A nonclassical MHC class I U lineage locus in zebrafish with a null haplotypic variant

Hayley Dirscherl[1,2] and **Jeffrey A. Yoder**[1,3,*]

[1]Department of Molecular Biomedical Sciences, North Carolina State University, 1060 William Moore Drive, Raleigh, NC 27607 USA

[2]The Joint Biomedical Engineering Graduate Program, University of North Carolina-North Carolina State University, 911 Oval Drive, Raleigh, NC 27695 USA

[3]Center for Comparative Medicine and Translational Research, North Carolina State University, 1060 William Moore Drive, Raleigh, NC 27607 USA

Abstract

Three sequence lineages of MHC class I genes have been described in zebrafish (*Danio rerio*): U, Z, and L. The U lineage genes encoded on zebrafish chromosome 19 are predicted to provide the classical function of antigen presentation. This MHC class I locus displays significant haplotypic variation and is the only MHC class I locus in zebrafish that shares conserved synteny with the core mammalian MHC. Here we describe two MHC class I U lineage genes, *mhc1ula* and *mhc1uma*, that map to chromosome 22. Unlike the U lineage proteins encoded on chromosome 19, Ula and Uma likely play a nonclassical role as they lack conservation of key peptide binding residues, display limited polymorphic variation, and exhibit tissue-specific expression. We also describe a null haplotype at this chromosome 22 locus in which the *mhc1ula* and *mhc1uma* genes are absent due to a ∼30 kb deletion with no other MHC class I sequences present. Functional and non-functional transcripts of *mhc1ula* and *mhc1uma* were identified; however, *mhc1uma* transcripts were often not amplified or amplified at low levels from individuals possessing an apparently bona fide gene. These distinct U lineage genes may be restricted to the superorder Ostariophysi as similar sequences only could be identified from the blind cavefish (*Astyanyx mexicanus*), fathead minnow (*Pimephales promelas*), goldfish (*Carassius auratus*), and grass carp (*Ctenopharyngodon idellus*).

Keywords

Teleost; MHC class I; evolution; polymorphism; haplotype

Introduction

While the genes of the major histocompatibility complex (MHC) were named after the discovery that they encode the primary antigens responsible for determining transplant compatibility (Klein 2001), MHC molecules also play a central role in adaptive immunity through presenting antigens to T cells. MHC class I molecules primarily present cytosolic antigens to CD8$^+$ cytotoxic T cells while MHC class II molecules primarily present exogenous peptides to CD4$^+$ helper T cells (Horton et al. 2004). MHC class I molecules can be further classified as classical or nonclassical. Classical MHC class I molecules tend to be ubiquitously expressed and are highly polymorphic in their peptide binding groove which permits the presentation of a wide range of peptides to CD8$^+$ T cells. In contrast, nonclassical MHC class I molecules tend to display tissue specific expression, exhibit limited levels of polymorphic variation and many have been shown to bind peptides, lipids or vitamin metabolites (Kjer-Nielsen et al. 2012; Adams and Luoma 2013; Grimholt et al. 2015). Classical and nonclassical MHC class I molecules also can serve as ligands to receptors on natural killer (NK) cells (Raulet and Vance 2006; Halenius et al. 2014)

Genes encoding MHC class I molecules have been identified in all jawed vertebrate species examined, including cartilaginous fish, but are absent from jawless fish (Flajnik and Kasahara 2010). Zebrafish are an intriguing animal model in which to study the MHC class I genes as they have a complete reference genome (Howe et al. 2013) and provide a non-mammalian perspective on the origins of adaptive immunity. However, it has yet to be determined if the zebrafish MHC class I molecules carry out a similar repertoire of functions as the well-characterized human MHC class I (HLA) molecules or if these functions arose later in mammalian evolution. This question is especially relevant since the genomic organization of the genes encoding MHC molecules in zebrafish differs greatly from that of mammals - class I and class II loci are linked in tetrapods and cartilagionous fish but are not linked in zebrafish and other teleost species (Dijkstra et al. 2007).

Five phylogenetic lineages of MHC class I genes, termed U, Z, L, S, and P, have been identified from multiple bony fish (Dirscherl et al. 2014; Grimholt et al. 2015). The U and Z lineage proteins are predicted to bind peptides in a manner similar to classical MHC class I; however, only the U lineage genes share conserved synteny with the mammalian MHC core locus (Michalová et al. 2000). Although no functional data has been reported on the MHC class I Z lineage, there is direct experimental evidence of peptide presentation by U lineage MHC class I proteins (Chen et al. 2010). In contrast, the L, S and P lineage proteins are considered nonclassical due to their lack of a classical peptide binding groove, although virtually nothing has been reported on their function (Dijkstra et al. 2007; Lukacs et al. 2010; Dirscherl et al. 2014; Grimholt et al. 2015).

Zebrafish possess only the MHC class I U, Z and L lineages and have been reported to encode a single U lineage locus on chromosome 19 that displays significant haplotypic variation (McConnell et al. 2014). The zebrafish Z lineage genes are encoded on chromosomes 1 and 3 (Kruiswijk et al. 2002; Dirscherl and Yoder 2014) and the nonclassical L lineage genes are encoded on chromosomes 3, 8, and 25 (Dijkstra et al. 2007; Dirscherl et al. 2014).

Recent work characterizing the haplotypes of the U genes on chromosome 19 identified a putative U gene on chromosome 22 which was designated *mhc1ula* in accordance with nomenclature conventions (McConnell et al. 2014). McConnell et al. predicted that this chromosome 22 locus may display haplotypic variation due to the occasional presence of a band matching the predicted size of the fragment containing the *mhc1ula* gene in a Southern blot hybridized with a *mhc1uda* probe. We recently reported the presence of a second putative U gene adjacent to *mhc1ula* which has been designated *mhc1uma* (Dirscherl et al. 2014). This U locus also was depicted in a recent analysis of MHC class I sequences predicted from various teleost genomic databases (Grimholt et al. 2015).

Here, we report the cloning and characterization of full-length transcripts for both *mhc1ula* and *mhc1uma* and describe a null haplotype at this chromosome 22 locus with a ∼30 kb deletion that completely removes the *mhc1ula* and *mhc1uma* genes. In addition, we show that *mhc1ula* and *mhc1uma* differ from the other genes in the U lineage in that they exhibit many of the characteristics of nonclassical MHC class I genes including limited polymorphic variation, lack of key conserved residues, and tissue-specific expression. A better understanding of the recently identified MHC class I U lineage locus on chromosome 22 and its relation to the core MHC class I locus on chromosome 19 will inform the evolutionary history of this dynamic gene family and may have implications in the variable disease susceptibility of zebrafish.

Materials and Methods

Zebrafish and cell lines

All experiments involving live zebrafish were performed in accordance with relevant institutional and national guidelines and regulations and were approved by the North Carolina State University Institutional Animal Care and Use Committee. AB zebrafish were obtained through the Zebrafish International Resource Center (ZIRC; http://zebrafish.org). Tübingen (TU) zebrafish were provided by John Rawls (Duke University). Wild-type zebrafish (EKW and FS) were purchased from EkkWill Waterlife Resources (Ruskin, FL) and from Doctors Foster and Smith (Rhinelander, WI). Individual zebrafish from the LSB and HSB lines that display low and high stationary behavior, respectively, and are three generations away from wild-caught zebrafish from West Bengal, India were provided by John Godwin, North Carolina State University (Wong et al. 2012). The CG2 double haploid, clonal golden zebrafish line was provided by Sergei Revskoy, University of Illinois at Chicago (Mizgirev and Revskoy 2014). The zebrafish ZF4 fibroblast cell line was purchased from ATCC and grown at 28°C with 5% CO_2 in DMEM:F-12 medium (ATCC) with 10% fetal bovine serum (Atlanta Biologicals).

Transcriptome analyses

A single CG2 zebrafish was euthanized and the kidney, intestine, gills and spleen were dissected and combined for RNA extraction (TRIzol Reagent, Life Technologies). RNA was prepared for sequencing with the TruSeq RNA kit (Illumina) and sequenced (2×100 bp paired end reads) on a single lane of a HiSeq2000 (Illumina). In an effort to detect rare transcripts, a normalization procedure also was employed. Normalized cDNA was achieved

with the MINT-Universal cDNA synthesis and TRIMMER-DIRECT cDNA normalization kits (Evrogen) which was then prepared for sequencing with the TruSeq DNA kit (Illumina). The normalized library was sequenced (2×100 bp paired end reads) on a single HiSeq2000 lane. The Trinity Assembler (Grabherr et al. 2011; Haas et al. 2013) was used to create a *de novo* transcriptome assembly from the raw RNA-seq data, and the transcriptome was formatted into a BLAST database using the formatdb command for the stand-alone NCBI BLAST software. Raw sequence data has been deposited in the NCBI short read archives (SRA) with accession number SRP057116, and the assembled transcriptomes are available for BLAST searches (http://www4.ncsu.edu/~jayoder/).

Data mining and sequence analyses

Protein domains were identified using SMART software (Letunic et al. 2012). The hydrophobicity of the peptide binding region (α1 and α2 domains) was determined by calculating the GRAVY (grand average of hydropathy) value using the hydropathy scale determined by Kyte and Doolittle (Kyte and Doolittle 1982). Ula and Uma α domains were used as queries to search (BLASTp) the NCBI non-redundant protein database as well as to search (tBLASTn) the non-redundant nucleotide database, the EST database, the *de novo* genomic assemblies from double haploid AB and TU individuals generated by the Wellcome Trust Sanger Institute (www.sanger.ac.uk), and the recently published grass carp (*Ctenopharyngodon idellus*) genome (Wang et al. 2015). Protein sequences were aligned by Clustal Omega (Sievers et al. 2011) and phylogenetic trees constructed using MEGA4 software (Tamura et al. 2007). MHC class I sequences were from the following species: blind cavefish (*Asme, Astyanyx mexicanus*), goldfish (*Caau, Carassius auratus*), grass carp (*Ctid, Ctenopharyngodon idellus*), zebrafish (*Dare, Danio rerio*), Atlantic cod (*Gamo, Gadus morhua*), rainbow trout (*Onmy, Oncorhynchus mykiss*), fathead minnow (*Pipr, Pimephales promelas*), Atlantic salmon (*Sasa, Salmo salar*), pufferfish (*Taru, Takifugu rubripes*), and green pufferfish (*Teni, Tetraodon nigroviridis*). MHC class I sequences used for comparisons (and their GenBank accession numbers) are as follows: Caau-U (AM927065.1), Dare-Laa (NP_001017904.1), Dare-Lba (XP_001920787.3), Dare-Lca (XP_001340413.4), Dare-Lda (XP_001920290.4), Dare-Uba (AAH74095.1), Dare-Uca (XP_005159526.1), Dare-Uda (AAI28863.1), Dare-Uea (AAH53140.1), Dare-Ufa (AAH66754.1), Dare-Uga (AAH56726.1), Dare-Uha (AAI24258.1), Dare-Uia (AAH93149.1), Dare-Uja (AAH54592.1), Dare-Uka (AAI22402.1), Dare-Ula (KR086339), Dare-Uma (KR086346), Dare-Zaa (AHA37369.1), Dare-Zba (AHA37371.1), Dare-Zca (AHA37375.1), Dare-Zda (NP_001274026.1), Dare-Zea (AHA37389.1), Dare-Zfa (AHA37394.1), Dare-Zga and Dare-Zha (XP_001919291.4, two sets of α domains in one predicted transcript), Dare-Zia (XP_001919303.1), Dare-Zja (AHA37398.1), Dare-Zka (AHA37402.1), Dare-Zla (AHA37403.1), Gamo-Paa (GW844691.1), Onmy-Laa (ABI21842.1), Onmy-Lba (ABI21844.1), Onmy-Lca (ABI21845.1), Onmy-Lda (BAF37937.1), Onmy-Saa (AAB57877.1), Onmy-Uba (AAG25197.1), Onmy-Uca (BAD89552.1), Onmy-Uda (AAS93695.1), Onmy-Uea (BAD89553.1), Pipr-U (DT124856.1), Sasa-Saa (ACY30362.1), Sasa-Uba (AAN75116.1), Sasa-Uda (ACY30371.1), Sasa-Uga (ACX35601.1), Sasa-Uha1 (ACY30367.1), Sasa-Uha2 (ACY30368.1), Sasa-Ula (ABO13870.1), Sasa-Zaa (ACX35596.1), Sasa-Zba (ACX35613.1), and Sasa-Zca (ACX35618.1). Additional sequences include those predicted

in a recent publication characterizing teleost MHC class I sequences (Grimholt et al. 2015) as well as those identified in the grass carp genome. The sequence identifiers used by Grimholt et al. are included in the phylogenetic analysis in parentheses and the two grass carp sequences are encoded on scaffold CI01000031 (Wang et al. 2015).

Amplification of zebrafish *mhc1ula* and *mhc1uma* sequences

Adult zebrafish were euthanized, frozen in liquid nitrogen, and pulverized by mortar and pestle. Total RNA was purified from half of each individual as well as from ZF4 cells (TRIzol Reagent) and reverse transcribed into cDNA (SuperScript III Reverse Transcriptase, Life Technologies). Primer pairs were designed to amplify the entire coding sequences of *mhc1ula* and *mhc1uma* by RT-PCR. A primer pair that amplifies a portion of the β-*actin* transcript was used as a positive control for cDNA quality as previously described (Dirscherl and Yoder 2014).

Genomic DNA was extracted by proteinase K digestion and ethanol precipitation from the remaining half of the individuals used for RNA extraction as well as from ZF4 cells. Primer pairs were designed to amplify genomic sequences of *mhc1ula* and *mhc1uma* that span a single intron. The same β-*actin* primer pair described above was used as a positive control for genomic DNA quality.

Primer sequences and cycling parameters are provided in Online Resource 1, Table S1. All primer pairs were validated by sequencing of representative amplicons. New sequence data including non-canonical and non-functional transcripts have been deposited with GenBank under accession numbers: KR086339- KR086356.

Southern blot

Genomic DNA (5 μg) was digested with *PstI* (Thermo Scientific) according to the recommendations of Green and Sambrook (Green and Sambrook 2012) and separated on a 0.8% agarose gel at 30 V overnight. Following denaturation, DNA was transferred to a Zeta-Probe GT membrane (Bio-Rad) using the TurboBlotter rapid downward transfer system (Whatman) and alkaline transfer buffer. A 273 bp probe was generated by PCR amplification of the α3 domain of *mhc1ula* and a 430 bp control probe was generated by amplification of exon 2 from β-*actin (actb2)*. Primer sequences and cycling parameters are provided in Online Resource 1, Table S1. PCR products were purified and labeled using the DIG High Prime DNA Labeling and Detection Starter Kit II (Roche). Hybridization and detection were completed with a hybridization temperature of 42°C and an exposure time of 25 minutes using a ChemiDoc MP Imaging System (Bio-Rad).

Chromosome walking

PCR primer pairs were designed to amplify 150 to 250 bp flanking genomic amplicons (FGAs) at intervals upstream (–) and downstream (+) of *mhc1ula* and *mhc1uma* on chromosome 22 based on the reference genome (Zv9). Primer pairs were used with genomic DNA from different individual zebrafish to identify conserved FGAs. Using the Universal GenomeWalker 2.0 kit (Clontech), adapter-ligated libraries were generated with genomic DNA digested with four different blunt-end restriction endonucleases. Chromosome walking

was then accomplished by performing long-range nested PCR with gene-specific and adapter-specific primers starting from the genomic regions encoding the −5 kb and +13 kb FGAs. FGA and chromosome walking primer sequences are provided in Online Resource 1, Table S1.

Chromosome 22 U locus genotyping

Based on the sequence results from chromosome walking, primers were designed to genotype the chromosome 22 U locus by a duplex PCR assay (primer sequences in Online Resource 1, Table S1). In this assay a single forward primer upstream of the MHC class I locus is used with two different reverse primers that amplify a 634 bp fragment from the chromosome 22 haplotype in the reference genome encoding *mhc1ula* and *mhc1uma* (haplotype A) or a 747 bp fragment from a haplotype in which the *mhc1ula* and *mhc1uma* sequences have been deleted (haplotype B). The assay was applied to the same panel of genomic DNA from individual zebrafish described above. The assay was also used to genotype mating pairs of EKW and FS zebrafish using genomic DNA from fin-clips and then used to genotype representative offspring at three days post fertilization (dpf) in order to define inheritance ratios.

Evaluation of tissue expression

Additional primer pairs were designed to amplify transcript segments encoding the α1 to α3 domains of *mhc1ula* and *mhc1uma* from a panel of tissue-specific cDNAs derived from EKW individuals that had been genotyped as described above. The individuals were also genotyped by previously described methods (McConnell et al. 2014) and determined to by homozygous for haplotype A, B, C, or D at the U lineage locus on chromosome 19. Primers were designed to amplify the one, two, or three MHC class I sequences associated with each chromosome 19 haplotype from the tissue-specific cDNAs. β-*actin* primers were used as a control for tissue cDNA quality. Primer sequences and cycling parameters are provided in Online Resource 1, Table S1.

Results

Zebrafish *mhc1ula* and *mhc1uma* map to chromosome 22

In an effort to characterize novel MHC class I sequences, transcriptome (RNA-Seq) analysis was conducted from pooled tissues (spleen, kidney, intestine, and gills) of a single zebrafish from the double haploid, clonal CG2 line (Mizgirev and Revskoy 2014). BLASTn searches of this transcriptome database with *mhc1uga*, which is the only U lineage MHC class I gene encoded on chromosome 19 in CG2 zebrafish (McConnell et al. 2014), as a query identified *mhc1uga* as well as additional transcripts encoding two novel MHC class I sequences. The genes encoding these sequences have since been named *mhc1ula* and *mhc1uma* (discussed below). BLASTn searches of the zebrafish reference genome (Zv9) with *mhc1ula* and *mhc1uma* transcript sequences as queries revealed that both sequences map to a single scaffold on chromosome 22 (Zv9_scaffold3011; GenBank: NW_001878374.3). Using the predicted CG2 transcripts as a reference, primer pairs (Online Resource 1, Table S1) were designed to amplify transcripts encoding the entire open reading frames of *mhc1ula* and *mhc1uma* (discussed in detail below). Recovered transcripts indicate that each gene

possesses six exons and that the genes are separated by 248 bp in a head to head configuration in the reference genome (Fig. 1).

As all previously identified U lineage genes map to a single locus on zebrafish chromosome 19, it was unexpected that *mhc1ula* and *mhc1uma* map to a region of chromosome 22 with no apparent shared synteny with previously described MHC class I loci (Fig. 1). The nearest identifiable genes, *s1pr4* and *zgc:171566*, are located 211 kb and 273 kb from *mhc1ula* and *mhc1uma*, respectively.

Database analyses

No full-length transcripts encoding Ula or Uma can be identified in the current EST database. However, the complete coding sequence of *mhc1ula* can be accounted for by the joining of two ESTs (start codon to α3 domain, EH452054.1; α3 domain to stop codon, EH477854.1). The *mhc1uma* coding sequence is only supported by one EST that spans the start codon through part of the α1 domain (EB901873.1). In the *de novo* genomic assembly of a double haploid TU individual (www.sanger.ac.uk), contigs can be found that encode all three α domains for both *mhc1ula* and *mhc1uma*. In a similar *de novo* assembly from an AB individual (www.sanger.ac.uk), all three α domains of *mhc1uma* are present, but only the α1 domain, transmembrane, and cytoplasmic exons of *mhc1ula* are present. It is unclear if this represents an alternate haplotype in which a functional copy of only *mhc1uma* is present or if this is due to insufficient sequence coverage in the assembly.

Searching the NCBI non-redundant nucleotide and protein databases did not reveal any definitive orthologs of *mhc1ula* or *mhc1uma*. The most similar non-zebrafish sequences identified were MHC class I proteins from various carp species. However, when these carp sequences are compared to the zebrafish sequence database they more closely match U proteins encoded on chromosome 19. Searching the non-zebrafish EST database did identify two sequences, one from fathead minnow (*Pimephales promelas*, GenBank: DT124856.1) and one from goldfish (*Carassius auratus*, GenBank: AM927065.1), that show significant similarity to *mhc1ula* (61–62% identity). Sequences similar to *mhc1ula* were also identified from the recently published grass carp (*Ctenopharyngodon idellus*) genome (Wang et al. 2015). A comparison (tBLASTn) of the α1 to α3 domains of Ula to the genomic scaffolds of an adult gynogenetic female identified a single scaffold (CI01000031) that is predicted to encode two full-length Ula-like proteins in a head to head configuration. This scaffold was excluded from the linkage-map-anchored draft of the genome generated by Wang et al. due to a predicted cross-chromosome arrangement representing synteny between zebrafish chromosomes 10 and 22 and grass carp linkage group 24. However, the predicted U genes are supported by sequences identified in the published grass carp gene models database which takes into account RNA-seq data. A BLASTn alignment of this grass carp scaffold with the zebrafish scaffold encoding *mhc1ula* and *mhc1uma* suggests that these loci share conserved synteny (Online Resource 1, Fig. S1).

Phylogenetic analyses

A phylogenetic comparison of Ula and Uma sequences to those representing all known MHC class I lineages in bony fish confirms that they are most similar to genes of the MHC

class I U lineage (Fig. 2) and validates their nomenclature. The EST sequences from goldfish and fathead minnow as well as the sequences predicted from the grass carp genome group most closely with Dare-Ula. The tree includes MHC class I sequences of interest that were predicted from a recent analysis of various teleost genomic databases (Grimholt et al. 2015). Two U sequences predicted from a single scaffold (GenBank: KB882234.1) of the blind cavefish (*Astyanyx mexicanus*) group with Ula and Uma though they only share 30–36% identity. The genes that are reported to flank this predicted U locus in cavefish display conserved synteny with zebrafish chromosome 5 rather than chromosome 22 (Grimholt et al. 2015). A third MHC class I protein from the same cavefish scaffold (AM35) which was predicted to be of the U lineage by Grimholt et al. does not group with any of the five sequence lineages in our analysis.

An additional lineage of MHC class I genes has been described in salmonids and has thus been termed the S lineage. The Atlantic salmon *SAA* gene was originally named *UAA* but renamed to indicate its phylogenetic divergence from the other U lineage genes (Lukacs et al. 2010). The S lineage as well as a fifth MHC class I lineage designated P have recently been identified in other fish species but appear to be absent from zebrafish (Grimholt et al. 2015). Care was taken to ensure that *mhc1ula* and *mhc1uma* were correctly assigned to the U lineage rather than the S or P lineages. While the minimum evolution tree in Figure 2 shows that Ula and Uma are the most divergent members of the group, they do fall within the U lineage.

Transcript variants

Full-length transcripts of *mhc1ula* and *mhc1uma* were amplified and sequenced from multiple individual zebrafish. The amino acid sequences encoded by each functional transcript variant (tv) are included in Online Resource 1, Fig. S2. While almost no polymorphic variation was detected for either gene, both *mhc1ula* and *mhc1uma* are subject to alternative splicing as summarized in Online Resource 1, Fig. S3. All identified non-canonical splice variants of *mhc1ula* are presumably non-functional due to a frameshift caused by the retention of an intron (tv3) or by deletions at the beginning of the α1 domain and/or the α3 domain (tv4–7). A splice variant of *mhc1uma* was identified that lacks the α1 domain but with the reading frame maintained (tv4–5), possibly representing a functional isoform. All other recovered non-canonical *mhc1uma* transcripts retained one or two introns resulting in frameshifts (tv6–11) and were predicted to encode secreted (tv8, tv10 and tv11) or non-functional molecules.

Predicted nucleotide and protein sequences were identified from GenBank that correspond to some of the non-canonical transcripts detected by PCR. The zebrafish non-redundant protein database includes one predicted protein corresponding to Ula (XP_001336749.4) and three predicted protein isoforms corresponding to Uma (isoform X1, XP_005161940.1; isoform X2, XP_005161941.1; isoform X3, XP_009294249.1). The majority of the sequenced transcripts matched the predicted Ula sequence or the Uma isoform X1 sequence, both of which are designated here as tv1. The sequence encoding Uma isoform X2 has a three nucleotide deletion at the beginning of exon 5, a polymorphism that was observed in Uma_tv5 and Uma_tv7. The sequence encoding Uma isoform X3 has a six nucleotide

deletion at the beginning of exon 6, a polymorphism that was observed in Uma_tv3. A non-coding predicted transcript corresponding to *mhc1ula* (XR_659472.1) is present in the non-redundant nucleotide database with a deletion of 22 nucleotides at the beginning of the α3 domain, the same deletion that was observed in Uma_tv6 and Uma_tv7.

Divergent residues at locations of functional consequence

An alignment of representative sequences for all described zebrafish U proteins (Fig. 3) reveals that the divergence between the U genes encoded on chromosome 22 and chromosome 19 affects some key structural residues as well as residues thought to be important for protein-protein interactions. In both Ula and Uma the leader peptide and transmembrane regions differ greatly from those of the chromosome 19 molecules, but the cysteine pairs that form disulfide bonds in the α2 and α3 domains are conserved. Ula and Uma show conservation of one of the four tyrosine residues important for anchoring the N-terminus of bound peptide (position 156 in Fig. 3); however, a second conserved tyrosine (position 168 in Fig. 3) is replaced with the structurally similar phenylalanine (F) in Ula and Uma which may maintain peptide interactions. Ula shows conservation of four of five residues important for anchoring the C-terminus of bound peptide while Uma has three of five; however, a conserved F (position 119 in Fig. 3) is replaced by the structurally similar Y in Ula which also may contribute to peptide interactions. It has been shown in mammals that a loop of acidic residues in the α3 domain is the primary site of interaction with CD8 on cytotoxic T cells (Wang et al. 2009). In this region, Ula is slightly less acidic than the chromosome 19 U lineage genes while Uma has a drastically different sequence with only one acidic residue and a putative insertion of two residues. The peptide binding regions of Ula and Uma have hydrophobicity scores of −0.727 and −0.801, respectively. These scores are in line with the previously reported average hydrophobicity (−0.704) of the other zebrafish U lineage proteins (Grimholt et al. 2015).

Identification of an alternate haplotype lacking *mhc1ula* and *mhc1uma*

The expression of *mhc1ula* and *mhc1uma* was evaluated from a panel of eleven individual zebrafish representing a range of genetic backgrounds and a zebrafish cell line. The panel included standard laboratory strains (AB, TU), wild-type zebrafish (EKW, LSB, HSB), a clonal line (CG2), and a fibroblast cell line (ZF4). Transcripts of *mhc1ula* were detected from nine of eleven individual zebrafish and *mhc1uma* transcripts were detected from six of those individuals expressing *mhc1ula* (Fig. 4a). Genomic amplicons for *mhc1ula* and *mhc1uma* were detected from the nine individual zebrafish expressing *mhc1ula* transcripts. Amplicons for *mhc1ula* and *mhc1uma* were not detected from the cDNA or genomic DNA of the ZF4 cell line. These results indicate that two individual zebrafish (AB1 and EKW2) and the ZF4 cell line lack the *mhc1ula* and *mhc1uma* genes. To confirm the absence of *mhc1ula* and *mhc1uma* from the genomes of the these zebrafish and ZF4 cells, genomic DNA from these individuals and cells, as well as from five individuals that encode *mhc1ula* and *mhc1uma*, were analyzed by Southern blot with a probe designed to hybridize to the α3 domain of *mhc1ula* (Fig. 4b). The probe hybridized to a genomic fragment of the predicted size (2.7 kb based on *Pst*I restriction sites) from the five individuals shown to encode *mhc1ula* and *mhc1uma* but failed to hybridize to any genomic fragments from the AB1 individual or ZF4 cells. This supports the existence of an alternate haplotype in which

mhc1ula and *mhc1uma* are absent from the genome. The probe hybridized to a larger band in the EKW2 genomic DNA suggesting a third alternate haplotype in this individual. The larger band most likely represents cross-reactivity between the probe and another U gene, although it is not known if that gene would represent a third chromosome 22 haplotype or a haplotype of the chromosome 19 locus not present in the other individuals.

In order to determine the haplotype sequence when *mhc1ula* and *mhc1uma* are absent, chromosome walking was applied to the genome of the AB1 individual. Genomic PCR was performed to identify the region of chromosome 22 that was lost in the AB1 individual but present in CG2 zebrafish. As shown in ,Online Resource 1, Fig. S4 and summarized in Figure 5a, flanking genomic amplicons (FGAs) approximately 5 kb upstream (−5 kb) and 13 kb downstream (+13 kb) of the locus could be amplified from the genomes of both AB1 and CG2 while all primer pairs designed between these FGAs only amplified their target sequences from the CG2 genome. Complete chromosome walking between the FGAs at −5 kb and +13 kb of the AB1 individual revealed that a ∼30 kb region, which is present in the reference genome and includes both *mhc1ula* and *mhc1uma* genes, is absent from the AB1 genome: this sequence has been designated haplotype B for this locus (GenBank: KR086338). Figure 5b shows a dot plot generated from the alignment of haplotype B to the reference genome (Zv9) haplotype, haplotype A. While the primer sequences of the +12 kb FGA are present in haplotype B, the absence of a detectable amplicon from AB1 can be explained by the presence of a ∼1.7 kb insertion between the primer sequences. This insert includes a dinucleotide (TA) repeat that was refractory to high quality sequencing. The number of repeats at this site represented in haplotype B, $(TA)_{80}$, is based on direct sequencing and supported by the length of PCR amplicons spanning the repeat (not shown).

A genotyping strategy for the chromosome 22 U locus

Based on the ∼30 kb deletion on chromosome 22 in the AB1 individual, primers were designed to genotype the individuals in the original panel by a duplex PCR assay (Fig. 6a). As predicted, a haplotype A amplicon (634 bp) could be detected from all nine individuals that encode *mhc1ula* and *mhc1uma*. The haplotype B amplicon (747 bp) was detected from the AB1 individual and ZF4 cells but not from the EKW2 individual that also lacks *mhc1ula* and *mhc1uma*. This further supports the existence of a third haplotype in the EKW2 individual. Amplicons for both haplotypes A and B were detected from two individuals, EKW1 and LSB2, suggesting that these individuals are heterozygous at the chromosome 22 U locus. The same FGA primer pairs described above demonstrate that this region of chromosome 22 in the EKW2 individual differs from haplotypes A (CG2 zebrafish) and B (the AB1 individual) (Online Resource 1, Fig. S4). Chromosome walking was initiated to define the sequence of this third haplotype (haplotype C); however, repetitive regions were encountered that made it impossible to design unique primers for the subsequent "walking" steps. The sequence of haplotype C remains to be defined.

In order to demonstrate that haplotypes A and B represent the same locus and do not represent two different genomic loci, the ratios of inheritance were investigated (Fig. 6b). Male and female zebrafish of the FS background were identified that were homozygous for haplotype A or B. Crosses between males and females of the same genotype were performed

and eight offspring were selected for genotyping at 3 dpf. All offspring genotyped from each cross had the same genotype as their respective parents. A male and female of the EKW background that were heterozygous for haplotypes A and B also were crossed: the genotyping results of 23 of their offspring approximate the 1:2:1 ratio that is expected in the case of simple Mendelian inheritance (7 A/A : 12 A/B : 4 B/B).

Tissue-specific expression

Individual zebrafish of the EKW background that were homozygous for haplotype A on chromosome 22 and homozygous for haplotype A, B, C, or D on chromosome 19 were identified by genotyping from fin clip DNA and euthanized for tissue dissection to create a panel of tissue-specific cDNAs. All MHC class I genes encoded on chromosome 19 could be amplified from the cDNA of all tissues tested except for *mhc1uca* and *mhc1uka* (Fig. 7). This corresponds with recent qPCR data showing *mhc1uca* and *mhc1uka* to be expressed at much lower levels than the other MHC class I U genes encoded on chromosome 19 (McConnell et al. 2014). In contrast, amplicons spanning the α1-α3 domains of *mhc1ula* and *mhc1uma* were amplified most strongly from intestine and gill cDNA in all four individuals.

Discussion

This is the first report of full-length transcripts for two new zebrafish MHC class I genes of the U lineage, *mhc1ula* and *mhc1uma*. These are the first U genes identified in zebrafish that map to a locus other than the core MHC on chromosome 19. The detailed characterization of this locus suggests that *mhc1ula* and *mhc1uma* represent a distinct subset of U lineage genes that have a nonclassical function and may be limited to the superorder Ostariophysi.

Three features of *mhc1ula* and *mhc1uma* lead to their characterization as nonclassical MHC class I genes. First, whereas classical MHC class I genes are typically polymorphic which increases the diversity of antigens that can be presented, *mhc1ula* and *mhc1uma* display very little polymorphism suggesting that they present a small number of antigens or have a function other than antigen presentation. Second, classical MHC class I genes are ubiquitously expressed but *mhc1ula* and *mhc1uma* display tissue-specific expression (Fig. 7). Third, although Ula and Uma possess conserved residues important for structural features (i.e. disulfide bonds), they lack certain conserved residues that are involved in protein-protein interactions in other U lineage proteins (i.e. peptide anchor residues, CD8-interacting residues): this suggests that Ula and Uma exhibit structures similar to the classical U molecules but likely have a function other than antigen presentation to CD8[+] T cells.

Evidence from mammals indicates that nonclassical MHC class I molecules can carry out functions such as specialized antigen presentation, IgG transport, or immunoregulation via interaction with NK-cell receptors (Watanabe et al. 2004; Rodgers and Cook 2005). It has been predicted that the relatively high average hydrophobicity of the nonclassical L lineage proteins in zebrafish (−0.377) indicates that they may bind lipids as the mammalian CD1 proteins have similarly high hydrophobicity scores (Grimholt et al. 2015). However, Ula and Uma likely do not bind lipids as their hydrophobicity values (−0.727 and −0.801, respectively) are similar to those of the other zebrafish U lineage proteins as well as HLA-

A2 with a reported value of −0.902. The fact that *mhc1uma* transcripts were sometimes amplified at low levels or not at all from whole animal cDNA (Fig. 4) provides further support for its nonclassical function, perhaps with a specialized stimulus required to induce expression. The divergent signal peptide and transmembrane sequences present in Ula and Uma might also indicate that the molecules undergo specialized trafficking rather than ubiquitous expression at the cell surface.

In searching the NCBI non-redundant protein database, the non-zebrafish proteins with the greatest similarity to Ula and Uma are actually more similar to the proteins encoded by the MHC class I U genes on zebrafish chromosome 19. However, EST sequences from goldfish and fathead minnow were identified that are most similar to *mhc1ula*. Additionally, U sequences recently predicted from the genomic database of the blind cavefish are more similar to Ula and Uma than they are to the U proteins encoded on chromosome 19. Taken together, these database and phylogenetic analyses suggest that these distinct U lineage genes are unique to the teleost superorder Ostariophysi. However, the evolutionary origin of these genes remains unclear. Analyzing the flanking genes of the zebrafish chromosome 22 U locus revealed no shared synteny with previously described MHC class I loci. While the grass carp scaffold encoding similar U sequences shows some evidence of conserved synteny with the zebrafish locus (Online Resource 1, Fig. S1), the cavefish scaffold encoding AM36 and AM37 shares conserved synteny with zebrafish chromosome 5 rather than chromosome 22 (Grimholt et al. 2015). It is possible that over the ∼150 million years of divergent evolution between the zebrafish and the cavefish (Carlson et al. 2014) that this locus has undergone interchromosomal translocations as seems to have recently occurred with the MHC class I sequences in the tammar wallaby (Deakin et al. 2007).

We define here an alternate haplotype for the chromosome 22 U locus, haplotype B, in which *mhc1ula* and *mhc1uma* are absent and no other U genes are present. A similar null haplotype with a large-scale deletion has been described for the MHC class I related genes MIC-A and MIC-B in humans (Komatsu-Wakui et al. 1999). This haplotypic variation is distinctly different from the chromosome 19 U locus in zebrafish in which the six different described haplotypes encode one to three unique MHC class I genes (McConnell et al. 2014). The fact that individuals that are homozygous for haplotype B are present in stocks of wild type zebrafish and appear healthy suggests that *mhc1ula* and *mhc1uma* are not essential for survival. It is noted that one individual recently derived from the wild is heterozygous for haplotypes A and B (LSB2 in Fig. 6) demonstrating that the null haplotype is not a derived feature only present in laboratory zebrafish strains. Even when a zebrafish appears to encode bona fide *mhc1ula* and *mhc1uma* genes, they are often expressed as nonfunctional transcripts (Online Resource 1, Fig. S3) supporting the idea that these genes may be transitioning to pseudogenes due to lack of selective pressures. It will be of interest to determine if similar null alleles can be identified in other cyprinid species.

The potential differential susceptibility to disease between individuals homozygous for haplotype A or B remains to be evaluated. Nonclassical MHC class I genes have been associated with resistance to bacterial and viral diseases in salmonid species (Johnson et al. 2008). Research characterizing tuberculosis susceptibility in mice suggests that nonclassical MHC class I molecules may contribute to disease resistance via TAP-dependent or -

independent antigen presentation to distinct populations of T cells such as CD8⁻ T cells (Sousa et al. 2000). The detailed description of the locus encoding *mhc1ula* and *mhc1uma* presented here provides a framework for future experiments that may be able to determine if these nonclassical MHC class I genes contribute to zebrafish immunity through similar interactions with distinct T cell populations or via activating/inhibitory regulation of NK cells.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

Adams EJ, Luoma AM. The adaptable major histocompatibility complex (MHC) fold: structure and function of nonclassical and MHC class I-like molecules. Annu Rev Immunol. 2013; 31:529–561. [PubMed: 23298204]

Carlson BM, Onusko SW, Gross JB. A High-Density Linkage Map for Astyanax mexicanus Using Genotyping-by-Sequencing Technology. G3. 2014; 5:241–251. [PubMed: 25520037]

Chen W, Jia Z, Zhang T, et al. MHC class I presentation and regulation by IFN in bony fish determined by molecular analysis of the class I locus in grass carp. J Immunol. 2010; 185:2209–2221. [PubMed: 20660353]

Deakin JE, Siddle HV, Cross JGR, et al. Class I genes have split from the MHC in the tammar wallaby. Cytogenet Genome Res. 2007; 116:205–211. [PubMed: 17317961]

Dijkstra JM, Katagiri T, Hosomichi K, et al. A third broad lineage of major histocompatibility complex (MHC) class I in teleost fish; MHC class II linkage and processed genes. Immunogenetics. 2007; 59:305–321. [PubMed: 17318646]

Dirscherl H, McConnell SC, Yoder JA, de Jong JLO. The MHC class I genes of zebrafish. Dev Comp Immunol. 2014; 46:11–23. [PubMed: 24631581]

Dirscherl H, Yoder JA. Characterization of the Z lineage Major histocompatability complex class i genes in zebrafish. Immunogenetics. 2014; 66:185–198. [PubMed: 24287892]

Flajnik MF, Kasahara M. Origin and evolution of the adaptive immune system: genetic events and selective pressures. Nat Rev Genet. 2010; 11:47–59. [PubMed: 19997068]

Grabherr MG, Haas BJ, Yassour M, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011; 29:644–652. [PubMed: 21572440]

Green, MR.; Sambrook, J. Molecular Cloning: A Laboratory Manual. Fourth Edi. Cold Spring Harbor Laboratory Press; 2012.

Grimholt U, Tsukamoto K, Azuma T, et al. A comprehensive analysis of teleost MHC class I sequences. BMC Evol Biol. 2015

Haas BJ, Papanicolaou A, Yassour M, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 2013; 8:1494–512. [PubMed: 23845962]

Halenius A, Gerke C, Hengel H. Classical and non-classical MHC I molecule manipulation by human cytomegalovirus: so many targets—but how many arrows in the quiver? Cell Mol Immunol. 2014:139–153. [PubMed: 25418469]

Horton R, Wilming L, Rand V, et al. Gene map of the extended human MHC. Nat Rev Genet. 2004; 5:889–899. [PubMed: 15573121]

Howe K, Clark MD, Torroja CF, et al. The zebrafish reference genome sequence and its relationship to the human genome. Nature. 2013; 496:498–503. [PubMed: 23594743]

Johnson, Na; Vallejo, RL.; Silverstein, JT., et al. Suggestive association of major histocompatibility IB genetic markers with resistance to bacterial cold water disease in rainbow trout (Oncorhynchus mykiss). Mar Biotechnol. 2008; 10:429–437. [PubMed: 18274824]

Kjer-Nielsen L, Patel O, Corbett AJ, et al. MR1 presents microbial vitamin B metabolites to MAIT cells. Nature. 2012; 491:717–723. [PubMed: 23051753]

Klein J. George Snell's first foray into the unexplored territory of the major histocompatibility complex. Genetics. 2001; 159:435–439. [PubMed: 11606523]

Komatsu-Wakui M, Tokunaga K, Ishikawa Y, et al. MIC-A polymorphism in Japanese and a MIC-A-MIC-B null haplotype. Immunogenetics. 1999; 49:620–628. [PubMed: 10369920]

Kruiswijk CP, Hermsen TT, Westphal AH, et al. 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. 2002; 169:1936–1947. [PubMed: 12165519]

Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol. 1982; 157:105–132. [PubMed: 7108955]

Letunic I, Doerks T, Bork P. SMART 7: recent updates to the protein domain annotation resource. Nucleic Acids Res. 2012; 40:D302–5. [PubMed: 22053084]

Lukacs MF, Harstad H, Bakke HG, et al. Comprehensive analysis of MHC class I genes from the U-, S-, and Z-lineages in Atlantic salmon. BMC Genomics. 2010; 11:154. [PubMed: 20205726]

McConnell SC, Restaino AC, De Jong JLO. Multiple divergent haplotypes express completely distinct sets of class i MHC genes in zebrafish. Immunogenetics. 2014; 66:199–213. [PubMed: 24291825]

Michalová V, Murray BW, Sültmann H, Klein J. A contig map of the Mhc class I genomic region in the zebrafish reveals ancient synteny. J Immunol. 2000; 164:5296–5305. [pii]. [PubMed: 10799891]

Mizgirev IV, Revskoy S. A new zebrafish model for experimental leukemia therapy. Cancer Biol Ther. 2014; 9:895–902. [PubMed: 20339318]

Raulet DH, Vance RE. Self-tolerance of natural killer cells. Nat Rev Immunol. 2006; 6:520–531. [PubMed: 16799471]

Rodgers JR, Cook RG. MHC class Ib molecules bridge innate and acquired immunity. Nat Rev Immunol. 2005; 5:459–471. [PubMed: 15928678]

Schwartz S, Zhang Z, Frazer Ka, et al. PipMaker--a web server for aligning two genomic DNA sequences. Genome Res. 2000; 10:577–586. [PubMed: 10779500]

Sievers F, Wilm A, Dineen D, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011

Sousa AO, Mazzaccaro RJ, Russell RG, et al. Relative contributions of distinct MHC class I-dependent cell populations in protection to tuberculosis infection in mice. Proc Natl Acad Sci U S A. 2000; 97:4204–4208. [PubMed: 10760288]

Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol. 2007; 24:1596–1599. [PubMed: 17488738]

Wang R, Natarajan K, Margulies DH. Structural basis of the CD8 alpha beta/MHC class I interaction: focused recognition orients CD8 beta to a T cell proximal position. J Immunol. 2009; 183:2554–2564. [PubMed: 19625641]

Wang Y, Lu Y, Zhang Y, et al. The draft genome of the grass carp (Ctenopharyngodon idellus) provides insights into its evolution and vegetarian adaptation. Nat Genet. 2015; 47:625–631. [PubMed: 25938946]

Watanabe Y, Maruoka T, Walter L, Kasahara M. Comparative genomics of the Mill family: A rapidly evolving MHC class I gene family. Eur J Immunol. 2004; 34:1597–1607. [PubMed: 15162429]

Wong RY, Perrin F, Oxendine SE, et al. Comparing behavioral responses across multiple assays of stress and anxiety in zebrafish (Danio rerio). Behaviour. 2012; 149:1205–1240.
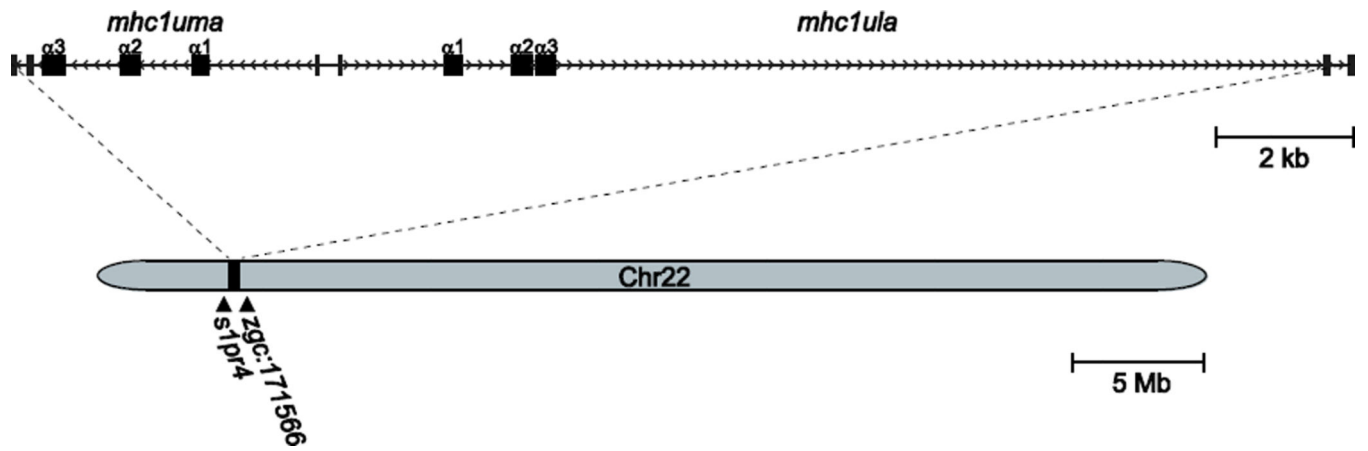
Fig. 1.
Genomic annotation of *mhc1ula* and *mhc1uma* on chromosome 22. Exons are represented as rectangles and introns are represented as arrows pointing in the direction of transcription. Black triangles indicate the position of the genes flanking *mhc1ula* and *mhc1uma*. Annotation is based on Zv9_scaffold3011 (GenBank: NW_001878374.3).
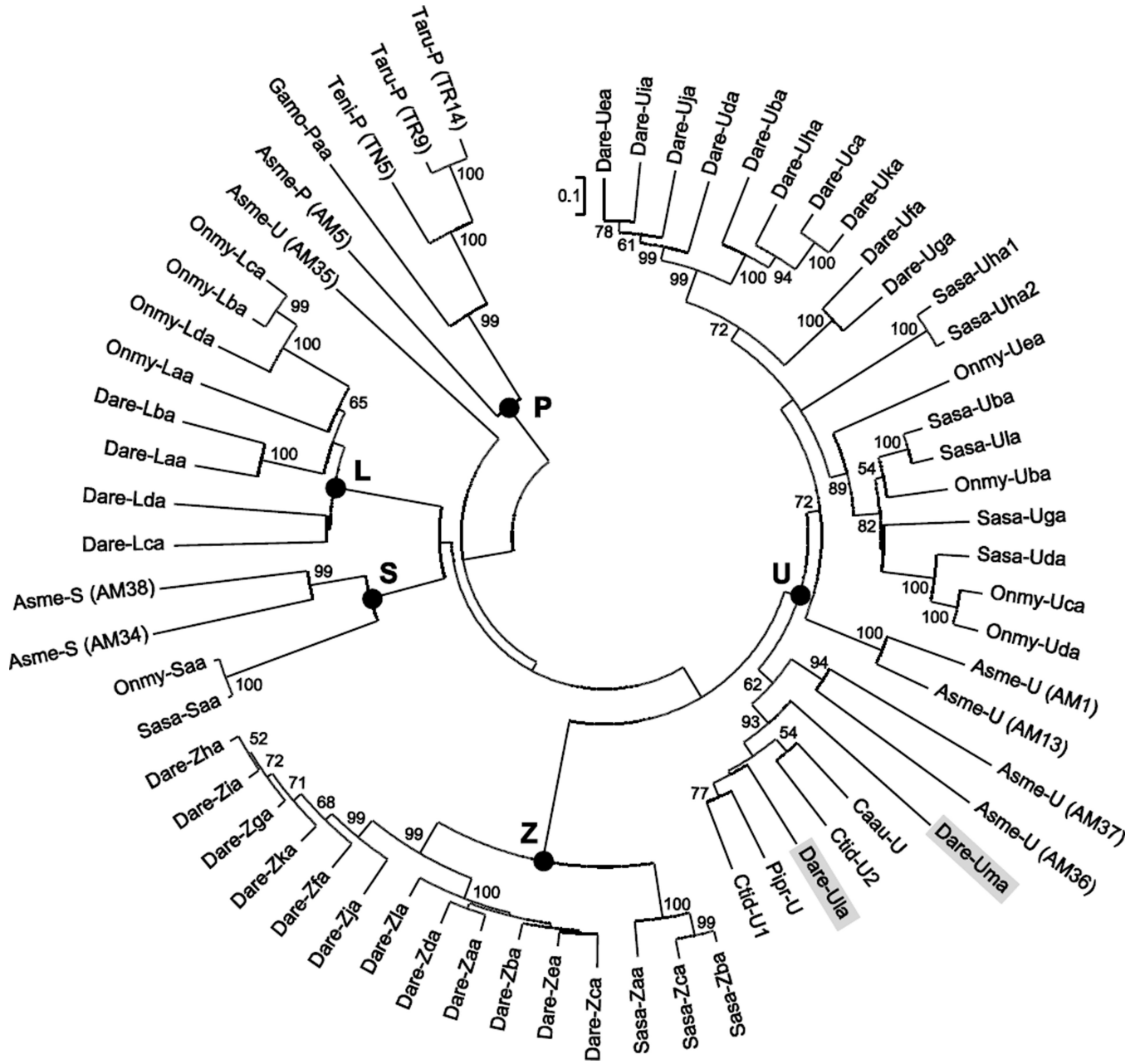
Fig. 2.
Phylogenetic comparison of Ula and Uma to representative MHC class I sequences of teleosts. A minimum evolution tree was constructed from the alignment of the α1-α3 domain sequences from the different MHC class I lineages of blind cavefish (*Asme, Astyanyx mexicanus*), goldfish (*Caau, Carassius auratus*), grass carp (*Ctid, Ctenopharyngodon idellus*), zebrafish (*Dare, Danio rerio*), Atlantic cod (*Gamo, Gadus morhua*), rainbow trout (*Onmy, Oncorhynchus mykiss*), fathead minnow (*Pipr, Pimephales promelas*), Atlantic salmon (*Sasa, Salmo salar*), pufferfish (*Taru, Takifugu rubripes*), and green pufferfish (*Teni, Tetraodon nigroviridis*). The consensus tree was based on 2000

bootstrap replications. Bootstrap values below 50% are not shown. Sequence identifiers are provided in Materials and Methods.

Leader peptide

```
Uba   -------VCS-ILIGLAAIVVCIQILSG
Uca   MADRITQLKMEYIIAIVVCIPLVDG
Uda   -------IQ-IVILALGABISIL--
Uea   -------IQ-VILLILLGABIALA-
Ufa   -------RQ-LVLIIALGVBIALA-
Uga   -------IQ-VILFLALGABIALA-
Uha   -------VCS-ITGIAAIVVCIQILSG
Uia   -------IQT-IIILIALGABIANL-
Uja   -------IRA-IILIFLIGLNITNLV-
Uka   ---------MEYIIAIVVCIPLVDG
Ula   ----------KIIVVFICIPLLVCS
Uma   -------ISI-IKFIIFFFNILPLLVYS
```

α1

```
        1                         30                         60
Uba   VIHSWAIIIGTIGITIFPPEIVALNLIDDQLIGYIDSIINIFKSQFQILIEDNICKLYDIQQINILLCYPEVIKNNIKVVMERFNQTQ
Uca   VIHSWNAIIVVATGISQFPEIVAIQNLIDGQLIGYIDSIINCFKSQFQILKLIEDKICTAIDIQILITILQRHTAKIKNNIKVVAIERFNQTQ
Uda   CIHSLIVIITCVSGDIDFPEITAIGIVDCFCFYMIDSNKMISVPKTIWIIQNECADYWRLIQIIAIGYEQIYIVDEITLKIRFNQSQ
Uea   CGTHSLIVIIITCVSGDIDFPEITVVSLIVDCFCFVMIDSNLMISIPKTIWMIQNVCEIYWIRETQIFTCAIPIVIKNNIQVIKLRFNQTQ
Ufa   CKHSHTVIIIVLWGIPLDFPEIVZATNVDCMQVMIIDSEIKEVIPFQIWVIGALDEQILQRIKIQRSNMIQIIKNNINIAMERFNQTQ
Uga   CGTHSLIVIIITATGITIFPPEIVAMNVIDDQLIGYIDSIKMEAIVPKTIWIIQNECADYWDIINIQRLIITEQATKNNIQVAKERFNQSQ
Uha   VIHSWAIIATIGITIFPPEIVALNVIDDQLIGYIDSIINIFKSQFQILKLIEDNICKLYDIQQEITILQCHTASIKNNVKVAMERFNQTQ
Uia   VIHSMAYIILCVELDSFIPEITSVGTVDDIPFVILIDSNIKRMIPKTIWIILNECAIKLWIREIQRESIVPIVIKNNIQLAIQRFNQTA
Uja   VIHHSEYIITTATSGIIEGIFPPEIVIVGMVDGQQINIYIDSTHSLMIVIPKTIWITDIVISGAIDPDIINRIGQHASCTILNIKNNIDEVAIKQRFNQTG
Uka   VIHSWNAIIVVATGISQFPEIVALNLIDGQLIGYIDSIINIFKSQFQILKLIEDKICTAIDIQILITILQRHTATIKNNIKVAIERFNQTQ
Ula   CYHSLIITIICIRCASETIIFIAVAIIDCEQIDYIDSVIINILIPIQDMMEYASEIIKEDCKIRGDVEQIYISIHIVLVQRFNQSN
Uma   ELHSFVVTIMMG----NQTIFIAATIDCQQIGHILIDKNIKTLVPIQDWNEQIKSKTPIKVDILIIE-TVQRTIYNINIHDLILCQIPQSN
                   n                         n                         c
```

α2

```
        90                        120                        150
Uba   GVETFCPIIYCCEIMDDI-GNKCVHWCIGYDGEDFISLDIKITLTITAANSQAIITKVIIDSTGAEANYWIGILENICHIIVQKIVGYGKDITLER
Uca   GVETFLQEIIYCCEWDDIQTCNLNAIPIRQYGYNGEDFILTIDFMEIMRWITFVQQAMITIQKWNNDRGFIESIDMNYIRSECIEIWLQKYIEYGKSSIMIKT
Uda   GIEHIILDEIMYCCEWDDITRATNGEIYCDSMDGEDIVYLDLIKEMRYISFVQQGLITIQNWNNDKIFIAQQINILSICIEIWLQKYIQGKSSILKIT
Uea   GVETFTCQIIYCCEWDDITEARNGEICCIGYDGEDFLSLDILKEIRTISFVQQGIITIQNWNNNRCFIQSYGGIISIVCIEIWLQKYIQGKSSIKKIT
Ufa   GVETFTFIVYCCEIMDD--CTTRCHWIQYGYNGEDFISIFDKNTITNIVANPQIVIIKNTIKDANIIQNIYRIQYLENQCIIIRIQIVIGYKGDSIGCK
Uga   GVETFCVIMYCCEILDDI-CSTRCIWIQYGYDGEDFLSLDIKSTITLITATKPQAVIIKNIIRIDADNIDRQYTISILENICHIIWKIVIDYGKDILKIT
Uha   GVEANQRIIYCCEWDDITCAINGEICQIGYDGEDFLSLDILKELRWISFVQQGLMITIQKILNNDRANTIYYICNYITIECIEIWLQKYIQGKSSIQKIT
Uia   GVRIVIVQNIIYCCEWDDTICAKNGEIDQYGYDGEDIILTLIKELRWISFVQQGIITIQNWNNKNRQIYIYNYISIECIEIWLQKYIQGKSSIQKIT
Uja   GVETFVCAIMYCCEWDDTICAKSIIKAFCGYDGEDFLSVDFIKELRYISFVQQGIPIIVQKWNNDRGMIESIDINYIGIVCIEIWLQKYIQGKSSILKIT
Uka   GVETFCQEIIYCCEWDDTICNLNAIFIRQYGYNGEDFIIILDFRIVWRIIIFVQQAMIITIQKWNNDKCFIESIDINYIRSECIEIWIFKIAIMICKSSILKIT
Ula   GVEIVQRIIYCCEWDDISNREQIKAFCGYDGEDIVIMLDMREIRYIIIFVQQGEIIVQNWNDRQIKILQBYIKYICIVITIPKQIGLRIADILIIR
Uma   CAHIYICGRHCCSIWEDKINYHDEIIHDYIYDGDDFITIDVKRIKYIRASISKIKYIVKIINNDREQNKILIEIIKLGCIIYILIIEIJVITINAFGGT
            *                        c                         c cc              n    *    n
```

α3

```
        180                       210                       240                       270
Uba   VSPQVSLLQKSSSSPVVCHVTGFYPSGIKISWQRNGQDHDED--VELGELIPNEDGTYIRTSTINVKPEIVLKDI--FSCVVEHCSKIINSITITEDEIRTNN-------
Uca   VSPQVSVLQRYYFISPVVCHATGCFYPSGIKISWQKNGQDHDED--VELGELLPNADGTFQKRSTILNVKPEEWKNSI--FSCVVEHCQCERILT---EDEIRTNN-----
Uda   VSPQVSLLQKSSSSPVVCHATGFYPSGIKISWQRNGQDHDED--VEIPEHIIAFPIEDGTIQKSSTILVTPEEIWKNNKI--LRCVVEHCQCKIKTE---K-IIRTNE----
Uea   VSPQVSLLQKSSSSPVVCHATGFYPSGIKISWQRNGQDHDED--VELGELLPNADGTFQKTSTILNVKPEEWKNNK--FSCVVEHCQCKERAV---DDIPDLINE----
Ufa   DAIIEVFHILQKIDFISSPIVCOATGFYPSNIMITWTWQDKIKEIFFED--VDVGATLTNADGTFQKLVISVKPEEVWKNNKEAIRCVVQEVCAKDDVILIVKDIRSIG---
Uga   DAIIEVFHILQKIDFISSPVVCOATGFYPSNIMITWTWQDKIKEIFFED--VDVGATLTNADGTFQKLVISVKPEEVWKNNKEAIRCVVQEVCAKNDVILIVKDIRSIE---
Uha   VSPEIVSLLQKSI--SPVVCHATGFYPSGIKISWQRNGQDHDED--VDLGEHIIAFPNADGTFQRASTIKVINSDPILQNNKI--FICVVEHCQCEIETRAIE---------
Uia   AISPQVSLLQKSSSSPVVCHATGFYPSAIKISWQRNGQDHDED--VDLGELLPNADGTFQRTSTILNVKPEEWKNNK--FSCVVEHCQEIRAA---DDKIILIN--------
Uja   VSPQVILLQKSSSSPVVCHATGFYPRGIHISWQKNGQDIYDED--VDLGELLPNADGTFQKTSTILNVKPEEWKNNK--FSCVVEHCQCKIRAA---DEITLTNE------
Uka   VSPQVEVLQKSSSSPVVCHATGFYPSGIKISWQRNGQDHDED--VEIPSEILLPNADGTFQRASIPIVIPEEIWERNKI--LRCVVEHCQCVIINSIITIAIVIIRTNN---
Ula   -DPEIVFIAIQNPISSPVCEPATGFYPSGVIITLIPNGQDEGEN--VELPEIKPIEDGTFQRSSIIINVSQEDWLKSLI--YICVVEHCRMIIQKS--GDEIKSN------
Uma   -ALKIXIFIAIQKSIPNSDVKCHVACIIFISNNIIVELIMINICQEVSNSAIIKSCIIIAIPIEDGTYIQRLVIYIRIJSLQYIMRKEQ--FICTVRYMEKIIRRNITISELKIKSLIYSEMSITA
                          *                                 a  aaa   a                                  *
```

TM/Cyt

```
                300                       330                       360
Uba   -PLIAPLGIIHGIVVIAVIAIV--AIIAVAGFVVYRFIKIC--------IFIIVPQNTSDGCGSDNSSIIT----------------------------
Uca   -PSATICIIIGIVLSAVIAII--FIIAVAGTVVVWKIKIKIC--------IFIIVPINPSDSVAEKIKSQTSVNSDNDSGCGSGGSDSVSHKSSDSTSDISSVCSDNSSDSTTALLP
Uda   -VSIVFFGIIIAVIIA-VVLLII--AIGITGTMVWKKIQIC--------IFKFVSGSDI-DGSNSLTIIAIVPQA--------------------------
Uea   -DSLIPIGLIHGVVVAIIVLLV--VIIAVAGTVYIVQIKKKIC--------IFKFVSGSDI-DGSNSSAIFAIVPQA------------------------
Ufa   GSDNTHIIIVVGCVAIAVAIIA--VI--ILACGILYIWKIISNIC--------IGRASSKITDSEQSD-PPTIQVK----------------------------
Uga   GSDNTHIIIIVCCVAIVAIIA--VI--ILACGILYIWRIISNIC--------IGKASSKITDSEQSD-PIIAIVEVK----------------------------
Uha   -DSVSIVGLIIACGVAVAVLLVI--VIIAVAGFMMYRKKKIC--------ILIIANSSII-DGSRSIAQDGPRA----------------------------
Uia   -DSIAPTGLIIAGVVAILVLLV--VIIAVAGTMVYIQKKKIC--------IFKFVSGSDI-DGSNSSAIFAIVPQA------------------------
Uja   -ESAPIIAIIAGVVAILIALIV--VIIVVAGTMVIYQIKKKIC--------IFKFVSGSDI-DGSNSSAIIAIVPQA----------------------------
Uka   -ESAPIIGLIIGIVVIAVVLLV--FIIAVAFVVYRKKKIC--------ILIBVAIN--------AVNSDKNSGCGSGSGSDSVSHKS-------------
Ula   --SSSEVFITPERVLAIIAIAALV--ICLCICIYIVLIPICLLK------P-ICQIVNQNIEMYEYI-------VKNE-------------------
Uma   -SDSSTSIIILVPILIV--------VVLIVSILIIMWMSCHQNCRLGQCIHLENIIN-----------------------------------
       ///////////////////////////////////////
```

```
        390
Uca   ENSDKNSGCDSGSDSVSHKSSDSTSDISSVCSDNSSDSTTALF
```

Fig. 3.

Zebrafish MHC class I U lineage protein sequence alignment. The zebrafish U proteins included in Figure 2 were organized into domains and aligned. The leader and α domains are each encoded by a single exon while the transmembrane (TM) and cytoplasmic (Cyt) domains are encoded by a variable and sometimes unknown number of exons. Numbering of amino acids is based on the Uba sequence starting with the first residue of the α1 domain. Positions that are at least 50% identical are shaded in black and similar residues are shaded in gray. Conserved features are indicated with the following symbols below the alignment:

cysteine residues likely involved in the Ig-fold (*), putative anchor residues implicated in binding the amino- (n) or carboxy-terminus (c) of peptides, acidic residues in the α3 domain predicted to bind CD8 (a), and the putative transmembrane domains (/). Note: the indicated transmembrane region does not accurately portray the borders of all transmembrane domains but represents the best fit based on the alignment. All of the proteins in the alignment have a transmembrane domain predicted by SMART software (Letunic et al. 2012).
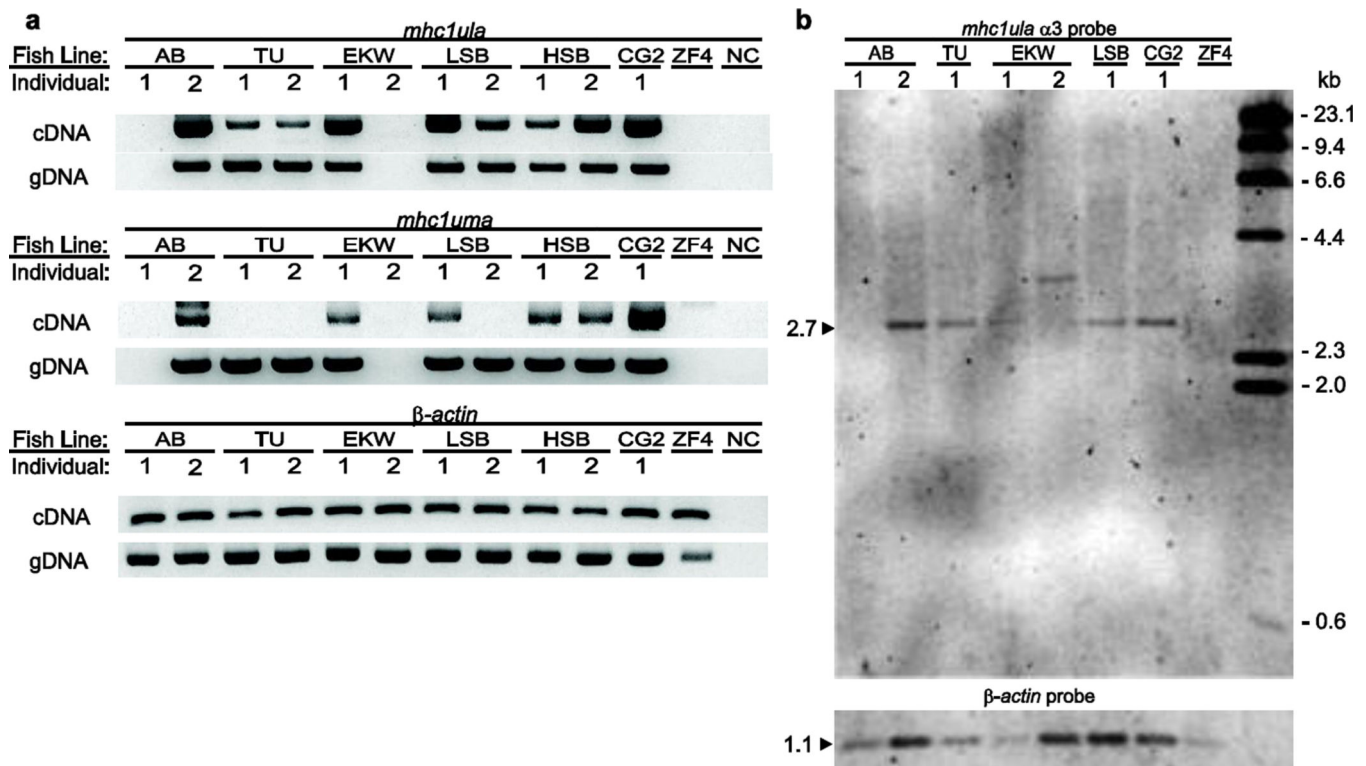
Fig. 4.

Presence or absence of *mhc1ula* and *mhc1uma* in individual zebrafish. **a** PCR was performed on a panel of matching cDNA and genomic DNA samples from individual zebrafish representing different genetic backgrounds: AB, TU, EKW, LSB, HSB, and CG2 zebrafish lines and a zebrafish fibroblast cell line, ZF4. A negative control (NC) with no template was included for each primer pair. Full-length sequences of *mhc1ula* and *mhc1uma* were amplified from cDNA while intron-spanning segments (α2-α3 and α1-α2, respectively) were amplified from genomic DNA. The same primer pair, spanning exon 5 to exon 6, was used to amplify β-*actin* from cDNA and genomic DNA. Representative bands were cloned and sequenced to validate all primer pairs. **b** Southern blot analysis was performed on *Pst* I-digested genomic DNA from seven of the individuals used in (**a**) as well as ZF4 genomic DNA with probes designed to detect the α3 domain of *mhc1ula* (2.7 kb band) or exon 2 of β-*actin* (1.1 kb band). The larger band detected by the *mhc1ula* probe in EKW2 likely represents an alternate haplotype in this individual.
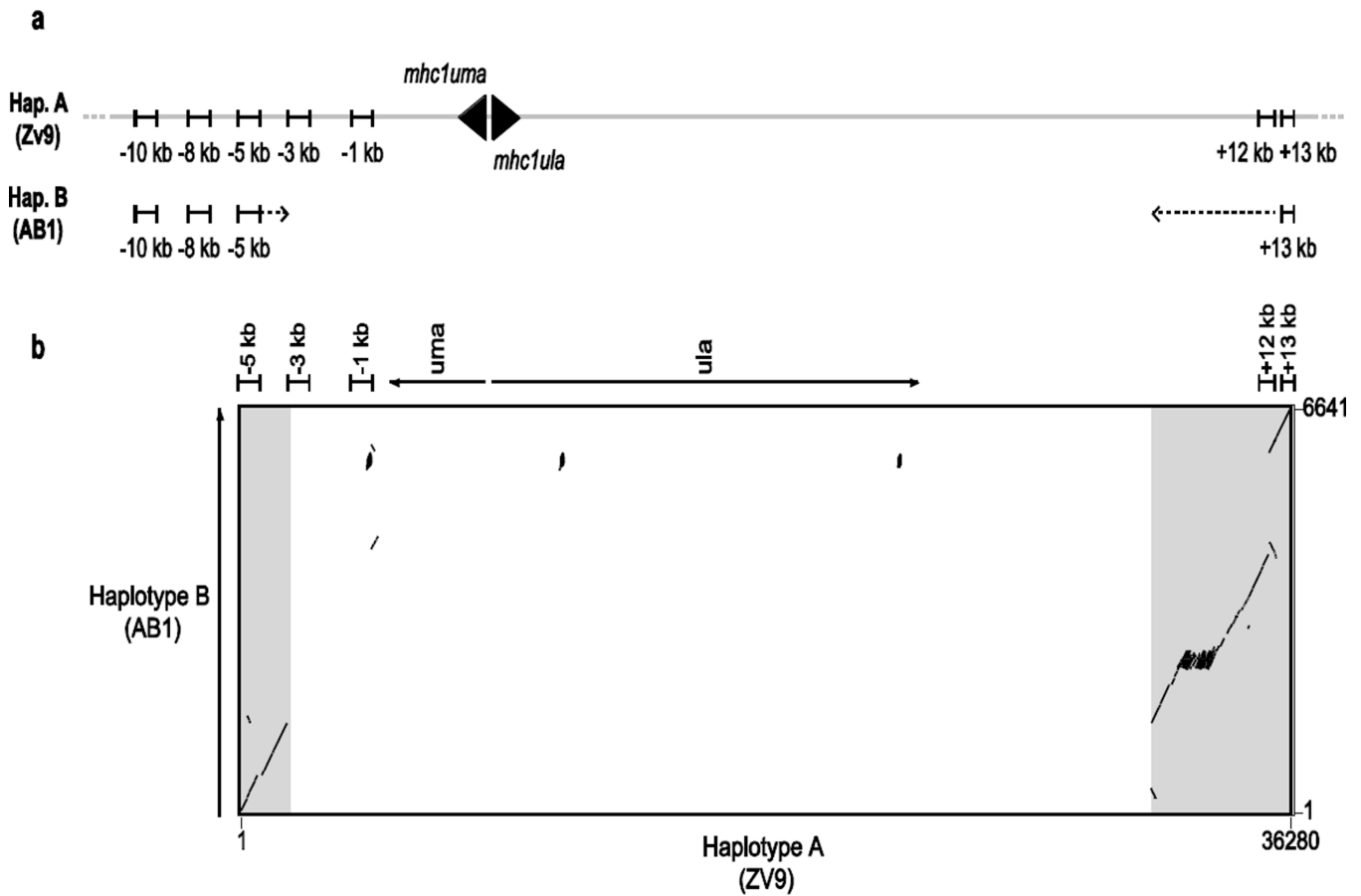
Fig. 5.
Sequencing an alternate haplotype. **a** The approximate locations of the sequences encoding the flanking genomic amplicons (FGAs) are shown relative to *mhc1ula* and *mhc1uma* on chromosome 22 (top). The −5 kb and +13 kb FGAs could be amplified from the AB1 individual encoding haplotype B (without *mhc1ula* and *mhc1uma*) while the −3 kb, −1 kb, and +12 kb FGAs could not (bottom). Chromosome walking was carried out from the AB1 individual between the −5 kb and the +13 kb FGAs. **b** The dot plot generated using the PipMaker program (Schwartz et al. 2000) shows the alignment of the resulting sequence of haplotype B on the y-axis with the sequence of haplotype A from the reference genome (Zv9) on the x-axis. The gray shading emphasizes the regions of similarity between the two haplotypes which does not include any of *mhc1ula* or *mhc1uma*.
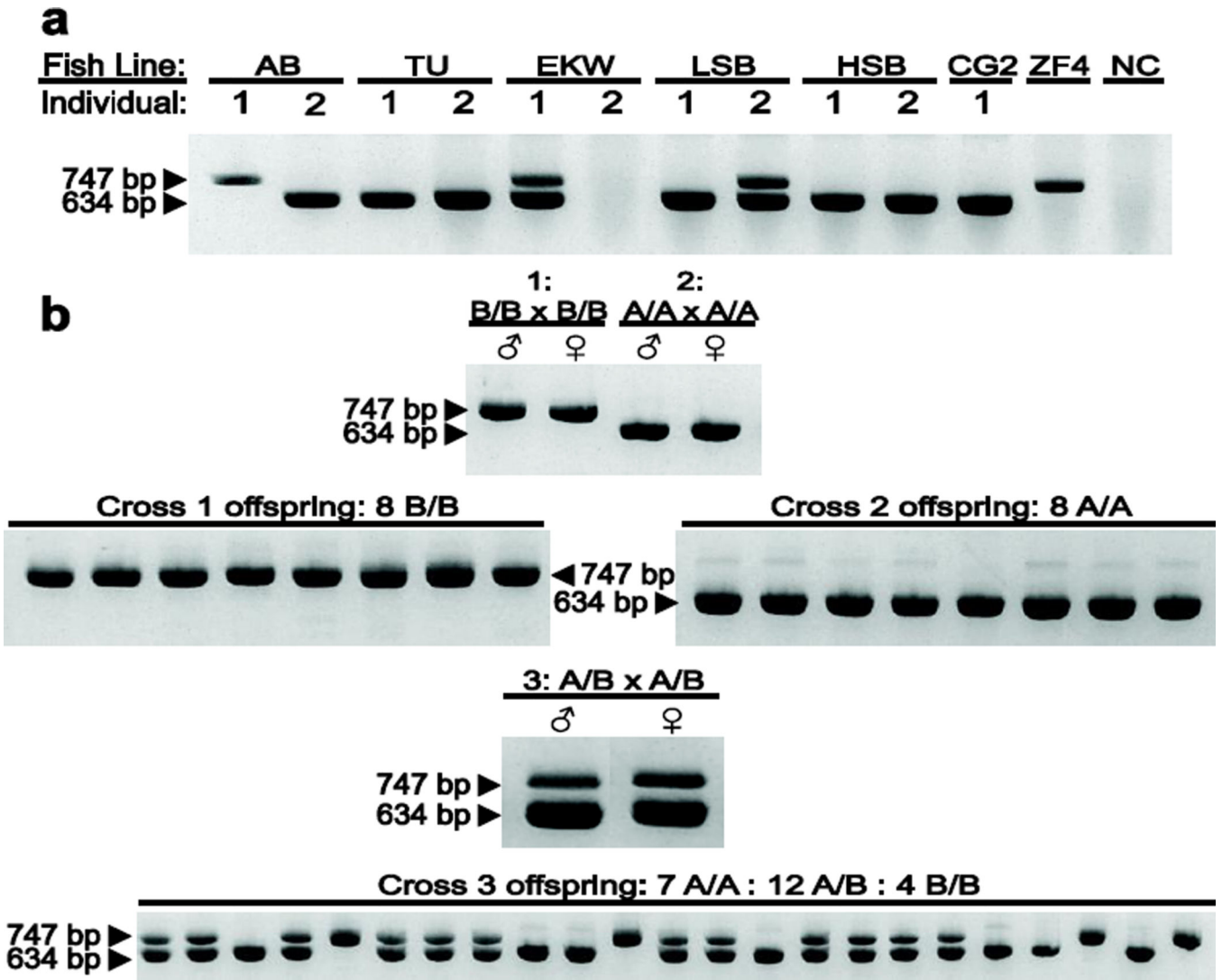
Fig. 6.
Genotyping strategy for the chromosome 22 locus. **a** A duplex PCR assay was used to genotype the same individuals evaluated in Figure 4a. The 634 bp amplicon represents haplotype A and the 747 bp amplicon represents haplotype B. The presence of both amplicons in EKW1 and LSB2 suggests that these individuals are heterozygous at this chromosome 22 locus. **b** The offspring from two zebrafish homozygous for haplotype B (cross 1) and from two zebrafish homozygous for haplotype A (cross 2) were genotyped. The offspring from two zebrafish heterozygous for haplotypes A and B (cross 3) were genotyped revealing 7 offspring homozygous for haplotype A (A/A), 12 offspring heterozygous for haplotypes A and B (A/B), and 4 offspring homozygous for haplotype B (B/B). This approximates the ratio of 1:2:1 that would be expected in the case of simple Mendelian inheritance at a single locus.
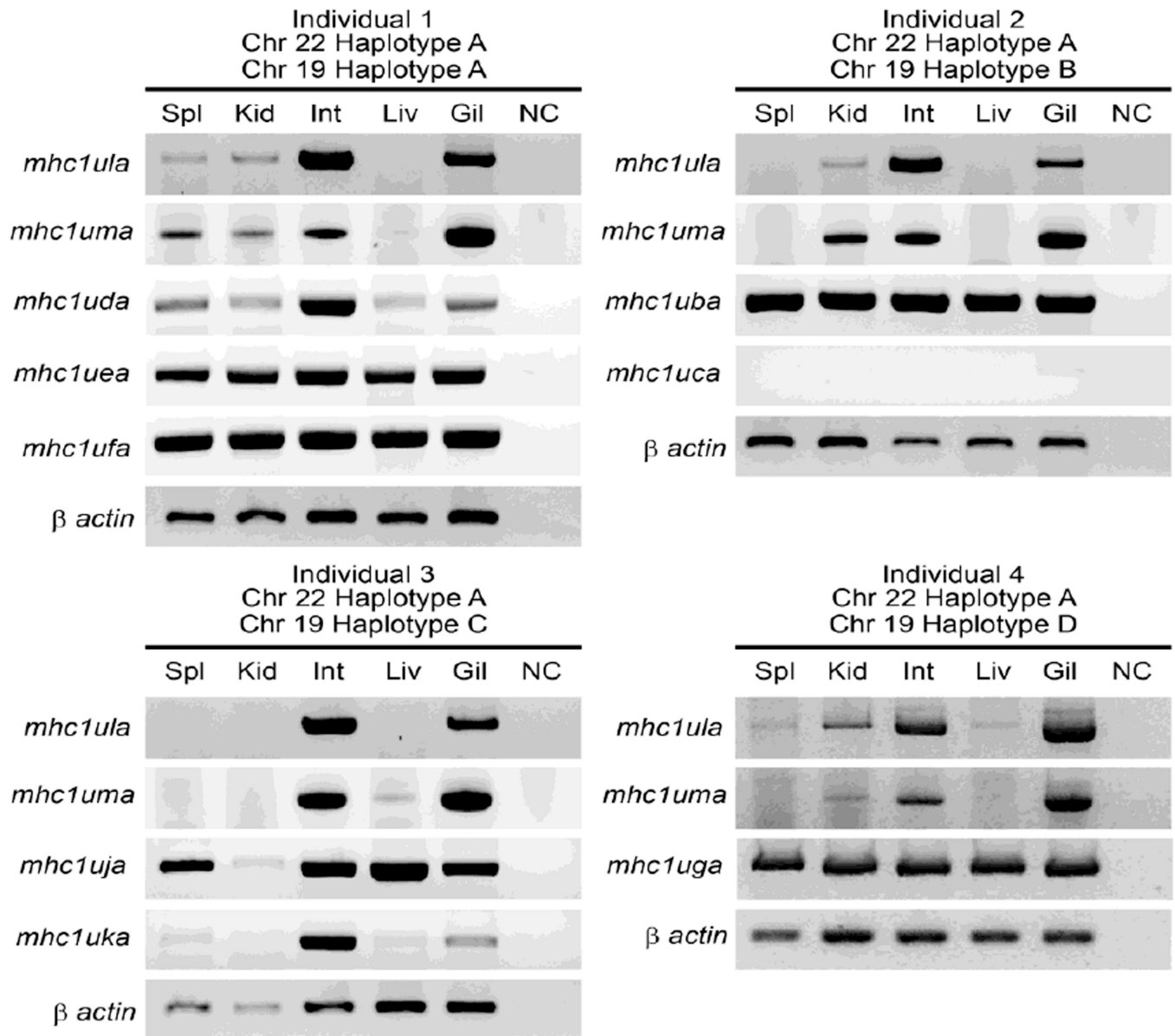
Fig. 7.
Tissue-specific expression of *mhc1ula* and *mhc1uma*. Primer pairs amplifying the α1-α3 domains of *mhc1ula* and *mhc1uma* were used to detect transcripts from cDNA derived from the tissues (Spl - spleen, Kid - kidney, Int - intestine, Liv - liver, Gil - gill) of genotyped EKW individuals. *mhc1ula* and *mhc1uma* show tissue-specific expression while most of the MHC class I genes of the U lineage encoded on Chromosome 19 and β-*actin* are ubiquitously expressed. A negative control (NC) with no cDNA template was included for each sequence.