbreviations are indicated between brackets following the sequence designation. Morphotype abbreviations are: acute (ac), barbel (ba), bighead (bh), bigmouth mini-eye (me), bigmouth small-eye (se), bigmouth big-eye (be), carplike (ca), dark (da), intermedius (in), lip (li), shorthead (sh), score-complex (sc), troutlike (tr), white hunch (wh), zurki (zu)

*Phylogenetic relationship of African 'large' barb DAB*01 intron-1 and exon-2 sequences*

Analyses of the phylogenetic relationship of the 69 African 'large' barb *DAB*01* partial exon-2 sequences revealed a clustering pattern similar observed by DIXON and co-workers[70]. They were able to identify ten clusters supported by high bootstrap values (above 70%) that comprised several different African 'large' barb species. Eleven out of the 29 novel *DAB*01* sequences clustered within one of these ten clusters (*fig. 3*). Three additional clusters (*fig. 3*: 1-3), supported by high bootstrap values, were formed that comprised in total 5 novel sequences and 3 sequences identified by DIXON and co-workers[70]. Cluster 3 was formed by identical partial exon-2 sequences, *Bain-DAB*0101(sc)* and *Bain-DAB*0106(li)*, but their intron sequences differed (see *fig. 4*). Possibly the missing 51 nucleotides at the end of exon-2 of those sequence differ. The African 'large' barb clusters contained up to a maximum of six different species per cluster, while the clusters VIII, IX, and X contained sequences from a single species.

A phylogenetic tree constructed using all *DAB*01* intron-1 sequences of Blue Nile system 'large' barb (*B. intermedius*) and Lake Tana African 'large' barb species, and common carp *DAB*01* and *DAB*02* intron-1 sequences showed a specific clustering separating the two genera. The clustering in several groups deviated only slightly from phylogenetic analyses of *DAB*01* introns performed by DIXON and co-workers[70] (*fig. 4*). The nine previously observed clusters, A to I, were also observed when novel intron sequences were included in the analyses. However, the clusters B, E, and I, now showed bootstraps value below 50. Only eight novel intron sequences clustered together with one of the previous identified clusters, including A, B, E, F or G, with up to three novel sequences in a single cluster. Two additional clusters (J, K) supported by high bootstraps were formed that each comprised two intron sequences. Each cluster contained up to a maximum of four different species. Four clusters (C, D, I, J) only contained intron sequences from a single African 'large' barb species. The clusters G and D together comprise intron sequences belonging to the exon-2 sequences of cluster I in the phylogenetic analysis of exon-2 sequences, with the exception of *Batr-DAB*0102(tr)*. The intron clusters A, B, C, E, F, H, and I comprised similar sequences to the exon clusters VI, IV, X, II, V, III, and VIII, respectively, with some exceptions (*Baac-DAB*0107(ac), Bane-DAB*0107(li), Bats-DAB*0104(in), -06(in),* and *-08(in)*). No intron cluster was formed that corresponded to exon cluster VII (*fig. 3, fig. 4*).

Nucleotide diversity in African 'large' barb class II *DAB*01 intron-1 and exon-2*

Clusters in phylogenetic analyses of MHC genes usually represent lineages of alleles. On the basis of the phylogeny of African 'large' barb *DAB*01* exon-2 sequences twelve such lineages could be identified (*fig. 3*). The phylogenetic clusters comprising at least three sequences (I, IV, V, VI, VII), were used for comparison of nucleotide diversity within and between lineages (*table 2*). Two clusters (1 and 2), previously not identified by DIXON and co-workers[70], were also included in this analyses since these clusters are supported by high bootstrap values in the phylogenetic analyses (*fig. 3*). Overall synonymous (d_s) and nonsynonymous (d_n) substitution rates of pairwise comparisons between all *DAB*01* exon-2 sequences revealed a significantly higher mean d_n value of more than two times the mean d_s. Similar, comparisons within and between phylogenetic lineages revealed mean d_n and d_s values essentially the same for within lineage comparisons, while a significantly higher mean d_n value of almost twice the mean d_s was observed for between lineage comparisons (*table 2*).

The function of class II molecules, that is binding of peptides by domains encoded by exon-2 sequences of class IIA and IIB genes, predicts that synonymous and nonsynonymous substitution rates are expected to differ for amino acid residues involved in peptide binding (PBR) and non-peptide binding residues (non-PBR). However, residues involved in peptide binding have not yet been identified for fish class II molecules by crystallographic analysis. Therefore, we used the model of BROWN and co-workers[12] to assign PBR and non-PBR amino acid residues, since DAB*01 variability corresponded well with HLA polymorphism[70]. Overall and between pairwise comparison of PBR residues of phylogenetically defined DAB lineages revealed significantly higher mean d_n values of almost five times the mean d_s. Similar comparisons revealed a mean d_n value more than twice the value of d_s for within lineage calculations, although this was not significant. Mean d_n and d_s values were essentially similar for overall, within, and between lineage pairwise comparison of non-PBR residues (*table 2*). Mean d_s values of overall, within and between lineage pairwise comparisons, calculated for PBR, non-PBR or all exon-2 codons, were not significantly higher with respect to mean d values in introns. Differences in substitution rates are expected between exons and introns of MHC genes. HUGHES[197] showed that in the case of *HLA-DRB1*01*, recombination events

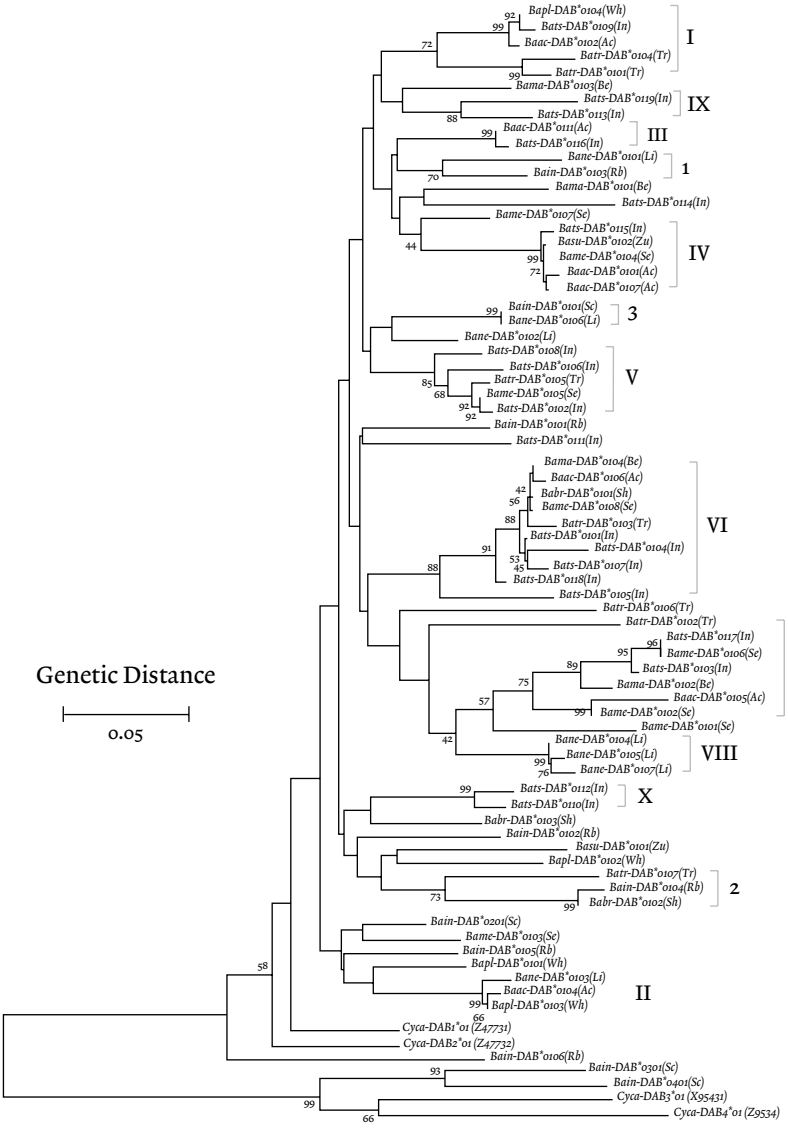

Bapl-DAB*0104(Wh)
Bats-DAB*0109(In)
Baac-DAB*0102(Ac)
Batr-DAB*0104(Tr)
Batr-DAB*0101(Tr)
Bama-DAB*0103(Be)
Bats-DAB*0119(In)
Bats-DAB*0113(In)
Baac-DAB*0111(Ac)
Bats-DAB*0116(In)
Bane-DAB*0101(Li)
Bain-DAB*0103(Rb)
Bama-DAB*0101(Be)
Bats-DAB*0114(In)
Bame-DAB*0107(Se)
Bats-DAB*0115(In)
Basu-DAB*0102(Zu)
Bame-DAB*0104(Se)
Baac-DAB*0101(Ac)
Baac-DAB*0107(Ac)
Bain-DAB*0101(Sc)
Bane-DAB*0106(Li)
Bane-DAB*0102(Li)
Bats-DAB*0108(In)
Bats-DAB*0106(In)
Batr-DAB*0105(Tr)
Bame-DAB*0105(Se)
Bats-DAB*0102(In)
Bain-DAB*0101(Rb)
Bats-DAB*0111(In)
Bama-DAB*0104(Be)
Baac-DAB*0106(Ac)
Babr-DAB*0101(Sh)
Bame-DAB*0108(Se)
Batr-DAB*0103(Tr)
Bats-DAB*0101(In)
Bats-DAB*0104(In)
Bats-DAB*0107(In)
Bats-DAB*0118(In)
Bats-DAB*0105(In)
Batr-DAB*0106(Tr)
Batr-DAB*0102(Tr)
Bats-DAB*0117(In)
Bame-DAB*0106(Se)
Bats-DAB*0103(In)
Bama-DAB*0102(Be)
Baac-DAB*0105(Ac)
Bame-DAB*0102(Se)
Bame-DAB*0101(Se)
Bane-DAB*0104(Li)
Bane-DAB*0105(Li)
Bane-DAB*0107(Li)
Bats-DAB*0112(In)
Bats-DAB*0110(In)
Babr-DAB*0103(Sh)
Bain-DAB*0102(Rb)
Basu-DAB*0101(Zu)
Bapl-DAB*0102(Wh)
Batr-DAB*0107(Tr)
Bain-DAB*0104(Rb)
Babr-DAB*0102(Sh)
Bain-DAB*0201(Sc)
Bame-DAB*0103(Se)
Bain-DAB*0105(Rb)
Bapl-DAB*0101(Wh)
Bane-DAB*0103(Li)
Baac-DAB*0104(Ac)
Bapl-DAB*0103(Wh)
Cyca-DAB1*01 (Z47731)
Cyca-DAB2*01 (Z47732)
Bain-DAB*0106(Rb)
Bain-DAB*0301(Sc)
Bain-DAB*0401(Sc)
Cyca-DAB3*01 (X95431)
Cyca-DAB4*01 (Z9534)
I
IX
III
1
IV
3
V
VI
VII
VIII
X
2
II
Genetic Distance
0.05

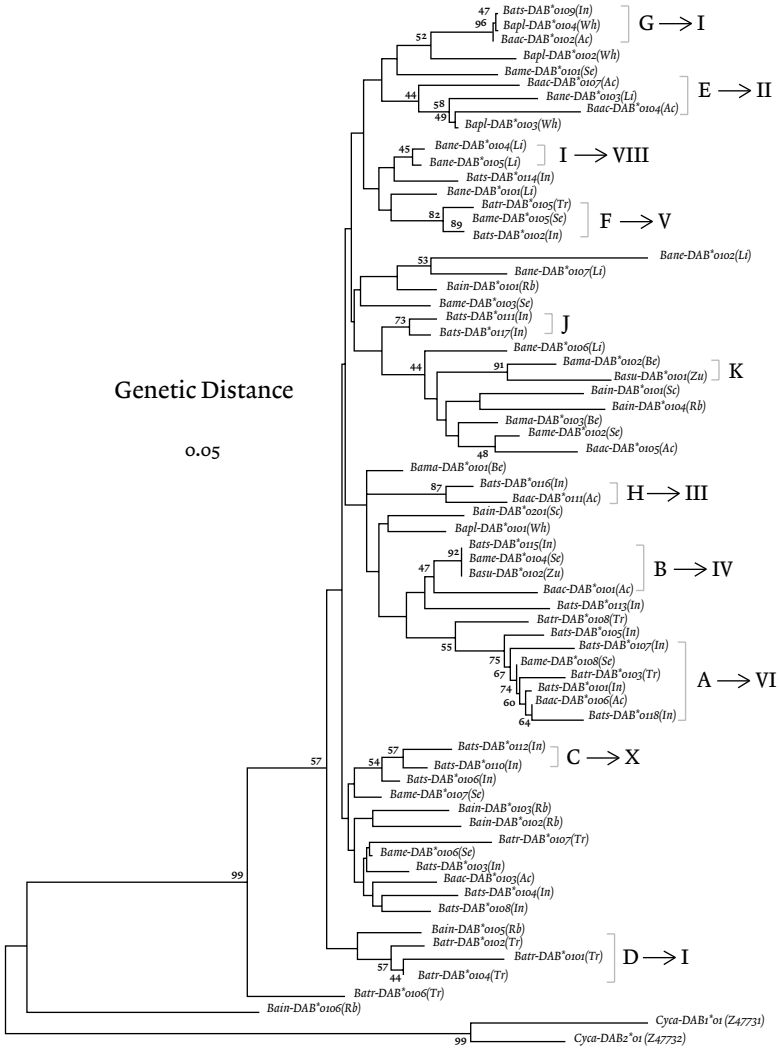
Table 2: Mean ± SE of nucleotide substitutions per site (d) in intron-1 and synonymous (d_S) and non-synonymous (d_N) substitutions per site in exon-2 within and between groups of African 'large' barb *DAB*01* alleles as defined by phylogeny (Fig. 2)

	Exon-2		Intron-1
	d_N ± SE	d_S ± SE	d ± SE
Overall[1]	[4]0.167 ± 0.030	0.074 ± 0.016	0.0676 ± 0.0084
Within[2]	0.048 ± 0.015	0.040 ± 0.021	0.0500 ± 0.0120
Between[3]	[4]0.183 ± 0.041	0.093 ± 0.036	0.0675 ± 0.0132

	PBR		Non-PBR	
	d_N ± SE	d_S ± SE	d_N ± SE	d_S ± SE
Overall[1]	[4]0.650 ± 0.128	[5]0.137 ± 0.053	0.071 ± 0.015	0.062 ± 0.016
Within[2]	0.139 ± 0.060	0.063 ± 0.054	0.026 ± 0.011	0.035 ± 0.023
Between[3]	[4]0.833 ± 0.286	0.170 ± 0.101	0.070 ± 0.022	0.073 ± 0.038

1 Mean ± SE of pairwise comparison of 69 barbus class II *DAB*01* sequences
2 Mean ± SE of pairwise comparison within groups of barbus class II *DAB*01* sequences
3 Mean ± SE of pairwise comparison between groups of barbus class II *DAB*01* sequences
4 d_N significantly higher than d_S; Tested hypothesis: $d_N > d_S$; $p < 5\%$
5 d_S significantly higher than d; Tested hypothesis: $d_S > d$; $p < 10\%$

‹ *Figure 3*: Neighbour-joining tree of African 'large' barb class II *DAB*01* exon-2 nucleotide sequences. Trees were constructed using the method of SAITOU and NEI[161] based on JUKES-KANTOR method in MEGA software. *Numbers* at branch nodes indicate bootstrap confidence levels of 1000 bootstrap replications. Only bootstrap values over 40% are shown. African 'large' barb sequences starting with *Bats-, Baac-, Bane-, Batr-* are taken from DIXON *et al.* [70]. Previously, the Lake Tana African 'large' barb species were indicated by the morphotype nomenclature. To facilitate recognition, the morphotype abbreviations (see *fig. 2*) are indicated between brackets following the sequence designation. The accession numbers of several common carp (*Cyca*) and zebrafish (*Dare*) class IIB sequences, which were included in the analyses, are shown in brackets following the sequences assignments. Brackets indicated denote clusters (I to XI and 1 to 3) supported by bootstrap values over 70%. The clusters I to XI correspond to the clustering pattern and nomenclature of DIXON and co-workers[70], while those numbered 1 to 3 represent novel clusters identified in this study.


Genetic Distance
0.05
Bats-DAB*0109(In)
Bapl-DAB*0104(Wh)
Baac-DAB*0102(Ac)
G → I
Bapl-DAB*0102(Wh)
Bame-DAB*0101(Se)
Baac-DAB*0107(Ac)
Bane-DAB*0103(Li)
Baac-DAB*0104(Ac)
E → II
Bapl-DAB*0103(Wh)
Bane-DAB*0104(Li)
Bane-DAB*0105(Li)
I → VIII
Bats-DAB*0114(In)
Bane-DAB*0101(Li)
Batr-DAB*0105(Tr)
Bame-DAB*0105(Se)
Bats-DAB*0102(In)
F → V
Bane-DAB*0102(Li)
Bane-DAB*0107(Li)
Bain-DAB*0101(Rb)
Bame-DAB*0103(Se)
Bats-DAB*0111(In)
Bats-DAB*0117(In)
J
Bane-DAB*0106(Li)
Bama-DAB*0102(Be)
Basu-DAB*0101(Zu)
K
Bain-DAB*0101(Sc)
Bain-DAB*0104(Rb)
Bama-DAB*0103(Be)
Bame-DAB*0102(Se)
Baac-DAB*0105(Ac)
Bama-DAB*0101(Be)
Bats-DAB*0116(In)
Baac-DAB*0111(Ac)
H → III
Bain-DAB*0201(Sc)
Bapl-DAB*0101(Wh)
Bats-DAB*0115(In)
Bame-DAB*0104(Se)
Basu-DAB*0102(Zu)
B → IV
Baac-DAB*0101(Ac)
Bats-DAB*0113(In)
Batr-DAB*0108(Tr)
Bats-DAB*0105(In)
Bats-DAB*0107(In)
Bame-DAB*0108(Se)
Batr-DAB*0103(Tr)
Bats-DAB*0101(In)
Baac-DAB*0106(Ac)
Bats-DAB*0118(In)
A → VI
Bats-DAB*0112(In)
Bats-DAB*0110(In)
C → X
Bats-DAB*0106(In)
Bame-DAB*0107(Se)
Bain-DAB*0103(Rb)
Bain-DAB*0102(Rb)
Batr-DAB*0107(Tr)
Bame-DAB*0106(Se)
Bats-DAB*0103(In)
Baac-DAB*0103(Ac)
Bats-DAB*0104(In)
Bats-DAB*0108(In)
Bain-DAB*0105(Rb)
Batr-DAB*0102(Tr)
Batr-DAB*0101(Tr)
D → I
Batr-DAB*0104(Tr)
Batr-DAB*0106(Tr)
Bain-DAB*0106(Rb)
Cyca-DAB1*01 (Z47731)
Cyca-DAB2*01 (Z47732)

homogenised intron-1 and intron-2 sequences relative to exon-2 sequences within lineages of alleles, but not between lineages.

In the case of cyprinid DAB*01, comparisons of African 'large' barb intron-1 and exon-2 sequences revealed that these evolved at a similar rate, while between genera comparisons revealed that intron-1 and exon-2 sequences evolved at different rates[197]. The African 'large' barb *DAB*01* sequences used in the analyses performed by HUGHES[197] may represent unique lineages or may be on their way of becoming unique lineages. Therefore, pairwise comparisons of within phylogenetic lineage sequences may reveal a correlation between d_S in exons and d in introns with higher d_S rates than d rates.

Plots of d in intron-1 vs d_S in exon-2 of pairwise comparisons within and between phylogenetic lineages did not reveal such a correlation (*fig.5A, fig. 5B*). However, both plots showed some pairwise comparisons with relatively high d values in introns while d_S in exons were low. In other comparisons the opposite was observed (*fig. 5A, fig. 5B*). Mean values of pairwise comparisons per individual sequence revealed significantly higher d_S rates in exons of some sequences, namely *Bats-DAB*0111(in), Bane-DAB*0101(li), Bats-DAB*0101(be), Bama-DAB*0104(rb), -05(rb)* and *-06(rb)* (*fig. 5D*). A plot of d in intron-1 vs d_S in exon-2 including only pairwise

‹ *Figure 4*: Neighbour-joining tree of African 'large' barb class II *DAB*01* intron-1 nucleotide sequences. Trees were constructed using the method of SAITOU and NEI[161] based on JUKES-KANTOR method in MEGA software. *Numbers* at branch nodes indicate bootstrap confidence levels of 1000 bootstrap replications. African 'large' barb sequences starting with *Bats-, Baac-, Bane-, Batr-* are taken from DIXON and co-workers[70]. Previously, the Lake Tana African 'large' barb species were identified as morphotypes. To facilitate recognition, the morphotype abbreviations (see *fig. 2*) are indicated between brackets following the sequence assignments. The accession numbers of several common carp (*Cyca*) and zebrafish (*Dare*) class IIB sequences, which were included in the analyses, are shown in brackets following the sequences name. Brackets, indicated with the letters A to I, correspond to the clustering pattern and nomenclature of DIXON and co-workers[70]. Brackets, indicated with the letters I, J and K represent novel clusters observed in this study which were supported by bootstrap values over 70%. *Arrows* point to cluster designation showing a similar clustering pattern as identified in phylogeny of class II *DAB*01* exon-2 (*fig. 3*).

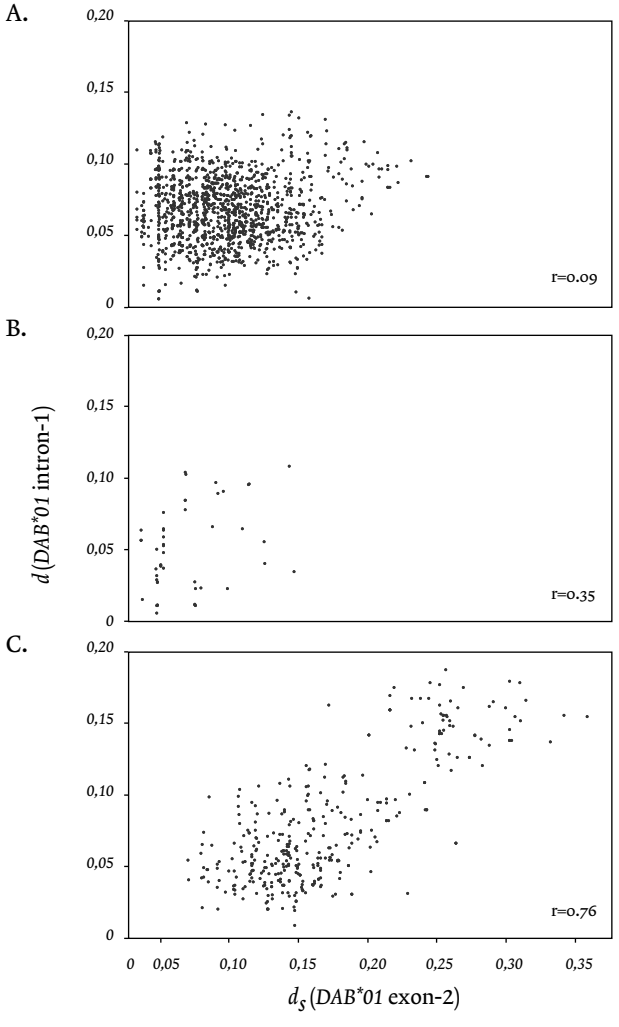


Figure 5: Pairwise comparisons of substitution rates in DAB*01 intron-1 (d) and in DAB*01 exon-2 (d_S). Plots of between lineage pairwise comparisons (A) within lineage pairwise comparisons (B) and only pairwise comparisons *Bats-DAB*0111(in), Bane-DAB*0101(li), Bats-DAB*0101(be), Bama-DAB*0104(rb), -05(rb)* and *-06(rb)* with all other Lake Tana 'large' barb class II *DAB*01* sequences (y=0.5435x) (C). Correlation coefficients are indicated in boxes. D) Pairwise comparisons of a single *DAB*01* intron-1 or *DAB*01* exon-2 sequences with all other Lake Tana 'large' barb class II *DAB*01* sequences. Test $H_0 : d_S = d$; $H_1 : d_S > d$; * = $p < 0.05$, using the Z-test.

D.

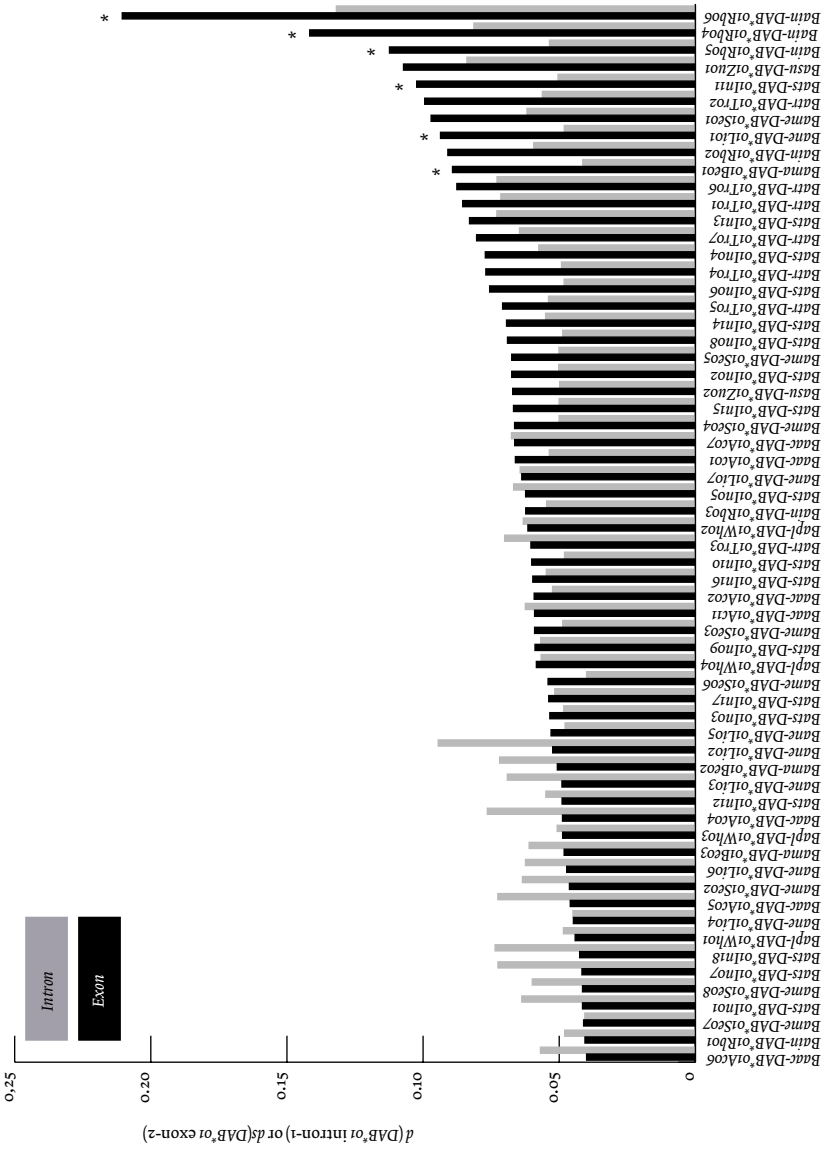

Intron
Exon
0
0.05
0.10
0.15
0.20
0.25
d (DAB*or intron-1) or dS(DAB*or exon-2)
Baac-DAB*olAc06
Bain-DAB*olRb01
Bame-DAB*olSc07
Bats-DAB*olIn01
Bame-DAB*olSc08
Bats-DAB*olIn07
Bats-DAB*olIn18
Bapl-DAB*olWh01
Bame-DAB*olLi04
Baac-DAB*olAc05
Bame-DAB*olSc02
Bame-DAB*olLi06
Bama-DAB*olBc03
Bapl-DAB*olWh03
Baac-DAB*olAc04
Bats-DAB*olIn12
Bame-DAB*olLi03
Bama-DAB*olBc02
Bame-DAB*olLi02
Bame-DAB*olLi05
Bats-DAB*olIn03
Bats-DAB*olIn17
Bame-DAB*olSc06
Bapl-DAB*olWh04
Bats-DAB*olIn09
Bame-DAB*olSc03
Baac-DAB*olAc01
Baac-DAB*olAc02
Bats-DAB*olIn16
Bats-DAB*olIn10
Bapl-DAB*olWh02
Bain-DAB*olRb03
Bats-DAB*olIn05
Bame-DAB*olLi07
Baac-DAB*olAc01
Baac-DAB*olAc07
Bame-DAB*olSc04
Bats-DAB*olIn15
Basu-DAB*olZu02
Bats-DAB*olIn02
Bame-DAB*olSc05
Bats-DAB*olIn08
Bats-DAB*olIn14
Bats-DAB*olIn06
Bats-DAB*olIn04
Bapl-DAB*olWh01
Bame-DAB*olLi04
Baac-DAB*olAc05
Basu-DAB*olZu01
Bain-DAB*olRb05
Bain-DAB*olRb04
Bain-DAB*olRb06

Table 3: Mean ± SE of ratios of non-synonymous (d_n) per synonymous (d_s) substitutions per site and (d_n - d_s) values of PBR and non-PBR regions within African 'large' barb *DAB*01* alleles and between groups of African 'large' barb *DAB*01* alleles as defined by phylogenetic relationship (*fig. 3*).

	PBR			non-PBR		
	d_n/d_s ± SE	d_n-d_s ± SE	Sig[1]	d_n/d_s ± SE	d_n-d_s ± SE	Sig[1]
Overall	4.74 ± 0.139	0.5141 ± 0.1080	*	1.15 ± 0.022	0.0092 ± 0.0225	«
I	1.10 ± 0.124	0.0128 ± 0.1184	«	0.88 ± 0.028	-0.0028 ± 0.0220	«
IV		0.0098 ± 0.0102	«		0.0030 ± 0.0032	«
V	2.00 ± 0.080	0.0525 ± 0.0444	«	1.09 ± 0.014	0.0151 ± 0.0077	«
VI		0.0193 ± 0.0148	«	0.89 ± 0.013	-0.0008 ± 0.0117	«
VII	10.00 ± 0.063	0.1263 ± 0.0496	*	3.43 ± 0.021	0.0341 ± 0.0205	«
1	1.21 ± 0.157	0.0326 ± 0.1123	«	0.52 ± 0.061	-0.0387 ± 0.0597	«
2	4.09 ± 0.144	0.2844 ± 0.1537	**	0.27 ± 0.042	-0.0527 ± 0.0369	«
I - IV	2.66 ± 0.431	0.4830 ± 0.3465	«	1.21 ± 0.037	0.0125 ± 0.0346	«
I - V	5.98 ± 0.347	0.6821 ± 0.3019	*	0.38 ± 0.047	-0.0427 ± 0.0446	«
I - VI	3.33 ± 0.195	0.3310 ± 0.1893	**	2.25 ± 0.025	0.0439 ± 0.0250	«
I - VII	5.03 ± 0.393	0.8671 ± 0.3582	*	2.83 ± 0.038	0.0748 ± 0.0370	«
I - 1	4.31 ± 0.250	0.4703 ± 0.2071	*	0.65 ± 0.034	-0.0195 ± 0.0309	«
I - 2	3.50 ± 0.156	0.5389 ± 0.2501	*	0.74 ± 0.049	-0.0215 ± 0.0485	«
IV - V	8.84 ± 0.192	0.4475 ± 0.1731	*	0.57 ± 0.056	-0.0298 ± 0.0559	«
IV - VI	4.84 ± 0.409	0.6304 ± 0.3412	**	1.74 ± 0.035	0.0364 ± 0.0323	«
IV - VII	4.75 ± 0.290	0.5978 ± 0.2870	*	2.87 ± 0.050	0.0878 ± 0.0504	«
IV - 1	4.14 ± 0.311	0.4803 ± 0.2411	**	0.55 ± 0.045	-0.0305 ± 0.0428	«
IV - 2	4.76 ± 0.521	1.0385 ± 0.4946	*	0.61 ± 0.060	-0.0417 ± 0.0581	«
V - VI	6.05 ± 0.258	0.5097 ± 0.2190	*	0.72 ± 0.047	-0.0020 ± 0.0483	«
V - VII	8.55 ± 0.290	0.7026 ± 0.2524	*	2.26 ± 0.039	0.0635 ± 0.0411	«
V - 1	7.42 ± 0.179	0.4559 ± 0.1609	*	0.48 ± 0.046	-0.0341 ± 0.0455	«
V - 2	4.65 ± 0.327	0.7005 ± 0.3150	*	0.41 ± 0.070	-0.0789 ± 0.0707	«
VI - VII	23.35 ± 0.192	0.5812 ± 0.1906	*	3.07 ± 0.031	0.0653 ± 0.0350	«
VI - 1	2.88 ± 0.247	0.4019 ± 0.1847	*	1.40 ± 0.036	0.0314 ± 0.0379	«
VI - 2	6.69 ± 0.338	0.8707 ± 0.2944	*	0.39 ± 0.047	-0.0480 ± 0.0492	«
VII - 1	3.54 ± 0.399	0.7146 ± 0.3107	*	2.98 ± 0.043	0.0860 ± 0.0457	«
VII - 2	9.71 ± 0.392	1.7506 ± 0.3763	*	0.72 ± 0.053	-0.0261 ± 0.0517	«
1- 2	3.33 ± 0.331	0.6778 ± 0.2885	*	0.61± 0.060	-0.0475 ± 0.0598	«

1 Z-test for selection H_0: $d_n = d_s$: neutral H_1: $d_n =/= d_s$: « = p ‹ 5%; positive H_1: $d_n › d_s$: * = p ‹ 2.5%, ** = p ‹ 5%

comparisons of these sequences with all others did reveal a correlation with d_s rates in exons of twice the d rates in introns (*fig. 5C*).

*Selection on African 'large' barb class II DAB*01 sequences*

As a result of the immunological function of MHC molecules, amino acid residues of PBRS are expected to be under positive Darwinian selection. It is, therefore, expected that in the case of positive selection the ratio of nonsynonymous sub-

stitutions per site to synonymous substitutions per site is higher than unity. NEI and HUGHES[198] concluded that in cases where positive Darwinian selection is acting upon PBR residues a value of at least 3 is expected. Most d_n/d_s ratios calculated for PBR residues of all DAB*01 sequences and for between lineage comparisons were above 3, with the exception of between lineage comparisons of I vs IV, and VI vs 1 (*table 3*). Within lineage d_n/d_s ratios were above 1, with a maximum value of 10. Some calculations were hampered by the lack of synonymous substitutions. However, to establish whether d_n was significantly higher than d_s it is necessary to conduct a statistical test of the difference between d_n and d_s.

Performing such a test revealed that most d_n values were significantly higher than d_s values for PBR residues. Exceptions were almost all within lineage calculations (I, IV, V, VI, and 1) and a single between lineage calculation (I vs IV). Similar analyses for non-PBR residues revealed d_n/d_s ratio's ranging from 0.2 up to 3. However, d_n values calculated for non-PBR residues of overall, within or between lineage comparisons were not significantly higher or lower than d_s values (*table 3*).

Analyses of African 'large' barb class I UA genes

The genomic DNA isolated from 34 Lake Tana African 'large' barb species flock or Blue Nile *Barbus intermedius* individuals that served as template in PCRs for *DAB*01* analyses, was also used to amplify class I *UA* exon-3 sequences, with the exception of individuals 29, 43 and 62. The class I *UA* analyses included 4 individuals, that were not covered in the *DAB*01* analyses (*B. intermedius* No. 08971, No. 32972; *B. megastoma* No. 74 and *B. platydorsus* No. 23). Analyses of 35 individuals by PCR using primers matching the start and the end of class I exon-3 yielded products of approximately 250 bp. In total 72 different genomic sequences (*fig. 6*) could be identified by sequence analyses. Analyses of the deduced amino acid sequences confirmed that these sequences were equivalent to fish class I *UA* genes. Most of these sequences possessed the conserved amino acid residues important for MHC function as determined for HLA class I molecules[7]. These residues included the cysteine residues (C98 and C162) that form a disulphide bridge stabilising the peptide binding groove, a residue involved in salt bridge formation (D115) and evolutionary conserved peptide binding residues (F119, T139, K142, W143, Y157, Y172). The 72 sequences could be grouped on the basis of motifs at variable positions.

African 'large' barb class I *UA* exon-3 amino acid sequences

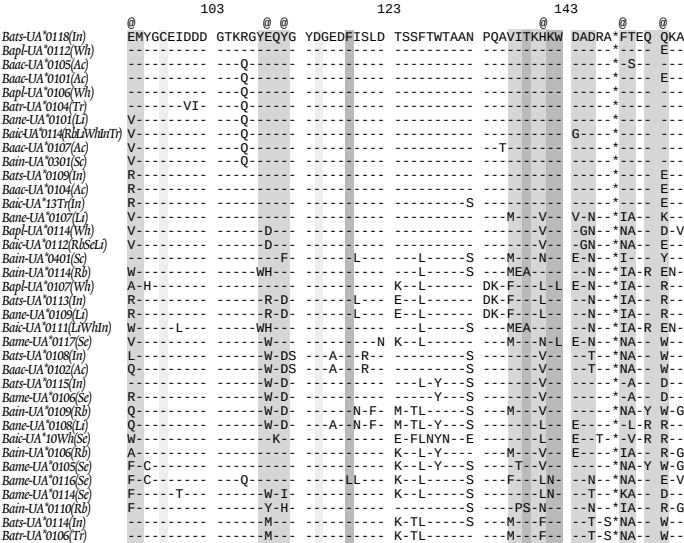


Figure 6: The deduced African 'large' barb class I UA exon-3 amino acid sequences. The *numbers* above denote the amino acid position relative to the mature protein. *Dashes* indicate identity to the *Bats-UA*0109(in)* sequence and *asterisks* indicate gaps. Highly conserved cysteine residues and residues involved in salt-bridges are marked with *light grey boxes*. Conserved amino acid residues important for peptide binding by mammalian class Ia molecules are marked with *deep grey boxes*. Residues known to be involved in peptide binding in mammalian class Ia molecules[7] are indicated by @ below numbering. Putative recombination sites (*table 8*) are marked with *grey boxes*. Previously, the Lake Tana African 'large' barb species were indicated as morphotypes. To facilitate recognition, the morphotype abbreviations (see *fig. 2*) are indicated between brackets following the sequence assignments. *All* indicates all morphotypes studied.

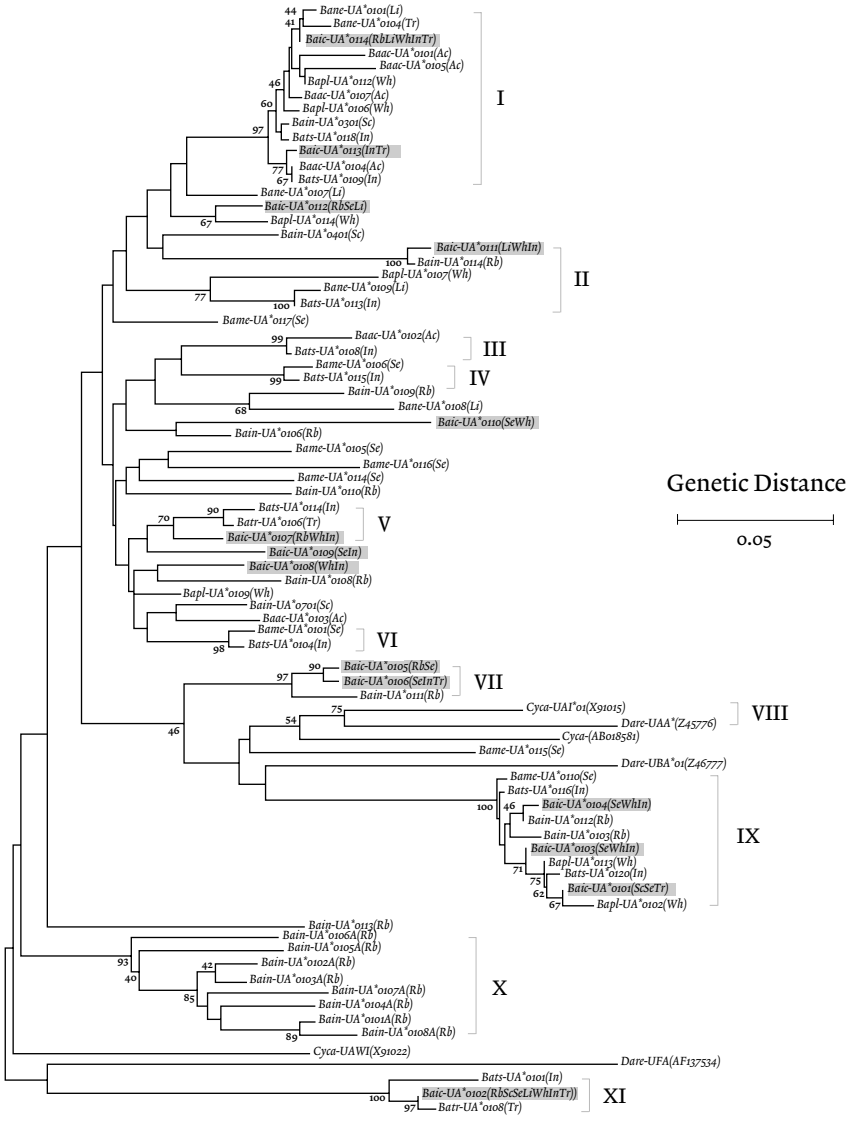


Figure 7: Neighbour-joining tree of African 'large' barb class I *UA* exon-3 nucleotide sequences. Trees were constructed using the method of SAITOU and NEI[161] based on JUKES-KANTOR method in MEGA software. *Numbers* at branch nodes indicate bootstrap confidence levels of 1000 bootstrap replications. Accession numbers of common carp (*Cyca*) and zebrafish (*Dare*) sequences included in the analyses are shown in brackets following the sequences designation. Previously, the Lake Tana African 'large' barb species were indicated by the morphotype nomenclature. To facilitate recognition, the morphotype abbreviations (see *fig. 2*) are indicated between brackets following the sequence assignments. Shared class I *UA* exon-3 sequences are marked by *grey boxes*. *Numbered brackets* denote clusters with bootstrap values over 70%.

Table 4: Presence of shared class 1 sequences in African 'large' barb individuals

Individuals	**Location**	**Baic-UA*:**												
		01	02	03	04	05	06	07	08	09	10	11	12	13	14
B. intermedius River barbs	Blue Nile system														
No 08971	Gibe river														
No 08973	Gibe river														
No 10971	Didessa river		x					x							x
No 30971	Bulbula river		x										x		
No 09973	Loko river		x			x							x		
No 30973	Bulbula river		x												
No 32972	Gilgel Gibe river		x										x		
B. intermedius Shore-complex	Lake Tana														
Flip		x	x												
B. acutirostris Acute Bob	Lake Tana														
B. megastoma Bigmouth small-eye	Lake Tana														
No 16	Gerima	x	x			x				x	x		x		
No 17	Gerima		x	x		x					x		x		
No 74	Gumara river		x			x	x						x		
No 75	Gumara river	x	x		x						x				
No 84	Gumara mouth				x								x		
No 85	Gumara mouth	x			x						x				
B. nedgia Lip	Lake Tana														
No 55	Bahar Dar gulf		x									x	x		
No 87	Bahar Dar gulf											x			x
B. platydorsus White hunch	Lake Tana														
No 22	Kibran		x	x				x				x			
No 23	Kibran		x	x	x			x				x			x
No 24	Kibran		x		x						x				x
No 37	Kibran		x		x						x				
No 65	Ura								x		x				
B. tsanensis Intermedius	Lake Tana														
No 58	Bahar Dar gulf				x		x	x							
No 60	Bahar Dar gulf		x		x			x	x			x			
No 67	Bahar Dar gulf			x								x			
No 76	Bahar Dar gulf		x		x							x			
No 78	Bahar Dar gulf			x	x					x					x
No 438	Bahar Dar gulf			x										x	
No 481	Bahar Dar gulf		x		x										
B. truttiformis Troutlike	Lake Tana														
No 457	Ura	x												x	
No 458	Ura	x	x											x	x
No 471	Gumara river	x	x				x								

Table 5: Presence of species specific class I *UA* sequences in African 'large' barb individuals

Individuals	Sequence number																		
	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20
B. intermedius																				
River barbs																				
No 08971	rba				rba		rba	rba												
No 08973		rba	rba	rba		rba														
No 10971																				
No 30971			rb			rb							rb	rb						
No 09973								rb				rb								
No 30973									rb	rb				rb						
No 32972											rb	rb								
B. intermedius																				
Shore-complex																				
Flip			sc	sc			sc													
B. acutirostris																				
Acute																				
Bob	ac	ac	ac	ac	ac		ac													
B. megastoma																				
Bigmouth small-eye																se				
No 16										se										
No 17	se									se										
No 74					se															
No 75						se														
No 84														se			se			
No 85															se	se				
B. nedgia																				
Lip																				
No 55	li																			
No 87							li	li	li											
B. platydorsus																				
White hunch																				
No 22		wh					wh		wh	wh*										
No 23						wh	wh		wh	wh*										
No 24						wh	wh		wh	wh*										
No 37									wh	wh*										
No 65							wh					wh	wh	wh						
B. tsanensis																				
Intermedius																				
No 58	in			in																in
No 60								in												
No 67									in				in							
No 76														in	in					
No 78																in				
No 438								in					in							
No 481								in								in		in		
B. truttiformis																				
Troutlike																				
No 457								tr												
No 458				tr																
No 471						tr														

* pseudogene

Table 6: Mean ± SE of ratios of nonsynonymous (d_n) per synonymous (d_s) substitutions per site and (d_n - d_s) values of PBR and non- PBR regions within African 'large' barb *UA* alleles and between groups of African 'large' barb *UA* alleles as defined by phylogenetic relationship (Fig. 6)

	PBR			Non- PBR		
	d_n/d_s ± SE	d_n-d_s ± SE	Sig[1]	d_n/d_s ± SE	d_n-d_s ± SE	Sig[1]
all	3.714± 0.146	0.897 ± 0.123	*	2.395 ± 0.098	0.015 ± 0.028	«
I	23.204± 0.121	0.304 ± 0.112	*	1.194 ± 0.009	0.003 ± 0.008	«
II	4.965± 0.327	0.496 ± 0.253	*	1.505 ± 0.032	0.030 ± 0.029	«
III		0.062 ± 0.068	«	0.633 ± 0.024	-0.010 ± 0.019	«
IV		0.133 ± 0.165	«		0.007 ± 0.007	«
V		0.087 ± 0.064	«	1.119 ± 0.019	0.003 ± 0.018	«
VI		0.128 ± 0.159	«		0.007 ± 0.006	«
VII	1.101± 0.264	0.017 ± 0.282	«	0.380 ± 0.018	-0.014 ± 0.018	«
VIII		0.383 ± 0.192	*	0.586 ± 0.070	-0.079 ± 0.072	«
IX		0.078 ± 0.049	«	0.196 ± 0.017	-0.026 ± 0.018	«
X	3.917± 0.206	0.534 ± 0.181	*	0.690 ± 0.026	-0.019 ± 0.023	«
XI		0.221 ± 0.254	«	0.253 ± 0.023	-0.031 ± 0.022	«
I vs II	8.022± 0.214	0.669 ± 0.169	*	1.175 ± 0.036	0.015 ± 0.037	«
I vs III	36.323± 0.535	1.455 ± 0.539	*	1.794 ± 0.034	0.035 ± 0.029	«
I vs IV	23.091± 0.379	0.800 ± 0.372	*	1.301 ± 0.030	0.012 ± 0.031	«
I vs V	15.710± 0.586	1.667 ± 0.552	*	1.580 ± 0.034	0.029 ± 0.034	«
I vs VI	15.000± 0.682	1.644 ± 0.640	*	1.633 ± 0.041	0.036 ± 0.043	«
I vs VII	15.158± 0.444	1.154 ± 0.432	*	1.427 ± 0.051	0.038 ± 0.050	«
I vs VIII	2.116± 0.717	0.804 ± 0.868	«	0.799 ± 0.071	-0.045 ± 0.067	«
I vs IX	3.837± 1.025	1.617 ± 0.732	*	1.036 ± 0.063	0.005 ± 0.064	«
I vs X	6.296± 0.319	0.988 ± 0.269	*	0.524 ± 0.066	-0.100 ± 0.062	«
I vs XI	50.832± 0.525	1.605 ± 0.663	*	0.720 ± 0.085	-0.077 ± 0.081	«
II vs III	6.811± 0.314	0.809 ± 0.238	*	1.135 ± 0.043	0.014 ± 0.040	«
II vs IV	7.330± 0.348	0.853 ± 0.339	*	1.209 ± 0.038	0.017 ± 0.039	«
II vs V	8.299± 0.323	1.358 ± 0.266	*	0.760 ± 0.043	-0.026 ± 0.044	«
II vs VI	7.583± 0.453	1.327 ± 0.478	*	1.020 ± 0.044	0.002 ± 0.047	«
II vs VII	12.939± 0.573	2.043 ± 0.567	*	1.013 ± 0.053	0.002 ± 0.050	«
II vs VIII	1.444± 0.731	0.473 ± 0.704	«	0.767 ± 0.069	-0.056 ± 0.064	«
II vs IX	3.237± 0.897	1.050 ± 0.719	**	0.837 ± 0.064	-0.030 ± 0.064	«
II vs X	5.198± 0.305	1.107 ± 0.262	*	0.688 ± 0.058	-0.057 ± 0.056	«
II vs XI	15.602± 0.273	1.323 ± 0.235	*	0.704 ± 0.085	-0.087 ± 0.080	«
III vs IV	8.286± 0.440	0.435 ± 0.297	**	3.607 ± 0.030	0.056 ± 0.027	*
III vs V	5.623± 0.497	0.604 ± 0.418	**	1.544 ± 0.032	0.025 ± 0.030	«
III vs VI	4.598± 0.507	0.484 ± 0.338	**	2.311 ± 0.036	0.052 ± 0.031	«
III vs VII	36.215± 0.642	1.464 ± 0.651	*	1.479 ± 0.052	0.042 ± 0.047	«
III vs VIII	1.320± 0.741	0.267 ± 0.729	«	0.994 ± 0.063	-0.001 ± 0.057	«
III vs IX	2.792± 0.857	0.986 ± 0.586	*	1.194 ± 0.061	0.027 ± 0.060	«
III vs X	5.688± 0.325	0.984 ± 0.413	*	0.707 ± 0.061	-0.055 ± 0.054	«
III vs XI		0.934 ± 0.521	*	0.782 ± 0.077	-0.051 ± 0.069	«
IV vs V	8.814± 0.847	1.508 ± 0.719	*	2.049 ± 0.032	0.037 ± 0.030	«
IV vs VI	7.412± 0.722	1.277 ± 0.610	*	4.853 ± 0.032	0.065 ± 0.031	*
IV vs VII	22.326± 0.992	2.182 ± 0.954	*	1.598 ± 0.053	0.049 ± 0.053	«
IV vs VIII	4.170± 0.700	0.980 ± 0.796	«	1.101 ± 0.064	0.018 ± 0.063	«
IV vs IX	3.924± 0.757	1.316 ± 0.534	*	1.780 ± 0.058	0.077 ± 0.057	«
IV vs X	4.416± 0.351	0.839 ± 0.419	*	0.735 ± 0.056	-0.042 ± 0.048	«
IV vs XI	33.217± 0.618	1.585 ± 0.597	*	0.798 ± 0.081	-0.050 ± 0.075	«
V vs VI		0.449 ± 0.302	**	2.847 ± 0.023	0.033 ± 0.022	«
V vs VII	2.783± 0.395	0.312 ± 0.453	«	1.188 ± 0.046	0.016 ± 0.042	«
V vs VIII	0.247± 0.813	-1.627 ± 0.634	p	0.895 ± 0.057	-0.018 ± 0.045	«
V vs IX		-0,709 ± 0,622	«	0.790 ± 0.059	-0.031 ± 0.057	«
V vs X	1.241± 0.417	0.116 ± 0.533	«	0.571 ± 0.053	-0.070 ± 0.049	«
V vs XI	2.424± 0.434	0.326 ± 0.487	«	0.609 ± 0.080	-0.103 ± 0.072	«
VI vs VII	5.434± 0.400	0.798 ± 0.494	**	1.147 ± 0.052	0.013 ± 0.051	«
VI vs VIII	0.579± 0.847	-1.063 ± 0.725	«	0.845 ± 0.055	-0.024 ± 0.047	«
VI vs IX	2.900± 1.005	1.281 ± 0.614	*	1.084 ± 0.057	0.010 ± 0.055	«
VI vs X	2.020± 0.465	0.509 ± 0.652	«	0.814 ± 0.053	-0.026 ± 0.049	«
VI vs XI	3.363± 0.546	0.555 ± 0.639	«	0.527 ± 0.078	-0.120 ± 0.066	«
VII vs VIII	0.453± 0.645	-0.589 ± 0.565	«	0.823 ± 0.057	-0.031 ± 0.054	«
VII vs IX	3.718± 1.007	1.567 ± 0.763	*	0.875 ± 0.058	-0.017 ± 0.060	«
VII vs X	2.454± 0.380	0.467 ± 0.380	«	0.796 ± 0.068	-0.041 ± 0.064	«
VII vs XI	4.259± 0.237	0.461 ± 0.233	*	0.905 ± 0.072	-0.019 ± 0.068	«
VIII vs IX	6.414± 1.084	2.224 ± 0.857	*	0.775 ± 0.061	-0.043 ± 0.057	«
VIII vs X	1.909± 0.503	0.475 ± 0.578	«	0.763 ± 0.067	-0.053 ± 0.059	«
VIII vs XI	1.277± 0.677	0.195 ± 0.583	«	0.674 ± 0.084	-0.104 ± 0.075	«
IX vs X	1.544± 0.646	1.045 ± 0.509	*	0.769 ± 0.070	-0.045 ± 0.071	«
IX vs XI		-0,653 ± 0,691	«	0.761 ± 0.081	-0.063 ± 0.078	«
X vs XI	4.872± 0.225	0.661 ± 0.227	*	0.679 ± 0.087	-0.100 ± 0.078	«

1 Z-test for selection H_0: $d_n = d_s$: neutral H_1: $d_n =/= d_s$: « = p ‹ 5%; positive H_1: $d_n \rangle d_s$: * = p ‹ 5%, ** = p ‹ 7,5%

In addition to the 72 African 'large' barb sequences, an identical single pseudogene was detected in four *Barbus platydorsus* individuals (*table 5*). This pseudogene has two in frame stop codons at amino acid position W129 and W143 (see tryptophans *fig. 6*) and was not included in the analyses.

Fourteen out the 72 sequences identified were shared by at least two species (*table 4*). All individuals analysed possessed sequences that were shared between species with exception of three individuals, (BOB, No. 08971 and No. 08973) that did not possess shared sequences. The number of shared sequences per individual varied from one to six. The remaining 48 sequences were species specific and were only shared by individuals of the same species (*table 5*). An individual possessed up to six species specific *UA* sequences, while up to ten different sequences (shared and species specific) could be identified in a single individual (*table 5, table 6*: No 23)

Phylogenetic relationship of African 'large' barb class I UA sequences

Phylogenetic analyses included the African 'large' barb class I *UA* sequences and class I *UA* sequences from zebrafish and common carp (*fig. 7*). The sequences clustered in a *trans*-species manner in eleven sub-clusters supported by bootstrap levels of at least 70%.

Ten clusters contained only African 'large' barb sequences and one cluster comprised a common carp and a zebrafish sequence (cluster VIII). Clustering of sub-clusters into larger clusters was present. Bootstrap levels, however, were below 40%. The African 'large' barb sub-clusters contained sequences from at least two African 'large' barb species with the exception of cluster X. This cluster comprised only sequences from two Blue Nile system individuals (No. 08971 and No. 08973) both collected in the Gibe river (*fig. 1*). All sequences of these individuals clustered together in a single clade, separate from all other clusters.

Selection on African 'large' barb class I UA sequences

Similar to the situation for class II molecules, the amino acid residues of PBRS of the class I (UA) molecules are expected to be under positive Darwinian selection with d_n/d_s ratios higher than 1. Although, residues involved in peptide binding have not yet been identified for fish class I molecules, variability of the African 'large' barb class I UA exon-3 sequences corresponds well to that of the HLA class I molecules (*fig. 8*). We have used the model of SAPER[7] to assign PBR and non-PBR amino acid residues.

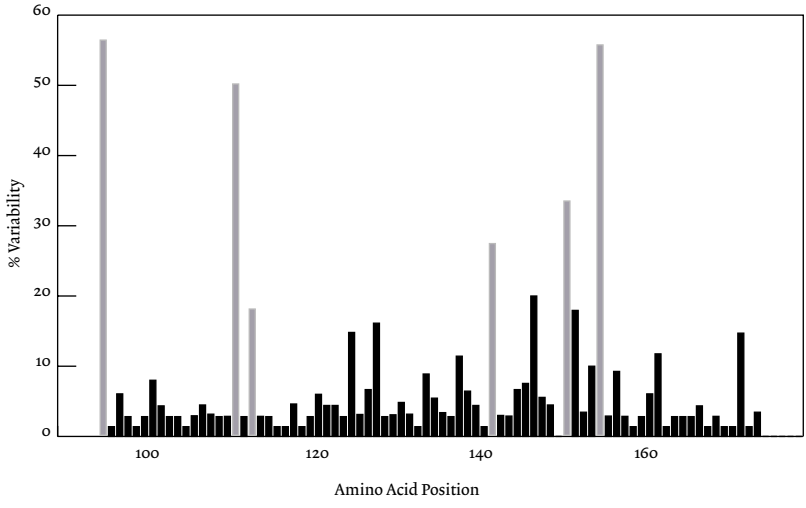


Figure 8: A WU-KABAT variability plot of the deduced amino acid sequences of all African 'large' barb class I UA exon-3 sequences. *Grey bars* indicate amino acid positions known to interact with peptides in human class I molecules[7]. *Numbers* along the X-axis denote the amino acid position relative to the mature protein.

The d_n/d_s ratio's calculated for PBR residues of all UA sequences, some within lineage comparisons (I, II, and X) and several between lineage comparisons, were above 3 (*table 6*). A number of within lineage d_n/d_s ratio's calculations (III-VI, VIII, IX, XI) and some between lineage d_n/d_s ratio's calculations (III vs XI, V vs VI, V vs IX, IX vs XI) were hampered by the lack of synonymous substitutions. Performing a statistical test of the difference between d_n and d_s for PBR residues revealed that d_n values calculated for all UA sequences, some within lineage comparisons (I, II, VIII and X) and several between lineage comparisons, were significantly higher than d_s. All others were subject to neutral selection, or in the case of between lineage comparison, V vs VIII, under purifying selection. The d_n/d_s ratio's calculated for non-PBR residues were below 3 with two exceptions, (III vs IV, IV vs VI). Performing statistical tests revealed that all d_n values were not significantly higher or lower than d_s values.

Table 7: Divergence times estimates based on synonymous substitution (*ds*) in *DAB*01* exon-2 or UA exon-3 sequence or nucleotide substitution (*d*) in *DAB*01* intron-1 sequence

		d_s exon-2	*d* intron-1
	Versus	T ± SE million year	T ± SE million year
*Bain-DAB*0101(Rb)*	[1]*all others*	7,7 ± 3,1	9,4 ± 1,9
*Bain-DAB*0102(Rb)*	*all others*	17,7 ± 3,4	11,7 ± 2,4
*Bain-DAB*0103(Rb)*	*all others*	11,8 ± 4,5	10,9 ± 2,5
*Bain-DAB*0104(Rb)*	*all others*	27,5 ± 4,6	16,4 ± 2,5
*Bain-DAB*0105(Rb)*	*all others*	22,4 ± 3,4	10,6 ± 2,5
*Bain-DAB*0106(Rb)*	*all others*	41,7 ± 4,5	26,2 ± 2,4
*Bain-DAB*0101(Rb)*	*Bats-DAB*0118(In)*	1,0 ± 1,0	13,8 ± 3,6
*Bain-DAB*0102(Rb)*	*Bain-DAB*0113(In)*	7,1 ± 4,6	17,7 ± 4,0
*Bain-DAB*0103(Rb)*	*Bain-DAB*0101(Rb)*	11,8 ± 4,5	10,9 ± 2,5
*Bain-DAB*0103(Rb)*	*Bain-DAB*0101(Rb)*	27,5 ± 4,6	16,4 ± 2,5
*Baac-DAB*0104(Ac)*	*Bane-DAB*0104(Li)*	4,7 ± 2,8	15,7 ± 3,9
*Baac-DAB*0105(Ac)*	*Bane-DAB*0102(Li)*	4,3 ± 3,8	27,0 ± 5,0
*Bame-DAB*0102(Se)*	*Batr-DAB*0103(Tr)*	6,6 ± 4,7	17,6 ± 4,0
*Bats-DAB*0118(In)*	*Bats-DAB*0104(In)*	7,4 ± 5,4	17,7 ± 3,9
*Bame-DAB*0102(Se)*	*Bats-DAB*0111(In)*	19,8 ± 8,8	8,8 ± 2,7
*Basu-DAB*0101(Zu)*	*Batr-DAB*0102(Tr)*	30,1 ± 9,9	17,2 ± 4,1
*Bats-DAB*0106(In)*	*Bama-DAB*0101(Be)*	26,4 ± 9,8	5,5 ± 2,2
*Bats-DAB*0111(In)*	*Batr-DAB*0102(Tr)*	30,9 ± 9,9	8,3 ± 2,7

		d_s exon-3
	Versus	T ± SE million year
*Bain-UA*0101A(Rb)*	[1]*all others*	53,4 ± 6,6
*Bain-UA*0102A(Rb)*	*all others*	40,4 ± 7,7
*Bain-UA*0103A(Rb)*	*all others*	44,1 ± 7,4
*Bain-UA*0104A(Rb)*	*all others*	46,3 ± 6,9
*Bain-UA*0105A(Rb)*	*all others*	36,8 ± 6,9
*Bain-UA*0106A(Rb)*	*all others*	33,7 ± 6,7
*Bain-UA*0107A(Rb)*	*all others*	57,2 ± 8,4
*Bain-UA*0108A(Rb)*	*all others*	63,5 ± 8,1
Cyca-(AB018581)	*all others*	45,2 ± 9,9
*Cyca-UA1*01(X91015)*	*all others*	51,8 ± 8,5
Cyca-UAW1(X91022)	*all others*	51,4 ± 7,1
Dare-UAA(Z45776)	*all others*	54,9 ± 6,8
*Dare-UBA*01(Z46777)*	*all others*	58,2 ± 6,9
Dare-UFA(AF137534)	*all others*	82,3 ± 9,9

1 only barbus sequences

Divergence time estimates

In the phylogenetic analyses of *DAB*01* intron-1 and exon-2 sequences, the six riverine African 'large' barb sequences (*Bain-DAB*0101(rb)* to *Bain-DAB*0106(rb)*) showed to be more distantly related to the Lake Tana African 'large' barb sequences, as reflected in the genetic distance separating these sequences (*fig. 3, fig. 4*). Only one exception is observed, *Bain-DAB*0104(rb)* exon-2 (*fig. 3*) which seems closely related to *Babr-DAB*0102(sh)* exon-2. Divergence time calculations based on synonymous nucleotide substitutions (d_S) in exons and the nucleotide substitutions (d) in introns might provide information on the origin of these sequences.

Using a substitution rate of $2.85*10^{-9}$ per site per year[70], divergence time calculations estimated mean values ranging from 7,7 to 41,7 million year for the six river barb *DAB*01* exon-2 sequences when compared to all Lake Tana African 'large' barb sequences (*table 7*). Similar calculation based on intron-1 sequences revealed mean values ranging from 9,4 to 26,2 million year for the six river barb *DAB*01* intron sequences (*table 7*). Remarkably, in several cases intron and exon values differed substantially. For example divergence time calculations between *Bain-DAB*0106(rb)* and all Lake Tana African 'large' barb *DAB*01* exon-2 sequences suggested that exons diverged on average 41,7 million years ago, while similar calculation based on *DAB*01* intron-1 revealed mean values of 26,2 million years. Similar differences were calculated for *Bain-DAB*0102(rb), Bain-DAB*0104(rb)* and *Bain-DAB*0105(rb)* with exons estimated to be 1.5 times older than introns. This is in contrast with divergence time calculations for *Bain-DAB*0101(rb)* and *Bain-DAB*0103(rb)* for which *DAB*01* exon-2 and intron-1 sequences were estimated to have a similar age (*table 7*).

Such differences were not only observed between comparisons of river barb sequences with Lake Tana African 'large' barb sequences, but also between Lake Tana African 'large' barb sequences (*table 7*). The reverse situation, with exons estimated to be much younger than introns, was also observed between some pairwise comparisons of Lake Tana African 'large' barb *DAB*01* sequences and between some pairwise comparisons of Lake Tana African 'large' barb *DAB*01* sequences and river barb *DAB*01* sequences (*table 7*).

Out of all 1880 pairwise comparisons of African 'large' barb *DAB*01* exon-2 sequences, 86.7% diverged more than 5 million years. The sequences that had

diverged less than 5 million years frequently lacked synonymous substitutions, although they did possess nonsynonymous substitutions (data not shown). Most pairwise comparisons that lacked synonymous substitutions are within phylogenetic lineage comparisons. Calculations based on *DAB*01* intron-1 rates revealed that 96% of intron-1 sequences were estimated to be older than 5 million years.

Divergence time estimates based on synonymous substitutions in exon-3 of African 'large' barb class I *UA* sequences, using a rate of $2.85*10^{-9}$ substitutions per site per year[70], yielded invariably divergence time estimates above 5 million year for all pairwise comparisons. Only a minority (5.8%) of all pairwise comparisons yielded divergence time estimates below 5 million year. Almost half of this minority was hampered by a lack of synonymous substitutions. Although these sequences lacked synonymous substitutions, they all possessed nonsynonymous substitutions (data not shown). Most pairwise comparisons that lacked synonymous substitutions are within phylogenetic lineage comparisons.

Eight sequences unique to two individuals, collected in the Gibe river, clustered together in the phylogenetic analyses. Their position seems to suggest that these sequences represent an ancient lineage. Divergence time estimates using a rate of $2.85*10^{-9}$ substitutions per site per year[70] revealed values between 33,7 and 63,5 million years. The common carp class I *UA* sequences diverged from the African 'large' barb sequences between 45,2 and 51,8 million years.

These estimates are higher than the estimated divergence time of 30 million years separating the genera *Cyprinus* and *Barbus* as determined by others[69-71]. The zebrafish class I sequences diverged between 51,4 and 82,3 million year from the African 'large' barb sequences, which is also longer than the estimated divergence time separating the genera *Danio* and *Barbus*[69-71].

Recombination

In the case of pairwise comparisons per individual sequence, significantly higher d_S rates in exon-2 of *Bats-DAB*0111(in), Bane-DAB*0101(li), Bats-DAB*0101(be), Bama-DAB*0104(rb), -05(rb)* and *-06(rb)* with respect to intron-1 (*fig. 5D*) strongly suggest frequent recombination events particularly close to the exon-intron boundary. The algorithm of GRASSLY and HOLMES[196] (PLATO) identifies such putative tracts involved in recombination that influence reconstruction of phylogeny.

Table 8: Putative regions of recombination present in nucleotide sequences of class II *DAB*01* intron 1 and exon 2 and class I UA exon 3 sequences with corresponding amino acid positions referring to *fig. 1* and *fig. 5*, respectively

class II *DAB*01*

	HKY				REV				Amino acid position	
	both		separate		both		separate		both	
	Start	End	Start	End	Start	End	Start	End	Start	End
intron-1	7	11			7	11				
	60	75	60	65	60	75	60	75		
	145	152	145	152	146	150	145	149		
	204	208	204	208	204	208	204	208		
exon-2	256	272	256	260	256	269	256	270	3	8
			262	269						
	310	318	310	318	310	318	310	318	21	24
	370	374	370	375	370	374	370	374	41	42
	389	393	389	393	389	393	389	467	47	49
	412	417			411	417			55	57
	431	435			431	435			61	63
	442	446			442	446			65	66
	454	458			454	458			69	70

class I *UA*

	HKY		REV		Amino acid position	
	Start	End	Start	End	Start	End
exon-3	1	5	1	5	94	95
	48	56	48	56	109	112
	130	144	130	144	137	141
	153	158			144	146
	169	173	169	173	150	151
	178	182	178	183	153	154
	200	204	200	204	160	161

Running PLATO using the model of HASEGAWA and co-workers[199] (HKY) or markov general reversible model (REV) of nucleotide substitutions, in the analyses with default settings, predicted 12 tracts, comparable for both methods, when the maximum likelihood phylogeny was based on the sequence spanning DAB*01 intron-1 and exon-2 (*table 8*). Using a similar approach, but now with the maximum likelihood phylogeny based on class II *DAB*01* intron or exon sequences separately, predicted comparable tracts. Some of the predicted recombination tracts were located close to exon-intron boundaries. Furthermore, the predicted tracts in *DAB*01* exon-2 were all located in the variable regions (*fig. 2*).

Similar analyses based on class I *UA* exon-3 sequences using the HKY method predicted seven putative tracts involved in recombination and six when using the REV method. These predicated tracts were all located in the variable regions (*fig. 6*).

DISCUSSION

DIXON and co-workers[70] observed that each of the four Lake Tana African 'large' barb species under study possessed their own set of species specific class II *DAB*01* alleles. These observations support the proposed species hypothesis of NAGELKERKE and SIBBING[75,132] that was postulated on cumulating evidence like the presence of 15 distinct morphotypes that also differ considerably in their ecology and spawning patterns. In this study we found compelling evidence that indeed each Lake Tana African 'large' barb species seems to possess a specific set of class II *DAB* alleles. Comparison of class II *DAB*01* sequences from the nine Lake Tana African 'large' barb species and Lake Tana *Barbus intermedius* and their riverine counterparts, *Barbus intermedius*, showed that no class IIB sequences were shared between species but were shared within species. In striking contrast, a similar analyses of class I *UA* exon-3 sequences revealed, beside species specific sequences, that the six Lake Tana African 'large' barb species studied possessed identical class I *UA* exon-3 sequences. In addition, class I sequences were also shared between Lake Tana African 'large' barb species and Blue Nile system barbs.

*African 'large' barb class II DAB*01 genes*

It is hypothysed that an ancestral *Barbus intermedius* population present in the Blue Nile system and in Lake Tana gave rise to the present-day Lake Tana African 'large' barb species flock. During 5 million years of isolation of the lake, individuals of an ancestral population adapted to different ecological niches in the lake. As a result, they underwent speciation resulting in 15 novel African 'large' barb species. During such speciation events, selection pressure on peptide binding residues of major histocompatibility (MHC) class I and class II molecules is expected to be extreme since organisms move to new environments and as a consequence their MHC molecules encounter new pathogens[39].

Dixon and co-workers[70] provided evidence that the Lake Tana African 'large' barb class II *DAB* sequences were subject to selection. Analyses of nonsynonymous and synonymous nucleotide substitutions rates calculated for within and between phylogenetic lineages revealed that d_n/d_s ratio's in PBR codons were usually higher than d_n/d_s ratio's in non-PBR codons. Values calculated for PBR codons were mostly above 3, indicating strong positive Darwinian selection.

In addition, the four Lake Tana African 'large' barb species (*B. acutirostris, B. nedgia, B. tsanensis, B. truttiformis*) studied possessed their own set of species specific class II *DAB*01* sequences. Several pairwise comparisons of species specific sequences yielded divergence time estimates within the age of the lake, indicating that these sequences diverged after isolation of the lake. Our study on class II *DAB*01* sequences isolated from six additional Lake Tana African 'large' barb species (*B. brevicephalus, B. macrophtalmus, B. megastoma, B. platydorsus, B. surkis, B. intermedius*) strongly confirmed their observations and thorough discussion[70].

In this study we were able to include class II *DAB*01* intron-1 and exon-2 sequences obtained from Blue Nile system barbs. Divergence time estimates based on synonymous substitution rates in class II *DAB* exon-2 sequences indicated that the riverine sequences predated those obtained from the Lake Tana. Phylogenetic analyses (*fig.3, fig 4*) suggested that one of the riverine sequences (*Bain-DAB*0106(rb)*) emerged before the separation of the genera *Cyprinus* and *Barbus*, 30 million years ago[69-71]. The maximum divergence time estimate of approximately 42 million years supports this observation (*table 7*).

Comparisons of pairwise divergence time estimates of *DAB*01* intron-1 and exon-2 revealed remarkable differences. Obviously, differences in divergence time estimates are reflected in nucleotide substitutions (d) in *DAB*01* intron-1 and synonymous substitution rates (d_s) in *DAB*01* exon-2. In most genes, d_s in exons and d in introns are essentially identical[200]. However, higher d_s rates in exons are hypothesised to occur as a consequence of recombination and subsequent genetic drift of a gene on which balancing selection acts to maintain polymorphism in an exon[197]. The putative recombination tracts located at exon-intron boundaries of the isolated African 'large' barb *DAB*01* sequences and within PBRs of exon-2 suggests that recombination played a role in creating the observed differences in substitution rates. Clearly, *DAB*01* intron-exon shuffling took place as observed in comparisons of the phylogenies of intron and exon sequences separately (*fig.3, fig. 4*).

Such homogenisation of introns is suggested to occur within allelic lineages[197,201]. In the analyses performed by HUGHES[197], comparisons of African 'large' barb *DAB*01* intron-1 and exon-2 sequences revealed that they evolved at similar rates, while between genera (*Barbus and Cyprinus*) comparisons revealed that intron-1 and exon-2 sequences evolved at different rates. The African 'large' barb *DAB*01* may represent unique allelic lineages and thus only within lineage comparisons may reveal a correlation between d_S in exons and d in introns with higher d_S rates than d rates. However, plots of d in introns versus d_S in exons for within and between pairwise comparisons of phylogenetic lineages did not reveal significantly different substitution rates in exons and introns (*fig. 5A*, *fig. 5B*). Such differences between substitution rates in exons and introns were observed for pairwise comparisons of some relatively older sequences (*fig. 5C*, *fig. 5D*) comparable to the between genera (*Barbus and Cyprinus*) comparisons presented by HUGHES[197].

It should be taken into account that analyses of substitution rates may be hampered by the fact that the sequences used were amplified from genomic DNA. This does not allow differentiation between functional and pseudogenes. In functional genes polymorphism would be concentrated in peptide binding regions, while sequence variation in pseudogenes is expected to occur at random[36]. Non-functional MHC class II sequences might be present among those isolated. However, with only one exception, the sequences amplified were not pseudogenes, although it can not be excluded that other parts of these genes may have in-frame stops due to point mutations, deletions/insertions, or complete deletion of exons rendering them pseudogenes.

African 'large' barb class I UA genes

All of the 72 unique putative class I *UA* exon-3 sequences encoded *bona fide* MHC class I UA alpha-2 domains. Remarkably, unlike class II *DAB*01*, class I *UA* sequences were shared between Lake Tana African 'large' barb species. Fourteen out of 72 genomic class I *UA* exon-3 sequences were shared by at least two African 'large' barb species, while the remaining 58 were species specific. However, the sequences designated as species specific in this study might not be species specific, but may turn out to be shared when additional African 'large' barb individuals would be analysed. In addition, it should be taken into account that

sharing of sequences is based on exon-3 sequences only. The complete *UA* sequences may differ in exon-1, which in general is the most divergent exon. However, sharing of exon-3 sequences was not only observed between Lake Tana African 'large' barb species, but also between Lake Tana species and African 'large' barbs from the Blue Nile system.

Five out of fourteen shared sequences were also present in at least one African 'large' barb individual from the Blue Nile system, strongly suggesting maintenance of class I sequences for more than 5 million years. Such sharing of alleles between closely related species is rarely seen. Comparisons of MHC class I genes between closely related teleostean or mammalian species[130,202-207] did not reveal sharing of alleles, except for two cases. COOPER and co-workers[208] isolated an MHC class I sequence shared between two species of chimpanzee, while EVANS and co-workers[209] identified an identical expressed class I allele in two new world primate species.

The total number of class I *UA* sequences, shared plus specific, isolated from a single African 'large' barb individual varied from three to ten. Since African 'large' barbs are considered to be hexaploid[68] isolation of more than six sequences per individual may reflect isolation of different loci with the primer set used or the presence of duplicated genes of a single locus. Individuals possessing fewer than six sequences may be homozygous for the loci examined. It is also possible that not all sequences were amplified by the primer set used. Alternatively, the class I *UA* locus may show haplotypic variation as observed in other bony fishes[176-178]. The presence of multiple class I *UA* sequences in the Lake Tana African 'large' barb population, phylogenetic clustering of these sequences in eleven sub-groups, including sequences of some species but usually not all species (*fig. 7*), may indicate such haplotypic variation.

In the phylogenetic analyses (*fig. 7*) all African 'large' barb clusters, except cluster XI, comprised sequences of multiple African 'large' barb species indicating that these sequences have evolved in a *trans*-species manner as reported for other teleosts[176] and mammals[42]. These clusters usually represent allelic lineages present in a common ancestral population, except in some cases in which severe recent bottlenecks have occurred[136]. A recent study revealed that highly divergent class I *UA* sequences in rainbow trout were derived from a single locus[182]. In this study, sequences within a cluster, in general, seem to be more related as short genetic distances separated them. Sequences outside these

clusters, on the other hand, seem to have relatively longer branch lengths. These included sequences from riverine barbs that were collected in the Blue Nile system. This may indicate that some class I lineages have been maintained in several African 'large' barb species, most likely by a process of natural selection.

Under positive Darwinian selection, the rate of nonsynonymous substitutions per nonsynonymous (d_n) site will be higher than the rate of synonymous substitutions per synonymous site (d_s) with d_n/d_s ratios exceeding a value of one. However, in the case of positive Darwinian selection on PBRS this ratio is expected to be at least three[198]. Average values of all pairwise comparisons indicated that positive selection acted upon PBRS while non-PBRS were subject to neutral selection (*table 6*). Several within lineage comparisons were hampered by a lack of synonymous substitutions, although nonsynonymous substitutions were present. This indicates that selection pressure seems to be in favour of nonsynonymous substitutions. The lack of significant d_n-d_s values for positive selection on PBRS of within lineage comparisons might be the result of the limited number of sequences in most clusters. In general, positive selection seems to diversify PBRS between phylogenetic lineages, while non-PBRS are maintained by neutral selection. Variability plots clearly indicated that most variation was located in PBRS (*fig. 8*). The generality of this conclusion is supported by analyses of the divergent phylogenetic clusters I to XI (*table 6*). Non-PBRS seem to be under neutral selection while calculations for PBRS were hampered by the lack of synonymous substitutions which suggest positive selection on PBRS.

The Lake Tana African 'large' barbs species population is estimated to be isolated from their riverine counterparts in Blue Nile system in the order of 5 million years. Within this time span, class I alleles present in the ancestral African 'large' barb population are expected to have diversified. Pairwise comparison of all sequences using a synonymous substitution rate of $2.85^{*}10^{-9}$ substitutions per site per year[70] yielded invariably divergence time estimates above 5 million year. This indicated that most sequences were already diversifying within the ancestral African 'large' barb population before the isolation of Lake Tana. Only a minority of 5.8% of all pairwise comparisons yielded divergence time estimates below 5 million year of which almost half were hampered by a lack of synonymous substitutions. Although these sequences lacked synonymous substitutions, they all possessed nonsynonymous substitutions. Interestingly,

these sequences belonged to a single phylogenetic cluster (*fig. 7,* I), with only a few exceptions (*Baic-UA*07(rbwhin)* vs *Bame-UA*0101(se), Bapl-AU*0109(wh)* or *Bats-UA*01049(in); Bain-UA*0106(rb)* vs *Batr-UA*0106(tr); Bapl-UA*0109(wh)* vs *Bats-UA*0104(in)* or *Bame-UA*0101(se)*). Furthermore, most clusters comprise sequences that are also present in riverine barbs.

Data presented in this study clearly indicate that phylogenetic lineages are maintained in multiple species, most likely by purifying selection. Diversification seems to be located in PBRS since nonsynonymous substitutions are mainly restricted to these regions. Diversification of PBRS is likely to be a result of positive Darwinian selection[36]. Recombination is another mechanism that is hypothesised to play a role in diversifying MHC sequences. Inter- and intralocus recombination in mammals has been suggested to be the driving force for the generation of new class I sequences[129] and new allelic lineages[30,31,33,205].

Several observations indicate that class I *UA* sequences are also subject to a mechanism of recombination. First, several putative recombination tracts were predicted in class I *UA* exon-3 sequences[196] (*fig. 1, table 8*). These predicted recombination tracts were all located in the PBRS of the class I *UA* exon-3 sequences suggesting that diversity is also generated in PBRS as a result of recombination.

Secondly, analyses of the deduced amino acid clearly showed sharing of motifs between African 'large' barb class I *UA* exon-3 clusters while the remaining part of the sequences has diverged. For example cluster VII shares the motif DST (aa 144-146) with cluster XI, while it also shares the motif YMQY with cluster V. Both motifs are located in putative recombination tracts which are located in PBRS. This suggests exchange of motifs by non-reciprocal recombination, although, convergent evolution might be an alternative explanation. Third, some sequences have overlapping regions with other sequences. For example, Bain-UA*0109(rb) is similar to Bane-UA*0108(li) from aa 94 to 136, while the remaining part of the sequence is similar to cluster VI sequences, suggesting a recombination event.

Similarities and differences between African 'large' barb class I *UA and class* II *DAB*

The data presented in this study suggests that diversity of the Lake Tana African 'large' barb class I and class II sequences is generated by positive Darwinian selection on PBRS. Alleles within an allelic lineages seem to be maintained by purifying selection or neutral selection on non-PBRS, whereas sequen-

ces belonging to different lineages have been under positive selection. Diversification of class I and class II sequences is not only the result of point mutations, but also of recombination events.

The most striking observation is that the Lake Tana African 'large' barb species investigated in this study showed no sharing of class II *DAB*01* sequences while they do share class I *UA* sequences. Similar, the riverine African 'large' barbs studied did not share class II *DAB*01* sequences with the Lake Tana African 'large' barbs while they do share the same class I *UA* sequences. These data indicate that class I and class II sequences were subject to different evolutionary pressures. This resulted in maintenance of class I sequences that were present in the African 'large' barb population before the isolation of the lake. Class II sequences, on the other hand, seem to have diversified within a time span of 5 million years. This is in contrast to observations in mammals in which MHC class I genes seem to be more divergent and more rapidly evolving than class II genes. HLA class I lineages are only recognised in great apes and thus maintained up to 6 million years, while certain HLA class II lineages were recognised in prosimians that diverged from human approximately 35 million years ago[43,45].

In mammals MHC genes are found in a complex and haplotypes of class I and class II alleles show a confounding pattern of linkage. In contrast, MHC class I and class II genes in teleosts[191] are found in different linkage groups. The functional class I locus in fish is closely linked to *LMP* and *TAP* (rainbow trout[53], zebrafish[210], Atlantic salmon[52]), whereas class II loci can be found on several linkage groups, either as functional genes or as pseudogenes[61,89]. This seems to suggest that the birth-and-death process as proposed by KLEIN[128] is valid for teleost class II genes but not class I genes. The turn-over of class II alleles in the Lake Tana 'large' barb species is much higher than that of class I alleles. Although it is tentative to speculate, the tight linkage of the class I genes with genes encoding proteins involved in generating and loading of peptides may constrain the evolution of class I genes. Alternatively, class I molecules might play a less important role in the immune system of fish compared to the class II molecules. Hence, class II molecules are probably more prone to evolutionary changes compared to class I molecules.

ACKNOWLEDGMENTS

The authors would like to thank Martin de Graaf, Leo Nagelkerke and Nand Sibbing for providing tissue samples of Lake Tana African 'large' barb species flock individuals and Darrell Siebert and Elizabeth Barratt for providing tissue samples of riverine African 'large' barb individuals.

CHAPTER 5

A NOVEL FUNCTIONAL CLASS I LINEAGE IN ZEBRAFISH (*DANIO RERIO*), COMMON CARP (*CYPRINUS CARPIO*) AND AFRICAN 'LARGE' BARB (*BARBUS INTERMEDIUS*), SHOWING AN UNUSUAL CONSERVATION OF THE PEPTIDE BINDING DOMAINS.

Corine P. Kruiswijk*, Trudi Hermsen*, Adrie H. Westphal**, Huub F.J. Savelkoul*, René J.M. Stet*

*Cell Biology and Immunology Group,
Department of Animal Sciences, Wageningen University,
PO box 338, 6700 AH Wageningen, The Netherlands

**Laboratory of Biochemistry, Department of Agrotechnology
and Food Sciences, Wageningen University,
PO box 8128, 6700 ET Wageningen, The Netherlands

Published in: The Journal of Immunology 169 (4) p 1936-1947 (2002)
Printed with kind permission of the American Association of Immunologists

---

ABSTRACT

Species from all major jawed vertebrate taxa posses linked polymorphic class I and II genes located in an MHC. The bony fish are exceptional with class I and II genes located on different linkage groups. Zebrafish (*Danio rerio*), common carp (*Cyprinus carpio*) and African 'large' barb (*Barbus intermedius*) represent highly divergent cyprinid genera. The genera *Danio* and *Cyprinus* diverged 50 million years ago while *Cyprinus* and *Barbus* separated 30 million years ago. In this study, we report the first complete protein-coding class I *ZE* lineage cDNA sequences with high similarity between the three cyprinid species. Two unique complete protein-coding cDNA sequences were isolated in zebrafish, *Dare-ZE*0101* and *Dare-ZE*0102*, one in common carp, *Cyca-ZE*0101* and six in African 'large' barb, *Bain-ZE*0101, Bain-ZE*0102, Bain-ZE*0201, Bain-ZE*0301, Bain-ZE*0401* and

*Bain-ZE*0402*. Deduced amino acid sequences indicate that these sequences encode *bona fide* class I proteins. In addition, the presence of conserved potential peptide anchoring residues, exon-intron organisation, ubiquitous expression and polymorphism generated by positive selection on putative peptide binding residues support a classical nature of class I ZE lineage genes. Phylogenetic analyses revealed clustering of the ze lineage clade with non-classical cyprinid class I Z lineage clade away from classical cyprinid class I genes, suggesting a common ancestor of these non-classical genes as observed for mammalian class I genes. Data strongly support the classical nature of these ZE lineage genes that evolved in a *trans*-species fashion with lineages being maintained for up tot 100 million years as estimated by divergence time calculations.

INTRODUCTION

Cartilaginous fish are the most ancient group of vertebrate species possessing major histocompatibility complex (MHC) genes. In all species studied to date, the MHC is a large chromosomal region, containing many genes encoding proteins of immunological importance[47,54,211] with exception of the bony fish that represent about half of all vertebrate species[162]. Unlike all other jawed vertebrates, bony fish class I and class II genes are located on different linkage groups[48,53,191,192]

MHC class I and II genes encode structurally similar proteins that present peptides to T lymphocytes. The class I genes can be subdivided into classical class I and non-classical class I molecules based on structural and functional differences and expression patterns[212]. The MHC classical class I genes are involved in antigen presentation, presenting endogenous derived peptides to CD8 positive T-cells. They have been shown to be highly polymorphic and co-dominantly expressed on cells in almost all tissues. Class I molecules are composed of a large alpha-chain, encoded in the MHC, non-covalently associated with a much smaller β_2-microglobulin (β_2m) molecule, encoded outside the MHC. The class I alpha chain consists of three extra-cellular domains with two membrane-distal domains that form the peptide binding region. Polymorphic residues within this peptide binding region interact with peptides and are under positive Darwinian selection[35]. MHC polymorphism evolves in a *trans*-species fashion[44]. In general MHC class I genes seem to be more divergent and more rapidly evolving than class II genes. HLA class I lineages are only recognised

in great apes and thus maintained up to 6 million years, while certain HLA class II lineages were recognised in prosimians that diverged from human approximately 35 million years[43,45].

The MHC non-classical class I genes encode molecules with a typical MHC class I structure but do not have the function and tissue distribution of the classical genes. They exhibit low polymorphism, are often expressed in a tissue specific fashion and are encoded either in the MHC or outside this complex[212]. Functions of non-classical class I molecules like CD1, HLA-E, HFE, MICA and MICB are now emerging and the presence and conservation of non-classical molecules among species underline the importance of their roles (reviewed[22]). CD1 molecules, an extensively studied group of non-MHC-encoded genes, were shown to present lipid structures and therefore play an important role in defence against bacterial infections. The MHC-encoded non-classical HLA-E molecule modulates NK cell function by presenting peptides derived from classical class I leader sequences, while the MHC-encoded HFE molecule plays a role in iron metabolism and does not bind peptides. Unlike CD1, HLA-E, and HFE, the MICA molecule does not associate with β_2m.

To date, non-classical class I Z lineage genes have only been identified in the teleost (bony fish) species[6,106,107], ginbuna crucian carp (*Carassius auratus langsdorfii*) and common carp (*Cyprinus carpio*). Despite extensive searches for Z lineage sequences in genomic DNA and cDNA libraries of the cyprinid species zebrafish (*Danio rerio*), no evidence was obtained for the existence of such genes[48,86,213].

Previously, Southern blot hybridisation performed on restriction-enzyme digested high molecular weight DNA of common carp of different geographical origins, using a probe to class I Z exon-4 detected 9-12 hybridising fragments at extremely low stringency[81]. These data suggested the existence of additional class I Z sequences in common carp. A recent attempt to identify novel class I sequences in common carp revealed the partial coding sequence of the extracellular domains of new class I Z lineage sequences[74], *Cyca-Zr2* and *Cyca-Zr3* . In phylogenetic analyses, these sequences clustered with other cyprinid class I Z lineage sequences, but formed a separate clade and are therefore renamed to *Cyca-ZE*0101* and *Cyca-ZE*0201*. That study also revealed two other unique class I Z lineage sequences, *Cyca-Zr1* and *Cyca-Zr4*. However, these sequences formed a clade[106] with *Cyca-ZB* and therefore renamed to *Cyca-ZB*0201* and *Cyca-ZB*0301*. In this study, we identified the complete coding sequence of *Cyca-ZE*0101*. Further,

we analysed the presence of these class I ZE molecules in zebrafish[66] (2n=50) and African 'large' barb (*Barbus intermedius*[68]; 2n=150), representing highly divergent cyprinid genera that separated from common carp[67] (2n=100) 50 and 30 million years ago[69-71], respectively. Analyses of the complete protein-coding cyprinid sequences indicated a more classical nature of the *ZE* lineage genes. Therefore, we investigated their expression, intron-exon organisation and the characteristics of polymorphic residues for peptide binding and evidence for positive selection.

MATERIALS AND METHODS

Fish

A λzap cDNA library (Cyca-λzap) prepared from PMA-stimulated phagocytes from *Cyprinus carpio* was available[214] to characterise the full-length cDNA sequences. Genomic DNA was extracted from a blood sample of a *Cyprinus carpio* R3 x R8 F1 hybrid individual[215] and total RNA was extracted from thymus, head-kidney, spleen, kidney and intestine samples to study gene expression.

A liver sample from a Lake Tana *Barbus intermedius* individual was collected to extract genomic DNA and total RNA for identification of class I Z sequences. Samples of muscle, kidney, liver and thymus from a Lake Tana *Barbus acutirostris* individual were collected for extraction of total RNA to study gene expression.

A muscle tissue sample of one *Danio rerio* wild type individual (Tübingen, Germany; hatched according to standard procedures[216]) was collected for total RNA extraction to identify expressed class I Z sequences. In addition, muscle samples of four *Danio rerio* F1 hybrid individuals (Dimamma, Brakel, Netherlands) were collected to extract genomic DNA to study positive selection.

Genomic DNA and total RNA extraction

Genomic DNA was isolated using a Wizard® genomic DNA purification kit (Promega, Madison, WI, USA) according to the protocol provided. Total RNA extraction from tissue samples was performed according to the protocol described by DIXON *et al.*[70]. DNA and RNA concentrations were determined using the GeneQuant system (Amersham Pharmacia Biotech, Roosendaal, Netherlands).

PCR and Expand™ long template PCR conditions

Standard PCR reaction conditions were 1 x reaction buffer, 1.5 mM $MgCl_2$, 1 unit of Taq polymerase (Goldstar; Eurogentec, Seraing, Belgium), 0.2 mM dNTPs,

Table 1: Primers used for amplification of genomic DNA, cDNA and gene expression studies

Amplification of genomic DNA and cDNA					
Primer		Nucleotide sequence (5'-3')			
A		TGTGTGGACTGGCTCAACAAAT			
B		ACTTCTGGGGTAGAAACCAGT			
C		ATGCGAACAAGCGTCACGAT			
D		CATTTCTTCTGCTCTGTGGTTGT			
E		AGACATGATGACCATTAGGA			
F		GACCACGAGTTTTGCTGTGCTT			
G		AATGGCTGTTTTTGCCGTTTTGTTCTC			
H		AGTTTTGCTCTTGGTGTTTTTATTCTC			
I		CCATCTTCAGTCACTACTTTTCCTGCTTC			
J		TCCCTGTATTACATTTACAC			
Gene expression					
Gene	Acc. no	Primer sequence (5'-3')		Bp	Location
		Cyprinus carpio			
		AGACATCAGGGTGCTATGGTTGGT	s[a]		
ß-actin	M24113	GATACCGCAAGACTCCATACCCA	a[b]	648	
		GAACACACTCTCTGAGATAC	s		alpha-1
*Cyca-UA1*01*	AJ007901	GAAGCGTTCCTTTGCTTTCTG	a	282	alpha-2
		GATGGAACAGAGAGTCTGAGAGC	s		alpha-2
*Cyca-ZB*0201*	AJ007848	ATCTGGTGGAGTCTTTTGTG	a	288	alpha 2-3-border
		GTTGTGAAATTGAGAAAAATGGAG	s		alpha 2
*Cyca-ZE*0101*	AJ007849	GCTTGGTTTTGTCCCTGGTGC	a	295	alpha 3
		Barbus intermedius			
		AGACATCAGGGTGCTATGGTTGGT	s		
ß-actin	M24113	GATACCGCAAGACTCCATACCCA	a	648	
		TGTGTGGACTGGCTCAACAAAT	s		alpha-2
*Bain-ZE*0401*		GTTCCCGCTTCATCATCAGGC	a	219	alpha-3
*Bain-ZE*0101*		AGTTTTGCTGTGCTTGGTGTTTTTATTCTC	s		leader
*Bain-ZE*0102*		CCACACACTCTTTCTCCAGGTAG	a	567	alpha-2

[a] sense; [b] antisense

0.2 μM of each primer and 5 μl phage suspension or 100 ng genomic DNA. The cycling profile was 1 cycle at 94°C for 5 min followed by 30 cycles consisting of denaturing at 94°C for 30 sec, annealing at 50°C or 55°C for 30 sec, polymerisation at 72°C for 1 min, and a final cycle of 10 min at 72°C. Expand™ long template PCR was performed according to the protocol described for amplification of cDNA (Boehringer Mannheim, Ingelheim, Germany). The standard and Expand™ long template PCR was performed using a GeneAmp PCR system 9700 (PE Applied Biosystems, Foster City, CA, USA).

Amplification of expressed class I ZE genes in Cyprinus carpio

To amplify the missing 3' end of class I ZE sequences, anchored PCR was performed on a common carp Cyca-λzap library. A lambda specific anti-sense primer was used (T7) in combination with a class I ZE lineage specific sense primer A matching the end of exon-3 of known cyprinid class I ZE sequences (*table 1*).

Amplification of expressed class I ZE genes in Barbus intermedius and Danio rerio

Expressed class I ZE lineage sequences in African 'large' barb and zebrafish were isolated using the Generacer™ kit for full-length, RNA ligase-mediated rapid amplification 5' and 3' ends (Invitrogen, Carlsbad, CA, USA). Full-length African 'large' barb liver cDNA and zebrafish muscle cDNA was synthesised according to the protocol described. The 5' ends were amplified by PCR using a anti-sense primer B (*table 1*) based on a conserved part of exon-4 of several common carp and ginbuna crucian carp class I Z nucleotide sequences in combination with the Generacer™ 5' primer. The 3' ends of African 'large' barb and zebrafish class I ZE sequences were amplified by Expand™ long template PCR using the generacer™ 3' primer in combination with a sense primer C, D, E, F (African 'large' barb) or G (zebrafish) matching exon-1 of the amplified 5' end sequences (*table 1*).

Amplification of genomic cyprinid class I ZE genes

To amplify common carp genomic class I ZE sequences, Expand™ long template PCR was performed using sense primer H matching exon-1 of *Cyca-ZE*0101* in combination with a specific anti-sense primer I matching the 3'-untranslated region of this sequence (*table 1*).

Class I *ZE* sequences from zebrafish genomic DNA were amplified using a sense primer J matching amino acid 7 to 13 of the alpha-1 domain and anti-sense primer matching amino acid 163 to 171 of the alpha-2 domain (*table 1*).

Cloning and DNA sequencing

PCR products were ligated and cloned using the PGEM® T-easy kit (Promega, Madison, WI, USA) following the manufacturer's description. Plasmid DNA was isolated from cells using the QIAprep spin miniprep kit (QIAGEN, Valencia, CA, USA) according to the protocol provided. Subsequently, plasmid DNA was sequenced using the ABI prism bigdye™ terminator cycle sequencing ready reaction kit and analysed using an ABI 377 sequencer (PE Applied Biosystems, Foster City, CA, USA).

Accession numbers and nomenclature

The new sequences reported here were deposited in the EMBL database under the following accession numbers: *Bain-ZE*0101-0501*: AJ420274-AJ420284; *Cyca-ZE*0101*: AJ420951, AJ420952, AJ420957, AJ420958; *Dare-ZE*0101-Dare-ZE*1401*: AJ420953-AJ420956 and AJ420975-AJ420984.

The nomenclature used to assign new sequences or rename existing sequences adheres to the recommendations described in the HLA facts book[165]. Based on phylogenetic clustering of the identified sequence in a separate clade together with known common carp *Cyca-ZA, -ZB, -ZC* and goldfish *Caau-ZD*, the locus designation *ZE* was given. An asterisk and four digits follow the locus name. The first two digits describe the lineage and the third and fourth digits that follow describe alleles.

Expression of Z lineage genes in Barbus intermedius and Cyprinus carpio

To prepare cDNA of several tissues of common carp and African 'large' barb, 20 µg total RNAwas reverse transcribed using the Universal riboclone cDNA synthesis system (Promega, Madison, WI, USA) according to the protocol. Subsequently, 2.5 µl cDNA was used to study gene expression by PCR using gene specific sense and anti-sense primers (*table 1*). Two positive controls were included, classical MHC class I gene *Cyca-UA1*01* gene expression and *ß*-actin expression. The results were analysed by agarose gel electrophoresis.

Nucleotide sequence, amino acid sequence and phylogenetic analyses

Genomic and cDNA cyprinid class I *ZE* nucleotide sequences were represented by at least two identical clones. Sequence data obtained using the ABI sequencer were analysed with sequencer 4.1 software (Genes CodeS, Ann Arbor, Michigan, USA). Multiple alignments were done using the program CLUSTAL-W[166,217] version 1.8. Signal peptide prediction analyses[167] were performed using PSORT II (http://psort.nibb.ac.jp).

Phylogenetic analyses and synonymous and nonsynonymous distance estimations were performed using MEGA 2.1 software[168]. Phylogenetic relationships using p-distances for amino acid sequences were constructed by the neighbour-joining algorithm[161]. Synonymous and nonsynonymous distances were estimated by the modified NEI-GOJOBORI method[218] with p-distances or JUKES-CANTOR correction.

*Protein modelling of the Dare-ZE*0101 amino acid sequence*

The mouse MHC class I H-2LD model was predicted to be a suitable modelling template by the swiss-model blast tool. (The expasy proteomic server of the Swiss institute of bioinformatics: http://www.expasy.ch). The sequence of Dare-ZE*0101 was aligned to the sequence mouse MHC H-2LD with CLUSTALX [166] using the pam 350 matrix. Model building of Dare-ZE*0101 was performed with modeller[219] using the CVFF forcefield[220]. The mouse H-2LD structure (PDB entry: 1LPD) was used as a template. The model was verified after several rounds of energy minimisation. The stereochemical quality of the homology model was verified by PROCHECK[221], and the protein folding was assessed with PROSAII[222], which evaluates the compatibility of each residue to its environment independently.

RESULTS

Highly similar class I Z lineage genes in three teleost genera

This is the first report of complete protein-coding *Z* lineage sequences identified in three different cyprinid species including zebrafish. The complete protein-coding sequences were all generated by PCR using cDNA as template, ensuring that all sequences were transcribed. Anchored PCR on a cDNA library of common carp revealed the missing membrane proximal, transmembrane

and cytoplasmic regions of a common carp class I Z gene[74], *Cyca-ZE*0101*. The complete deduced amino acid sequence of *Cyca-ZE*0101* encoded a putative cleavable signal peptide of 24 N-terminal amino acids, three extra-cellular domains similar in length to other class I molecules, a transmembrane and cytoplasmic region (*fig. 1*).

Using the generacer™ kit for full-length, RNA ligase-mediated rapid amplification 5' and 3' ends sequences highly similar to *Cyca-ZE*0101* were identified in a African 'large' barb and a zebrafish individual. We determined six complete protein-coding sequences in African 'large' barb, *Bain-ZE*0101, Bain-ZE*0102, Bain-ZE*0201, Bain-ZE*0301, Bain- ZE*0401* and *Bain-ZE*0501* and two complete protein-coding sequences in zebrafish, *Dare-ZE*0101* and *Dare-ZE*0201* (*fig. 1*). The extra-cellular domains of the complete protein-coding African 'large' barb and zebrafish sequences exhibited 75% to 87% amino acid sequence identity to Cyca-ZE*0101. The alpha-1 domains of the zebrafish and African 'large' barb sequences showed 84% to 94% identity to the Cyca-ZE*0101 alpha-1 domain. The alpha-2 and 3 domains of the zebrafish and African 'large' barb ZE sequences were 77% to 86% and 65% to 78% identical to the Cyca-ZE*0101 alpha-2 and 3 domains. Signal peptide prediction analyses identified putative cleavable signal peptide of 26, 19 and 19 N-terminal amino acids in length for Bain-ZE*0201, Dare-ZE*0101 and Dare-ZE*0102, respectively. Bain-ZE*0101, Bain-ZE*0102 and Bain-ZE*0301 sequences only showed putative cleavage sites between the glutamine (Q-4) and threonine (T-3) residues. Bain-ZE*0401 and Bain-ZE*0501 sequences possess a leader peptide of 100 amino acids. However, a putative cleavages site is only predicted between the alanine (A-29) and glutamic acid (E-28) residues, suggesting an extension of the alpha-1 domain with 28 amino acids (*fig. 1*).

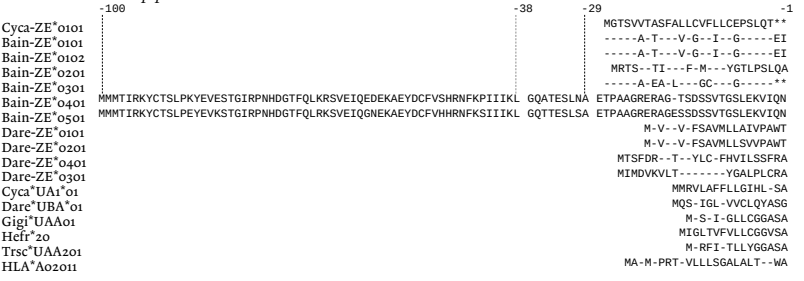
Analyses of the presence of transmembrane and cytoplasmic regions in the complete protein-coding of Bain-ZE*0301, Bain-ZE*0401 and Bain-ZE*0501 cDNA sequences revealed an in-frame termination codon a few codons down stream of the alpha-3 domain, resulting in absence of these regions.

Cyprinid class I ZE sequences encode bona fide class I molecule

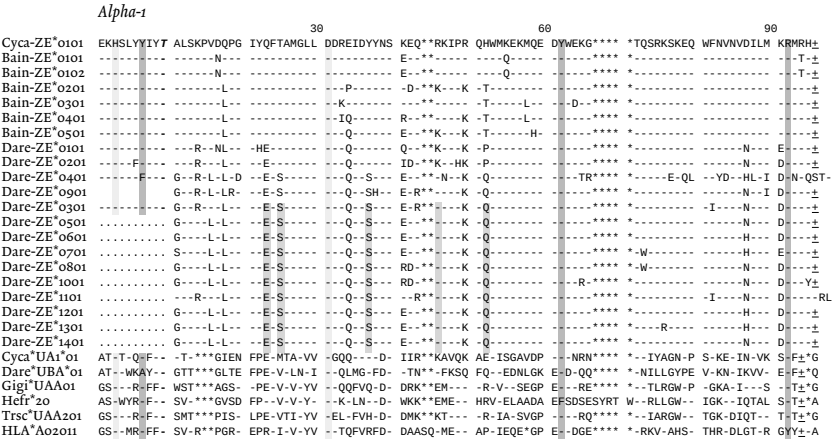
Deduced amino acid sequence of all ZE lineage genes were aligned with common carp, zebrafish, shark and human classical class I sequences (*fig. 1*). Most features known to be conserved in classical and non-classical class I molecules[7,64]

Amino acid sequence of class I heavy chains

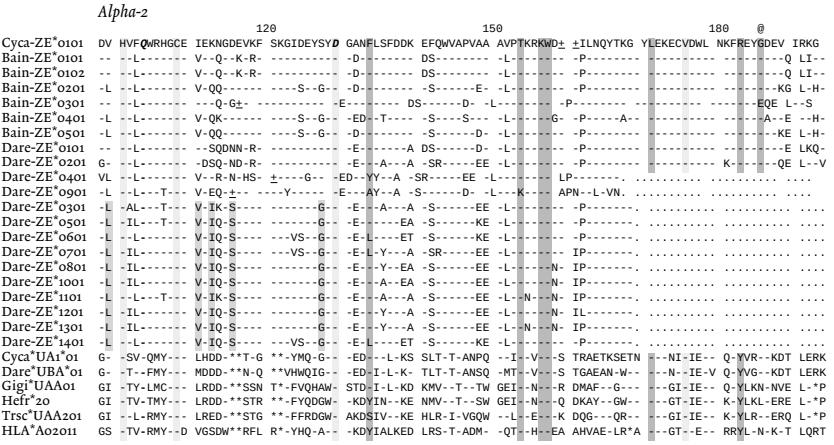
Leader peptide



Alpha-1



Alpha-2



Alpha-3

```
                    210                                 240                                 270
Cyca-ZE*0101  SPPEVH VFAKRCTRDK TKLELTCFAT GFYPKDVILS IRKYRSPLPE KEIESSGVRP NHDGTYQLKK TVIIQED**E KADYDCVVSH STL±IITK W
Bain-ZE*0101  ------ -------S-- ---K-S-L-- ---------T ------S--- A-------- ---E-F---- S-E-PQ-**- ------F-F- R--TKP--SR -
Bain-ZE*0102  ------ -------S-- ---K-S-L-- ---------T ------S--- A-------- ---E-F---- S-E-PQ-**- ------F-F- R--TKP--SR -
Bain-ZE*0201  ---A-- L----SIK-R S--K---M-- ----R--T-G ---S-TF--- D-T--T-I-- ----S--MR- S-ET---**- -TN---F--- R--K----VT -
Bain-ZE*0301  ------ -----SININ S--K-S-L-- --------M-L -----T---- D-------I-- -----F---- --E-PQ-**- ------F-F- R--S-PV-SR -
Bain-ZE*0401  -S-T-- -----SIK-- S--K---M-- ----R--T-L ---SHTS--D D-***T-T-- -----F-V-- S-E----**- --E---F--- R--K----VT -
Bain-ZE*0501  --LA-- -----SIK-R S--K---M-- ----R--T-G ---S-TF-S- D-T--T-I-- ----S--MR- S-ET---**- -TN---F--- R--K----VT -
Dare-ZE*0101  -A-K-- -----YVNG- A--K---L-- --------Y-T -----TA-SD S-V------- -----F--R- STY-L-E**- --E---Y-A- R---APVV-T -
Dare-ZE*0201  --RD-- -----YING- D--K---L-- --------Y-T -----TA-SD NDL------- -----F--R- STY-N--**- --E---Y-N- R--K-PV-IR -
Cyca*UA1*01   VS-Q-S LLQ-DPL*** **SPV--HT- ----SG-TIT WQ-NGQDHD- DVDLGE*LII -E---F-RAS -LNVKPEEWK NNKFS---E- QGKTIRE-LT E
Dare*UBA*01   VS-Q-S LLQ-SSS*** **SPVV-HV- ----SGLKI- WQRNGQDHD- DVELGE*LI- -E-----RTS -LNVKPEEWK -DKFS---E- QSKTINS-LT E
Gigi*UAA01    VK-S-T FTSV-GN*** **KQ-S-V-- ----HSIEVN LFRDSAKID- T-STG**--- ----S--IHR STEFDPN**S Q-K-S---D- DG-GQQLVVF Y
Hefr*20       VA-S-S FTRRGDS*** **NR-S-VT- ----QAIEVT L-RDGVS-D- SHFDG**IL- -------TH- WTD-DPA**D Q-KFS-E-D- -G-K-KMVVF Y
Trsc*UAA201   AA-T-S FTRLGDS*** **KK-S-V-- ----QSIE-K LHRDQAADV PYSSG**--- -------ME- QTDFEPS**D P-KFS-E-D- AG-SQKLVMF Y
HLA*A02011    DA-KT- MTHHAVSD** HEAT-R-W-L S---AEIT-T WQRDGEDQTQ DTELVE*T-- AG---F-KWA A-VVPSG**Q EQR-T-H-Q- EG-PKPLTLR -
```

Transmembrane

```
              300                                 330
Cyca-ZE*0101  DGTCPDC** ******GSSL VIGIVIGAVL VIVVVAAVCL YLFMTKRLTC
Bain-ZE*0101  ----TN-NE GLILVN-PAI GV-VGVVV-G L*--A--I-- --Y-----VF
Bain-ZE*0102  ----TN-NE GLIPVN-TAI GV-VGVVV-G L*--A--I-- --Y-----VF
Bain-ZE*0201  --KYSPVTI°
Bain-ZE*0301  ----TN-NK GTTLVN-NAI GV-VAVAV-V -GL-----F- --Y--E--VF
Bain-ZE*0401  --E°
Bain-ZE*0501  --KYSPVTI°
Dare-ZE*0101  --K-S--** ******SKES A--LIV--II GA-----IVV VA-IL-KNNK
Dare-ZE*0201  N-E-LS*** ********EP P-AMIA-IIG -LILLG-IGV TVWIL-KKNI
Cyca*UA1*01   -KIKTNY** *******APF P---I--I-V AVLLLIVIGV AGYKVYQKKK
Dare*UBA*01   -EIRTNN** ********PTA LL--I--I-V AA-LLV-IAV AG-VVY-RHK
Gigi*UAA01    EPKTSS*** ********** MLPVI--V-V ---L-IIAVI AGVVWYKKKA
Hefr*20       ERKSDS*** ********** LLP-IV-VLV -A-LLIIAVI VGVVIYKKKA
Trsc*UAA201   EPKADS*** ********** MLPVI--I-I AVL-LV-LAV VGVVLY-KKA
HLA*A02011    EPSSQPT** *******IP* IV--IA-L-- FGA-ITGAVV AAV-WR-KSS
```

Cytoplasmic region

```
                    360                                 390
Cyca-ZE*0101  RQQSATNGQV QG**AYEVNE NLLPGYN*SN GHAPTAVNG* ********** GPNGVIIIAN
Bain-ZE*0101  TR--N---A- --NN-NG-KD T----K-F-- ---ASQWRT
Bain-ZE*0102  TR--N---A- --NN-NG-KD T----K-F-- ---ASQWRT
Bain-ZE*0301  T--IN--VT- N-ND-NG--D T----H-HT- ---AANGGHN HTNGHAAANG -S--K-D-V
Dare-ZE*0101  FCFRT-QEPS EENGRVLMKD PVFKEE-GAG KDPSVVPLTN GNH
Dare-ZE*0201  IGNGDEKHF- VSTVSGNKD- -GFVVSTV-G NKDEKRSMPD GS******** ADS-KGSS-
Cyca*UA1*01   GFKP**V**N GSDDGSNSSA HTD-KA
Dare*UBA*01   GFKP**VP-N TSDGGSDNSS RT
Gigi*UAA01    G-KTGY-PAK TSDKGDSSSS S-ATA
Hefr*20       GEKTGY-PAK TSDRGESSSN SSANA
Trsc*UAA201   G-KTGY-PAK TSDK-ESSSN SSATA
HLA*A02011    DRKGGSYS-A ASSDSAQGSD VS-TACKV
```

Figure 1: Comparison of the deduced class I ZE amino acid sequences identified in common carp (Cyca), African 'large' barb (Bain) and zebrafish (Dare) with MHC classical class I sequences of common carp, zebrafish, nurse shark (Gici), horned shark (Hefr) and human (accession numbers are given in the legend of *figure 5*).

Numbering is based on the Cyca-ZE*0101 including gaps. *Dashes* indicate identity to the Cyca-ZE*0101 sequence, *asterisks* indicate gaps and *dots* absence of sequence information. Highly conserved residues, cysteine residues, residues involved in salt-bridges and the FYP motif are marked by *light grey boxes*. Nine putative evolutionary conserved residues important for anchoring peptide termini are marked with *deep grey boxes*. The conserved tyrosine residue (Y186) located three amino acids downstream in all cyprinid ZE sequences is marked by an @. The putative CD8 binding loop is shown in *bold* in the *grey box*, putative N-linked glycosylation sites are *underlined* and conserved putative β_2m contact residues are marked in *bold* and in *italic*. Amino acid motifs within the zebrafish ZE sequences, characterising a group of sequences used to calculate positive selection (*table 3*) are marked by *grey boxes*. Premature termination codons are indicated by °.

Table 2: Conserved peptide binding residues in classical class I molecules

*Amino Acid No. correspond to HLA-A*02011 (31)*			*Peptide N-terminus*					*Peptide C-terminus*				
			7	59	159	171	84	123	143	146	147	Diff[e]
	HLA-A, -B, -C[a]		Y[d]	Y[d]	Y[d]	Y[d]	Y[d]	Y[d]	T[d]	K[d]	W[d]	
Human	consensus	n=312	-	-	-	-	-	-	-	-	-	0
Human		n=30	-	-	-	H	-	-	-	-	-	1
Human		n=1	-	H	-	-	-	-	-	-	-	1
Human		n=9	-	-	-	-	-	-	-	-	L	1
Human		n=1	H	-	-	H	-	-	-	-	-	2
Human		n=11	-	-	-	-	-	-	s	-	L	2
	H-2 K, D, L[a]											
Mouse	consensus	n=17	-	-	-	-	-	-	-	-	-	0
Mouse		n=6	-	-	-	H	-	-	-	-	-	1
Fish & other[b]		n=11	-	-	-	-	R	F	-	-	-	2
Leopard shark[a]	Trsc-UAA*101		-	-	-	-	R	L	-	-	-	2
Rainbouw trout[a]	Onmy-UBA*C32		-	-	-	-	R	F	-	-	L	3
Cyprind ZE lineage		n=21	-	-	-	F/Y	R	F	-	-	-	2
Human[a]	HLA-E		-	-	-	-	-	-	S	-	S	2
Human[a]	HLA-F		-	-	-	-	-	-	S	-	C	2
Human[a]	HLA-G		-	-	-	-	R	-	-	F	Y	3
Human[c]	CD1d		L	Q	L	L	M	I	I	V	L	9
Human[a]	HFE		-	M	-	L	N	H	-	E	-	5
Human[a]	FcRN		-	M	F	H	K	F	A	R	-	7
Human[a]	MICA		-	T	A	-	H	F	V	F	L	7
Mouse[a]	H-2 M3		-	-	-	F	-	-	-	R	L	3
Frog[a]	Xela-XNC1.1		F	-	F	-	V	F	V	L	-	6
Rainbouw trout[a]	Onmy -UAA*0101		H	F	H	L	H	F	Y	-	R	8
Coelacanth[a]	Lach -UB*01		E	S	D	-	E	F	R	V	C	8
Common carp[c]	Cyca ZA1		?	D	F	F	C	C	-	-	R	6
Common carp[c]	Cyca ZA2		?	D	F	F	C	C	-	-	R	6
Common carp[c]	Cyca ZB1		?	D	-	F	C	F	-	-	-	4
Common carp[c]	Cyca ZB2		?	D	-	F	C	F	-	-	-	5
Common carp[c]	Cyca ZB3		?	D	-	F	C	F	-	R	-	5
Common carp[c]	Cyca ZC1		?	A	F	F	C	F	-	N	H	7

[a] Data adapted from Shum et al (41). [b] See legend figure 5, all sequences marked by #;. [c] See legend figure 5 for accession numbers. [d] Identity to the HLA motif is indicated by dashes. [e] Number of differences compared to the HLA motif. ? not determined.

are present in ZE sequences. All ZE sequences possessed the conserved cysteine residues (C109, C176, C217, C276) in the alpha-2 and 3 domains to form disulfide bonds within these domains, conserved residues (H3, D31, H101, D130) in alpha-1 and -2 to form two salt bridges within these domains and the FYP (222-224) motif in alpha-3 domain.

Three acidic residues in an exposed loop in the alpha-3 domain (aa: 237-243) form a major CD8 binding site in mammals. The ZE lineage sequences all possess at least three acidic residues within this region. Four residues (T10, Q104, D130, Q257) known to be involved in β_2m binding of human class I molecules are conserved in the cyprinid ZE sequences. Like classical class I sequences, a putative N-linked glycosylation site is present at the end of alpha-1 domain in ZE sequences with two exceptions (Dare-ZE*0401, Dare-ZE*1101). Four sequences, Cyca-ZE*0101, Bain-ZE*0301, Dare-ZE*0301 and Dare-ZE*0901 possessed additional putative N-linked glycosylation sites in the alpha-2 or -3 domains. Only small insertions (aa: 15/16; 116/117; 122; 219/220; 247) and deletions (aa: 44/45) were present in the alignments of alpha-1, -2 and -3 domains of cyprinid class I ZE sequences compared to human classical class I sequences. Hydrophobicity plots of cyprinid ZE amino acid sequences are comparable to the plot for Cyca-UA1*01 with exception of a hydrophilic region in alpha-1 domain (data not shown). The African 'large' barb sequences Bain-ZE*0201, ZE*0401 and ZE*0501 lack the transmembrane and cytoplasmic region. In addition, the Bain-ZE*0401 and ZE*0501 sequences possess an extended leader peptide which is hydrophilic in nature as indicated by hydrophobicity plots (data not shown).

Cyprinid class I ZE sequences possess an evolutionary conserved peptide binding motif

The presence of nine evolutionarily conserved putative peptide-anchoring residues in the amino acid sequence of classical class I genes has shown to be a useful criterion in discriminating classical class I genes and non-classical class I genes in many vertebrates[79,170-173]. In classical mammalian class I molecules, this motif of nine amino acids is YYYYTKWYY, while in non-mammalian vertebrates it slightly changed to YYRFTKWYY (*table 2*). Alignment of cyprinid ZE sequences with a human classical class I sequence revealed the presence of the conserved non-mammalian motif (*table 2*; *fig.1*: Y7, Y62, R92, F134, T154, K157, W158, Y171, F183/Y186) with a minor difference. The tyrosine residue (Y186) was located three amino acids downstream in all cyprinid ZE sequences,

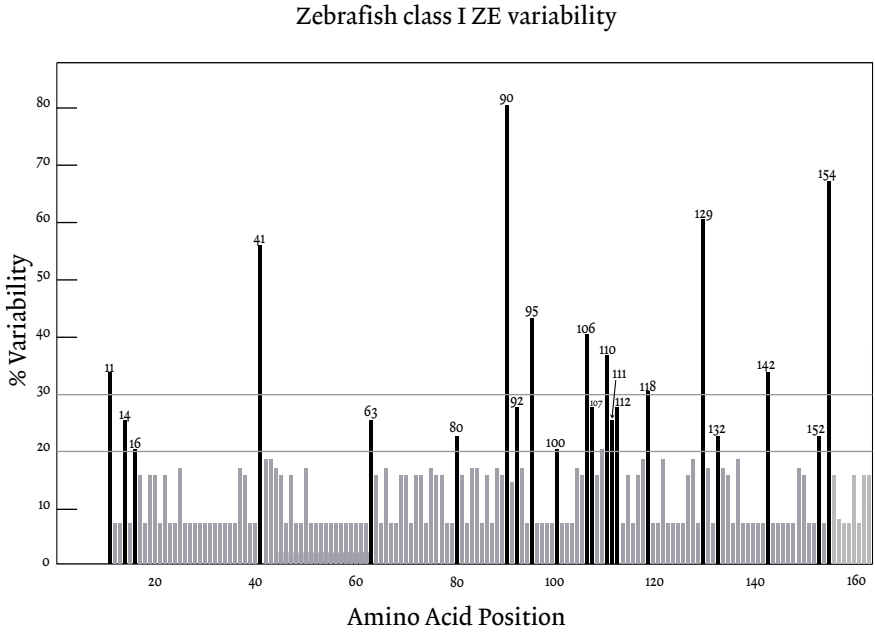


Figure 2: A Wu-Kabat variability analysis[223] for the deduced amino acid sequences of alpha-1 and 2 from all Dare-ZE sequences. Residues with above 20 % variability are marked by numbers. The *numbers* along the x-axis denote the amino acid position starting with the first amino acid of the alpha-1 domain of the Dare-ZE molecule. Variability of the first ten amino acids is not available.

when aligned with human classical class I (*fig. 1*). All other carp Z lineage sequences differ at least at four positions and up to eight when compared to the conserved motif in mammalian class I sequences.

Polymorphism of class I ZE *sequences in Danio rerio*

Analyses of five zebrafish individuals revealed 14 unique *Dare-ZE* sequences (*fig. 1*). Four zebrafish ZE sequences, *Dare-ZE*0101*, *Dare-ZE*0201*, *Dare-ZE*0301* and *Dare-ZE*0401* were identified in mRNA of a zebrafish individual. Ten zebrafish ZE lineage sequences (*Dare-ZE*0501* to *Dare-ZE*1401; fig. 1*) were generated by PCR

on genomic DNA of four zebrafish individuals using primers designed to the start of alpha-1 and the end of alpha-2 of Dare-ZE sequences identified in mRNA. Agarose gel electrophoreses of the genomic PCR products revealed three fragments,~650 bp,~800 bp and~1000 bp in length in each individual. Subsequent cloning of these PCR products and sequence analyses revealed ten unique sequences only representing the 650 bp fragment (intron data not shown).

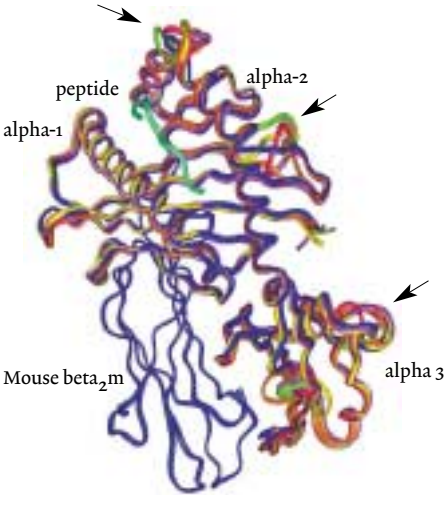
A WU-KABAT variability plot[223] based on 162 amino acids of the alpha-1 and alpha-2 domain from 14 unique Dare-ZE sequences showed the presence of putative polymorphic residues. Twenty-one amino acid residues within the alpha-1 and alpha-2 domains showed 20% or higher variability, while nine of these 21 residues (aa: 11, 41, 90, 95, 106, 110, 129, 142, 154) showed 30% or higher variability (*fig. 2*).

*Protein modelling of Dare-ZE*0101*

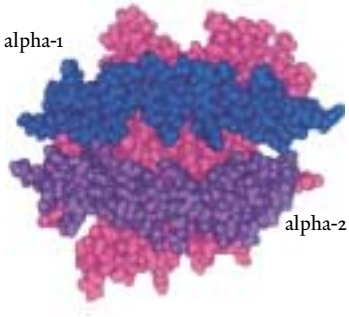
Protein model construction of the Dare-ZE*0101 amino acid sequence using the mouse H-2LD crystal structure as template resulted in several putative protein models. Figure 3A shows the mouse template (blue) in complex with a peptide mainly containing alanine residues (APAAAAAAM) and the β_2m superimposed on the constructed models of the Dare-ZE*0101 sequence. Only four major putative deviations from the mouse crystal structure were observed in four different loops (*fig. 3A*, arrows). Two putative loop structures were located in the alpha-3 domain and two in the alpha-2 domain. The flexible loop structure between de two alpha-helixes of the alpha-2 domain, due to an amino acid insertion in the Dare-ZE*0101 sequence compared to the mouse sequence, may have major implication for peptide binding. The position of this loop, either pointing towards the peptide binding pocket, away from the pocket or a position between these two, may result in different peptide binding characteristics.

A space filled (*fig. 3B*) and ribbon structures models (*fig. 3C, 3D*) of the peptide binding domains a of constructed protein model show that the ends of the pocket are more occluded due to the two alpha-helixes of the alpha-2 domain which each point toward one ends of the alpha-1 alpha-helix. The ribbon structure shown in figure 3C shows the position of nine conserved residues in the ZE molecule (*fig. 3C*, light blue). These residues were located at similar position when compared to the mouse MHC class I model and known to be involved in binding of peptide of mammalian class I molecules. The tyrosine residue (*fig.*

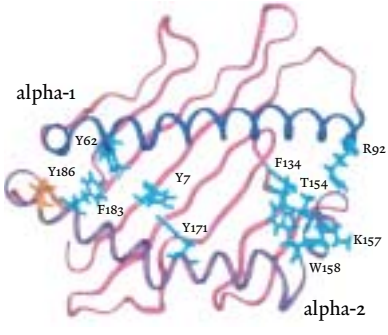
A.


peptide
alpha-2
alpha-1
Mouse beta$_2$m
alpha 3

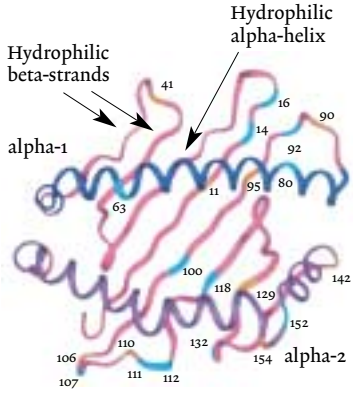
B.


alpha-1
alpha-2

C.


alpha-1
Y62
R92
Y186
Y7
F134
T154
F183
Y171
K157
W158
alpha-2

D.


Hydrophilic
alpha-helix
Hydrophilic
beta-strands
41
16
90
14
alpha-1
92
11
95
80
63
100
118
142
129
152
110
132
106
154
alpha-2
111
112
107

3C, orange) located three amino acid down stream of the phenylalanine residue is located near the end of one of the alpha-2 domain alpha-helixes. The position of the polymorphic residues exhibiting between 20% and 30% variability (*fig. 3D*, light blue) and more then 30% variability (*fig. 3D*, orange) in WU-KABAT plots are shown to be located in the beta-strands, alpha-helixes and loops.

Exon-intron organisation of cyprinid class I ZE sequences

Sequence specific primers were designed to the leader peptide and to the 3'-untranslated region of *Bain-ZE*0401, Bain-ZE*0501* and *Cyca-ZE*0101* to generate genomic PCR products. Two unique genomic African 'large' barb sequences similar to the CDNA nucleotide sequence of *Bain-ZE*0401* and *Bain-ZE*0501* and one genomic sequence similar to the CDNA nucleotide sequence of *Cyca-ZE*0101* were identified (*fig. 4*). Both genomic African 'large' barb sequences consisted of 7 exons and 6 introns. The leader peptide is encoded by exon-1a, -1b and -1c, the alpha-1, alpha-2, and alpha-3 domains are encoded by exon-2, exon-3, and exon-4, respectively. The connecting peptide, the transmembrane and the cytoplasmic tail were not present in these genomic sequences. Exon-5 encoded three (*Bain-ZE*0401*) or nine (*Bain-ZE*0501*) in frame amino acids followed by a termination codon. The remainder of exon-5 contained the 3'-untranslated region.

‹ *Figure 3*: Putative protein models of Dare-ZE*0101 using the mouse H-2LD structure as template. A) Several putative protein models of Dare-ZE*0101 superimposed on the mouse H-2LD structure (PDB entry: 1LDP). The mouse template is shown in blue and the alpha-1, alpha-2 and alpha-3 domains are indicated. Arrows in figure 3A indicate putative loop structures deviating from the mouse model. Figures 3B, 3C and 3D are space filled or ribbon structure models of the peptide binding domains, corresponding to the violet model in 3D. The beta-strands are shown in the original colour (violet). The alpha-1 domain alpha-helix is shown in blue and the alpha-2 domain alpha-helixes are shown in indigo. C) Evolutionary conserved residues are shown in light blue and tyrosine residue located three residues down stream of the phenylalanine at position 183 is shown in orange. D) Residues showing a variability of 20 to 30% are shown in light blue and residues showing 30% or higher variability are shown in orange.

Exon-Intron organization class I ZE genes

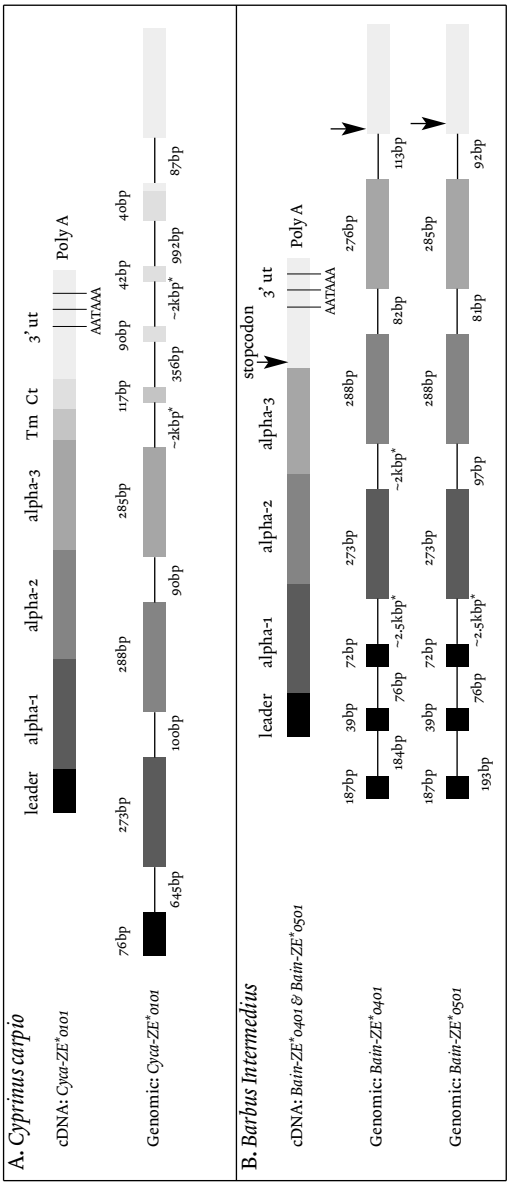


Figure 4: The genomic organisation of common carp *Cyca-ZE*0101* and African 'large' barb *Bain-ZE*0401* and *Bain-ZE*0501*. Exons encoding for the leader peptide, extra-cellular domains, alpha-1, alpha-2, and alpha-3, the transmembrane (TM) and cytoplasmic region (CT) and the 3' untranslated region (3' ut) the are indicated by ■■■■■■■■ respectively, Exon lengths are placed above the different protein domains while intron lengths are placed below. *Asterisks* indicated intron lengths estimated by PCR.

All introns start with GT and end with AG and are all phase 1 introns, a codon split between the first and the second base[10].

The genomic common carp sequence consists of nine exons and eight introns (*fig. 4*). The leader peptide and the three extra-cellular domains, alpha-1, alpha-2 and alpha-3 are encoded by separate exons (exons 1-4). The connecting peptide and the transmembrane region are encoded by exon-5 and the cytoplasmic tail by exon-6, exon-7 and exon-8. Exon-8 also encoded the start of the 3' untranslated region, while exon-9 contained for the remainder of the 3'-untranslated region. All introns start with GT and end with AG and are all phase 1 introns.

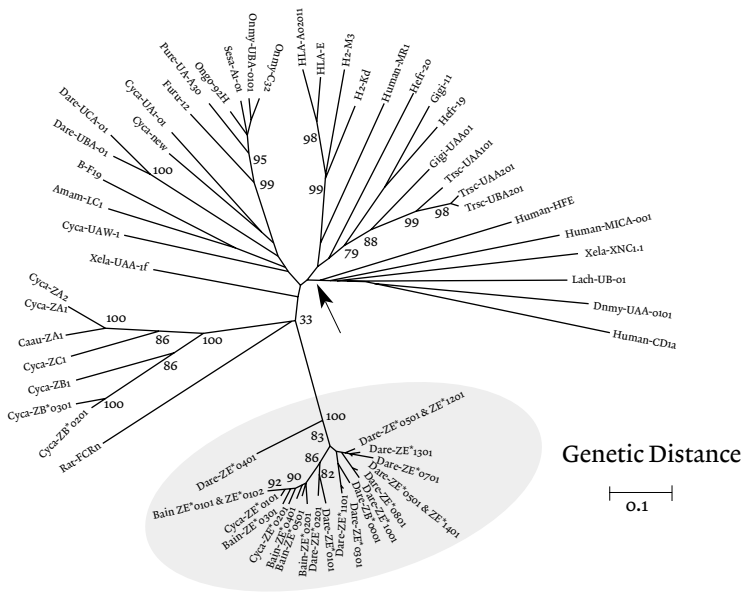
Phylogenetic analyses

Phylogenies were constructed separately for the alpha-1, -2 and -3 domains of the cyprinid ZE sequences, with representatives of class I genes from other vertebrate taxa. The branching order and major groupings are similar to those documented in previous studies[74,107,171,224]. In trees constructed of the alpha-1, alpha-2 or alpha-3 domain sequences, all the cyprinid ZE sequences cluster together in single clades supported by a high bootstrap values (*fig. 5*) and the cyprinid ZE clades clustered on a single branch with the clades comprising all other cyprinid Z lineage sequences, ZA, ZB, ZC and ZD. However, the latter topology is only supported by a high bootstrap value in tree of alpha-3 domain, while medium and low bootstrap values are observed in trees of alpha-2 and alpha-1 domains, respectively.

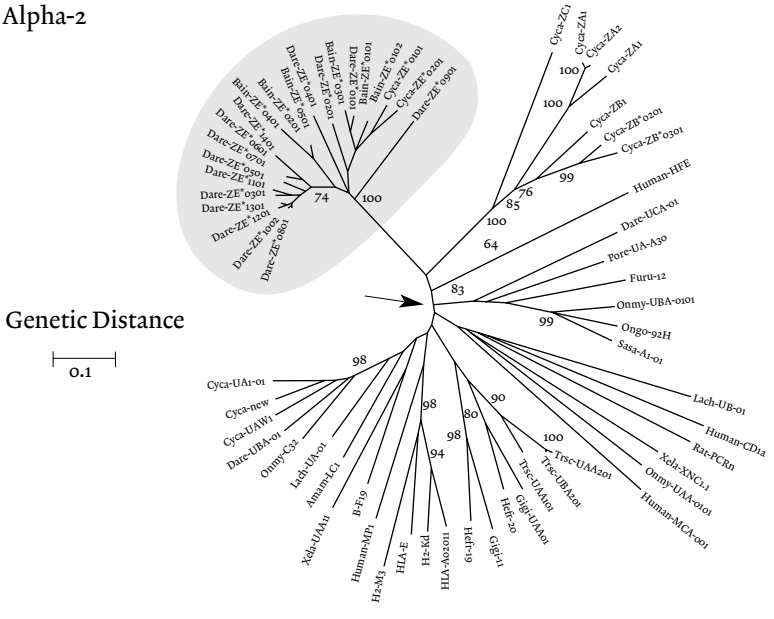
STET and co-workers[74] extensively studied non-classical fish class I ZA, ZB, ZC and ZD sequences in comparison with classical ones. In that study, comparison of classical fish class I sequences, including Cyca-UA, showed a very high degree of diversity in the alpha-1 domains, a moderate diversity in the alpha-2 domains and conservation of the alpha-3 domains. This diversity is reflected in the phylogeny, with longer branch lengths indicating higher diversity. Branch lengths in the cyprinid ZE clade of the alpha-1 domain tree are remarkable short in relation to branch length of the alpha-2 and -3 domain trees. This observation is in stark contrast to the paradigm of highly divergent class I alpha-1 and alpha-2 domains.

Both the alpha-1 and alpha-2 trees show a large zebrafish ZE sub-cluster supported by high bootstrap values. The sub-cluster in the tree of alpha-1 comprised all ten genomic zebrafish ZE sequences and one CDNA sequence. These

A. Alpha-1



B. Alpha-2



C. Alpha-3

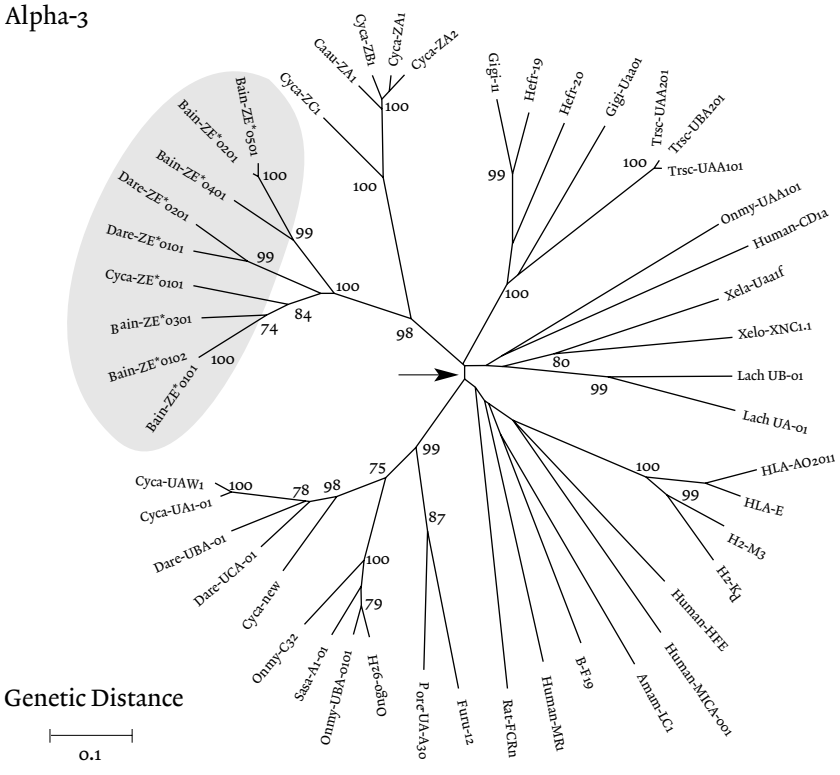


Figure 5: Neighbour-joining trees of the alpha-1, alpha-2 and alpha-3 domains of class I sequences. Representative classical and non-classical class I sequences of the major vertebrate classes were included. *Numbers* at branch nodes represent the confidence level of 1000 bootstrap replications. *Arrows* indicate the cluster formation starting point. The accession number of the sequences used in *figure 6* and *table 2* are as follows: cichlid fish[176]: Auha-517$^{\#}$; lizard: Amam-LC1$^{\#}$(M81094), chicken: B-F19$^{\#}$ (M84766); goldfish: Caau-ZA1 (L10418); common carp: Cyca-new (AB018581), Cyca-UAW1 (X91022), Cyca-UA1*01$^{\#}$ (X91015), Cyca-ZA1 (M37107), Cyca-ZA2 (L10419), Cyca-ZB1 (L10420), Cyca-ZB2 (AJ007848), Cyca-ZB3 (AJ007851), Cyca-ZC1 (L10421); zebrafish: Dare-UAA*01$^{\#}$ (Z46776), Dare-UBA01 (Z46777), Dare-UCA*01 (Z46778); pufferfish: Furu-12$^{\#}$ (AF001216); mouse: H2-M3 (U18797), H2-K^{d} (J00402); human: HLA-E (M20022), HLA-F (X17093), HLA-G (X17273), human-CD1d (X14974) human-HFE (U60319), human-MICA-001 (L14848), human-FcRn (AF220542), HLA-A*02011 (K02883), human-CD1a (M28825), human-MR1 (U22963); coelacanth: Lach-UB*01 (U08034), Lach-UB*01 (U08034); pink salmon: Ongo-92H$^{\#}$ (D58386); rainbow trout: Onmy-C32 (U55380), Onmy-UAA*0101 (L63541), Onmy-UBA*0101$^{\#}$ (AF091785), guppy: Pore-UA*A30$^{\#}$ (Z54085); Atlantic salmon: Sasa-A1-01$^{\#}$ (L07606); leopard shark: Trsc-UAA*101$^{\#}$ (AF034316); shark: Gici-11 (AF028557), Gici-UAA*01 (AF220063); horned shark: Hefr-19 (AF028558), Hefr-20 (AF028559); leopard shark: Trsc-UAA*201 (AF034335), Trsc-UBA*201 (AF034345); rat: rat-FcRn (X14323); frog: Xela-UAA1f^{r} (l20733), Xela-XNC1.1 (M58019).

mRNA expression of class I Z lineage genes

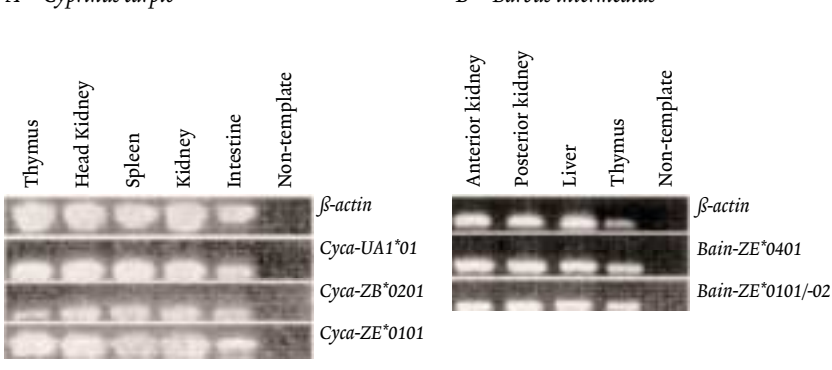


Figure 6: MRNA expression of cyprinid class I *Z* lineage. The presence of mRNAof *Cyca-UA1*01, Cyca-ZE*0101* and *Cyca-ZB*0201* was studied in thymus, head kidney, spleen, kidney and intestine of a common carp individual. Expression of *Bain-ZE*0401* and *Bain-ZE*0101 & Bain-ZE*0102* was studied in muscle, anterior and posterior kidney, liver and thymus of a African 'large' barb individual. *ß*-actin was included as a positive control.

sequences clustered also together in the alpha-2 tree with exception of Dare-ZE*0901. The Dare-ZE cDNA sequences, Dare-ZE*0101, Dare-ZE*0201 formed a second, and all African 'large' barb and common carp ZE sequences a third sub-cluster in the alpha-1 tree supported by high bootstrap values. However, this topology was dissolved in the tree of the alpha-2 domain. Three clear sub-clusters in the cyprinid ZE clade, supported by high bootstrap values, were formed in the tree of alpha-3 domain. African 'large' barb ZE sequences formed two clusters comprising Bain-ZE*0201, Bain-ZE*0401 and Bain-ZE*0501 or Bain-ZE*0101, Bain-ZE*0102 and Bain-ZE*0401 with the latter cluster including Cyca-ZE*0101. The two cDNA zebrafish ZE sequences formed a third sub-cluster.

Class I Z mRNA expression in tissues of cyprinid fish

Expression of class I Z mRNAwas studied in several tissues of a common carp and African 'large' barb individual by RT-PCR with sequence specific primers

(*fig. 6*). Both *ß*-actin mRNA and *Cyca-UA1*01* mRNA are expressed in all tissues studied.

Expression of mRNA of *Cyca-ZB*0201* and *Cyca-ZE*0101* was detected in thymus, head kidney, spleen, kidney, intestine. *Bain-ZE*0401* mRNA and *Bain-ZE*0101* and *Bain-ZE*0102* mRNA are expressed in anterior and posterior kidney, liver, thymus, and muscle (data not shown).

Amplification specificity of the class I *Z* lineage genes was verified by sequencing and analysing the amplified PCR products. Analysing the amplified RT-PCR products revealed the presence of both *Bain-ZE*0101* and *Bain-ZE*0102* sequences in all tissues investigated. The amplified region of the nucleotide sequences of *Bain-ZE*0101* and *Bain-ZE*0102* sequences differ by two synonymous nucleotide substitutions.

Positive selection acting on zebrafish class I ZE sequences

Positive selection plays an important role in generating polymorphism in the peptide binding region (alpha-1 and alpha-2 domains) of classical class I molecules. Particularly, residues involved in peptide binding are under positive Darwinian selection. To search for evidence of positive selection[35] (d_N/d_S ratio > 1), the ratio between the rates of synonymous (d_S) and nonsynonymous substitutions (d_N) was calculated using two different distance methods (*table 3*). Although the evolutionary conserved peptide anchor residues of HLA class I sequences are superimposable on cyprinid ZE sequences, HLA polymorphic peptide binding residues are, however, not superimposable on zebrafish ZE variability. Therefore, putative residues involved in peptide binding were identified based on the variability they possessed in the WU-KABAT analysis (*fig. 2*). Positions, possessing variability higher than 20% were assigned as putative peptide binding residues and all others as non-peptide binding residues.

The cyprinid ZE sequences represent an ancient lineage, therefore noise between different loci accumulated over a long period of time may obscure the signal for positive selection. However, when the genomic organisation of loci in multigene families is unknown, their identity can only be inferred using multiple criteria. FIGUEROA and co-workers[213] used five criteria to estimate a minimum number of loci in swordtail fishes. Applying three of the five criteria identified a group of ten sequences, *Dare-ZE*0301, Dare-ZE*0501 to -0801, Dare-ZE*1001* to *-1401* belonging to a putative similar locus. These ten sequences formed a distinct

Table 3: Calculation of d_n/d_s ratios

	dn/ds	*S.E.*	*p value*[a]	*dn/ds*	*S.E.*	*p value*[a]
Distance method: p-distance						
	Non-peptide binding residues					
variability	< 20%			< 30%		
14 taxa	0.3686	0.0871	1.0000	0.4589	0.1008	1.0000
10 taxa	1.1141	0.8105	0.4167	1.4323	0.8810	0.1689
	Peptide binding residues					
variability	≥ 20%			≥ 30%		
14 taxa	1.4051	0.4478	0.0447	2.4264	1.3371	0.0007
10 taxa	2.8884	1.8698	0.0002	3.4734	3.1293	0.0019
Distance method: Jukes-Cantor						
	Non-peptide binding residues					
variability	< 20%			< 30%		
14 taxa	0.3050	0.0866	1.0000	0.3794	0.0993	1.0000
10 taxa	1.1192	0.8309	0.4187	1.4337	0.9008	0.1711
	Peptide binding residues					
variability	≥ 20%			≥ 30%		
14 taxa	1.3569	0.6206	0.0164	2.9866	2.3672	0.0022
10 taxa	3.2646	2.4318	0.0004	4.3297	5.2554	0.0218

a Test for positive selection using Mega software: H_0: $d_n = d_s$, H_1: $d_n > d_s$, reject H_0 when p-value < 0.05

clade in phylogenetic analyses of alpha-1 and -2 domains (*fig. 6*), they shared characteristic motifs in exon sequences (*fig. 1*) and possessed comparable intron-2 sequences of 92 to 99 bp in length (data not shown). Therefore d_n/d_s ratios were calculated for 14 zebrafish sequences that may represent multiple loci and for 10 sequences, that most likely represents only one locus.

D_n/d_s ratio based on 14 zebrafish ZE sequences (*table 3*, 14 taxa) revealed purifying or neutral selection for non-peptide binding and positive selection for putative peptide binding residues for both distance methods. Although ratios for positive selection were slightly above 1 (1.4051, 1.3569) at putative PBR positions, they were supported by p-values below 0.05. Using only the ten sequences (*table*

3, 10 taxa), substantially increased the d_n/d_s ratios of the putative peptide binding and non-peptide binding residues (*table 3*). Similar evidence was found for assigning residues exhibiting a variability of 30% or higher as putative peptide binding residues and all others as non-peptide binding.

DISCUSSION

In the past, extensive search for non-classical class I sequences and additional classical class I sequences in zebrafish failed to provide evidence for existence of such loci in this species[48,86,213]. However, this study clearly shows the presence of a novel *bona fide* class I ZE lineage in zebrafish that is also present in African 'large' barb and common carp. The class I ZE lineage sequences identified in zebrafish and in two other cyprinid species exhibit four important features in favour to consider them as classical class I molecules. First, they possess the conserved amino acid residues involved in peptide binding of mammalian classical class I molecules. Second, they show ubiquitous expression in African 'large' barb and common carp. Third, class I ZE sequence variability is observed among sequences from several zebrafish individuals. Fourth, the variability seen among zebrafish class I ZE sequences is generated by positive selection acting upon putative peptide binding residues. However, unlike mammalian classical class I genes, the cyprinid class I ZE alpha-1 domains show extraordinary high amino acid conservation between and within the three divergent species studied, while less conservation is observed in the alpha-2 and -3 domains. Phylogenetically, the class I ZE lineage seems to be more related to the common carp (*Cyca*-ZA, ZB, ZC) and goldfish (*Caau*-ZD) class I Z lineage sequences which were assigned as non-classical class I genes[74]. The presence of this additional class I ZE lineage in zebrafish and in two other cyprinids, either classical or non-classical in function, may have implications for the observation that in bony fish class I and class II genes identified to date, are located in different linkage groups[48,53,191,192]. The novel zebrafish ZE lineage may be linked to one of the class II regions identified in three different linkage groups. This linkage would implicate a complex of class I and class II genes like the MHC observed in all other jawed vertebrate species studied.

To date, unlike the mammalian counterparts, the function of class I molecules in fish has not been formally demonstrated. Their classical or non-classical nature is inferred from amino acid sequence analyses, expression patterns

and the extent of polymorphism and compared to their mammalian counterparts. The presence of conserved peptide anchoring residues[64,213] in the cyprinid class I ZE lineage favours assigning them as classical, although one of the tyrosine residues is replaced by a phenylalanine residue. The substitution of a tyrosine residue by a phenylalanine residue is seen in most non-mammalian classical class I sequences at position Y123. This indicates that such a replacement may not have major implications for the ability to bind peptide termini. Another possibility might be that the tyrosine residue located three amino acid residues downstream (*fig. 1,* Y186) functions as the conserved peptide anchoring residue in cyprinid class I ZE molecules instead of the phenylalanine residue (*fig. 1*, F183). However, protein modelling suggests that this extend the peptide binding groove. This may implicate binding of somewhat larger peptides or molecules similar to CD1 that bind lipid antigens in a substantially larger binding groove[225]. The binding groove of cyprinid class I ZE molecules possess hydrophilic residues in the alpha-1 domain as indicated by hydrophobicity plots (data not shown). Protein modelling indicates that the hydrophilic residues in this domain comprise two beta-strands and the alpha-helix that follows (*fig. 3*), suggesting that one side of the peptide binding groove is extremely hydrophilic.

Ubiquitous expression is another feature in favour of the classical nature of cyprinid class I ZE sequences. However, ubiquitous expression is also seen for another non-classical class I Z sequence (*Cyca-ZB*0201*) that does not possess the conserved peptide binding motif. In the past, characteristics of non-classical class I genes led to the hypothesis that these gene are non-functional relics of ancient classical class I genes, whose ultimate extinction is inevitable[20,21]. However, many non-classical class I genes have been reported in mammals since and several studies revealed the functionality of these molecules (reviewed[22]). A possible explanation for ubiquitous expression of this non-classical class I Z gene in common carp might be that fish posses a variety of non-classical functional class I-molecules similar to the situation seen in mammals. This is supported by the fact that non-classical class I sequences are not limited to common carp and goldfish (reviewed[74]) as previously suggested[213]. To date non-classical fish class I genes have also been identified in coelacanth[120], shark[54,171,226], salmonids[224], and catfish[227].

With the variety of class I-like genes that now have been identified in different mammalian species, the non-classical label has become ambiguous in mammals.

Thus it is suggested that classical class I genes, presenting peptides to cytotoxic T lymphocytes, are only those highly expressed MHC encoded loci that are subject to balancing selection which favours polymorphism at the positions that function as peptide binding residues[23,29]. The evidence that zebrafish class I ZE loci are subject to balancing selection generating variability at specific positions in the alpha-1 and -2 domain, can be considered as the most important feature supporting the classical nature of these genes. In contrast to the evolutionary conserved peptide anchor residues of HLA class I sequences that are superimposable on cyprinid ZE sequences, HLA polymorphic peptide binding residues[228] are not superimposible on zebrafish class I ZE variability. However, binding of peptides or other small molecules of a different chemical nature might have favoured variability at positions other than those identified in HLA class I sequences.

Protein modelling indicate that only 10 out of the 21 variable residues are located in a beta-strand or an alpha-helix at a position that might be involved in ligand binding. The remaining 11 residues are located in loops where they might play a role in receptor binding such as NK receptor and TCR. Two polymorphic residues in the loop between the two alpha-2 helixes of the alpha-2 domain of the class I ZE molecules may have an undefined implication for peptide binding. This loop possesses high flexibility created through insertion of two amino acid residues compared to the mouse model. However, only co-crystallisation with what is bound into the groove and functional studies will provide data on the actual structure of the molecule and the biological role of this novel class I ZE lineage.

The genomic organisation the common carp class I *Cyca-ZE*0101* gene is similar to mammalian MHC classical class I genes 10 with exception of an additional intron which is located in the 3'-untranslated region. Such an intron is observed for common carp classical class I genes. However, in the case of *Cyca-UA1*01* the cytoplasmic region of this gene is encoded by two exons[107], exon-6 and exon-7. Although mammalian classical class I genes possess a characteristic intron-exon organisation it is not a criterion for distinguishing them from non-classical class I genes. They may have similar or different organisations from that of classical class I genes.

Remarkable is the presence of African 'large' barb sequences lacking a transmembrane and cytoplasmic region at the MRNA level. The exon-intron organisation

clearly showed the absence of these regions at the genomic level. The soluble class I Bain-ZE*0401 and ZE*0501 molecules are thus not due to alternative splicing. These sequences also exhibit three introns within the much longer hydrophilic leader peptide compared to other class I sequences. Although these soluble molecules are expressed, it remains to be seen whether they are functional.

The birth-and-death model of evolution assumes that over the long-term evolution of MHC molecules, new genes are created by repeated gene duplications. Some of the duplicated genes are maintained in the genome while others are deleted or become non-functional through deleterious mutations[198,229,230]. KLEIN *et al.* [128] described a similar mechanism designated the accordion model. The three aberrant *Bain-ZE* sequences, encoding soluble class I molecules may be the remnants of previous gene duplications that underwent one or two deletion events after duplication resulting in sequences coding for soluble molecules with normal or aberrant MHC leader sequences. The aberrant *Bain-ZE* sequences then, are duplicated genes that lost the exons coding for connecting peptide, transmembrane and cytoplasmic regions. This results in genes encoding soluble molecules like *Bain-ZE*0201* that possesses a putative signal peptide of 26 amino acids similar to the other class I ZE sequences. A second deletion event resulted in aberrant leader sequences like *Bain-ZE*0501*. This is supported by the observation that the *Bain-ZE*0201* and *Bain-ZE*0501* possesses a normal or an aberrant leader sequence, respectively, but the remainder of sequences shows 98% nucleotide similarity including de 3'-untranslated region and intron-6 (data not shown). Exon-1a of the aberrant leader peptide *Bain-ZE*0501* shows 75% nucleotide similarity to exon-4 of *Bain-ZE*0101* suggesting a deletion event in a region between two closely linked class I *ZE* genes that arose from gene duplication. The hexaploid status of African 'large' barb[68] may explain the presence of these aberrant class I *ZE* sequences in this species. Deletions, insertions and amino acid replacements in a single gene will not result in immediate lethality due to the presence of multiple gene copies in this species. Redundancy in the genome of polyploid species allows duplicated genes to persist as functional genes over a long period of time, although they may accumulate non-deleterious mutations or alternatively become pseudogenes[97].

The phylogeny of cyprinid fish class I genes is reviewed by STET *et al.*[74]. At that time we only identified two partial class I *ZE* lineage sequences (*Cyca-ZT2*

and *Cyca-Zr3*) in common carp. These sequences were considered to be non-classical based on clustering of these sequences with common carp and goldfish non-classical class I Z genes. With more data available now, the classical nature of these sequences becomes clear. The relatively close relationship of the ZE lineage with common carp and goldfish non-classical class I Z genes (ZA, ZB, ZC and ZD) suggests a common ancestor. However, non-classical Z lineage genes have not been identified in zebrafish suggesting that the non-classical Z lineage arose after divergence of the genera *Danio* and *Cyprinus*. This would suggest that these Z lineage genes should be present in African 'large' barb. However, this species has not been studied to that extent. An alternative for the absence of these Z lineage genes in zebrafish may be that they are simple not yet identified. An explanation for the lack of evidence of these genes might be the approach that assumes a conservation of the alpha-3 domain as is observed for mammalian class I sequences. In this study we demonstrated that the class I lineage ZE sequence show considerable divergence in the alpha-3 domains and there fore the above approach is flawed.

The cyprinid ZE sequences evolved in a *trans*-species fashion suggesting an orthologous relationship. The number of ZE loci within a species and whether some ZE loci arose from gene duplications within species after divergence of the genera *Danio*, *Cyprinus* and *Barbus* is unclear from these data. However, two zebrafish sub-clusters within the ZE clade suggest at least two loci in zebrafish. The four different ZE sequences identified in single zebrafish individuals and the presence of two is consistent with the fact that this species is diploid[66].

Several studies estimated that the genera, *Danio, Cyprinus* and *Barbus* diverged 50 and 30 million years ago, respectively[69-71]. The synonymous mutation rate at primate MHC loci seems to be similar to other loci with an average rate of 2.3 $*10^{-9}$ synonymous substitutions per synonymous site per year for HLA loci A, B and C[231]. Divergence time calculations using this mutation rate and the rate of synonymous substitution per synonymous site between cyprinid ZE lineage sequences estimated~100 million year for zebrafish and African 'large' barb or common carp and~40 million years for common carp and African 'large' barb. Divergence time calculation using only exon-2, exon-3 or exon-4 sequences all resulted in comparable divergence time estimates. Analyses of intron sequences, which may provide more reliable divergence time estimates, also resulted in comparable values when a substitution rate of 2.3 $*10^{-9}$ substitutions per site

was used. These time estimates suggest the presence of an ancestral class I *ZE* gene before separation of the three cyprinid genera, *Danio, Cyprinus* and *Barbus*.

What distinguishes the class I *ZE* genes from the counterparts in primates is presence of this lineage in three teleost genera being maintained for up to 100 million years. In contrast HLA class I lineages can only be recognised in great apes, which diverged from humanoids ~6 million years ago[45]. Ancient classical class I lineage maintained up to 20 million years old are also described for other bony fish[232]. Bony fish class I and class II genes, unlike all other jawed vertebrate species, are located on different linkage groups[48,53,191,192]. The lack of linkage between the class I and II genes must have influenced the evolution of these genes. Imperative in this respect is the fact whether the class I *ZE* sequences are linked to other class I genes or class II genes. This will be clarified within the near future by the zebrafish genome project that will reveal the linkage group of these class I *ZE* lineage genes

The maintenance of the *ZE* lineage for up to 100 million years and the unusual conservation of the peptide binding domains not only within species but also across species highlight the importance of their function. Although these domains show an unusual high conservation at the amino acid level, each domain exhibited a high degree of nucleotide diversity as shown by divergence time estimates based on the level of synonymous substitutions. The conservation of the alpha-1 and alpha-2 domains may relate to recognition of highly conserved molecular patterns derived from pathogens common to the three cyprinid species. Recognition of these conserved molecular structures might be the driving force to conserve the alpha-1 and alpha-2 domains in cyprinids.

ACKNOWLEDGEMENTS

The authors would like to thank MARTIN DE GRAAF and NAND SIBBING for providing tissue samples of a Lake Tana *Barbus intermedius* individual and JOHAN KOELMAN for his assistance in identifying the novel sequences in common carp.

CHAPTER 6

GENERAL DISCUSSION

Corine P. Kruiswijk

Cell Biology and Immunology Group,
Wageningen Institute of Animal Sciences, Wageningen University,
PO box 338, 6700 AH Wageningen, The Netherlands.

INTRODUCTION

The history of major histocompatibility complex (MHC) molecules starts at the beginning of the previous century with the discovery of their role in transplant rejection[1]. However, their biological function remained unknown until the end of the sixties[2,3]. In the eighties with the introduction of new molecular biology techniques and the improvement of others, the typing, identification and structural data of MHC molecules were boosted. MHC genes were isolated from representatives of major vertebrate taxa, including cartilaginous fish, bony fish, amphibians, birds, and mammals. In fact all of the genes that define the adaptive immune system are present from the oldest jawed vertebrates (*fig. 1A*), the cartilaginous fish onward (reviewed[233]). However, extensive scrutiny in older jawless vertebrates (Agnathans; hagfish and lamprey) has yielded no such genes that together make up an adaptive immune system. Furthermore, these jawless fish seem to lack the primary and secondary lymphoid organs (thymus and spleen) found in all jawed vertebrates. Thus, the adaptive immune system, including the MHC as we know it, seemingly arose rather abruptly in its entirety over an evolutionary short period of time.

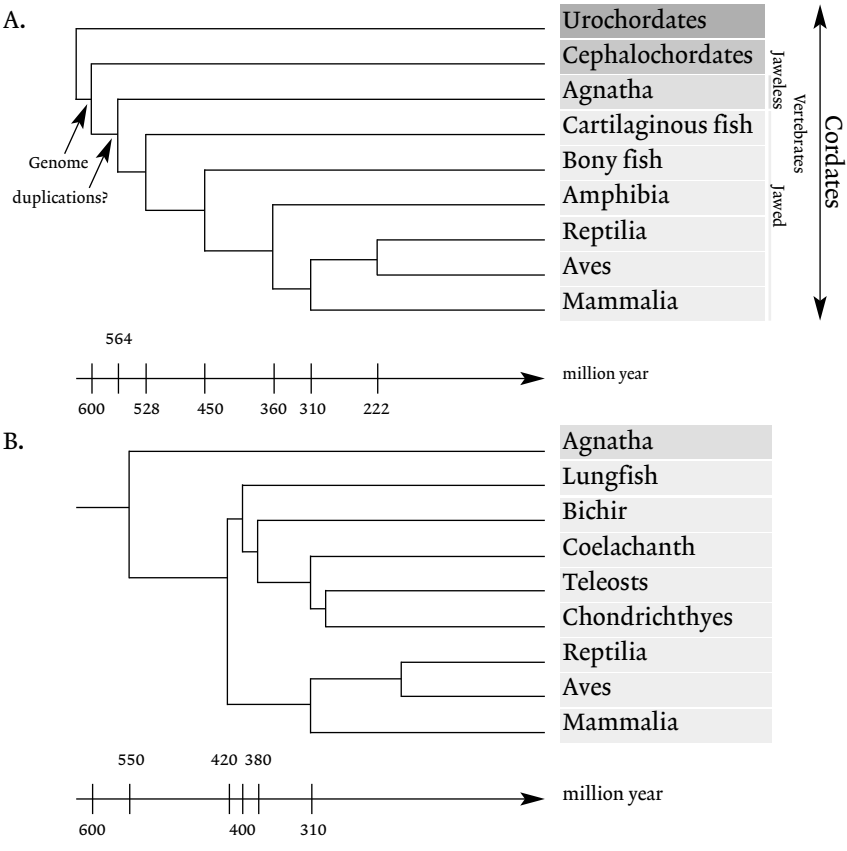


Figure 1: Phylogenetic relationship among extant Chordates. A) divergence times are based on molecular data compiled by KUMAR and HEDGES[250]. Two rounds of genome-wide duplications as indicated are according to the 2R hypothesis[238]. B) divergence times are based on molecular data compiled by RASMUSSEN and ARNASON[56].

Genome-wide duplications have been suggested to contribute to the abrupt emergence of the adaptive immune system. Genes from invertebrates often have up to four copies in vertebrates[234,235]. Such copies are often found in syntenic regions of the human genome[57,236]. The most widely accepted hypothesis of genome-wide duplications is the 2R hypothesis (reviewed[237,238]), which assumes two

rounds of genome duplication. The first round in a common ancestor of all vertebrates and a second round in a common ancestor of jawed vertebrates after its separation from the jawless fishes (*fig. 1A*). However, whether the 2R hypothesis is correct is still subject to intense controversy (reviewed[238]). Two rounds of genome-wide duplication of a proto-MHC region would give rise to the MHC and its three paralogous regions as observed in the vertebrate genome[236,239,240]. Recent observations in the amphioxus genome revealed a region in this species, which shows some level of synteny with the MHC[239].

Questions on the origin and evolution of MHC molecules can only be addressed by examining the molecular evolution and comparative genomics of the MHC of jawed vertebrates and putative molecules in jawless vertebrates and invertebrates that gave rise to the structure of class I and II molecules as we know them. Comparing difference and similarities relating to the MHC of species representing different vertebrate taxa will provide insight into its origin. Here we focus on the implications of recent findings in three cyprinid fishes, the African 'large' barb, common carp and zebrafish.

THE TELEOSTEAN UNMHC

All jawed vertebrates are equipped with full-fledged MHC genes. However, studies of the MHC organisation in vertebrates indicated that the genes are not organised in the same way in all animals. Organisation can differ among and between mammalian as well as non-mammalian vertebrates. The oldest jawed vertebrates, cartilaginous fishes, possess linked class I and class II genes similar to the tetrapods. Remarkably, teleosts do not possess linked class I and class II genes. The lack of linkage of class Ia and class II loci in teleosts has been demonstrated in zebrafish[84], common carp[241], Atlantic salmon[50], platyfish[242] medaka[49] and pufferfish[243]. In addition, class IIB genes can be found on different linkage groups in several teleost species[49,84].

The cartilaginous fishes (Chondrichthyes: sharks, skates and rays) are commonly accepted as being sister group to the other extant Gnathostomata (jawed vertebrates). Teleost fishes have estimated to be diverged within 450 million year from cartilaginous fish, while tetrapods arose approximately 360 million year ago from a bony fish ancestor (*fig. 1A*). The fact that cartilaginous fishes possess linked class I and class II genes[54] suggests that the lack of linkage between class I and class II genes in teleosts is a derived character.

Several possibilities have been suggested for the absence of linkage of class I and class II loci in teleostean fishes. First, two rounds of genome-wide duplications have been suggested to result in two or four duplicates of the ancestral MHC-bearing region[57-60] that in the teleostean lineage over time has lost either class I loci or class II loci. Second, block duplication of parts of a chromosome bearing an MHC region followed by translocation and subsequent loss of loci. Third, translocation of class II loci from a prototype MHC to other chromosomes in the ancestor of the teleosts[61].

However, although the Chondrichthyes are commonly accepted as being sister group to the other extant Gnathostomata (jawed vertebrates), the basal position of the Chondrichthyes relative to the other Gnathostomes is still under debate[55,56]. Phylogenetic analyses of MHC class Ia exon-4, class IIA, and class IIB exon-3 sequences of representatives of the major jawed vertebrate species support the opposite, namely that the Actinopterygii are basal relative to the Chondrichthyes (*fig. 2*). Phylogenies of MHC class Ia exon-4, class IIA and class IIB exon-3 sequences suggest that the teleosts shared a common ancestor with a lineage that gave rise to all other Gnathostomes. Such ancestral position of the teleosts may implicate that the unlinked class I and class II loci is the ancestral organisation, not the observed linkage of class I and class II seen in other Gnathostomes. Teleost species do possess class Ia loci closely linked to *LMP* and *TAP* genes in rainbow trout[53], zebrafish[210], medaka[49], and Alantic salmon[52] (reviewed[60]). Translocation of class II loci into this region may have given rise to an MHC. Alternatively, class I and class II loci were initially linked and located, most likely, in four paralogous regions. Each region lost either the class Ia or the class II loci after the separation of other Gnathostomes. All other Gnathostomes, in contrast, lost all class Ia and class II loci located in three of the four paralogous regions, thus leaving only a single MHC. These observations illustrate the importance to clarify the Gnathostome relationships in particular the position of the Chondrichthyes, in order to shed light on the origin of the MHC as we know it in mammals.

IMPLICATIONS OF POLYPLOIDY FOR EVOLUTION, INHERITANCE AND USAGE OF MHC GENES

In cyprinids diploid, tetraploid, and hexaploid species have been identified. Zebrafish (*Danio rerio*[66], 2n=50), common carp (*Cyprinus carpio*[67], 2n=100) and

the African 'large' barb (*Barbus intermedius*[68], 2n=150) including the Lake Tana African 'large' barb species flock, represent such different ploidy status. These cyprinid species, representing highly divergent cyprinid genera, *Danio, Cyprinus* and *Barbus*, are estimated to have diverged 50 and 30 million years ago, respectively[69-71]. The different ploidy status may have implications for evolution, inheritance and usage of MHC genes (Chapter 2).

In species with higher ploidy genomes, an increased number of MHC genes, known to be co-dominantly expressed in mammals, are expected. However, expression of too many MHC alleles in a single individual is thought to be deleterious. A high number of expressed class I sequences has been demonstrated in a single Atlantic cod individual[186]. However, there seems to be a penalty on expression of too many alleles in this species. Serum of Atlantic cod has an unusual abundance of natural antibodies that react with a multitude of different antigens[244], whereas other teleost species in general show a lower serum immunoglobulin concentration[245]. The fact that cod has a high level of serum immunoglobulins may be a compensatory mechanism of having too many class I alleles expressed, compensating the cellular arm of the immune response.

In other studies only a limited number of expressed class I sequences, with a maximum of four, are found[107,117,119,171,176,246]. In zebrafish and common carp the norm of four expressed sequences seems to be the case for class IIA and class IIB sequences[50,104,105]. These findings of only a limited number of expressed MHC genes is also observed in the African 'large' barb. A study on MHC class Ia and class IIA and class IIB of a hexaploid African 'large' barb individual showed that five class I, four class IIA and four class IIB genes were shown to be expressed (Chapter 3).

In the African 'large' barb individual studied, the number of expressed class IIA is similar to the number of expressed class IIB sequences. In addition, phylogenetic analyses indicate the presence of two class IIA and two class IIB loci. This suggests that an alpha or beta chain from a single class II locus combine with one another. This contrasts the situation in cichlids in which the number of class IIB exceeds the number of class IIA loci, suggesting that an alpha chain combines with more than one beta chain[184,247]. Such discrepancy between the number of class IIA and class IIB chain genes, are also observed in humans[165].

In the African 'large' barb, studied in chapter 3, only five class I, four class IIA, and four class IIB genes present at the genomic level were shown to be expressed,

indicating that the ploidy status does not correlate with the presence and expression of MHC genes. This suggested that gene silencing is most likely a result of physical exclusion of complete genes. However, functional silencing also seems to have played a role in reducing the number of expressed class II genes, as class IIB pseudogenes were identified in the genome of this African 'large' barb individual. These observations, that both mechanisms are operative, are corroborated by another study in which it has been shown that a *Barbus acutirostris* individual expressed only three out of the six identified class IIB sequences present in its genome[70].

Silencing mechanisms acting upon MHC genes have been observed in other species. Physical silencing reduced the number of class I and class II genes in polyploid *Xenopus* species[79,80]. Another mechanism to reduce the number of expressed MHC molecules is functional silencing. For example, in the zebrafish genome only two genomically *bona fide* loci, *DAA/DAB* and *DEA/DEB*, are found, while several class IIB loci are present. The *DEA/DEB* genes are only present in some strains and expression of the *DEA/DEB* locus has never been demonstrated, suggesting functional silencing perhaps due to promoter disruption. The remaining loci were truncated or corrupted in their coding regions and thus functionally silenced pseudogenes[127]. An alternative explanation could be translocation of the class IIB genes[61]. The loss of physical linkage to the class IIA gene may render the class IIB genes non-functional and over-time turn them into pseudogenes.

TELEOST CLASS I AND II EVOLUTION DIFFERS FROM THE MAMMALIAN PARADIGM

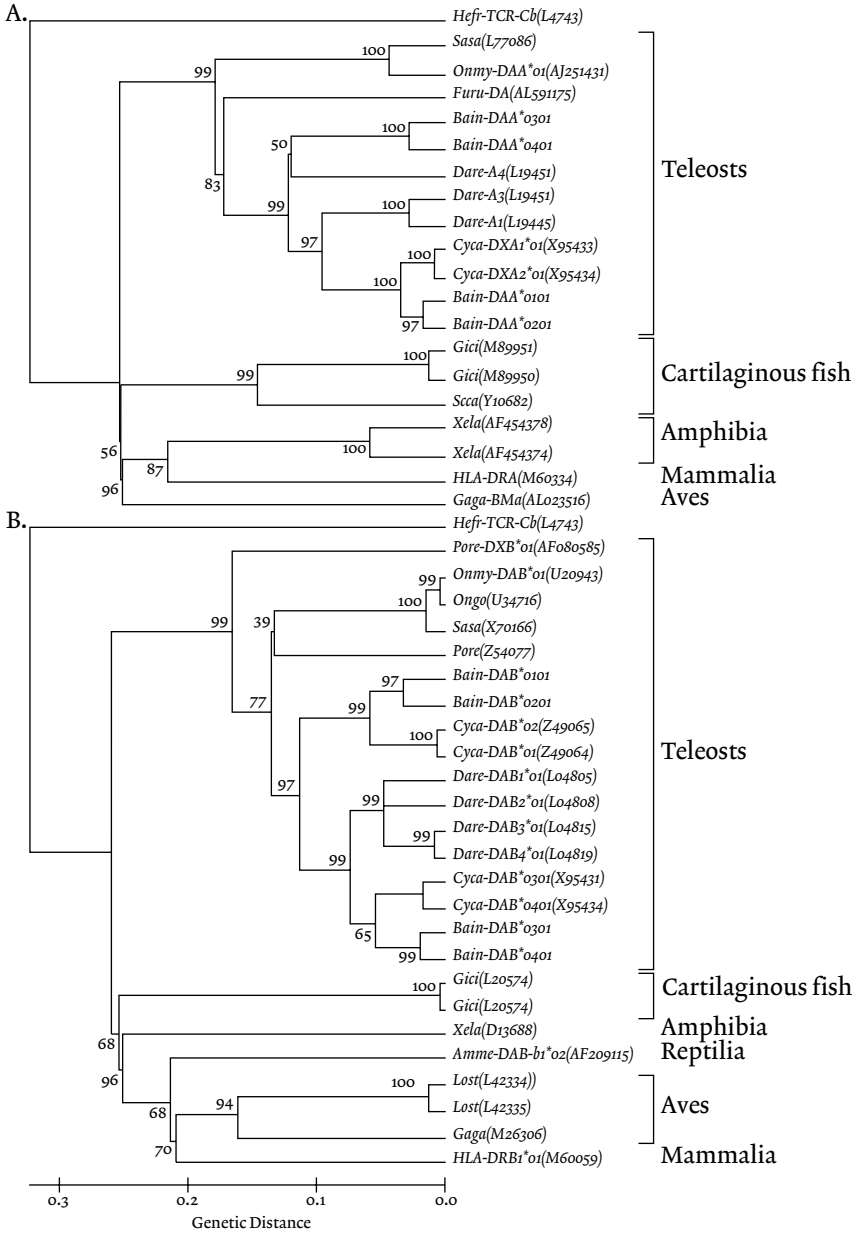
In the teleost genome, the location of class I and class II genes in different linkage groups allows them to evolve independently, presumable under different selection pressures. In a study on the evolution of class I and II genes in a natural model, the Lake Tana African 'large' barb species flock revealed such different selection pressures. The most striking observation is that the Lake Tana African 'large' barb species and the riverine African 'large' barb species showed no sharing of class II *DAB*01* intron-1-exon-2 sequences. In contrast, class I *UA* exon-3 sequences were shared by different African 'large' barb species (Chapter 4). Although sharing of class II *DAB*01* exon-2 sequences was not observed

between the African 'large' barb species studied, it can not be excluded with absolute certainty that sequences occasionally are shared between Lake Tana African 'large' barb species.

Lake Tana is an isolated system thought to be 5 million years old. This aquatic system is created by a tectonic event that blocked the outlet of the Blue Nile with a 40 m high waterfall. During isolation, 15 novel African 'large' barb species emerged within the lake as a result of adaptive radiation of the ancestral *Barbus intermedius* population. The Lake Tana African 'large' barb species occupy different ecological niches and exploit different food sources, which provide different selection pressures on MHC genes. These selection pressures seem to have resulted in a lack of sharing of class IIB alleles between species, while they do share class I alleles. Due to partly overlapping ecological niches and food sources between the Lake Tana African 'large' barb species, some sharing of class IIB alleles may be present among the species. However, individuals belonging to the species *B. acutirostris* and *B. megastoma*, both real piscivores, which is rare among cyprinids, did not share class IIB intron-1 and exon-2 sequences. Also, interspecies hybridisation of species sharing temporal and spatial spawning areas[248] may contribute to the possibility of shared class IIB sequences. For example, *B. acutirostris, B. megastoma, B platydorsus* spawn during august and september in the Gumara river, a tributary of Lake Tana[248]. Again, we did not find sharing of class IIB genes between these species (chapter 4), reiterating the fact that each Lake Tana African 'large' barb has their own set of class IIB genes.

Class IIB alleles were not studied in the species *B. crassibarbis, B. gorgorensis, B. gorguari, B. longissimus,* and *B. dainellii*. Therefore, it can not be excluded that these species might share class IIB alleles. These species spawn during august and september within the lake in contrast to most others that spawn in the tributaries of the lake. However, these species use different food sources, with *B. crassibarbis* being a real benthivore, *B. gorgorensis* a specialised moluscivore, *B. gorguari* a partial piscivore, and *B. longissimus* and *B. dainellii* real piscivores. The possibility of shared of class IIB sequences by these species does not undermine the general observation that the turn-over of class II alleles in the Lake Tana African 'large' barb species is much higher than that of class I alleles.

Although it is tentative to speculate, the tight linkage of the class I genes with genes encoding proteins involved in generating and loading of peptides (*TAP,*


A.
Hefr-TCR-Cb(L4743)
Sasa(L77086)
Onmy-DAA*01(AJ251431)
Furu-DA(AL591175)
Bain-DAA*0301
Bain-DAA*0401
Dare-A4(L19451)
Dare-A3(L19451)
Dare-A1(L19445)
Cyca-DXA1*01(X95433)
Cyca-DXA2*01(X95434)
Bain-DAA*0101
Bain-DAA*0201
Gici(M89951)
Gici(M89950)
Scca(Y10682)
Xela(AF454378)
Xela(AF454374)
HLA-DRA(M60334)
Gaga-BMa(AL023516)
Teleosts
Cartilaginous fish
Amphibia
Mammalia
Aves
B.
Hefr-TCR-Cb(L4743)
Pore-DXB*01(AF080585)
Onmy-DAB*01(U20943)
Ongo(U34716)
Sasa(X70166)
Pore(Z54077)
Bain-DAB*0101
Bain-DAB*0201
Cyca-DAB*02(Z49065)
Cyca-DAB*01(Z49064)
Dare-DAB1*01(L04805)
Dare-DAB2*01(L04808)
Dare-DAB3*01(L04815)
Dare-DAB4*01(L04819)
Cyca-DAB*0301(X95431)
Cyca-DAB*0401(X95434)
Bain-DAB*0301
Bain-DAB*0401
Gici(L20574)
Gici(L20574)
Xela(D13688)
Amme-DAB-b1*02(AF209115)
Lost(L42334))
Lost(L42335)
Gaga(M26306)
HLA-DRB1*01(M60059)
Teleosts
Cartilaginous fish
Amphibia
Reptilia
Aves
Mammalia
0.3
0.2
0.1
0.0
Genetic Distance

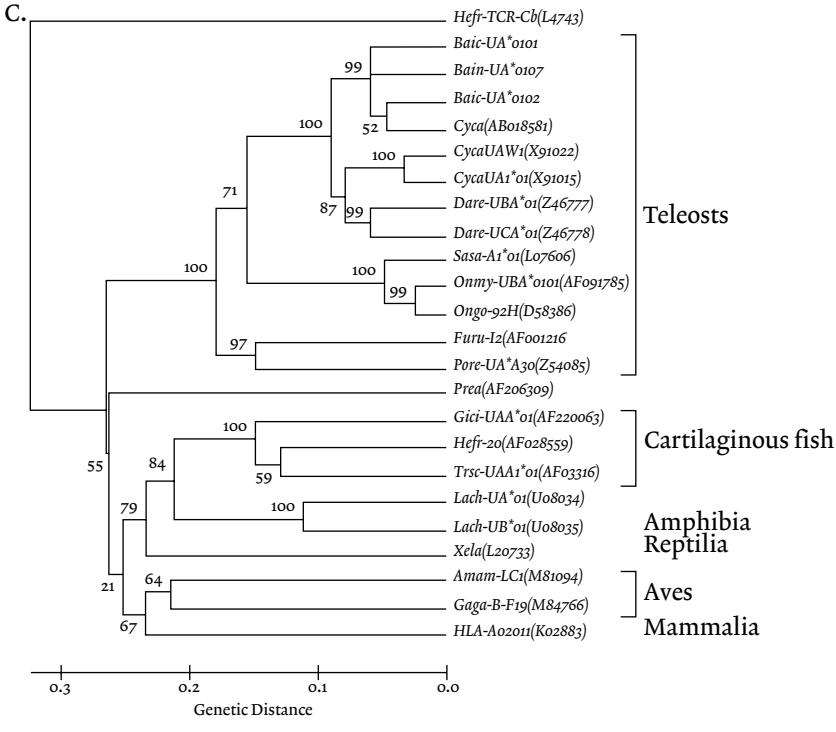


Figure 2: Neighbour-joining trees based on nucleotide sequences of class IIA (A) and class IIB (B) exon-3 genes and class Ia exon-4 genes (C). Representative of classical class I sequences of the major vertebrate classes are included. The horned shark TCR gene was used as an outgroup. The *numbers* at branch nodes represent the confidence level of 1000 bootstrap replications. Accession number are indicated between *brackets*.

LMP) may constrain the evolution of class I genes. Alternatively, class I molecules might play a less important role in the immune system of fish compared to the class II molecules. Hence, class II molecules are probably more susceptible to evolutionary changes compared to class I molecules.

A more rapid evolution of MHC class II is also observed in brown trout and rainbow trout. Although these species shared class I lineages, class II genes, on the other hand, do not form such old lineages[115,232]. Other observations suggesting a rapid evolution of teleostean class II genes is the variable numbers of class II genes found in different haplotypes in cichlids[184]. This is in contrast to mammalian MHC class I genes which seem to evolve more rapidly than class II genes. Allelic HLA class I lineages were shown to be maintained for up to 6 million years, while certain hla class II lineages were shown to be maintained for up 35 million years[43-45].

A NOVEL CLASS I ZE LINEAGE IN CYPRINID FISH

Recently, an ancient class I lineage was found in three different teleost species, zebrafish, common carp, and African 'large' barb (chapter 2 and 5). The presence in representative species of three distantly related genera, *Danio, Cyprinus* and *Barbus* indicate that this novel class I *ZE* lineage is maintained at least 80 million years. Imperative with respect to the fact that teleosts have unlinked class I and class II genes is whether the recently discovered class I *ZE* lineage is linked to other class Ia genes or class II genes.

Although this novel lineage seems to be phylogenetically more related to other non-classical class I Z lineages, to date only identified in common carp (*Cyprinus carpio* L.) and ginbuna crucian carp (*Carassius auratus* L.), the ZE lineage has multiple features in favour of a classical nature. They possess the conserved amino acid residues involved in peptide binding of mammalian classical class I molecules and were ubiquitous expressed. Variability among *Dare-ZE* sequences from several zebrafish individuals seems to be generated by positive selection acting upon putative peptide binding residues.

In contrast to classical class I molecules which are only found within the MHC, non-classical class I genes are also found at some distance from the MHC or on a different chromosome. In non-mammalian vertebrates, a 'true class Ia region', has been proposed that comprises closely linked class Ia, *LMP* and *TAP* genes. Such a linkage is hypothesised to provide advantageous co-evolution of these

genes[60] and is observed in several teleost species. With respect to the classical nature of the novel zebrafish class I *ZE* lineage, it seems likely that it resides in close vicinity of this true class Ia region in linkage group 19 of zebrafish[48,191,249]. However, recent studies on the class I region and the flanking regions in zebrafish genome did not indicate the presence of such an additional class I locus[191,210,249]. This suggests that the ZE locus resides on the same linkage group at some distance from the class Ia region or on a different linkage group. A classical class I gene not linked to the *TAP* and *LMP* genes is observed in medaka. In the medaka genome the class I *UCA* gene resides on a different linkage group than the *TAP* and *LMP* genes. In addition the *UCA* locus is not linked to one of the class II genes indicating that a classical class I locus can be found separate from both *TAP*/*LMP* and class II loci. However, the position of this gene might be a derived character in medaka fishes. A similar dispersion of loci is observed for some class III equivalents of human MHC-encoded genes in the zebrafish genome[191] and in the medaka genome[49], were they are found on several linkage groups.

Another putative location for this novel class I *ZE* locus might be in close vicinity of one of the class II genes residing on three different linkage groups in zebrafish[191]. Such an organisation of a classical class I gene in close vicinity of a class II gene may represent an ancient synteny, hypothesised to be present in a jawed vertebrate ancestor[60]. The location of this novel class I *ZE* gene in the zebrafish genome, either residing in close vicinity of class II or class III or on an altogether different linkage group has implications for syntenic relationships and co-evolution of MHC loci. In the near future the zebrafish sequencing consortium most likely will reveal the location of this novel class I *ZE* gene. Similarly, this will reveal the presence of non-classical class I *Z* loci sequences in zebrafish and their location in the zebrafish genome. Such observation may shed some light on questions addressing the origin of class Ia and Ib molecules. To date numerous functional non-classical class I genes have been identified in human and other mammals[23]. Such non-classical MHC class I molecule were shown to be present in a number of non-mammalian vertebrates studied including the oldest jawed vertebrates, cartilaginous fishes[171]. In addition, non-classical class I *Z* lineage loci are present in common carp (*Cyprinus carpio* L.) that diverged from the genera *Danio* approximately 80 million years. The phylogeny of *Z* loci suggested a common ancestor of these *Z* lineages. The novel classical *ZE* lineage, which is present in common carp and zebrafish suggests that

non-classical Z lineage occurred after the split of the two genera, most likely as a result of genetic drift of a duplicated Z lineage gene. Characterisation of the non-classical Z loci in zebrafish will in the future shed new light on the evolution of classical and non-classical class I genes.

UNUSUAL CONSERVATION OF THE CYPRINID CLASS I ZE PEPTIDE BINDING DOMAINS

Although the novel class I ZE lineage possesses all features of classical class I molecules, some differences are observed. The most enigmatic feature of this lineage is their unusual conservation of the peptide-binding domains within and between three distantly related species. Although overall conserved, in zebrafish this lineage was characterised by sequence variability located within the peptide binding pocket, which seemed to be generated by positive Darwinian selection. Remarkably, polymorphism was not superimposable on mammalian class Ia variability. Such aberrant polymorphism is also observed for *MICA* and *MICB* genes[15]. In addition, 3D modelling indicated that many variable positions were located in loop structures instead of in the alpha-helices and beta-strands of the peptide binding domains. With respect to classical class I function it is odd that most variable positions are located in loop structures since it is well established that amino acids at variable positions in the alpha-helices and beta-strands of the peptide binding domains interact with the peptides. Variability shapes the specificity of the peptide binding pocket. The lack of variability in the alpha-helices and beta-strands of the peptide binding pocket of ZE sequences indicate binding of highly conserved molecular patterns derived from pathogens common to the three cyprinid species. Recognition of these conserved molecular structures might be the driving force to conserve the alpha-1 and alpha-2 domains in cyprinids. The discovery of several class I-like molecules in human and their structural and functional extremes revealed differences in the underlying class I-like fold and showed that this type of folding can interact with diverse ligands through binding sites that together cover much of the membrane distal parts of the molecule[23]. In this respect, variability observed in loop structures of ZE peptide binding pocket, which seemed to be generated by positive Darwinian selection, may interact with the ligand bound. Alternatively, the polymorphic residues could have a

function in interacting with the receptor recognising this class I molecule when complexed with its ligand.

In conclusion, the data presented in this study highlight the importance of studying MHC genes both in a molecular and phylogenetic perspective. Detailed molecular analyses of the novel *ZE* genes revealed that these genes encode molecules which are classical in nature, although their phylogenetic position suggested that they were orthologous to non-classical *Z* lineage genes. The conservation of the peptide binding domains of ZE molecules is unusual, and impinge on the fact that MHC class I molecules in teleosts do not conform to the mammalian paradigm. This is also substantiated by the phylogenetic analyses of the class I and class II genes of the African 'large' barb species presented in this study. Although the *trans*-species hypothesis is valid for both class I and class II genes, class II alleles behave differently compared to class I alleles. Speciation events seems to result in species-specific sets of class II alleles, whereas several class I alleles are shared among different species. This suggests that class II molecules are subject to intense selection during speciation events, leading to non-overlapping partitioning of class II alleles among African 'large' barb species, occupying different ecological niches. Class I alleles on the other hand seem to be less receptive to intense selection favouring extensive polymorphism. This could probably be due to the fact that class I molecules need to be conserved to perform their function. This seems to suggest that the peptide binding specificities of class I molecules are much broader than those of class II molecules are. Alternatively, the role of the cellular immune response in fish is of less importance compared to the humoral response. The first is controlled by antigen presentation in the context of class I molecules, whereas the latter by class II antigen presentation. It should be taken into account that these tentative conclusions are based on studies in cyprinid fishes. The abundance of teleost species and their long evolutionary history may have resulted in different characteristics of the immune system among different classes. The Actinopterygii are by far the most diversified group of vertebrates. They can be found in almost every conceivable aquatic habitat, ranging from the ocean floors to the headwaters of mountain streams and from hotsprings to subzero polar waters. Certainly, this global radiation must have shaped the immune system in various ways.

REFERENCES

1. Little, C. C., E. E. Tyzzer (1916) Further experimental studies on the inheritance of susceptibility to transplantable tumour, carcinoma (jwa) of the Japanese waltzing mouse. *J med res 33: 393-453.*
2. McDevitt, H. O., A. Chinitz (1969) Genetic control of the antibody response: relationship between immune response and histocompatibility (h-2) type. *Science 163: 1207-1208.*
3. Shevach, E. M., W. E. Paul, I. Green (1972) Histocompatibility-linked immune response gene function in guinea pigs. Specific inhibition of antigen-induced lymphocyte proliferation by alloantisera. *J exp med 136: 1207-1221.*
4. Ploegh, H. L., H. T. Orr, J. L. Strominger (1980) Molecular cloning of a human histocompatibility antigen cdna fragment. *Proc natl acad sci USA 77: 6081-6085.*
5. Bourlet, Y., G. Behar, F. Guillemot, N. Frechin, A. Billault, A. M. Chausse, R. Zoorob, C. Auffray (1988) Isolation of chicken major histocompatibility complex class ii (b-l) beta chain sequences: comparison with mammalian beta chains and expression in lymphoid organs. *Embo j 7: 1031-1039.*
6. Hashimoto, K., T. Nakanishi, Y. Kurosawa (1990) Isolation of carp genes encoding major histocompatibility complex antigens. *Proc natl acad sci USA 87: 6863-6867.*
7. Saper, M. A., P. J. Bjorkman, D. C. Wiley (1991) Refined structure of the human histocompatibility antigen hla-a2 at 2.6 Å resolution. *J mol biol 219: 277-319.*
8. Madden, D. R., J. C. Gorga, J. L. Strominger, D. C. Wiley (1992) The three-dimensional structure of hla-b27 at 2.1 Å resolution suggests a general mechanism for tight peptide binding to mhc.

Cell 70: 1035-1048.

9. Venkatesh, B., Y. Ning, S. Brenner (1999) Late changes in spliceosomal introns define clades in vertebrate evolution. *Proc natl acad sci USA 96: 10267-10271.*

10. Dixon, B., S. H. M. van Erp, P. N. S. Rodrigues, E. Egberts, R. J. M. Stet (1995) Fish major histocompatibility complex genes: an expansion. *Dev comp immunol 19: 109-133*.

11. Shum, B. P., P. M. Mason, K. E. Magor, L. R. Flodin, R. J. M. Stet, P. Parham (2002) Structures of two major histocompatibility complex class I genes of the rainbow trout (*Oncorhynchus mykiss*). *Immunogenetics 54: 193-199*.

12. Brown, J. H., T. S. Jardetzky, J. C. Gorga, L. J. Stern, R. G. Urban, J. L. Strominger, D. C. Wiley (1993) Three-dimensional structure of the human class II histocompatibility antigen hla-dr1. *Nature 364: 33-39*.

13. Vilches, C., P. Parham (2002) Kir: diverse, rapidly evolving receptors of innate and adaptive immunity. *Annu rev immunol 20: 217-251.*

14. Margulies, D. H. (1999) The major histocompatibility complex. *In: W. E. Paul (ed), Fundamental Immunology*. Lippincott-Raven Publishers, Philadelphia. *p 263-285.*

15. Stephens, H. A. (2001) Mica and micb genes: can the enigma of their polymorphism be resolved? *Trends immunol 22: 378-385.*

16. Feder, J. N., A. Gnirke, W. Thomas, Z. Tsuchihashi, D. A. Ruddy, A. Basava, F. Dormishian, R. Domingo, Jr., M. C. Ellis, A. Fullan, L. M. Hinton, N. L. Jones, B. E. Kimmel, G. S. Kronmal, P. Lauer, V. K. Lee, D. B. Loeb, F. A. Mapa, E. McClelland, N. C. Meyer, G. A. Mintier, N. Moeller, T. Moore, E. Morikang, R. K. Wolff (1996) A novel mhc class I-like gene is mutated in patients with hereditary haemochromatosis. *Nat genet 13: 399-408.*

17. Driscoll, J., M. G. Brown, D. Finley, J. J. Monaco (1993) Mhc-linked lmp gene products specifically alter peptidase activities of the proteasome. *Nature 365: 262-264.*

18. Elliott, T. (1997) Transporter associated with antigen processing. *Adv immunol 65: 47-109.*

19. Kropshofer, H., G. J. Hammerling, A. B. Vogt (1999) The impact of the non-classical mhc proteins hla-dm and hla-do on loading of mhc class II molecules. *Immunol rev 172: 267-278.*

20. Klein, J., F. Figueroa (1986) Evolution of the major histocompatibility complex. *Crit rev immunol 6: 295-386.*

21. Howard, J. C. (1987). *In: Klesoe, G., D. H. Schulze (eds), Evolution and vertebrate immunity*. University of Texas Press, Texas.

22. Braud, V. M., D. S. Allan, A. J. McMichael (1999) Functions of non-classical mhc and non-mhc-encoded class I molecules. *Curr opin immunol 11: 100-108.*

23. Strong, R. K. (2000) Class (I) will come to order-not. *Nat struct biol 7: 173-176.*

24. Sanchez, L. M., A. J. Chirino, P. Bjorkman (1999) Crystal structure of human zag, a fat-depleting factor related to mhc molecules. *Science 283: 1914-1919.*

25. Kennedy, M. W., A. P. Heikema, A. Cooper, P. J. Bjorkman, L. M. Sanchez (2001) Hydrophobic ligand binding by Zn-alpha-2-glycoprotein, a soluble fat-depleting factor related to major histocompatibility complex proteins. *J biol chem 276: 35008-35013.*

26. Li, P., D. L. Morris, B. E. Willcox, A. Steinle, T. Spies, R. K. Strong (2001) Complex structure of the activating immunoreceptor nkg2d and its mhc class I-like ligand mica. *Nat immunol 2: 443-451.*

27. Lebron, J. A., M. J. Bennett, D. E. Vaughn, A. J. Chirino, P. M. Snow, G. A. Mintier, J. N. Feder, P. J. Bjorkman (1998) Crystal structure of the hemochromatosis protein hfe and characterisation of its interaction with transferrin receptor. *Cell 93: 111-123.*

28. Martin, W. L., A. P. West, Jr., L. Gan, P. J. Bjorkman (2001) Crystal structure at 2.8 Å of an FcRn/heterodimeric Fc complex: mechanism of pH-dependent binding. *Mol cell 7: 867-877.*

29. Hughes, A. L., M. Yeager, A. E. Ten Elshof, M. J. Chorney (1999) A new taxonomy of mammalian mhc class i molecules. *Immunol today 20: 22-26.*

30. Kenter, M., N. Otting, J. Anholts, M. Jonker, R. Schipper, R. E. Bontrop (1992) Mhc-drb diversity of the chimpanzee (*Pan troglodytes*). *Immunogenetics 37: 1-11.*

31. Gaur, L. K., G. T. Nepom (1996) Ancestral major histocompatibility complex drb genes beget conserved patterns of localised polymorphisms. *Proc natl acad sci USA 93: 5380-5383.*

32. Gyllensten, U. B., M. Sundvall, H. A. Erlich (1991) Allelic diversity is generated by intraexon sequence exchange at the drb1 locus of primates. *Proc natl acad sci USA 88: 3686-3690.*

33. Brunsberg, U., I. Edfors-Lilja, L. Andersson, K. Gustafsson (1996) Structure and organisation of pig mhc class ii drb genes: evidence for genetic exchange between loci. *Immunogenetics 44: 1-8.*

34. Doherty, P. C., R. M. Zinkernagel (1975) Enhanced immunological surveillance in mice heterozygous at the h-2 gene complex. *Nature 256: 50-52.*

35. Hughes, A. L., M. Nei (1988) Pattern of nucleotide substitution at major histocompatibility complex class i loci reveals overdominant selection. *Nature 335: 167-170.*

36. Hughes, A. L., M. Nei (1989) Nucleotide substitution at major histocompatibility complex class ii loci: evidence for overdominant selection. *Proc natl acad sci USA 86: 958-962.*

37. Hughes, A. L., M. K. Hughes (1995) Natural selection on the peptide-binding regions of major histocompatibility complex molecules. *Immunogenetics 42: 233-243.*

38. Takahata, N., M. Nei (1990) Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics 124: 967-978.*

39. Klein, J. (1987) Origin of major histocompatibility complex polymorphism: the *trans*-species hypothesis. *Hum immunol 19: 155-162.*

40. Gyllensten, U. B., H. A. Erlich (1989) Ancient roots for polymorphism at the hla-dq alpha locus in primates. *Proc natl acad sci USA 86: 9986-99890.*

41. Lawlor, D. A., F. E. Ward, P. D. Ennis, A. P. Jackson, P. Parham (1988) Hla-a and b polymorphisms predate the divergence of humans and chimpanzees. *Nature 335: 268-271.*

42. Mayer, W. E., M. Jonker, D. Klein, P. Ivanyi, G. van Seventer, J. Klein (1988) Nucleotide sequences of chimpanzee mhc class i alleles: evidence for *trans*-species mode of evolution. *Embo j 7: 2765-2774.*

43. Bontrop, R. E., N. Otting, N. G. de Groot, G. G. Doxiadis (1999) Major histocompatibility complex class ii polymorphisms in primates. *Immunol rev 167: 339-350.*

44. Boyson, J. E., C. Shufflebotham, L. F. Cadavid, J. A. Urvater, L. A. Knapp, A. L. Hughes, D. I. Watkins (1996) The mhc class i genes of the rhesus monkey. Different evolutionary histories of mhc class i and ii genes in primates. *J immunol 156: 4656-4665.*

45. Vogel, T. U., D. T. Evans, J. A. Urvater, D. H. O'Connor, A. L. Hughes, D. I. Watkins (1999) Major histocompatibility complex class i genes in primates: co-evolution with pathogens. *Immunol rev 167: 327-337.*

46. Guillemot, F., A. Billault, O. Pourquie, G. Behar, A. M. Chausse, R. Zoorob, G. Kreibich, C. Auffray (1988) A molecular map of the chicken major histocompatibility complex: the class ii beta genes are closely linked to the class i genes and the nucleolar organiser. *Embo j 7: 2775-2785.*

47. Flajnik, M. F., Y. Ohta, C. Namikawa-Yamada, M. Nonaka (1999) Insight into the primordial mhc from studies in ectothermic vertebrates. *Immunol rev 167: 59-67.*

48. Bingulac-Popovic, J., F. Figueroa, A. Sato, W. S. Talbot, S. L. Johnson, M. Gates, J. H.

Postlethwait, J. Klein (1997) Mapping of mhc class i and class ii regions to different linkage groups in the zebrafish, *Danio rerio*. *Immunogenetics 46: 129-134.*

49. Naruse, K., A. Shima, M. Nonaka (2000) Mhc gene organisation in the bony fish, medaka. *In: M. Kasahara (ed), Major histocompatibility complex: Evolution, structure, and function* Springer-Verlag, Tokyo. *p 91-109.*

50. Stet, R. J. M., R. Johnestone, P. Parham (1997) The unmhc of teleostean fish: segregation analyses in common carp and Atlantic salmon. *Hereditas 127: 169-170.*

51. Stet, R. J. M., B. D. Vries, K. Mudde, G. T. Hermsen, J. v. Heerwaarden, B. P. Shum, U. Grimholt (2002) Unique haplotypes of co-segregating major histocompatibility class iia and iib alleles in Atlantic salmon (*Salmo salar*) give rise to diverse class ii genotypes. *Immunogenetics 54: 320-331.*

52. Grimholt, U., F. Drabløs, J. S. M., b. Høyheim, R. J. M. Stet (2002) The major histocompatibility class i locus in Atlantic salmon (*Salmo salar* l.): polymorphism, linkage and protein modelling. *Immunogenetics, in press.*

53. Hansen, J. D., P. Strassburger, G. H. Thorgaard, W. P. Young, L. Du Pasquier (1999) Expression, linkage, and polymorphism of mhc-related genes in rainbow trout, *Oncorhynchus mykiss*. *J immunol 163: 774-786.*

54. Ohta, Y., K. Okamura, E. C. McKinney, S. Bartl, K. Hashimoto, M. F. Flajnik (2000) Primitive synteny of vertebrate major histocompatibility complex class i and class ii genes. *Proc natl acad sci USA 97: 4712-4717.*

55. Rasmussen, A. S., U. Arnason (1999) Phylogenetic studies of complete mitochondrial dna molecules place cartilaginous fishes within the tree of bony fishes. *J mol evol 48: 118-123.*

56. Rasmussen, A. S., U. Arnason (1999) Molecular studies suggest that cartilaginous fishes have a terminal position in the piscine tree. *Proc natl acad sci USA 96: 2177-2182.*

57. Kasahara, M. (1998) What do the paralogous regions in the genome tell us about the origin of the adaptive immune system? *Immunol rev 166: 159-175.*

58. Kasahara, M., M. Hayashi, K. Tanaka, H. Inoko, K. Sugaya, T. Ikemura, T. Ishibashi (1996) Chromosomal localisation of the proteasome z subunit gene reveals an ancient chromosomal duplication involving the major histocompatibility complex. *Proc natl acad sci USA 93: 9096-90101.*

59. Abi-Rached, L., M. F. McDermott, P. Pontarotti (1999) The mhc big bang. *Immunol rev 167: 33-44.*

60. Flajnik, M. F., M. Kasahara (2001) Comparative genomics of the mhc: glimpses into the evolution of the adaptive immune system. *Immunity 15: 351-362.*

61. Kuroda, N., F. Figueroa, C. O'Huigin, J. Klein (2002) Evidence that the separation of mhc class ii from class i loci in the zebrafish, *Danio rerio*, occurred by translocation. *Immunogenetics: accepted.*

62. Hughes, A. L., M. Yeager (1997) Molecular evolution of the vertebrate immune system. *Bioessays 19: 777-786.*

63. Klein, J., A. Sato (1998) Birth of the major histocompatibility complex. *Scand j immunol 47: 199-209.*

64. Kaufman, J., J. Salomonsen, M. F. Flajnik (1994) Evolutionary conservation of mhc class i and class ii molecules-different yet the same. *Semin immunol 6: 411-424.*

65. Flajnik, M. F., C. Canel, J. Kramer, M. Kasahara (1991) Which came first, mhc class i or class ii? *Immunogenetics 33: 295-300.*

66. Sola, L., S. Cataudella, E. Capanna (1981) New developments in vertebrate cytotaxonomy iii, karyology of bony fishes: a review. *Genetica 54: 285-328.*

67. Ohno, J., J. Muramoto, L. Christian, N. B. Atkin (1967) Dipliod-tetrapliod relationship

amoung old world members of the fish family Cyprinidae. *Chromosoma 23: 1-9.*
68. Golubtsov, A. S., E. Y. Krysanov (1993) Karyological study of some cyprinid species from Ethiopia. The ploidy differences between large and small *Barbus* of Africa. *Journal of fish biology 42: 445-455.*
69. Zardoya, R., I. Doadrio (1999) Molecular evidence on the evolutionary and biogeographical patterns of European cyprinids. *J mol evol 49: 227-237.*
70. Dixon, B., L. A. Nagelkerke, F. A. Sibbing, E. Egberts, R. J. M. Stet (1996) Evolution of mhc class ii beta chain-encoding genes in the Lake Tana barbel species flock (*Barbus intermedius* complex). *Immunogenetics 44: 419-431.*
71. Cavender, T. M. (1991) The fossil record of the Cyprinidae. Chapman & Hall, London. *p 35-54.*
72. Wiegertjes, G. F., R. J. M. Stet, W. B. Van Muiswinkel (1994) Divergent selection for antibody production in common carp (*Cyprinus carpio* l.) using gynogenesis. *Anim genet 25: 251-257.*
73. Wiegertjes, G. F., A. B. Bongers, P. Voorthuis, B. Zandieh Doulabi, A. Groeneveld, W. B. Van Muiswinkel, R. J. M. Stet (1996) Characterisation of isogenic carp (*Cyprinus carpio* l.) lines with a genetically determined high or low antibody production. *Anim genet 27: 313-319.*
74. Stet, R. J. M., C. P. Kruiswijk, J. P. Saeij, G. F. Wiegertjes (1998) Major histocompatibility genes in cyprinid fishes: theory and practice. *Immunol rev 166: 301-316.*
75. Nagelkerke, L. A. J., F. A. Sibbing (2000) The large barbs (*Barbus* spp., Cyprinidae, Teleostei) of Lake Tana (Ethiopia), with a description of a new species, *Barbus osseensis*. *Netherlands journal of zoology 50: 179-214.*
76. Mohr, P. A. (1962) *The geology of Ethiopia*. University college of Addis Ababa press, Addis Ababa, Ethiopia.
77. Erp van, S. H. M. (1996) Major histocompatibility complex genes in the common carp (*Cyprinus carpio* l.). Wageningen Agricultural University, Wageningen. *Thesis: 1- 159.*
78. Rodrigues, P. N. S. (1996) Expression of major histocompatibility complex genes in carp (*Cyprinus carpio* l.). Wageningen Agricultural University, Wageningen, *Thesis: 1-143.*
79. Shum, B. P., D. Avila, L. Du Pasquier, M. Kasahara, M. F. Flajnik (1993) Isolation of a classical mhc class i cdna from an amphibian. Evidence for only one class i locus in the *Xenopus* mhc. *J immunol 151: 5376-5386.*
80. Sato, K., M. F. Flajnik, L. Du Pasquier, M. Katagiri, M. Kasahara (1993) Evolution of the mhc: isolation of class ii beta-chain cdna clones from the amphibian *Xenopus laevis*. *J immunol 150: 2831-2843.*
81. Stet, R. J. M., S. H. M. van Erp, T. Hermsen, H. A. Sultmann, E. Egberts (1993) Polymorphism and estimation of the number of *mhcCyca* class i and class ii genes in laboratory strains of the common carp (*Cyprinus carpio* l.). *Dev comp immunol 17: 141-156.*
82. Goldsby, R. A., T. J. Kindt, b. A. Osborne, J. Kuby (2000) *Kuby immunology*. W H Freeman & Co.
83. Nusslein-Volhard, C. (1996) The identification of genes controlling development in flies and fishes (Nobel Lecture). *Angew chem int ed engl: 2176-2187.*
84. Postlethwait, J. H., S. L. Johnson, C. N. Midson, W. S. Talbot, M. Gates, E. W. Ballinger, D. Africa, R. Andrews, T. Carl, J. S. Eisen, *et al.* (1994) A genetic linkage map for the zebrafish. *Science 264: 699-703.*
85. Johnson, S. L., M. A. Gates, M. Johnson, W. S. Talbot, S. Horne, K. Baik, S. Rude, J. R. Wong, J. H. Postlethwait (1996) Centromere-linkage analysis and consolidation of the zebrafish genetic map. *Genetics 142: 1277-1288.*
86. Takeuchi, H., F. Figueroa, C. O'Huigin, J. Klein (1995) Cloning and characterisation of class

I MHC genes of the zebrafish, *Brachydanio rerio*. *Immunogenetics 42: 77-84.*

87. Ono, H., D. Klein, V. Vincek, F. Figueroa, C. O'Huigin, H. Tichy, J. Klein (1992) Major histocompatibility complex class II genes of zebrafish. *Proc natl acad sci USA 89: 11886-11890.*

88. Sultmann, H., W. E. Mayer, F. Figueroa, C. O'Huigin, J. Klein (1993) Zebrafish MHC class II alpha chain-encoding genes: polymorphism, expression, and function. *Immunogenetics 38: 408-420.*

89. Graser, R., C. O'Huigin, V. Vincek, A. Meyer, J. Klein (1996) *Trans*-species polymorphism of class II MHC loci in *Danio* fishes. *Immunogenetics 44: 36-48.*

90. Gudding, R., A. Lillehaug, P. Midtlyng, F. Brown, *Eds*. (1997) Fish vaccinology. Oslo, Norway, June 5-7, 1996. Symposium proceedings and abstracts. *Dev biol stand 90: 1-484.*

91. Parham, P., T. Ohta (1996) Population biology of antigen presentation by MHC class I molecules. *Science 272: 67-74.*

92. Klein, J., C. O'Huigin (1995) Class IIB MHC motifs in an evolutionary perspective. *Immunol rev 143: 89-111.*

93. Meyer, A., T. D. Kocher, P. Basasibwaki, A. C. Wilson (1990) Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. *Nature 347: 550-553.*

94. Kornfield, I., K. E. Carpenter (1984) Cyprinids of Lake Lanao, Philippines: taxonomic validity, evolutionary rates, and speciation scenarios. *In: A. A. Echelle, I. Kornfield (eds) Evolution of fish species flocks*. University of Maine, Orono Press, Orono, *p 69-83.*

95. Nagelkerke, L. A., F. A. Sibbing, G. M. van den Boogaart, E. H. H. R. Lammens, J. W. M. Osse (1994) The barbs (*Barbus* spp.) of Lake Tana: a forgotten species flock? *Environmental biology of fishes: 1-22.*

96. Nelson, J. S. (1976) *Fishes of the World*. John Wiley and Sons, New York.

97. Larhammar, D., C. Risinger (1994) Molecular genetic aspects of tetraploidy in the common carp, *Cyprinus carpio*. *Mol phylogenet evol 3: 59-68.*

98. Kirpchinikov, V. S. (1981) *Genetic bases of fish selection*. Springer-Verlag, Berlin.

99. Hubbs, C. L. (1955) Hybridisation between fish species in nature. *Syst zool: 1-20.*

100. Stet, R. J. M., E. Egberts (1991) The histocompatibility system in teleostean fishes: From multiple histocompatibility loci to a major histocompatibility complex. *Journal of fish and shellfish immunology 1: 1-16.*

101. Kaastrup, P., R. J. M. Stet, A. J. Tigchelaar, E. Egberts, W. B. van Muiswinkel (1989) A major histocompatibility locus in fish: serological identification and segregation of transplantation antigens in the common carp (*Cyprinus carpio* L.). *Immunogenetics 30: 284-290.*

102. Hildemann, W. H. (1970) Transplantation immunity in fishes: Agnatha, Chondrichthyes and Osteichthyes. *Transplant proc 2: 253-259.*

103. Kallman, K. D. (1970) Genetics of tissue transplantation in Teleostei. *Transplant proc 2: 263-271.*

104. Ono, H., C. O'Huigin, V. Vincek, R. J. M. Stet, F. Figueroa, J. Klein (1993) New beta chain-encoding MHC class II genes in the carp. *Immunogenetics 38: 146-149.*

105. Erp, S. H. van, E. Egberts, R. J. M. Stet (1996) Characterisation of class IIA and B genes in a gynogenetic carp clone. *Immunogenetics 44: 192-202.*

106. Okamura, K., T. Nakanishi, Y. Kurosawa, K. Hashimoto (1993) Expansion of genes that encode MHC class I molecules in cyprinid fishes. *J immunol 151: 188-200.*

107. Erp, S. H. M. van, B. Dixon, F. Figueroa, E. Egberts, R. J. M. Stet (1996) Identification and characterisation of a new major histocompatibility complex class I gene in carp (*Cyprinus carpio* L.). *Immunogenetics 44: 49-61.*

108. Silver, M. L., H. C. Guo, J. L. Strominger, D. C. Wiley (1992) Atomic structure of a human

MHC molecule presenting an influenza virus peptide. *Nature 360:367-369.*
109. Dixon, B., R. J. M. Stet, S. H. M. van Erp, B. Pohajdak (1993) Characterisation of β_2-microglobulin transcripts from two teleost species. *Immunogenetics 38: 27-34.*
110. Shawar, S. M., J. M. Vyas, J. R. Rodgers, R. R. Rich (1994) Antigen presentation by major histocompatibility complex class I-b molecules. *Annu rev immunol 12: 839-880.*
111. Salter, R. D., R. J. Benjamin, P. K. Wesley, S. E. Buxton, T. P. Garrett, C. Clayberger, A. M. Krensky, A. M. Norment, D. R. Littman, P. Parham (1990) A binding site for the T-cell co-receptor CD8 on the alpha 3 domain of HLA-A2. *Nature 345: 41-46.*
112. Fairhurst, R. M., C. X. Wang, P. A. Sieling, R. L. Modlin, J. Braun (1998) CD1-restricted t cells and resistance to polysaccharide-encapsulated bacteria. *Immunology today 19: 257-259.*
113. Erp, S. H. M. van, E. Egberts, R. J. M. Stet (1996) Evidence for multiple distinct major histocompatibility complex class I lineages in teleostean fish. *Eur j immunogenet 23: 371-381.*
114. Bjorkman, P. J., M. A. Saper, B. Samraoui, W. S. Bennett, J. L. Strominger, D. C. Wiley (1987) Structure of the human class I histocompatibility antigen, HLA-A2. *Nature 329: 506-512.*
115. Hansen, J. D., P. Strassburger, L. Du Pasquier (1996) Conservation of an alpha-2 domain within the teleostean world, MHC class I from the rainbow trout *Oncorhynchus mykiss*. *Dev comp immunol 20: 417-425.*
116. Shum, B. P., K. Azumi, S. Zhang, S. R. Kehrer, R. L. Raison, H. W. Detrich, P. Parham (1996) Unexpected β_2-microglobulin sequence diversity in individual rainbow trout. *Proc natl acad sci USA 93: 2779-2784.*
117. Grimholt, U., I. Hordvik, V. M. Fosse, I. Olsaker, C. Endresen, O. Lie (1993) Molecular cloning of major histocompatibility complex class I cDNAs from Atlantic salmon (*Salmo salar*). *Immunogenetics 37: 469-473.*
118. Katagiri, T., I. Hirono, T. Aoki, M. Sakai (1996) Isolation of major histocompatibility complex class I cDNA from pink salmon (*Oncorhynchus gorbuscha*). *Dev comp immunol 20: 217-228.*
119. Sato, A., F. Figueroa, C. O'Huigin, D. N. Reznick, J. Klein (1996) Identification of major histocompatibility complex genes in the guppy, *Poecilia reticulata*. *Immunogenetics 43: 38-49.*
120. Betz, U. A., W. E. Mayer, J. Klein (1994) Major histocompatibility complex class I genes of the coelacanth *Latimeria chalumnae*. *Proc natl acad sci USA 91: 11065-11069.*
121. Fremont, D. H., M. Matsumura, E. A. Stura, P. A. Peterson, I. A. Wilson (1992) Crystal structures of two viral peptides in complex with murine MHC class I H-2K^b. *Science 257:919-927.*
122. Salter-Cid, L., M. F. Flajnik (1995) Evolution and developmental regulation of the major histocompatibility complex. *Crit rev immunol 15:31-75.*
123. Klein, J. (1987) *Natural history of the major histocompatibility complex*. Wiley and Sons, New York.
124. Rodrigues, P. N. S, B. Dixon, J. Roelofs, J. H. Rombout, E. Egberts, B. Pohajdak, R. J. M. Stet (1998) Expression and temperature-dependent regulation of the β_2-microglobulin (Cyca-β_2m) gene in a cold-blooded vertebrate, the common carp (*Cyprinus carpio* L.). *Dev immunol 5: 263-275.*
125. Rodrigues, P. N., T. T. Hermsen, A. van Maanen, A. J. Taverne-Thiele, J. H. Rombout, B. Dixon, R. J. M. Stet (1998) Expression of MHCCyca class I and class II molecules in the early life history of the common carp (*Cyprinus carpio* L.). *Dev comp immunol 22: 493-506.*
126. Kaufman, J. (1993) Ontogeny and phylogeny of MHC-like molecules. *In: E. L., N. B. E. Cooper (eds) Developmental immunology*. Oxforduniversity Press, Oxford, *p 53-84.*
127. Sultmann, H., W. E. Mayer, F. Figueroa, C. O'Huigin, J. Klein (1994) Organisation of MHC class IIB genes in the zebrafish (*Brachydanio rerio*). *Genomics 23: 1-14.*
128. Klein, J., H. Ono, D. Klein, C. O'Huigin (1993) The accordion model of MHC evolution. *Prog.*

immunol. 8: 137-143.

129. Parham, P., E. J. Adams, K. L. Arnett (1995) The origins of hla-a,b,c polymorphism. *Immunol rev 143: 141-180.*

130. McAdam, S. N., J. E. Boyson, X. Lui, T. L. Garber, A. L. Hughes, R. E. Bontrop, D. I. Watkins (1994) A uniquely high level of recombination at the hla-b locus. *Proc natl acad sci USA 91: 5893-5897.*

131. Johnson, T. C., C. A. Scholz, M. R. Talbot, K. Kelts, R. D. Ricketts, G. Ngobo, K. Beuning, I. Semmanda, J. W. McGill (1996) Late Pleistocene desiccation of Lake Victoria and rapid evolution of cichlid fishes. *Science 273: 1091-1093.*

132. Nagelkerke, L. A., F. A. Sibbing (in press) A revision of the large barbs (*Barbus* spp., Cyprinidae, Teleostei) of Lake Tana, Ethiopia, with a description of seven new species. *Zoologische verhandelingen Leiden.*

133. Klein, D., H. Ono, C. O'Huigin, V. Vincek, T. Goldschmidt, J. Klein (1993) Extensive mhc variability in cichlid fishes of Lake Malawi. *Nature 364: 330-334.*

134. Grimholt, U., I. Olsaker, C. de Vries Lindstrom, O. Lie (1994) A study of variability in the mhc class ii beta-1 and class i alpha-2 domain exons of Atlantic salmon, *Salmo salar* l. *Anim genet 25: 147-153.*

135. Hughes, A. L., M. K. Hughes, C. Y. Howell, M. Nei (1994) Natural selection at the class ii major histocompatibility complex loci. *Phil. trans. r. soc. Lond. b 345: 359-367.*

136. Miller, K. M., R. E. Withler (1996) Sequence analysis of a polymorphic mhc class ii gene in pacific salmon. *Immunogenetics 43: 337-351.*

137. Slierendregt, b. L., N. Otting, M. Kenter, R. E. Bontrop (1995) Allelic diversity at the mhc-dp locus in rhesus macaques (*Macaca mulatta*). *Immunogenetics 41: 29-37.*

138. Nagelkerke, L. A. J., F. A. Sibbing, J. W. M. Osse (1995) Morphological divergence during growth in the large barbs (*Barbus* spp.) of Lake Tana, ethiopia. *Netherlands journal of zoology 45: 431-454.*

139. Streisinger, G., C. Walker, N. Dower, D. Knauber, F. Singer (1981) Production of clones of homozygous diploid zebrafish (*Brachydanio rerio*). *Nature 291: 293-296.*

140. Hyodo-Taguchi, Y., N. Egami (1985) Establishment of inbred strains of the medaka, *Oryzias laptipes* and the usefullness of the strains for biomedical research. *Zool sci: 305-316.*

141. Schartl, A., U. Hornung, I. Nanda, R. Wacker, H. K. Muller-Hermelink, I. Schlupp, J. Parzefall, M. Schmid, M. Schartl (1997) Susceptibility to the development of pigment cell tumors in a clone of the Amazon molly, *Poecilia formosa*, introduced through a microchromosome. *Cancer res 57: 2993-3000.*

142. Turner, b. J., J. F. Elder, Jr., T. F. Laughlin, W. P. Davis (1990) Genetic variation in clonal vertebrates detected by simple-sequence dna fingerprinting. *Proc natl acad sci USA 87: 5653-5657.*

143. Schartl, M., I. Nanda, I. Schlupp, B. Wilde, J. T. Epplen, M. Schmid, J. Parzefall (1995) Incorperation of subgenomic amounts of dna as compensation for mutational load in a gynogenetic fish. *Nature 373: 68-71.*

144. Nakanishi, T. (1987) Transferability of immune plasma and pronephric cells in isogeneic, allogeneic and xenogeneic transfer systems in crucian carp. *Dev comp immunol 11: 521-528.*

145. Ihssen, P. E., L. R. McKay, I. McMillan, R. B. Phillips (1990) Ploidy manipulation and gynogenesis in fishes: cytogenetic and fisheries applications. *Trans am fish soc 119: 698-717.*

146. Johnstone, R., R. J. M. Stet (1995) The production of gynogenetic Atlantic salmon, *Salmo salar* l. *Theor. appl. genet. 90: 819-826.*

147. Cherfas, N., B. Gomelsky, N. Ben-Dom, G. Hulata (1995) Evidence for the heritable nature of

spontaneous diploidisation in common carp, *Cyprinus carpio* L., eggs. *Aquacultural research 26: 289-292.*
148. Naruse, K., K. Ijiri, A. Shima, N. Egami (1985) The production of cloned fish in the medaka (*Oryzias latipes*). *J exp zool 236: 335-341.*
149. Han, H. S., N. Taniguchi, A. Tsujimura (1991) production of cloned ayu by chromosome manipulation and confirmation by isozome marker and tissue grafting. *Nippon suisan gakkaishi 57: 825-832.*
150. Komen, J., A. b. J. Bongers, C. J. J. Richter, W. B. van Muiswinkel, E. A. Huisman (1991) Gynogenesis in common carp (*Cyprinus carpio* L.) II. The production of homozygous gynogenetis clones and F1 hybrids. *Aquaculture 92: 127-142.*
151. Tabata, K., A. Mizuta (1997) The present situation and important issues in breeding by chromosomal set manipulation in Hirame *Paralichthys olivaceus*. *Bull natl res inst aquacult 3: 43-52.*
152. Komen, J., C. J. J. Richter (1993) Sex control in carp. *Rec adv aquacult 4: 78-86.*
153. Bongers, A. B. J., M. Z. Ben Ayed, b. Zandieh Doulabi, J. Komen, C. J. J. Richter (1997) Origin of variation in isogentic, gynogenetic, and androgenetic strains of common carp *Cyprinus carpio*. *J exp zool 277: 72-79.*
154. Crooijmans, R. P. M. A., V. A. F. Bierbooms, J. Komen, J. J. Poel van der, M. A. M. Groenen (1997) Microsatellite markers in common carp (*Cyprinus carpio* L.). *Anim genet 28: 129-134.*
155. Young, W. P., P. A. Wheeler, R. D. Fields, G. H. Thorgaard (1996) Dna fingerprinting confirms isogenicity of androgenetically derived rainbow trout lines. *J hered 87: 77-80.*
156. Bongers, A. B. J., E. P. C. Veld in't, K. Abo-Hashema, I. M. Bremmer, E. H. Eding, J. Komen, C. J. J. Richter (1994) Androgenesis in common carp (*Cyprinus Carpio* L.) using uv irradiation in a synthetic ovarian fluid and heat shocks. *Aquaculture 122: 119-132.*
157. Komen, J., P. J. Van den Dobbelsteen, W. J. Slierendrecht, W. B. Van Muiswinkel (1990) Skin grafting in gynogenetic common carp (*Cyprinus carpio*, L.). The development of histocompatible clones. *Transplantation 49: 788-793.*
158. Bandin, I., C. P. Dopoza, W. B. Van Muiswinkel, G. F. Wiegertjes (1997) Quantitation of antibody secreting cells in high and low antibody responder inbred carp (*Cyprinus carpio* L.) strains. *Fish and shellfish immunology 7: 487-501.*
159. Puel, A., P. C. Groot, M. G. Lathrop, P. Demant, D. Mouton (1995) Mapping of genes controlling quantitative antibody production in Biozzi mice. *J immunol 154: 5799-5805.*
160. Puel, A., J. C. Mevel, Y. Bouthillier, N. Feingold, W. H. Fridman, D. Mouton (1996) Toward genetic dissection of high and low antibody responsiveness in biozzi mice. *Proc natl acad sci USA 93: 14742-14746.*
161. Saitou, N., M. Nei (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol biol evol 4: 406-425.*
162. Nelson, J. S. (1994) *Fishes of the world.* Wiley, New York.
163. Fujiki, K., M. Booman, E. Chin-Dixon, B. Dixon (2001) Cloning and characterisation of cDNA clones encoding membrane-bound and potentially secreted major histocompatibility class I receptors from walleye (*Stizostedion vitreum*). *Immunogenetics 53: 760-769.*
164. Kruiswijk, C. P., T. Hermsen, H. J., D. Dixon, H. F. J. Savelkoul, R. J. M. Stet. (in preparation) Different modes of major histocompatibility class Ia and class IIb evolution in the Lake Tana African 'large' barb species flock.
165. Marsh, S. G. E., P. Parham, L. D. Barber (2000) *The HLA facts book*. Academic press, San Diego.
166. Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, D. G. Higgins (1997) The clustal-x windows interface: flexible strategies for multiple sequence alignment aided by quality

analysis tools. *Nucleic acids res 25: 4876-4882.*

167. Nakai, K., P. Horton (1999) Psort: a program for detecting sorting signals in proteins and predicting their subcellular localisation. *Trends biochem sci 24: 34-36.*

168. Kumar, S., K. Tamura, I. B. Jakobsen, M. Nei (2001) Mega2: molecular evolutionary genetics analysis software. *Bioinformatics 17: 1244-1245.*

169. Kruiswijk, C. P., T. T. Hermsen, A. H. Westphal, H. F. Savelkoul, R. J. Stet (2002) A novel functional class i lineage in zebrafish (*Danio rerio*), carp (*Cyprinus carpio*), and large Barbus (*Barbus intermedius*) showing an unusual conservation of the peptide binding domains. *J immunol 169: 1936-1947.*

170. Kasahara, M., M. F. Flajnik, T. Ishibashi, T. Natori (1995) Evolution of the major histocompatibility complex: a current overview. *Transpl immunol 3: 1-20.*

171. Bartl, S., M. A. Baish, M. F. Flajnik, Y. Ohta (1997) Identification of class i genes in cartilaginous fish, the most ancient group of vertebrates displaying an adaptive immune response. *J immunol 159: 6097-6104.*

172. Kaufman, J., R. Andersen, D. Avila, J. Engberg, J. Lambris, J. Salomonsen, K. Welinder, K. Skjodt (1992) Different features of the mhc class i heterodimer have evolved at different rates. Chicken b-f and $ß_2$-microglobulin sequences reveal invariant surface residues. *J immunol 148: 1532-1546.*

173. Flajnik, M. F., M. Kasahara, B. P. Shum, L. Salter-Cid, E. Taylor, L. Du Pasquier (1993) A novel type of class i gene organisation in vertebrates: a large family of non-mhc-linked class i genes is expressed at the rna level in the amphibian *Xenopus*. *Embo j 12: 4385-4396.*

174. Forey, P. (1984) *The living fossils*. Springer, New York.

175. Marshall, C., H. P. Schultze (1992) Relative importance of molecular, neontological, and paleontological data in understanding the biology of the vertebrate invasion of land. *J mol evol 35: 93-101.*

176. Sato, A., D. Klein, H. Sultmann, F. Figueroa, C. O'Huigin, J. Klein (1997) Class i mhc genes of cichlid fishes: identification, expression, and polymorphism. *Immunogenetics 46:63-72.*

177. Murray, B. W., P. Nilsson, Z. Zaleska-Rutczynska, H. Sultmann, J. Klein (2000) Linkage relationships and haplotype variation of the major histocompatibility complex class ia genes in the cichlid fish *Oreochromis niloticus. 2:437-448.*

178. Miller, K. M., K. H. Kaukinen, A. D. Schulze (2002) Expansion and contraction of major histocompatibility complex genes: a teleostean example. *Immunogenetics 53:941-963.*

179. Parnes, J. R., R. R. Robinson, J. G. Seidman (1983) Multiple mrna species with distinct 3' termini are transcribed from the $ß_2$-microglobulin gene. *Nature 302:449-452.*

180. Wen-Hsuing, L. (1997) *Molecular Evolution*. Sinauer Associates, inc., Publishers, Sunderland, Massachusetts.

181. Wang, J. H., R. Meijers, Y. Xiong, J. H. Liu, T. Sakihama, R. Zhang, A. Joachimiak, E. L. Reinherz (2001) Crystal structure of the human cd4 n-terminal two-domain fragment complexed to a class ii mhc molecule. *Proc natl acad sci USA 98: 10799-10804.*

182. Aoyagi, K., J. M. Dijkstra, C. Xia, I. Denda, M. Ototake, K. Hashimoto, T. Nakanishi (2002) Classical mhc class i genes composed of highly divergent sequence lineages share a single locus in rainbow trout (*Oncorhynchus mykiss*). *J immunol 168: 260-273.*

183. Nei, M. S. Kumar *Molecular evolution and phylogenetics* Oxford University Press, Oxford.

184. Malaga-Trillo, E., Z. Zaleska-Rutczynska, B. McAndrew, V. Vincek, F. Figueroa, H. Sultmann, J. Klein (1998) Linkage relationships and haplotype polymorphism among cichlid mhc class iib loci. *Genetics 149: 1527-1537.*

185. Ono, H., F. Figueroa, C. O'Huigin, J. Klein (1993) Cloning of the $ß_2$-microglobulin gene in the zebrafish. *Immunogenetics 38: 1-10.*

186. Persson, A. C., R. J. M. Stet, L. Pilstrom (1999) Characterisation of mhc class i and β_2-microglobulin sequences in Atlantic cod reveals an unusually high number of expressed class i genes. *Immunogenetics 50: 49-59.*

187. Lundqvist, M. L., P. Appelkvist, T. Hermsen, L. Pilstrom, R. J. M. Stet (1999) Characterisation of β_2-microglobulin in a primitive fish, the Siberian sturgeon (*Acipenser baeri*). *Immunogenetics 50: 79-83.*

188. Criscitiello, M. F., R. Benedetto, A. Antao, M. R. Wilson, V. G. Chinchar, N. W. Miller, L. W. Clem, T. J. McConnell (1998) β_2-microglobulin of ictalurid catfishes. *Immunogenetics 48: 339-343.*

189. Gussow, D., R. Rein, I. Ginjaar, F. Hochstenbach, G. Seemann, A. Kottman, H. L. Ploegh (1987) The human β_2-microglobulin gene. Primary structure and definition of the transcriptional unit. *J immunol 139: 3132-3138.*

190. Gould, S. J., N. Eldredge (1993) Punctuated equilibrium comes of age. *Nature 366: 223-227.*

191. Sultmann, H., A. Sato, B. W. Murray, N. Takezaki, R. Geisler, G. J. Rauch, J. Klein (2000) Conservation of mhc class iii region synteny between zebrafish and human as determined by radiation hybrid mapping. *J immunol 165: 6984-6993.*

192. Sato, A., F. Figueroa, B. W. Murray, E. Malaga-Trillo, Z. Zaleska-Rutczynska, H. Sultmann, S. Toyosawa, C. Wedekind, N. Steck, J. Klein (2000) Nonlinkage of major histocompatibility complex class i and class ii loci in bony fishes. *Immunogenetics 2: 108-116.*

193. Kruiswijk, C. P., T. Hermsen, K. Fujiki, B. Dixon, H. F. J. Savelkoul, R. J. M. Stet (in preparation) Analysis of genomic and expressed major histocompatibility class ia and class ii genes in a hexaploid Lake Tana African 'large' barb individual (*Barbus intermedius*).

194. Nei, M., L. Jin (1989) Variances of the average numbers of nucleotide substitutions within and between populations. *Mol biol evol 6: 290-300.*

195. Felsenstein, J., G. A. Churchill (1996) A hidden markov model approach to variation among sites in rate of evolution. *Mol biol evol 13: 93-104.*

196. Grassly, N. C., E. C. Holmes (1997) A likelihood method for the detection of selection and recombination using nucleotide sequences. *Mol biol evol 14: 239-247*.

197. Hughes, A. L. (2000) Evolution of introns and exons of class ii major histocompatibility complex genes of vertebrates. *Immunogenetics 51: 473-486.*

198. Nei, M., A. L. Hughes (1992) Balanced polymorphism and evolution by the birth-and death process in the mhc loci. I*n: Tsuji, K., M. Aizawa, T. Sasazuki (eds) Hla 1999, Proceedings of the elevent international histocompatibility workshop and conference*. Oxford university press, Oxford, *p 27-38*

199. Hasegawa, M., H. Kishino, T. Yano (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial dna. *J mol evol 22: 160-174.*

200. Hughes, A. L. (1999) *Adaptive evolution of genes and genomes*. Oxford University Press, New York.

201. Cereb, N., A. L. Hughes, S. Y. Yang (1997) Locus-specific conservation of the hla class i introns by intra-locus homogenisation. *Immunogenetics 47: 30-36.*

202. Ono, H., C. O'Huigin, H. Tichy, J. Klein (1993) Major-histocompatibility-complex variation in two species of cichlid fishes from Lake Malawi. *Mol biol evol 10: 1060-1072.*

203. Bontrop, R. E. (1994) Nonhuman primate mhc-dqa and -dqb second exon nucleotide sequences: a compilation. *Immunogenetics 39: 81-92.*

204. Bontrop, R. E., N. Otting, b. L. Slierendregt, J. S. Lanchbury (1995) Evolution of major histocompatibility complex polymorphisms and t-cell receptor diversity in primates. *Immunol rev 143: 33-62.*

205. Gyllensten, U. B., D. Lashkari, H. A. Erlich (1990) Allelic diversification at the class ii dqb

locus of the mammalian major histocompatibility complex. *Proc natl acad sci USA 87: 1835-1839.*

206. McAdam, S. N., J. E. Boyson, X. Liu, T. L. Garber, A. L. Hughes, R. E. Bontrop, D. I. Watkins (1995) Chimpanzee mhc class i a locus alleles are related to only one of the six families of human a locus alleles. *J immunol 154: 6421-6429.*

207. Watkins, D. I. (1995) The evolution of major histocompatibility class i genes in primates. *Crit rev immunol 15: 1-29.*

208. Cooper, S., E. J. Adams, R. S. Wells, C. M. Walker, P. Parham (1998) A major histocompatibility complex class i allele shared by two species of chimpanzee. *Immunogenetics 47: 212-217.*

209. Evans, D. T., M. S. Piekarczyk, L. Cadavid, V. S. Hinshaw, D. I. Watkins (1998) Two different primate species express an identical functional mhc class i allele. *Immunogenetics 47: 206-211.*

210. Murray, B. W., H. Sultmann, J. Klein (1999) Analysis of a 26-kb region linked to the mhc in zebrafish: genomic organisation of the proteasome component beta/transporter associated with antigen processing-2 gene cluster and identification of five new proteasome beta subunit genes. *J immunol 163: 2657-2666.*

211. Du Pasquier, L., M. F. Flajnik (1999) *Origin and evolution of the vertebrate immune system*. Lippincott, philadelphia.

212. Parham, P. (1994) The rise and fall of great class i genes. *Semin immunol 6: 373-382.*

213. Figueroa, F., W. E. Mayer, A. Sato, Z. Zaleska-Rutczynska, B. Hess, H. Tichy, J. Klein (2001) Mhc class i genes of swordtail fishes, *Xiphophorus*: variation in the number of loci and existence of ancient gene families. *Immunogenetics 53: 695-708.*

214. Saeij, J. P., R. J. Stet, A. Groeneveld, L. B. Verburg-van Kemenade, W. B. van Muiswinkel, G. F. Wiegertjes (2000) Molecular and functional characterisation of a fish inducible-type nitric oxide synthase. *Immunogenetics 51: 339-346.*

215. Irnazarow, I. (1995) Genetic variability of Polish and Hungarian carp lines. *Aquaculture 129: 215-219.*

216. Westerfield, M. (1993) *The zebrafish book*. University of Oregon Press, Eugene

217. Thompson, J. D., D. G. Higgins, T. J. Gibson (1994) Clustal w: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids res 22: 4673-4680.*

218. Nei, M., T. Gojobori (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol biol evol 3: 418-426.*

219. Sali, A., T. L. Blundell (1993) Comparative protein modelling by satisfaction of spatial restraints. *J mol biol 234: 779-815.*

220. Dauber-Osguthorpe, P., V. A. Roberts, D. J. Osguthorpe, J. Wolff, M. Genest, A. T. Hagler (1988) Structure and energetics of ligand binding to proteins: *Escherichia coli* dihydropholate reductase-trimethoprim, a drug-receptor system. *Proteins 4: 31-47.*

221. Laskowski, R. A., M. W. MacArthur, D. S. Moss, J. M. Thornton (1993) Procheck: a program to check the stereochemical quality of protein structures. *J. appl. cryst. 26: 283-291.*

222. Sippl, M. J. (1993) Recognition of errors in three-dimensional structures of proteins. *Proteins 17: 355-362.*

223. Wu, T. T., E. A. Kabat (1970) An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. *J exp med 132: 211-250.*

224. Shum, B. P., R. Rajalingam, K. E. Magor, K. Azumi, W. H. Carr, B. Dixon, R. J. M. Stet, M. A. Adkison, R. P. Hedrick, P. Parham (1999) A divergent non-classical class i gene conserved in

salmonids. *Immunogenetics 49: 479-490.*

225. Zeng, Z., A. R. Castano, B. W. Segelke, E. A. Stura, P. A. Peterson, I. A. Wilson (1997) Crystal structure of mouse CD1: An MHC-like fold with a large hydrophobic binding groove. *Science 277: 339-345.*

226. Ohta, Y., E. C. McKinney, M. F. Criscitiello, M. F. Flajnik (2002) Proteasome, transporter associated with antigen processing, and class I genes in the nurse shark, *Ginglymostoma cirratum*: evidence for a stable class I region and MHC haplotype lineages. *J immunol 168: 771-781.*

227. Antao, A. B., M. Wilson, J. Wang, E. Bengten, N. W. Miller, L. W. Clem, V. G. Chinchar (2001) Genomic organisation and differential expression of channel catfish MHC class I genes. *Dev comp immunol 25: 579-595.*

228. Bjorkman, P. J., P. Parham (1990) Structure, function, and diversity of class I major histocompatibility complex molecules. *Annu rev biochem 59: 253-288.*

229. Gu, X., M. Nei (1999) Locus specificity of polymorphic alleles and evolution by a birth-and-death process in mammalian MHC genes. *Mol biol evol 16: 147-156.*

230. Ota, T., M. Nei. (1994) Divergent evolution and evolution by the birth-and-death process in the immunoglobulin VH gene family. *Mol Biol Evol 11:469-482.*

231. Satta, Y., C. O'Huigin, N. Takahata, J. Klein (1993) The synonymous substitution rate of the major histocompatibility complex loci in primates. *Proc natl acad sci USA 90:7480-7484.*

232. Shum, B. P., L. Guethlein, L. R. Flodin, M. A. Adkison, R. P. Hedrick, R. B. Nehring, R. J. M. Stet, C. Secombes, P. Parham (2001) Modes of salmonid MHC class I and II evolution differ from the primate paradigm. *J Immunol 166:3297-3308.*

233. Laird, D. J., A. W. De Tomaso, M. D. Cooper, I. L. Weissman (2000) 50 million years of chordate evolution: seeking the origins of adaptive immunity. *Proc natl acad sci USA 97: 6924-6926.*

234. Spring, J. (1997) Vertebrate evolution by interspecific hybridisation-are we polyploid? *Febs lett 400: 2-8.*

235. Meyer, A., M. Schartl (1999) Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Curr opin cell biol 11: 699-704.*

236. Kasahara, M. (1999) The chromosomal duplication model of the major histocompatibility complex. *Immunol rev 167: 17-32.*

237. Ohno, S. (1999) Gene duplication and the uniqueness of vertebrate genomes circa 1970-1999. *Semin cell dev biol 10: 517-522.*

238. Wolfe, K. H. (2001) Yesterday's polyploids and the mystery of diploidisation. *Nat rev genet 2: 333-341.*

239. Abi-Rached, L., A. Gilles, T. Shiina, P. Pontarotti, H. Inoko (2002) Evidence of en bloc duplication in vertebrate genomes. *Nat genet 31: 100-105.*

240. Postlethwait, J. H., Y. L. Yan, M. A. Gates, S. Horne, A. Amores, A. Brownlie, A. Donovan, E. S. Egan, A. Force, Z. Gong, C. Goutel, A. Fritz, R. Kelsh, E. Knapik, E. Liao, B. Paw, D. Ransom, A. Singer, M. Thomson, T. S. Abduljabbar, P. Yelick, D. Beier, J. S. Joly, D. Larhammar, F. Rosa (1998) Vertebrate genome evolution and the zebrafish gene map. *Nat genet 18: 345-349.*

241. Wiegertjes, G. F., E. Egberts, R. J. M. Stet (1997) Segregation of mhc class I and II genes in the common carp (*Cyprinus carpio* L.). *Dev comp immunol 21: 105.*

242. McConnell, T. J., U. B. Godwin, S. F. Norton, R. S. Nairn, S. Kazianis, D. C. Morizot (1998) Identification and mapping of two divergent, unlinked major histocompatibility complex class IIB genes in *Xiphophorus* fishes. *Genetics 149: 1921-1934.*

243. Clark, M. S., L. Shaw, A. Kelly, P. Snell, G. Elgar (2001) Characterisation of the mhc class i region of the Japanese pufferfish (*Fugu rubripes*). *Immunogenetics 52: 174-185.*

244. Magnadottir, B., H. Jonsdottir, S. Helgason, B. Bjornsson, T. O. Jorgensen, L. Pilstrom (1999) Humoral immune parameters in Atlantic cod (*Gadus morhua* l.) ii. The effects of size and gender under different environmental conditions. *Comp biochem physiol b biochem mol biol 122: 181-188.*

245. Israelsson, O., A. Petersson, E. Bengten, E. J. Wiersma, J. Andersson, G. Gezelius, L. Pilstrom (1991) Immunoglobulin concentration in Atlantic cod, *Gadus morhua* l., serum and cross-reactivity between anti-cod-antibodies and immunoglobulins from other species. *J fish biol 39: 265-278.*

246. Timon, M., G. Elgar, S. Habu, K. Okumura, P. C. Beverley (1998) Molecular cloning of major histocompatibility complex class i cdnas from the pufferfish *Fugu rubripes*. *Immunogenetics 47:170-173.*

247. Murray, B. W., S. Shintani, H. Sultmann, J. Klein (2000) Major histocompatibility complex class iia genes in cichlid fishes: identification, expression, linkage relationships, and haplotype variation. *Immunogenetics 51: 576-586.*

248. Graaf de, M., E. Nentwich, F. A. Sibbing, J. W. M. Osse (in preparation) Spawning segregation along temporal and macro-spatial dimensions wihtin Lake Tana's cyprinid species flock (Ethiopia); implications for their evolution and management.

249. Michalova, V., B. W. Murray, H. Sultmann, J. Klein (2000) A contig map of the mhc class i genomic region in the zebrafish reveals ancient synteny. *J immunol 164: 5296-5305.*

250. Kumar, S., S. B. Hedges (1998) A molecular timescale for vertebrate evolution. *Nature 392: 917-920.*

SUMMARY / *SAMENVATTING*

All living organisms are exposed to pathogens in every life history stage and must prevent invasion or destruction by these pathogens to survive. The various phyla of the animal kingdom exploit different strategies to achieve defence against pathogens. Broadly speaking, these strategies fall in two categories: innate (or non-specific) immune responses, and adaptive (or specific) immune responses. During past evolutionary processes, both categories have not always co-existed. Defence strategies involving adaptive immunity appear to have arisen first in jawed vertebrates.

Two subfamilies of cell surface glycoproteins that play a crucial role in adaptive immunity are the polymorphic major histocompatibility complex (MHC) class I and class II molecules, which are similar in structure and function. Class I molecules present endogenous peptide to CD8 positive cytotoxic T lymphocytes, while class II molecules present exogenous peptide to CD4 positive helper T lymphocytes. In all jawed vertebrate species studied, genes encoding the MHC molecules are clustered in a single genomic region called the major

histocompatibility complex, with exception of the bony fish. These bony fishes arose after the cartilaginous fishes which suggests that the lack of linkage of MHC class I and class II molecules is a derived characteristic.

The group of bony fish comprises almost half of all 43,000 extant vertebrate species and includes the cyprinid fishes that play an important role in aquaculture. Among cyprinids polyploidisation of the genome is observed. The diploid zebrafish, the tetraploid common carp, and the hexaploid African 'large' barb represent such different genomes. In addition, these species represent three highly divergent cyprinid genera, *Danio*, *Cyprinus* and *Barbus*. The lack of linkage of MHC class I and class II genes, the differences in ploidy status, and their evolutionary distant relationship make these species interesting to study (*chapter 1*).

In *chapter 2*, three major aspects of cyprinid MHC research are reviewed. First, classical and non-classical class I sequences of zebrafish, common carp and African 'large' barb were analysed and compared. Two distinct lineages, *Z* and *U*, are found in common and ginbuna crucian carp, but only the *U* lineage is present in zebrafish, and other non-cyprinid species. The presence of the *Z* lineage is hypothesised to be the result of an allotetraploidisation event. Phylogenetic analyses and amino acid sequence comparisons suggests that *Cyca-Z* sequences are non-classical in nature. The comprehensive phylogenetic analyses of class I *Z* and *U* sequences revealed different phylogenetic histories of the exons encoding the extra-cellular domains.

Second, we addressed the evolution of the MHC genes in a natural model, the African 'large' barb species flock present in Lake Tana, Ethiopia. Analysis of class II sequences supported the species designation of the morphotypes present in Lake Tana, and as a consequence the *trans*-species hypothesis of MHC polymorphism. Third, the importance of common carp gynogenetic clones was established. These clones can be divergently selected for traits such as high and low antibody response. In addition, the role of MHC molecules in immune responses can be investigated using the panel of isogenic lines currently available.

The study on evolution of MHC genes in a natural model, the hexaploid African 'large' barb species flock of Lake Tana, indicated that their MHC genes might be prone to silencing mechanisms (*chapter 2*). Therefore, we studied the number of MHC genes present in the genome of an African 'large' barb individual (*Barbus*

intermedius) in relation to those expressed (*chapter 3*). Full-length cDNA sequences were generated from mRNA. Partial genomic class Ia and class IIB were subsequently obtained and compared to those from cDNA using the same primer set. In addition, Southern hybridisations were performed to obtain a verification of the number of class I and class IIB genes sequences as identified by PCR analyses and DNA sequencing.

The hexaploid Lake Tana African 'large' barb individual possessed five class Ia, four class IIA and four class IIB genes at genomic level, which were shown to be expressed. These data indicated that the ploidy status does not correlate with the presence and expression of MHC genes. This suggests that gene silencing is most likely a result of physical exclusion. However, functional silencing also played a role in reducing the number of expressed class II genes, as class IIB pseudogenes were identified in the genome of this African 'large' barb individual.

The study on evolution of MHC class IIB sequences in four Lake Tana African 'large' barb species (*B. acutirostris, B. nedgia, B. tsanensis, B. truttiformis*) performed by Dixon and co-workers supported the species designation of the morphotypes present in the lake (reviewed in *chapter 2*). We extended the study on class IIB genes with six Lake Tana African 'large' barb species (*B. intermedius, B. brevicephalus, B. macrophtalmus, B. megastoma, B. platydorsus,* and *B. surkis*) and African 'large' barbs (*B. intermedius*) from the Blue Nile and its tributaries, using PCR amplification and DNA sequencing. In addition, using a similar approach, we investigated the variability of class I *UA* genes in seven Lake Tana African 'large' barb species (*B. intermedius, B. megastoma, B. platydorsus, B. acutirostris, B. nedgia, B. tsanensis,* and *B. truttiformis*) and African 'large' barbs (*B. intermedius*), taken from the Blue Nile and its tributaries (*chapter 4*).

Comparisons of nonsynonymous and synonymous substitutions in peptide binding and non peptide binding regions revealed that diversity of Lake Tana African 'large' barb class I and class II is generated by positive Darwinian selection on peptide binding regions. In general, phylogenetic lineages are maintained by purifying or neutral selection on non-peptide binding regions. Positive selection on peptide binding regions between phylogenetic lineages and recombination events seem to have contributed to diversification between phylogenetic lineages. Remarkably, class II intron-1 and exon-2 sequences were not shared between African 'large' barb species. In contrast, class I *UA* exon-3

sequences were shared. This suggests that the turn-over of class II alleles in the Lake Tana African 'large' barb species is much higher than that of class I alleles.

A novel MHC class I lineage (*ZE*) was discovered in three cyprinid species (*chapter 5*). Two unique complete protein-coding cDNA sequences were isolated in zebrafish, *Dare-ZE*0101* and *Dare-ZE*0102*, one in common carp, *Cyca-ZE*0101* and six in African 'large' barb, *Bain-ZE*0101*, *Bain-ZE*0102, Bain-ZE*0201, Bain-ZE*0301, Bain-ZE*0401,* and *Bain-ZE*0501*. Several features support a classical nature of class I *ZE* lineage genes. Deduced amino acid sequences indicated the presence of conserved potential peptide anchoring residues. In addition, the class I *ZE* genes showed ubiquitous expression in common carp and in African 'large' barb and a classical class I exon-intron organisation in common carp. Furthermore, in zebrafish, polymorphism was generated by positive selection on putative peptide binding residues. Phylogenetic analyses revealed clustering of the *ZE* lineage cluster together with non-classical cyprinid class I *Z* lineage cluster away from classical cyprinid class I genes. The *ZE* lineage genes have evolved in a *trans*-species fashion with lineages being maintained for up to 100 million years, as estimated by divergence time calculations.

In *chapter 6* we focussed on the implications of recent findings in three cyprinid fishes, the African 'large' barb, common carp and zebrafish (*chapter 2 to 5*). Comparing differences and similarities relating to the MHC of species representing different vertebrate taxa will provide insight into its origin.

Phylogenetic analyses of cyprinid MHC class Ia exon-4, class IIA and class IIB exon-3 sequences and representatives of other major jawed vertebrate species, suggests that the Actinopterygii are basal relative to the Chondrichthyes. This implies that the teleosts shared a common ancestor with a lineage that gave rise to all other Gnathostomes. This is in contrast to the commonly accepted position of the Chondrichthyes relative to the other Gnathostomes, although this is still under debate. Such an ancestral position of the teleosts may imply that the unlinked class I and class II loci is the ancestral organisation and not the observed linkage of class I and class II seen in other Gnathostomes.

The turn-over of class II alleles in the Lake Tana African 'large' barb species is much higher than that of class I alleles. A more rapid evolution of MHC class II is also observed in other teleost species (brown trout, rainbow trout and cichlids). These findings indicate that teleost class I and II evolution differs from the mammalian paradigm. In mammals MHC class I genes seem to evolve more

rapidly than class II genes. The presence of a novel class I *ZE* lineage conserved in three distantly related cyprinid species, zebrafish, common carp, and the African 'large' barb supports this hypothesis (*chapter 5*).

Although the class I *ZE* lineage possesses all classical class I features it is phylogenetically more related to the non-classical class I *Z* lineage of carp. In addition polymorphism was not superimposable on mammalian class Ia variability. These findings indicate the intermediate position of the class I *ZE* lineage. In the near future the zebrafish sequencing consortium will shed light on the location of these *ZE* genes. They may reside in close vicinity to known class I genes or class II genes or alternatively, they may reside on a different linkage group.

SAMENVATTING

Een organisme wordt in al zijn ontwikkelingsstadia blootgesteld aan pathogenen. Om te overleven is het noodzakelijk dat het organisme zich tegen deze pathogenen beschermt. De verschillende phyla van het dierenrijk gebruiken hiervoor verschillende strategieën die kunnen worden verdeeld in twee categorieën: de niet-specifieke (aangeboren) afweer en specifieke (adaptieve) afweer. Alle diersoorten kunnen zich verdedigen tegen pathogenen met niet-specifieke afweer strategieën. De specifieke afweer is echter beperkt tot de gewervelde dieren met kaken. Bij deze specifieke afweer spelen de major histocompatibility complex (MHC) klasse I en klasse II eiwitten een zeer belangrijke rol.

De polymorfe MHC klasse I en II glycoproteines zijn verankerd in de celmembraan en hebben een vergelijkbare structuur en functie. Klasse I eiwitten presenteren endogene peptiden van ca. 9-12 aminozuren aan CD8 positieve cytotoxische T-cellen. Klasse II eiwitten presenteren exogene peptiden van ca. 12-25 aminozuren aan CD4 positieve helper T-cellen. Deze eiwitten worden gecodeerd door een cluster van klasse I en II genen. Dit cluster wordt major histocompatibility complex (MHC) genoemd. Met uitzondering van de beenvissen, bezitten alle gewervelde dieren met kaken zo'n cluster van klasse I en II genen. In beenvissen liggen de genen coderend voor klasse I eiwitten echter op een ander chromosoom dan de genen die coderen voor klasse II eiwitten. Omdat beenvissen evolutionair ontstaan zijn na de kraakbeenvissen wordt het ontbreken van een cluster van klasse I en II genen gezien als een nieuw verkregen eigenschap.

Bijna de helf van alle levende gewervelde diersoorten behoren tot de groep van de beenvissen. Tot deze groep behoren ook de karperachtigen die een belangrijke rol spelen in de aqua-

cultuur. Sommige soorten behorende tot de groep van karperachtigen vertonen polyploïdisatie van het genoom. Voorbeelden hiervan zijn de diploïde zebravis, de tetraploïde karper en de hexaploïde africaanse grote barbeel. Daarnaast behoren deze drie soorten tot drie verschillende genera namelijk, Danio (zebravis), Cyprinus (karper), en Barbus (Afrikaanse grote barbeel). Deze drie karperachtige soorten zijn door het ontbreken van een cluster van klasse I en II genen, de verschillen in polyploidisatie en de evolutionaire afstand tussen genera, Danio, Cyprinus, and Barbus zeer geschikt voor het bestuderen van het ontstaan en de evolutie van het major histocompatibility complex (hoofdstuk 1).

In hoofdstuk 2 worden drie belangrijke onderzoeksaspecten betreffende het MHC van karperachtigen besproken. Ten eerste, de analyse en vergelijking van klassieke en niet-klassieke MHC klasse I sequenties, geïdentificeerd in zebravis, karper en Afrikaanse grote barbeel. De karper en de goudvis bezitten twee groepen klasse I genen, Z en U terwijl in het genoom van de zebravis en andere niet-karperachtige soorten alleen de klasse I U genen voorkomen. Analyse van aminozuursequenties en de fylogenie suggeren dat de Z genen coderen voor niet-klassieke klasse I eiwitten. De aanwezigheid van dit type klasse I Z genen in de karper en de goudvis is waarschijnlijk het resultaat van allotetraploïdisatie. De fylogenie van partiele Z en U sequenties, exon-2, exon-3, of exon-4 coderend voor de extra-cellulaire eiwitdomeinen, toonden de verschillen in fylogenetische verwantschap.

Vervolgens werd de evolutie van MHC genen in een natuurlijk model: de Afrikaanse grote barbelen soortenzwerm in het Tana meer in Ethiopië, besproken. De analyse van klasse II sequenties ondersteunde de soortenbeschrijving van morfotypen en gekoppeld hieraan de trans-species MHC polymorfisme hypothese. Als laatste werd het belang van gynogenetische karper clonen besproken. Deze clonen kunnen worden geselecteerd op eigenschappen zoals een hoge en lage antistofrespons. Daarnaast kan de rol van MHC eiwitten worden bestudeerd door gebruik te maken van de verschillende beschikbare isogene karper lijnen.

De studie betreffende evolutie van MHC genen in een natuurlijk model, de Afrikaanse grote barbelen soortenzwerm in het Tana meer, suggereerden inactivatie van MHC genen (hoofdstuk 2). Dit gaf aanleiding tot het bestuderen van het aantal verschillende MHC sequenties in het genoom van een grote afrikaanse barbeel (B. intermedius) in relatie tot het aantal sequenties dat tot expressie wordt gebracht. Complete cDNA MHC klasse Ia en II sequenties werden gegenereerd van mRNA. Vervolgens werden partiële genomische klasse Ia en II sequenties gegenereerd en vergeleken met cDNA sequenties gegenereerd met dezelfde primerset. Southern blot analyse diende als verificatie van het aantal klasse Ia en IIB sequenties in het genoom van een grote afrikaanse barbeel geïdentificeerd met PCR amplificatie en DNA analyse. Het genoom van de hexaploïde Afrikaanse grote barbeel bevatte vijf

klasse Ia, vier klasse IIA en vier klasse IIB sequenties die ook tot expressie worden gebracht. Het aantal sequenties in het genoom van het bestudeerde individu komt niet overeen met het hexaploide genoom. Dit is een indicatie dat inactivatie van MHC genen het resultaat is van fysieke uitschakeling. Echter het genoom van het bestudeerde individu bevat ook een aantal MHC klasse II pseudogenen wat duidt op functionele inactivatie.

De studie betreffende evolutie van MHC klasse IIB genen in vier soorten Afrikaanse grote barbelen uit het Tana meer (B. acutirostris, B. nedgia, B. tsanensis en B. truttiformis) uitgevoerd door DIXON en medewerkers ondersteunde de soortenaanduiding van de verschillende morfotypen aanwezig in het meer. Deze studie werd uitgebreid met zeven andere soorten Afrikaanse grote barbelen uit het Tana meer (B. intermedius, B. brevicephalus, B. macrophtalmus, B. megastoma, B. platydorsus en B. surkis) en de Afrikaanse grote barbeel (B. intermedius) uit de Blauwe Nijl en andere zoetwatersystemen. De variatie tussen klasse IIB sequenties werden geïdentificeerd door PCR amplificatie van genomisch DNA en DNA sequencing. Op dezelfde wijze werd de variatie van klasse I UA sequenties in zeven soorten Afrikaanse grote barbelen uit het Tana meer (B. acutirostris, B. nedgia, B. tsanensis B. truttiformis, B. intermedius, B. megastoma, B. platydorsus) bepaald (hoofdstuk 4).

Vergelijking van synonieme en niet-synonieme substituties in peptide bindende en niet-peptide bindende regio's toonde aan dat de variatie tussen klasse I of tussen klasse II genen wordt gegenereerd door positive selectie op peptide bindende regio's van klasse I en II moleculen. In het algemeen worden fylogenetische groepen behouden door negatieve of neutrale selectie op niet-peptide bindende regio's van MHC moleculen. Positive selectie op peptide bindende regio's en recombinatie spelen een belangrijke rol bij de diversificatie tussen verschillende fylogenetische groepen. Geen van de bestudeerde Afrikaanse grote barbelen soorten bevatte identieke klasse IIB intron-1 en exon-2 sequenties. Dit ondersteunt de soorten aanduiding van de verschillende morfotypen aanwezig in het Tana meer. Echter identieke klasse I UA exon-3 sequenties komen wel voor tussen de verschillende Afrikaanse grote barbelen soorten. Dit geeft aan dat de turn-over van klasse IIB allelen veel hoger lijkt dan de turn-over van klasse I allelen (hoofdstuk 4).

*In hoofstuk 5 wordt een nieuwe groep klasse I genen (ZE) beschreven. Deze ZE genen werden geïdentificeerd in drie soorten karperachtigen, zebravis, karper en de Afrikaanse grote barbeel. Twee complete eiwit coderende cDNA sequenties werden geïsoleerd uit zebravis, Dare-ZE*0101 en Dare-ZE*0102, een uit karper, Cyca-ZE*0101 en zes uit de Afrikaanse grote barbeel, Bain-ZE*0101, -0102, -0201, -0301, -0401 en -0501. Deze ZE genen bezitten een aantal karakteristieke eigenschappen van de klassieke klasse I genen zoals de geconserveerde aminozuur residuen betrokken bij het binden van peptide. Daarnaast komen deze genen in*

verschillende weefsels van zowel de karper als de Afrikaanse barbeel tot expressie, vertonen deze genen in de karper dezelfde exon-intron organisatie als klassieke klasse I *genen en is het polymorfisme wat deze genen vertonen in zebravis het resultaat van positive selectie. Echter fylogenetisch zijn* ZE *genen meer verwant met de niet-klassieke klasse* I Z *genen in de karper en de goudvis.*

In hoofdstuk 6 worden de resultaten uit de vorige hoofdstukken gecombineerd en bediscussieerd. Fylogenetische analyses van MHC *klasse* Ia *exon-4, klasse* IIA *en* IIB *exon-3 sequenties van karperachtigen en andere representatieve soorten gewervelden met kaken suggereerden dat de Actinopterygii (beenvissen) evolutionair ontstaan zijn voor de Chondrichthyes (kraakbeenvissen). Een dergelijke voorouder positie van de beenvissen impliceert dat het ontbreken van een cluster van klasse* I *en* II *genen in beenvissen de oorspronkelijke organisatie weerspiegelt. Echter deze fylogenetische data contrasteren met de huidige opvatting die uitgaat van een voorouderlijke positie van Chondrichthyes ten opzichte van de andere Gnathostomates (gewervelde met kaken).*

De evolutie van klasse Ia *en* IIB *genen in de Afrikaanse grote barbelen soortenzwerm vertoonde een verschillend partoon. De turn-over van klasse* IIB *genen bleek veel hoger te zijn dan voor klasse* Ia*. Dergelijk observaties zijn ook in andere beenvissoorten geobserveerd. Dit suggereert dat evolutie van klasse* I *en* II *in beenvissen tegengesteld is aan de evolutie van deze genen in zoogdieren. Deze hypothese wordt ondersteunt door de identificatie van een groep klasse* I ZE *genen in drie karperachtige die evolutionair een oude verwantschap vertonen, zebravis, karper en Afrikaanse grote barbeel.*

De nieuwe groep klasse I *genen,* ZE*, bezit alle eigenschappen van klassieke klasse* I *genen. Echter deze groep is fylogenetisch meer verwant met de niet-klassieke klasse* I Z *genen van de karper en de goudvis. Daarnaast bevind het polymorfisme van de genen zich op een andere locatie dan het polymorfisme van klassieke klasse* I *moleculen van zoogdieren. Deze observaties suggereren dat deze klasse* I ZE *eiwitten een intermediare positie innemen. Het zebravis genoom sequencing project zal in de nabije toekomst de positie van deze genen ophelderen ten opzichte van de reeds bekende locaties van klasse* I *en* II *genen.*

CURRICULUM VITÆ

Corine (Cornelia Pia) Kruiswijk werd geboren op 6 april 1971 te Vinkeveen. In 1993 begon zij met de studie Biologie aan de toenmalige Landbouwuniveristeit te Wageningen. Tijdens de doctoraalfase heeft zij afstudeeronderzoek verricht bij de vakgroepen Virologie en Celbiologie & Immunologie van de Landbouwuniversiteit. De stage-periode werd doorgebracht in de Verenigde Staten bij de afdeling Biochemie & Moleculaire Biologie van de Medical University of South Carolina in Charlston. In juni 1998 studeerde zij af en in juli van hetzelfde jaar werd zij aangesteld als assistent in opleiding bij de leerstoelgroep Celbiologie & Immunologie. De resultaten van het daar uitgevoerde onderzoek staan beschreven in dit proefschrift. In januari 2000 begon zij met een driejarige deeltijdstudie Bedrijfskunde aan Universiteit Nyenrode te Breukelen, die zij begin 2003 hoopt te voltooien.

COLOFON

Ontwerp & opmaak: Coraline Korevaar, Atelier De Kazerne, Woudrichem.
Illustraties: Boudewijn van Schagen. Gebruikt lettertype: Lexicon 2a en 2c.
Papiersoort omslag: Job art ivoor 300 grs.; binnenwerk: Bioset 90 grs. Drukwerk:
Drukkerij Damen, Werkendam.