Immunogenetics (1999) 49:479–490

ORIGINAL PAPER

Benny P. Shum · Raja Rajalingam
Katharine E. Magor · Kaoru Azumi
William H. Carr · Brian Dixon · René J. M. Stet
Mark A. Adkison · Ronald P. Hedrick
Peter Parham

A divergent non-classical class I gene conserved in salmonids

Received: 22 September 1998 / Revised: 24 November 1998

Abstract Complementary DNA for two class I genes of the rainbow trout, *Oncorhynchus mykiss*, were characterized. *MhcOnmy-UBA*01* is similar to *Onmy-UA-C32* and the classical major histocompatibility complex class I genes of other fish species, whereas *Onmy-UAA*01* is divergent from all class I genes so far characterized. *Onmy-UAA*01* is expressed at lower levels than *Onmy-UBA*01*. Although *Onmy-UAA*01* exhibits restriction fragment length polymorphism on Southern blotting, the encoded protein is highly conserved. Two allotypes, which differ only by substitution at amino acid position 223 of the α_3 domain, have been defined. *Onmy-UAA*01* has an exon-intron organization like other class I genes and contains a Tc1-like transposon element in intron III. Orthologues of *Onmy-UAA*01* have been characterized in four other species of salmonid. Between four species of *Oncorhynchus,* UAA*01 proteins differ by only 2–6 amino acids, whereas comparison of *Oncorhynchus* with *Salmo trutta* (brown trout) reveals 14–16 amino acid differences. The *Onmy-UAA*01* gene has properties indicative of a particularly divergent non-classical class I gene.

Key words Comparative immunology · Evolution · MHC class I · Molecular biology · Salmonids

B.P. Shum · R. Rajalingam · K.E. Magor · K. Azumi[1]
W.H. Carr · B. Dixon[2] · P. Parham (✉)
Departments of Structural Biology and Microbiology & Immunology, Sherman Fairchild Building, Stanford University School of Medicine, Stanford, CA 94305-5126, USA
E-mail: peropa@leland.stanford.edu,
Tel.: +1-650-7237456,
Fax: +1-650-7238464

R.J.M. Stet
Department of Experimental Animal Morphology and Cell Biology, Wageningen Agricultural University, PO Box 338, 6700 AH Wageningen, The Netherlands

M.A. Adkison · R.P. Hedrick
Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California, Davis, CA 95616, USA

Present addresses:
[1] Department of Biochemistry, Faculty of Pharmaceutical Sciences, Hokkaido University, Sapporo 060, Japan
[2] Department of Biological Sciences, University of Windsor, 401 Sunset Avenue, Windsor, Ontario, N9B 3P4, Canada

Introduction

Comparison of the major histocompatibility complex (MHC) class I genes of mammalian species shows how dramatically these gene families can change under selection from pathogens. Such comparisons have led to different kinds of class I genes being distinguished (Parham 1994). Classical class I genes encode proteins that present diverse peptide antigens to CD8 $\alpha\beta$ T cells, and also select the CD8 $\alpha\beta$ T-cell repertoire in the thymus. The protein products of classical class I genes are expressed on the surfaces of most cells of the body, where they alert the immune system to infection of individual cells. In many species the classical class I genes are highly polymorphic, as is well illustrated by the human *HLA-A*, *-B* and *-C* genes (Lawlor et al. 1990; Pamer and Cresswell 1998; Parham et al. 1995). However, for species or populations which have passed through a bottleneck this is not necessarily the situation, as seen for example in strains of laboratory mice. Non-classical class I gene is a term used to embrace all genes with a typical class I structure that do not have the function and tissue distribution of classical class I genes. Non-classical class I genes are diverse both in structure and function and can either be encoded within the MHC or elsewhere in the genome (Hedrick 1992; Klein and O'Huigin 1994). Equally wide-ranging is the extent to which non-classical class I genes are shared by different species. Common to most classical and non-classical class I genes are functions to do with the immune system.

Although much has been learned from mammals, they represent only a small minority of the species possessing immune systems and class I gene families. To

develop more general understanding of the way class I gene families adapt to different circumstances requires the study of species that represent the range of vertebrate evolution. Here we report on an investigation of class I genes of the rainbow trout, *Oncorhynchus mykiss*, a salmonid species which originated in the waters of the northern Pacific coast of North America (MacCrimmon 1971). Rainbow trout are highly favored by anglers and are increasingly farmed as a source of food for humans (Hanfman 1993; Stickney 1996). To meet these demands rainbow trout have been introduced by humans into lakes and rivers throughout the world, where they come into contact with native species of trout and their pathogens. In Europe such contact led to rainbow trout becoming infected with *Myxobolus cerebralis,* a common parasite of brown trout (*Salmo trutta*) (Hoffman 1970). While brown trout tolerate the parasite, rainbow trout suffer an incapacitating and fatal whirling disease. Inadvertently, *M. cerebralis* has been introduced into North American waters where it now causes epidemic disease of trout populations (Potera 1997). A study of MHC and other genes determining immune responsiveness may reveal genetic factors conferring resistance to whirling disease.

Materials and methods

Animals

Rainbow trout (*Oncorhynchus mykiss*) were obtained from Lintt's Trout Farm (Half Moon Bay, Calif.), Lassen Hatchery (Lassen, Calif.), and Nimbus Hatchery (Rancho Cordova, Calif.). Genomic DNA samples of Australian steelhead trout (an anadromous form of *O. mykiss*) and lungfish (*Neoceratodus forsteri*) were generous gifts from R.L. Raison (University of Technology, Sydney, Australia). A kokanee salmon (the land-locked form of *O. nerka*) was collected from Lake Tahoe, California, with the help of W. Cox of the California Department of Fish and Game. Other salmonid samples were from the collection of the Fish Pathology Laboratory, University of California at Davis; these include chinook salmon (*O. tshawytscha*, also known as king salmon), coho salmon (*O. kisutch*, also known as silver salmon), and brown trout (*Salmo trutta*). Genomic DNA sample of a common carp (*Cyprinus carpio*) was obtained from the Wageningen Agricultural University, The Netherlands. A Pacific hagfish (*Eptatretus stouti*) was purchased from Pacific Biomarine (Los Angeles, Calif.). Fish were euthanized by immersion in 0.5% 3-aminobenzoic acid ethyl ester (MS-222; Sigma, St. Louis, Mo.) for organ harvest, or were anaesthetized in 0.025% MS-222 for peripheral blood collection via caudal vein puncture.

Primers and probes

The a3-200 DNA probe was described previously (Shum et al. 1996). Specific for the *Onmy-UAA* gene, the UA-a123 probe is the amplified product in a polymerase chain reaction (PCR) with the UA-a1f1 (5′-CCC TCA TTC TCT TCA TCG TCA-3′) and UA-a3r1 (5′-TCT TGC AGC TCG TAA AAT-3′) primers from plasmid containing the *Onmy-UAA*0101* cDNA (GenBank accession L63541), and it corresponds to most of exons II-IV (α_1–α_3 domains) of the gene. The UB-a1-3ut probe, encompassing most of the mature protein coding region and a short 3' untranslated region, is derived from PCR amplification of plasmid DNA containing the *Onmy-UBA*0101* cDNA with the UB-a1f1 (5′-CAC TCC CTG AAG TAT TTC TAC AC-3′) and UB-3r1 (5′-CCA CCA CCA ACA CAA CCA AC-3′) primers. The b2m-mp probe corresponds to the mature protein coding region of the trout β_2-microglobulin from the cDNA clone Jb-1 (Shum et al. 1996; accession L47354). The trout CK-1 probe is from the coding region of a trout chemokine gene isolated recently (Dixon et al. 1998). The carp *Cyca*-12 probe corresponds to a classical class I gene of the common carp (van Erp et al. 1996). The carp β-actin probe was a kind gift from P. Hackett (Liu et al. 1990; University of Minnesota, St. Paul). DNA probes were labeled by [γ-^{32}P]-dCTP with the Prime-it II Random Primer Labeling Kit (Stratagene, La Jolla, Calif.) according to the manufacturer's instructions.

Screening of cDNA libraries

Two rainbow trout libraries were screened for MHC class I clones – a spleen cDNA library constructed in the λgt10 bacteriophage and an anterior-kidney/spleen cDNA library constructed in the ZAP-Express vector as previously described (Shum et al. 1996; Stratagene). The a3-200 DNA probe was used to screen both libraries and yielded the *Onmy-UAA* clones (Sambrook et al. 1989; Shum et al. 1996). The *Cyca*-12 probe was used to screen the anterior-kidney/spleen cDNA library by cross-hybridization under reduced stringency conditions to obtain *Onmy-UBA*0101*. Briefly, replicate Colony/PlaqueScreen nylon filters (Du Pont NEN, Boston, Mass.) containing $\sim 5 \times 10^4$ plaque forming units (pfu) of the libraries were hybridized with probes. Pre-hybridization and hybridization under high-stringency conditions were performed in a 50% formamide hybridization buffer [50% deionized formamide, 5× standard saline sodium phosphate-ethylenediaminetetraacetate (SSPE) buffer (pH 7.4), 5× Denhardt's reagent, 5% dextran sulphate, 1% sodium dodecyl sulphate (SDS), 1 mg per ml of fragmented and denatured salmon sperm DNA] at 42 °C following standard protocols. Filters were washed subsequently to a final wash condition of 0.2× SSPE/0.5% SDS at 42 °C and exposed to X-ray films. For low-stringency conditions, 30% concentration was used to replace 50% formamide and washes were performed in 2× SSPE/0.5% SDS instead. Clones with positive hybridization signal were picked for plaque purification and then subcloned into pBluescript-derived vectors (Stratagene) for sequence analysis.

Construction and screening of a trout genomic library

Approximately 1 g of frozen liver tissue was shattered under liquid nitrogen and the cells dispersed in 5 ml of 1× standard sodium citrate buffer using a Dounce homogenizer. Cells were then lysed by addition of an equal volume of TNE buffer [100 mM Tris (pH 8), 100 mM NaCl, 10 mM disodium ethylenediaminetetraacetate (EDTA), and 1% Sarkosyl] and protein digestion was carried out overnight at 55 °C in the presence of 100 μg/ml proteinase K. The DNA was extracted gently with an equal volume of phenol, followed by phenol/chloroform and then chloroform/isoamyl alcohol (24:1). The DNA was dialyzed against 4 l of TE buffer (pH 8) containing 100 mM NaCl for 24 h, then against TE buffer for 24 h. DNA prepared in this manner was determined to be approximately 100 kilobase (kb) pairs in size and nuclease free. Approximately 300 μg of DNA was digested partially with *Sau* 3A and size-fractionated by density gradient ultra-centrifugation through a 10 to 40% sucrose gradient. Fractions in the range of 18–22 kb were ligated into the *Bam* HI site of λDASH II and packaged in Gigapack XL (Stratagene). A total of 300 000 plaques were screened directly without prior amplification. Since the trout haploid genome contains 2.7×10^9 base pairs (bp) of DNA (Tiersch et al. 1989), the library represented the equivalent of two haploid genomes. Approximately 300 000 plaques were lifted onto duplicate nylon filters and screened using a mixture of two MHC class I DNA probes, UA-a123, and UB-a1-3ut. Forty-six clones were isolated in the primary screen, and the DNA from

two of these (clones 18.4 and 18.6) hybridized with the UA-a123 probe under high-stringency conditions.

Southern and northern analysis

Genomic DNA was purified from either peripheral blood cells or livers of fish, and Southern analysis was performed following standard protocols (Marcadet et al. 1989; Sambrook et al. 1989). Briefly, aliquots containing 10–30 μg of genomic DNA was digested with >10 units/μg of the restriction enzymes *Taq* I or *Hin*dIII (Boehringer Mannheim, Indianapolis, Ind.) in 200 μl reaction volume of 1× manufacturer's buffer. Electrophoresis of DNA was performed under a constant voltage of 20–40 V through 0.7% agarose in 0.5× TBE buffer (pH 8.3). The separated DNA fragments were transferred to nylon membranes (S&S maximum strength Nytran; Schleicher and Schuell, Keene, N.H.) following standard alkaline transfer protocols and then immobilized by cross-linking under ultraviolet (UV) irradiation. Membranes were pre-hybridized and then hybridized with labeled DNA probe at 42 °C in 30% or 50% formamide-containing hybridization buffer as described above. After hybridization, the blots were washed to a final stringency of 2× SSPE/0.5% SDS at 42 °C for low-stringency conditions, or 0.2× SSPE/0.5% SDS at 50 °C for high-stringency conditions. Autoradiography was performed with intensifying screens at –80 °C for three to seven days.

For northern analysis, mRNA isolated from the spleen of one rainbow trout was enriched through two rounds of oligo-dT affinity column chromatography (Pharmacia Biotech, Piscataway, N.J.) following total cellular RNA extraction with RNAsol (TelTest, Friendswood, Tex.). Approximately 20 μg of mRNA were loaded into a single wide sample well and separated by electrophoresis through a 1% agarose/2.2 M formaldehyde gel in 1× standard MOPS Running Buffer under 150 V constant voltage. The separated RNA was transferred to a nylon membranes with the TurboBlotter system (Schleicher and Schuell) following the manufacturer's protocol, and was subsequently immobilized by UV cross-linking. The membrane was cut into strips, which were then pre-hybridized, and separately hybridized with various probes in 50% formamide hybridization solution. After hybridization, the filters were washed in 0.2× SSPE/0.5% SDS at 65 °C for 30 min. The washed membrane strips were subjected to autoradiography at –80 °C for 1 or 15 h.

Reverse transcriptase polymerase chain reaction amplification and subcloning of PCR products

Total cellular RNA was separately purified with RNAsol (TelTest) following the manufacturer's instructions. First-strand cDNA was synthesized from total RNA by priming with oligonucleotide-d$(T)_{16}$ and extending with Moloney Murine Leukemia Virus Reverse Transcriptase (Gibco BRL, Grand Island, N.Y.) following standard protocols (Sambrook et al. 1989). The first-strand cDNA was then used in the defined PCR amplification of the *UAA* transcript with either the UA-a1f1 (5′-CCC TCA TTC TCT TCA TCG TCA-3′)/UA-3ur2 (5′-TCG TTG GTC ATC CTC TTA-3′) or the UA-5UT1 (5′-GAG GAC ATG TTA TTG CAA ATA TG-3′)/UA-3UT1 (5′-TAG GGT TGT ATA GAG CAT CAA ATG-3′) primer pairs under the following conditions – 28 to 32 cycles: denaturation at 94 °C 30 s, primer-annealing at 50 °C/1 min, and extension at 72 °C/1 min. The PCR products were then cloned into the pCR-Script (Stratagene) or the pCR-2.1 (Invitrogen Corporation, Carlsbad, Calif.) plasmids, and selected clones were used for sequencing analysis.

PCR amplification of the UAA gene from genomic DNA

A nested-PCR method was used to amplify the *UAA* gene from genomic DNA samples of rainbow trout, coho salmon, and chinook salmon. The first primer pair is UAA3-5p (5′-GAG ACT GGG ACA CAT AAG G-3′)/UAA2-3p (5′-ACT AGA ATT TTC CAT GAA TGT AGT-3′) and the second internal primer pair is UA-ldf3 (5′-TTG AGG AGT AAT GGG GTT GAA-3′)/UA-3ur2 (5′-TCG TTG GTC ATC CTC TTA-3′). Both PCR were performed under the following conditions – 35 cycles: denaturation at 95 °C/30 s, primer-annealing at 60 °C/30 s, and extension at 72 °C/2 min; a final 10-min incubation at 74 °C was included at the end for complete DNA extension. A five percent volume of the PCR reaction was analyzed by electrophoresis through 1.5% agarose in 0.5× TBE buffer. For reactions that showed a single prominent band of the anticipated size, the PCR products were used as templates for sequencing.

DNA sequencing and analysis

Nucleotide sequences were determined on both DNA strands by "primer-walking" with the *Taq*-FS Dye-Terminator Sequencing Kit following the manufacturer's instructions (Applied Biosystems, Foster City, Calif.) and the sequencing reactions were analyzed on the 373 A Automated DNA Sequencer (Applied Biosystems). Direct sequencing of the PCR-amplified *UAA* gene was performed following purification of the PCR reaction with the QIAquick PCR Purification Kit (Qiagen, Valencia, Calif.). Plasmid DNA templates were prepared by following the modified alkaline lysis protocol provided by Applied Biosystems, or by using the QIAprep Miniprep Kit (Qiagen). Sequence analysis was performed using the Wisconsin Sequence Analysis Software Wisconsin Package [version 9.1; Genetics Computer Group (GCG), Madison, Wis.].

Accession numbers

The following sequences can be found in the GenBank database with the accession numbers in parentheses: *MhcOnmy-UAA*0101* genomic sequence from rainbow trout J (AF091779), *MhcOnts-UAA*0101* genomic sequence from chinook salmon 2 (AF091780), *MhcOnki-UAA*0101* genomic sequence from coho salmon 1 (AF091781), *Onmy-UAA*0101* cDNA variant (AF091784), *Onki-UAA*0101* cDNA (AF091782), *Onki-UAA*0101* cDNA variant (AF091783), *MhcOnne-UAA*0101* cDNA variant (AF091786), *MhcSatr-UAA*0101* cDNA (AF091787), and *Onmy-UBA*0101* cDNA (AF091785).

Phylogenetic analysis

Amino acid sequences of the α_3 domain were aligned with the Pileup program of the Wisconsin Sequence Analysis Software Package (GCG; version 9.1) and by visual inspection afterwards. Phylogenetic analysis was performed using the PAUP* program package (4.0 beta version; Sinauer Associates, Inc., Sutherland, Mass.). Genetic distances were calculated by mean-character differences and a phylogenetic tree was generated by the neighbor-joining method (Saitou and Nei 1987). The confidence of individual nodes were tested over 1000 bootstrap replications and only groups with a frequency over 60% were retained. A strict consensus tree was shown allowing polytomy.

Results

*MhcOnmy-UAA*01 and MhcOnmy-UBA*01 define two MHC class I genes of rainbow trout*

Complementary DNA corresponding to two distinct MHC class I genes of the rainbow trout, *Oncorhynchus mykiss*, were isolated using different strategies. The

first strategy used degenerate oligonucleotide primers based upon the regions encompassing invariant cysteines of the α_3 domain (Hashimoto et al. 1990). Reverse transcriptase (RT)-PCR amplification with these primers produced a 190 bp product, which when used to screen cDNA libraries yielded full-length cDNA clones of a class I gene. This gene has been named *MhcOnmy-UAA*01* (Shum et al. 1996) according to convention (Klein et al. 1990). The second strategy used the *Cyca*-12 cDNA probe, derived from a polymorphic classical class I gene of the common carp (van Erp et al. 1996), to screen a rainbow trout cDNA library. A full-length class I cDNA was isolated and sequenced. This cDNA encodes a protein having only ~30% amino acid sequence similarity in the extracellular domains with *Onmy*-UAA*01 (Fig. 1). This degree of divergence is unprecedented for alleles of an MHC class I locus and we concluded that the two cDNA correspond to different loci. The second gene was therefore named *MhcOnmy-UBA*01*. The *Onmy-UBA*0101* sequence is more closely related to the *Onmy-UA-C32* sequence as described by Hansen and co-workers (1996), particularly in the leader and the α_1 domain (Fig. 1). *Onmy-UBA*01* and *Onmy-UA-C32* may represent divergent alleles of the same locus or two independent class I genes.

Fig. 1 Comparison of the predicted amino acid sequences of *Onmy*-UAA*0101 with UBA*0101 and UA-C32. Numbering is based on *Onmy*-UAA*0101 with the beginning of the mature protein at position 1. Identity with *Onmy*-UAA*0101 is shown by a *dash*. The stop codon is denoted by an *asterisk*. Putative N-linked glycosylation sites are *underlined*. Nine conserved positions important for peptide binding by mammalian classical class I molecules are denoted by *Δ*. The *Onmy*-UAA*0101 sequence is described by Shum and co-workers (1996) and the *Onmy*-UA-C32 is described by Hansen and co-workers (1996)

*Onmy-UBA*01 is a classical class I gene and Onmy-UAA*01 a divergent class I gene*

The *Onmy-UBA*01* cDNA was isolated using a probe derived from a classical class I gene of carp and encodes a protein with characteristics of a classical MHC class I heavy chain. This can best be illustrated by a dendrogram constructed from the α_3 domains of various class I sequences (Fig. 2). The α_3 domain of *Onmy*-UBA*01 clusters with those of other classical MHC class I genes of salmonid fish. This includes the rainbow trout gene *Onmy-UA-C32* described by Hansen and coworkers (1996), which defines a different classical class I allele or gene from *Onmy-UBA*01*. By contrast, *Onmy*-UAA*01 is highly divergent from the α_3 domains of other MHC class I genes and forms its own separate lineage. These different affinities are also seen in a pairwise comparison of the nucleotide sequences for class I heavy-chain coding regions from other salmonid species: for example, *Onmy-UBA*0101* has ~91% nucleotide identity with the *Sasa-A1*01* from Atlantic salmon (Grimholt et al. 1993), whereas *Onmy-UAA*01* has only ~46% nucleotide sequence identity with *Sasa-A1*01*. In conclusion, the *Onmy-UAA*01* gene is very different from all other known classical and non-classical class I genes.

Further appreciation of the differences between *Onmy*-UAA*01 and classical class I heavy chains, including *Onmy*-UBA*01, came from an examination of the amino acid residues at a set of nine positions, comprising four in the α_1 domain and five in the α_2 domain, which are conserved in the peptide-binding sites of HLA-A, -B, and -C molecules (Kaufman et al. 1994; Madden 1995). This set of residues defines a motif "YYYYYYTKW" which is identical in 312 of the 364 classical HLA class I allotypes for which sequences have been determined (Fig. 3). In the 52 allotypes which do not conserve the motif, 48 have only one residue different from the motif and four have two residues

```
leader                                                                                    Amino-acid identity
                                                                                           with Onmy-UAA*0101
          .-15
Onmy-UAA*0101  MITTILISFMQFSI
Onmy-UBA*0101  -KGI--LVLGIGLLHTASA
Onmy-UA-C32    -KGI--LVLGIGLLHTASA

α1 domain
          .1        .10       .20       .30       .40       .50       .60       .70       .80
Onmy-UAA*0101  APHSL HRHCIATQGTLYPKNIQLVMIDDVIVYYYNSSAEQEAVVPEWLNHPEGI  EFWQEVHRNLKFNRYVMDTAVRVTSEHYNHSH
Onmy-UBA*0101  -T---KYFYTASSEVPNF-EFVVVA-V-GAQMVH-D-NSQRAVPKQD-M-RAAETLPQY-ESQTGIF-GAQQTFKANIDIVKQRF-Q-G                (18.1%)
Onmy-UA-C32    VT---KYFYTASSEVPNF-EFVVVG-V-G-QMVH-D-NSQRMVPRQD-M-KAAETLPQY-ESQTGI--GTQQTYKSSIDIVKQRF-Q-G                (19.3%)
                     Δ                                                  Δ                   Δ
α2 domain
             .90       .100      .110      .120      .130      .140      .150      .160      .170
Onmy-UAA*0101   DHFYQAHGRCGWKSD GTTEAFMSHAYDGKDFVSFDVSTRTWTAAVSHAVFYKRKRETDLEDLVRLVIHYESGCIRWLEKLLEFSVTVREPK
Onmy-UBA*0101  GV-IV-KMYG-E-DDEA-V--G-YQYG---E--IAL-LK-TK-I-PTPQ--IT-L-WDSNTANTEYWNNYLTQT--E--K-YVDYGKSTLMRT          (28.6%)
Onmy-UA-C32    GV-NR-LMYG-ELGDD -I-RGDFQYG---A--L-L-K--L-----NQKS-IT-L-LDATGAEANFQKNYL-NT--E--K-YVNYGKDTL-R-          (33.0%)
                                          Δ                  Δ  ΔΔ           Δ          Δ
α3 domain
                .180      .190      .200      .210      .220      .230      .240      .250      .260      .270
Onmy-UAA*0101  V PAVSLFERPPHGNSEVEVTCHVT GFYPRAVQVEWLGAEGLPMVDGVNSGEVLPNGDGSYQLRKSLTVPQEAQDTQSYSCLVLHSSIAGNITVTW
Onmy-UBA*0101  -P-S---LQKA-SS    P----A- -----D-M-S- QKD-QDHHED-EY--T---D--TF-KSSH---TP-DRKNSK-Q-V-QVKG-KKDFIEVLT          (41.1%)
Onmy-UA-C32    -R-S---IQKT-SS    P----ANQ----SG-M-F- QKD-QEQHGD-EH--I-Q-D--TF-KSTH---TP-EWKNNK-Q-V-QLAG-EDD--KVLIESEIQTN  (40.0%)

transmembrane domain                                    cytoplasmic domain
                    .280      .290      .300              .310
Onmy-UAA*0101       APKKNLANVLMAIV  IIVSVVLILTVLFKYLVW     RRAVAQNPRG*
Onmy-UBA*0101       D-DAANVVP   -IGGVVALLLV-VA-VVGVVI-     KKRSKKGFVPASTSDTDSENSGKAAPQI*
Onmy-UA-C32    FGKTNRGS-DP-TIGL-IGGV-A LLV-IA VVGVVI-     KKKNKKGFVPASTSDTDSENSGKGIQKI*
```

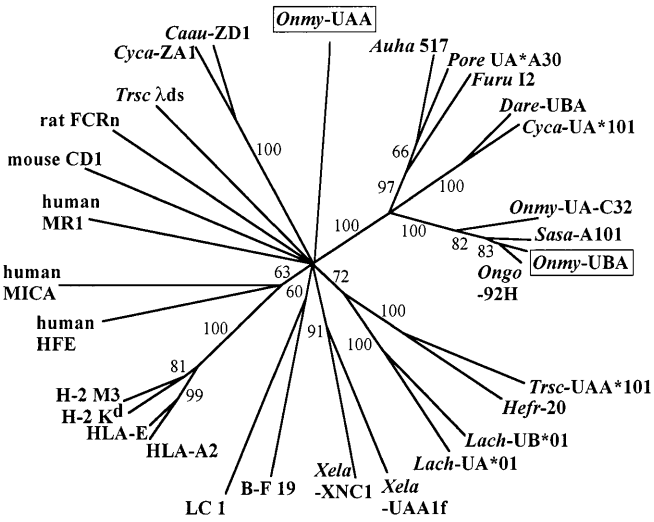


Fig. 2 A dendrogram of classical and non-classical class I. Amino acid sequences of the α_3 domain were used to construct a dendrogram by the neighbor-joining method. Representative classical and non-classical class I from all five vertebrate classes were included. The reliability of individual nodes were tested over 1000 bootstrap replicates and only groups with a frequency over 60% were retained. Shown is a strict consensus tree allowing polytomy. Percent bootstrap value is labeled *next to* the corresponding branch. The accession numbers of sequences used are as follows-human: HLA-A*0201 (K02883), HLA-E (M20022), MICA (A55739), MR1 (U22963), HFE (U60319); mouse: H-2 K^d (J00402), H-2 M3 (U18797), CD1.1 (M63695); rat: FCRn (X14323); chicken: B-F 19 (M84766); lizard: LC1 (M81094); frog: *Xela*-UAA1^f (L20733), *Xela*-XNC 1.1 (M58019); leopard shark: *Trsc*-UAA*101 (AF034316), *Trsc*-λDS-1 (M85291); horned shark: *Hefr*-20 (AF028559); coelacanth: *Lach*-UA*01 (U08043), *Lach*-UB*01 (U08034); common carp: *Cyca*-UA1*01 (X91015), *Cyca*-ZA1 (M37107); goldfish: *Caau*-ZD1 (M86747); zebrafish: *Dare*-UBA*01 (Z46777); cichild: *Auha*-517 (Sato el al. 1997); guppy: *Pore*-UA*A30 (Z54085); pufferfish: *Furu*-I2 (AF001216); Atlantic salmon: *Sasa*-A1*01 (L07606); pink salmon: *Ongo*-92H (D58386); rainbow trout: *Onmy*-UA-C32 (U55380), and *Onmy*-UAA*01 (L63541)

that are different. Moreover, the substitutions are of a conservative kind. Similarly, 17 of 23 known classical mouse class I allotypes have the motif, while the remaining six have one residue that is different. At these nine positions *Onmy*-UBA*01 differs from the human motif by two residues, and the same differences are seen in classical class I genes from other species of non-mammalian vertebrates. This similarity is consistent with *Onmy-UBA*01* being a classical class I gene (Fig. 3).

When *Onmy*-UAA*01 is examined in this way, eight of the nine residues are found to differ from the human motif and they include several non-conservative substitutions. Such divergence from the motif is characteristic of the non-classical class I genes of other species. For example, human MICA and MR1 also have seven differences from the consensus, while mouse CD1.1 has nine (Fig. 3). This analysis indicates that the binding groove of *Onmy*-UAA*01 is different from those of classical class I molecules and therefore likely to have a function different from that of trout classical class I molecules, such as *Onmy*-UBA*01 and *Onmy*-UA-C32. Further emphasizing the differences between the two genes is the much lower level of expression of *Onmy-UAA*01*, as compared with *Onmy-UBA*01*, when assessed by northern blotting analysis of mRNA isolated from spleen (Fig. 4A, B).

*The Onmy-UAA*01 gene contains a transposon-like element in intron III*

To determine the structure of the *Onmy-UAA*01* gene, a library made from genomic DNA obtained from a single rainbow trout was screened with the *Onmy-UAA*01* cDNA probe. Two hybridizing clones were isolated and subcloned into the pBluescript II plasmid vector for sequence analysis. A *Bam* HI fragment of ~12 kb was shown to contain the entire *Onmy-UAA* gene within a region of 4502 bp which was sequenced on both DNA strands. Identical nucleotide sequences were obtained from the two genomic clones and the coding-region sequences corresponded precisely to that of the *Onmy-UAA*0101* cDNA which had been isolated from the same trout used to make the genomic library, trout J.

The *Onmy-UAA*01* gene consists of six exons and five introns. The leader peptide and the three extracellular domains (α_1, α_2, and α_3) are encoded by separate exons as is typical of class I genes. Exon I specifies the 5' untranslated region and the signal peptide. Exons II, III, and IV encode the α_1, α_2, and α_3 domains, respectively. Exon V encodes the transmembrane region and part of the cytoplasmic tail, while exon VI specifies the remainder of the cytoplasmic tail and the 3' untranslated region including a polyadenylation signal. *Onmy-UAA*01* is one of the more compact class I genes due to the modest size of its introns (Fig. 5A).

Intron III of the *Onmy-UAA*01* gene contains a Tc1 transposon-like element similar to the inactive Tc1 transposons that occur throughout the genomes of Atlantic salmon and zebrafish (Goodier and Davidson 1994). Although >2% of the salmonid genome is believed to consist of transposons, the *Onmy-UAA*01* gene provides the first example of an expressed gene which contains a transposon. Following convention we refer to this sequence element as SALT1-*Onmy*1, signifying Salmonid Transposon of *Oncorhynchus mykiss* (Goodier and Davidson 1994). SALT1-*Onmy*1 is predicted to be inactive because in comparison with the sequence of active transposons it has a deletion of ~800 nucleotides, including ones that encode the amino (N)-terminal part of the transposase. The SALT1-*Onmy*1 sequence is flanked by two 39 bp inverted terminal repeats (ITR) as is typical for integrated Tc1 elements (Fig. 5B). These contain TA and CAGTGC sequence motifs which are conserved in the ITR of Tc1 family members. The TA dinucleotide in the beginning of the ITR is the target site for Tc1 transposons (Ro-

			Peptide N-terminus				Peptide C-terminus					
*Numbering in reference to HLA-A*02011*			7	59	159	171	84	123	143	146	147	Diff.
	HLA-A,-B,-C											
Human	(consensus)	n= 312	Y	Y	Y	Y	Y	Y	T	K	W	0
Human	(variant motif 1)	n= 30	-	-	-	H	-	-	-	-	-	1
Human	(variant motif 2)	n= 1	-	H	-	-	-	-	-	-	-	1
Human	(variant motif 3)	n= 9	-	-	-	-	-	-	-	-	L	1
Human	(variant motif 4)	n= 1	H	-	-	H	-	-	-	-	-	2
Human	(variant motif 5)	n= 11	-	-	-	-	-	-	S	-	L	2
	H-2 K, D, L											
Mouse	(consensus)	n= 17	-	-	-	-	-	-	-	-	-	0
Mouse	(variant motif 1)	n= 6	-	-	-	H	-	-	-	-	-	1
Human	Zn-α2		-	D	-	-	-	-	-	-	-	1
Chicken	B-F 19		-	-	-	-	R	F	-	-	-	2
Lizard	LC-1		-	-	-	-	R	F	-	-	-	2
Frog	*Xela* -UAA1f		-	-	-	-	R	F	-	-	-	2
Rainbow Trout	*Onmy* -UBA*0101		-	-	-	-	R	F	-	-	-	2
Pink Salmon	*Ongo* -92H		-	-	-	-	R	F	-	-	-	2
Atlantic Salmon	*Sasa* -A1*01		-	-	-	-	R	F	-	-	-	2
Common Carp	*Cyca* -UA1*01		-	-	-	-	R	F	-	-	-	2
Zebrafish	*Dare* -UAA*01		-	-	-	-	R	F	-	-	-	2
Cichlid fish	*Auha* -517		-	-	-	-	R	F	-	-	-	2
Guppy	*Pore* -UA-A30		-	-	-	-	R	F	-	-	-	2
Pufferfish	*Furu* -I2		-	-	-	-	R	F	-	-	-	2
Leopard Shark	*Trsc* -UAA*101		-	-	-	-	R	L	-	-	-	2
Rainbow Trout	*Onmy* -UA-C32		-	-	-	-	R	F	-	-	L	3
Human	HLA-E		-	-	-	-	-	-	S	-	S	2
Human	HLA-G		-	-	-	-	-	-	S	-	C	2
Human	HLA-F		-	-	-	-	R	-	-	F	Y	3
Mouse	H-2 M3		-	-	-	F	-	-	-	R	L	3
Human	HFE		-	M	-	L	N	H	-	E	-	5
Rat	FCRn		-	-	F	H	Q	F	V	L	-	6
Frog	*Xela* -XNC1.1		F	-	F	-	V	F	V	L	-	6
Common Carp	*Cyca* -ZA1		-	D	F	F	C	C	-	-	R	6
Human	MICA		-	T	A	-	H	F	V	F	L	7
Human	MR1		-	H	W	F	H	F	I	A	-	7
Rainbow Trout	*Onmy* -UAA*0101		H	F	H	L	H	F	Y	-	R	8
Coelacanth	*Lach* -UB*01		E	S	D	-	E	F	R	V	C	8
Mouse	CD1.1		F	L	L	L	M	V	N	D	Q	9

Fig. 3 Comparison of residues which anchor the peptide-terminal main-chain atoms and are highly conserved in HLA-A, -B, and -C allotypes. Nine conserved residues important for peptide binding in human classical class I molecules are shown along with the residues occupying similar positions in other class I family genes with defined or unknown functions. Identity with the HLA-A, -B, -C consensus motif is denoted by a *dash*. Number of differences from the consensus motif is shown on the *right*. HLA-A, -B, and -C sequences are from Mason and Parham (1998) and the H-2 K, D, and L sequences are from Pullen and co-workers (1992). The accession numbers of other sequences are found in the Fig. 2 legend, with the following exceptions: HLA-F (X17093), HLA-G (X17273), the human zinc-α_2-glycoprotein (D90427), and the zebrafish *Dare*-UAA*01 (Z46776)

senzweig et al. 1983). Another pair of 25 bp imperfect internal repeats (IIR) flank the original open reading frame of the putative transposase and they display partial similarity to SALT1-*Ssal*1 (Fig. 5B).

The 5' untranslated region of *Onmy-UAA*01* contains four tandem imperfect repeats of a 37 bp element (TTG TGG AAC AGG GAG AGT AAG GGA AGA CGT ATA AGA G), while the 3' untranslated region contains three tandem repeats of a 31 bp element (AAA AAA ACA CAT CAT TTA TTT AAA TAC AAT T). Repeated sequence elements are common in salmonid genomes and have also been observed in the T-cell receptor (De Guerra and Charlemagne 1997) and β_2-microglobulin (b_2m) genes of rainbow trout (K.E. Magor, unpublished data).

*The UAA *01 protein exhibits modest polymorphism in rainbow trout*

Southern blotting analysis revealed restriction fragment length polymorphism (RFLP) for both the *Onmy-UAA*01* and *-UBA*01* genes within small sample populations of rainbow trout. In this analysis rainbow trout from two hatcheries in California were examined as well as steelhead trout, the anadromous form of rain-

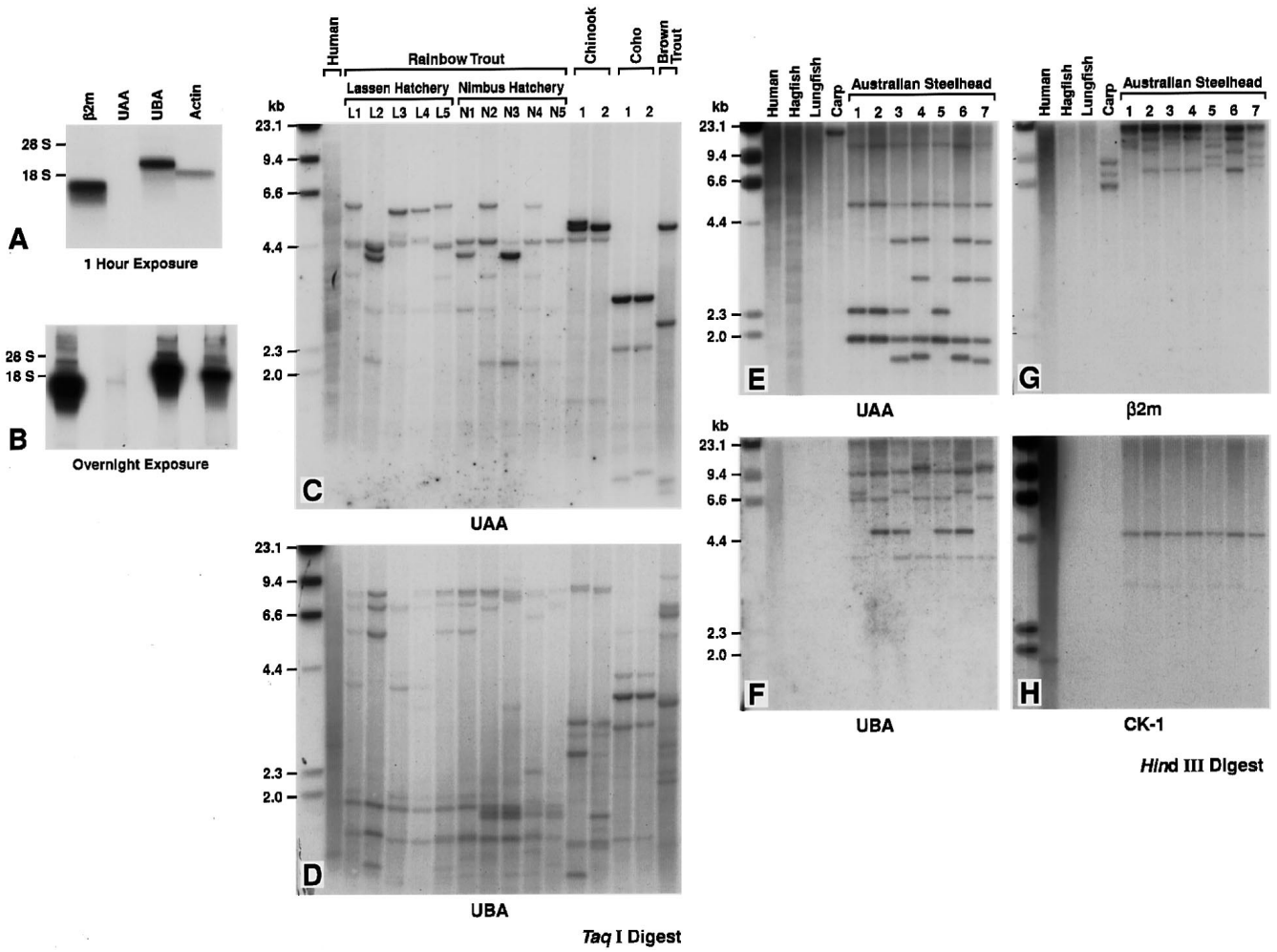
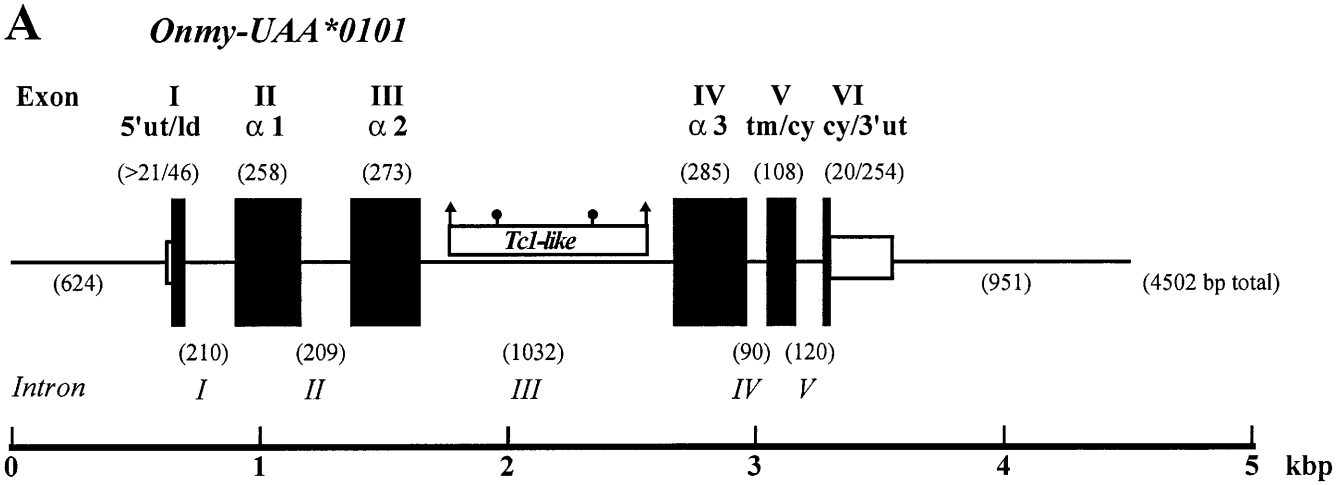


Fig. 4A–H Northern and Southern analysis of the *Onmy-UAA* and *-UBA* genes. The labels for DNA probes are as follow: UAA, UA-a123 probe; UBA, UB-a1-3ut probe; *β*2m, b2m-mp probe; Actin, carp *β*-actin probe; and CK-1, trout chemokine probe. The positions of the 18S and 28S ribosomal RNA band are noted. The sizes of the *Hin* dIII-digested lambda DNA fragments are included as markers. **A** Northern analysis of the transcription of *Onmy-UAA* and *-UBA* genes in the spleen of one fish. The membranes were exposed to film for 1 h. **B** The same membranes as in panel A were exposed to film overnight to detect the lower level of transcription of *Onmy-UAA*. **C, D** Southern analysis of genomic DNA digested with the restriction enzyme *Taq* I. Each lane represents sample obtained from one fish. The membrane was first hybridized with the UAA probe and subjected to autoradiography. The blot was then stripped and re-hybridized with the UBA probe. **E–H** Southern analysis of genomic DNA digested with the restriction enzyme *Hin*dIII. The same membrane was used in all four panels with successive stripping and re-hybridization

bow trout, from an Australian source (Fig. 4C–F). For comparison we used a probe for trout b_2m, which is known to be a polymorphic gene, and a probe for *CK-1*, a non-polymorphic trout gene encoding a chemokine (Fig. 4G–H). Multiple bands for both *Onmy-UAA*01* and *Onmy-UBA*01* were observed, indicating polymorphism of these genes.

To investigate further the polymorphism of *Onmy-UAA*01*, we isolated the gene from individual rainbow trout and determined their nucleotide sequences. Nested primers were used to PCR-amplify from genomic DNA a ~3 kb fragment containing the *Onmy-UAA*01* gene. The amplification products were used directly as templates for sequence analysis. To test the validity of the approach, we first analyzed genomic DNA from trout J, the fish from which the genomic library had been made. The nucleotide sequence obtained from genomic DNA of trout J was identical to that obtained from the individual clones isolated from the genomic DNA library. Subsequent analysis of 16 trout including nine analyzed by Southern blotting (Fig. 4C, D) revealed no *Onmy-UAA*01* sequence identical to that of fish J. A total of 14 different *Onmy-UAA*01* sequences were obtained from the 17 trout analyzed: three of the sequences being obtained from two individuals. However, the nucleotide sequences are all very similar to that obtained from trout J, with none differing by more than 13 nucleotide substitutions (Fig. 6). Of the 19 nucleotide positions exhibiting polymorphism, only one is in the coding region, a substitution that changes glycine to arginine at position 223 in the α_3 domain. Thus, 15 of the trout analyzed were



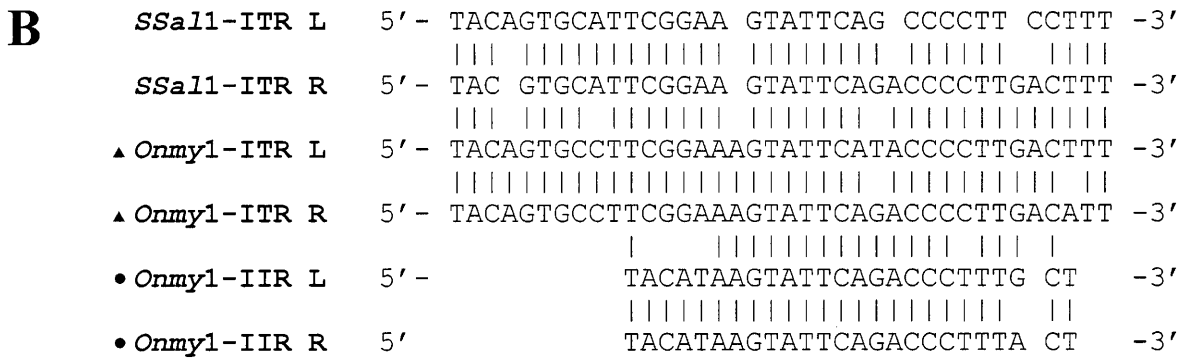


Fig. 5A, B The genomic structure of *Onmy-UAA*01*. **A** The exon-intron organization of *Onmy-UAA*0101* is depicted within the region of 4502 bp fully sequenced in both orientations. In *parentheses* are the numbers of base pairs in the corresponding region. The 5' untranslated region (5'ut) refers to the 21 bp upstream of the initiating ATG in the cDNA clone, although the actual region could be longer. Other labels are as follow: *ld* leader or signal peptide, *tm* transmembrane region, *cy* cytoplasmic tail, "3'ut" 3' untranslated region, and "▲●" represent the inverted repeat sequences in the Tc1-like transposon element. **B** Comparison of the inverted repeat sequences in the transposons of Atlantic salmon (*SSal*) and rainbow trout (*Onmy*)

shown to have the *Onmy-UAA*0101* allele which encodes glycine at position 223, while two of the trout have the second allele, named *Onmy-UAA*0102*, which has arginine at position 223. In conclusion, the coding region of the *Onmy-UAA*01* gene exhibits oligomorphism and the RFLPs seen on Southern analysis are largely due to substitutions within introns or non-coding regions that flank the gene.

*The UAA *01 gene and protein are conserved in salmonid species*

Genomic Southern analysis with the *Onmy-UAA*01* probe detected bands in all the salmonid species tested, but not in more distantly related species of fish (Fig. 4C, E). Using the PCR-amplification method that had been successfully applied to genomic DNA from rainbow trout, genes were isolated from two chinook, also called king, salmon (*Oncorhynchus tshawytscha*) and two coho, also called silver, salmon (*Oncorhynchus kisutch*), and their sequences determined. The exon-intron organization of the *UAA*01* genes in these species were found to be the same as that of *Onmy-UAA*01*, and their sequences are also similar (Fig. 8). Each fish gave a unique *UAA*01* sequence. The sequences of the chinook salmon genes, *Onts-UAA*01*, are very closely related to those of the *Onmy-UAA*01* alleles. Features that distinguish the chinook salmon sequences from those obtained from rainbow trout are two substitutions in exon I, two substitutions in intron III and three deletions in intron III. In addition the two chinook salmon sequences are distinguished by a single synonymous substitution in exon IV. By contrast the *Onki-UAA*01* sequences obtained from coho salmon are relatively divergent from those obtained from rainbow trout and chinook salmon. The differences consist of some 70 substitutions and various insertions and deletions. In particular, the differences are concentrated in intron III, exon VI and the 3' untranslated region. The Tc1-like transposon element, including the ~800 bp deletion, is also found in the introns III of *Onki-* and *Onts-UAA*01* with additional species-specific sequence

	Rainbow Trout	Position	-29	104	141	601	1007	1014	1173	1198	1204	1287	1428	2163	2381	2393	2394	2399 +1	2594	2681	2684	2689
		Region	5′	*i* -I		*i* -II	*i* -III							*e* -IV	*i* -IV				*i*- V	3′		
*Onmy-UAA*0101*	J		A	A	G	A	A	C	C	A	T	A	A	G (Gly)	T	T	C		A	T	T	A
	N3		-	-	-	-	-	-	-	-	-	-	-	- (-)	-	-	-		-	-	G	C
	L3		T	-	-	-	-	-	-	-	G	-	-	- (-)	-	-	-		-	?	?	?
	K		-	-	-	-	-	-	-	G	G	T	-	- (-)	-	-	-		-	-	-	-
	A, B		T	-	-	-	-	-	-	-	G	-	-	- (-)	-	-	-		-	-	G	C
	L2		T	-	-	-	-	-	-	G	G	-	-	- (-)	-	-	-		-	-	G	C
	N2		T	-	T	-	-	-	-	-	G	-	C	- (-)	-	-	-		-	-	-	C
	L4, D		T	-	-	-	-	-	A	-	G	-	-	- (-)	C	-	A	T*	-	-	-	-
	L		T	-	-	-	-	-	-	G	G	T	C	- (-)	-	-	-		-	-	G	C
	M		T	-	T	-	-	-	-	G	G	-	C	- (-)	-	-	-		-	-	G	C
	O		T	-	T	C	-	T	-	-	G	-	-	- (-)	-	-	-		-	-	G	C
	L5, N4		T	-	T	-	-	-	-	G	G	T	C	- (-)	-	-	-		-	-	G	C
202	N5		T	T	T	-	C	T	-	-	G	-	-	A (Arg)	-	C	-		C	-	G	C
	N1		T	T	T	C	C	T	-	-	G	-	-	A (Arg)	-	C	-		C	C	G	C

Fig. 6 *Onmy-UAA*01* gene polymorphism. Listed are positions of polymorphism in a comparison of *Onmy-UAA*01* gene sequences isolated from 17 rainbow trout. Comparisons are made with the *Onmy-UAA*0101* sequence of trout J. Nucleotide identity is denoted by a *dash*, insertion by an *asterisk*, and unknown sequence by a *question mark*. Position 1 is the first nucleotide of the initiation codon. A nucleotide insertion in the sequences from trout L4 and D is numbered 2399+1. Labels for gene regions are as follows: *e* is for exon, *i* is for intron, *5'* is for the region upstream of the initiation codon, and *3'* is for the region downstream of the stop codon. Rainbow trout with a single alphabet designation were collected from Lintt's Trout Farm, those with a numeric designation preceded by an *L* came from the Lassen hatchery, and those with a numeric designation preceded by an *N* came from the Nimbus hatchery

variations. The amino acid sequence of the UAA*01 heavy chain in chinook salmon differs from that of rainbow trout by just two amino acid substitutions, one in the leader peptide and a conservative substitution of valine for isoleucine at position 268 in the α_3 domain (Fig. 7B). Despite the extent of the difference in nucleotide sequences, the UAA*01 proteins of coho salmon only differ from those of rainbow trout by two substitutions: one at residue 43 in the α_1 domain, the other at residue 313 in the cytoplasmic domain caused by a frame-shift deletion in exon VI which changes the frame of translation (Fig. 7B).

To assess expression of the *UAA*01* genes in coho salmon we performed RT-PCR analysis on total RNA with *Onmy-UAA*-specific primers. Because of the small size of the fish available, the kidney and spleen tissues of five coho salmon fry were combined for total RNA extraction. Amplification product was cloned into a plasmid vector and two clones were sequenced completely. Demonstrating transcription of the *UAA* gene in coho salmon, one cDNA sequence was found to correspond to *Onki-UAA*0101* (from coho 1) as shown in Figs. 7 and 8. The other cDNA also corresponds to *Onki-UAA*0101* but contains a 95 bp deletion in the first third of exon VI (α_3 domain). Precisely the same deletion was also found in two clones isolated from the screening of rainbow trout cDNA libraries. This variant mRNA is probably generated from a splicing error with the donor site at the end of exon III (α_2) splicing into a cryptic splice site within exon IV (α_3); an AG dinucleotide highly conserved in acceptor sites can also be found at positions 671 and 672 in exon IV (α_3) of *Onmy-* and *Onki-UAA*0101* at the boundary of the deletion. When translated, this variant could produce a frame-shift mutation carboxyl (C)-terminal of the α_2 domain.

This type of analysis was extended to the spleen RNA extracted from a kokanee salmon (*Oncorhynchus nerka*), three PCR-amplified cDNA clones being fully sequenced. Two clones contain all exons with identical coding region sequences and are named *Onne-UAA*0101*, a third clone has the exon IV (α_3) splice error as described above for coho salmon and rainbow trout. Only four amino acid substitutions are found between the putative protein of *Onmy*-UAA*0101 and *Onne*-UAA*0101 (Fig. 7B). Probably due to a splicing error, a complete 120 bp intron V is found in all three sequences. The inclusion of intron V could potentially create a cytoplasmic tail with five amino acids substituted in the C-terminus GKSQS but would retain the same length due to a new stop codon. An allotype-defining, glycine to arginine substitution at position 223 of *Onmy*-UAA*0102 is also found in the kokanee salmon sequence.

Within the group of four species of *Oncorhynchus* studied (Fig. 7B), no two *UAA* sequences differ by more than six amino acids and most of these are conservative substitutions (Fig. 7B). The extant Pacific salmon and trout species of the genus *Oncorhynchus* are believed to have diverged for a minimum of six million years according to both fossil record and molecular sequence comparison (Devlin 1993; Stearley and Smith 1993), attesting to the relative conservation of the *UAA* gene.

The *UAA* gene of a brown trout (*Salmo trutta*) was also analyzed using the RT-PCR strategy. Four cDNA sequences were determined: one clone contains the aforementioned exon IV (α_3) splice error, another contains the intron V sequence, and the remaining two are

A

		leader		
		22	24	36
Rainbow	*Onmy-UAA*0101*	T	A	T
Rainbow	*Onmy-UAA*0102*	-	-	-
Chinook	*Onts-UAA*0101*	C	T	-
Coho	*Onki-UAA*0101*			-
Kokanee	*Onne-UAA*0101*	-	-	-
Brown	*Satr-UAA*0101*			c

		α1 domain												
		66	87	126	134	135	172	177	180	212	213	235	264	267
Rainbow	*Onmy-UAA*0101*	T	G	T	T	T	G	G	G	T	T	C	T	T
Rainbow	*Onmy-UAA*0102*	-	-	-	-	-	-	-	-	-	-	-	-	-
Chinook	*Onts-UAA*0101*	-	-	-	-	-	-	-	-	-	-	-	-	-
Coho	*Onki-UAA*0101*	c	a	-	-	-	A	a	a	-	-	-	-	-
Kokanee	*Onne-UAA*0101*	c	-	-	-	-	-	-	a	-	-	-	-	c
Brown	*Satr-UAA*0101*	c	-	c	C	G	-	-	a	C	C	t	g	-

		α2 domain															
		324	345	378	384	390	412	437	450	466	478	528	547	552	555	562	568
Rainbow	*Onmy-UAA*0101*	T	A	A	C	A	T	T	T	A	T	T	G	T	C	G	G
Rainbow	*Onmy-UAA*0102*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Chinook	*Onts-UAA*0101*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Coho	*Onki-UAA*0101*	c	-	g	-	g	-	-	-	-	-	g	-	-	-	-	-
Kokanee	*Onne-UAA*0101*	c	g	-	t	g	-	-	-	-	G	g	-	-	-	-	-
Brown	*Satr-UAA*0101*	-	-	-	-	g	G	G	a	c	-	-	C	G	t	T	A

		α3 domain												
		588	654	712	719	726	747	750	753	762	785	842	847	853
Rainbow	*Onmy-UAA*0101*	G	C	G	A	C	C	T	A	A	A	G	A	G
Rainbow	*Onmy-UAA*0102*	-	-	A	-	-	-	-	-	-	-	-	-	-
Chinook	*Onts-UAA*0101*	-	-	-	-	-	-	-	-	-	-	-	G	-
Coho	*Onki-UAA*0101*	-	t	-	-	g	t	-	-	-	-	-	-	-
Kokanee	*Onne-UAA*0101*	-	-	A	G	g	-	-	-	g	-	-	-	-
Brown	*Satr-UAA*0101*	c	-	-	G	g	-	c	g	g	T	A	-	C

		transmembrane / cytoplasmic							
		864	912	928	940	973	976	982-988	990
Rainbow	*Onmy-UAA*0101*	T	A	C	T	C	A	7 nt	A
Rainbow	*Onmy-UAA*0102*	-	-	-	-	-	-	7 nt	-
Chinook	*Onts-UAA*0101*	-	-	-	-	-	-	7 nt	-
Coho	*Onki-UAA*0101*	g	-	-	-	-	-	Deletion	-
Kokanee	*Onne-UAA*0101*	g	-	-	A	-	-	Deletion	-
Brown	*Satr-UAA*0101*	-	g	G	-	T	T	Deletion	G

B

		ld	α1			α2							α3						tm / cy				
		-8	30	43	56	123	131	145	168	169	173	175	223	225	247	266	268	270	294	299	310	313	314
Rainbow	*Onmy* -UAA*0101	S	I	V	F	S	V	L	E	F	V	E	G	N	Q	G	I	V	L	F	Q	R	G
Rainbow	*Onmy* -UAA*0102	-	-	-	-	-	-	-	-	-	-	-	R	-	-	-	-	-	-	-	-	-	-
Chinook	*Onts* -UAA*0101	P	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V	-	-	-	-	-	-
Coho	*Onki* -UAA*0101	-	-	I	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	E	*
Kokanee	*Onne* -UAA*0101	-	-	-	-	-	-	-	-	-	-	-	R	S	-	-	-	-	-	I	-	E	*
Brown	*Satr* -UAA*0101	-	T	-	S	A	G	V	Q	L	F	K	-	S	L	E	-	L	V	-	*		

Fig. 7A, B Differences in *UAA* coding regions from several species of salmonid. All comparisons are made with *Onmy-UAA*0101*. Identity is denoted by a *dash* and the stop codon by an *asterisk*. Nucleotide differences are shown in **A**. Position 1 is the first nucleotide of the initiation codon of the *Onmy-UAA*0101* cDNA. Non-synonymous substitutions are noted in *uppercase*, whereas synonymous substitutions are in *lowercase*. Amino acid differences are shown in **B**. Position 1 is the first amino acid residue of the α_1 domain of *Onmy*-UAA*0101

normal messages containing all the expected exons. Within the aligned regions of the four sequences are five amino acid positions that could be polymorphic within brown trout or could have been generated by PCR misincorporation; no two sequences are identical but they differ by no more than three amino acid substitutions (data not shown). The sequence of one full-length clone is named *Satr-UAA*0101* and is used for comparison in Fig. 7. As expected from the taxonomy of salmonids the *UAA*0101* coding-region sequence from brown trout is the most divergent, with *Satr-UAA*0101* differing from *Onmy-UAA*0101* by a total of 34 nucleotide substitutions and a 7 nucleotide deletion. These differences produce 14 amino acid substitutions, which are spread throughout the sequence, and also decrease the size of the cytoplasmic domain in brown trout by stopping translation at a position four codons before the stop codon in rainbow trout. The estimated divergence time of the Pacific (*Oncorhynchus*) and Atlantic (*Salmo*) salmonids is ~20 million years (Devlin 1993). Considering that rainbow trout and kokanee salmon are the most divergent *Oncorhynchus* species in this study (~6 million years) and that *Onmy*- and *Onne*-UAA*0101 differ by four amino acid residues, the 14 amino acid differences found between the rainbow and brown trout UAA are consistent with the greater separation of the genera *Oncorhynchus* and *Salmo*.

Discussion

In this paper we report on two class I genes of the rainbow trout: *Onmy-UAA*01* and *Onmy-UBA*01*. Analysis and comparison of their coding-region sequences shows that *Onmy-UBA*01* encodes a protein with characteristics typical of classical class I heavy chains, whereas *Onmy-UAA*01* encodes a heavy chain that lacks such characteristics and is divergent in sequence from all other known class I genes. Because of its

		5′ upstream of initiating ATG				*exon* I (46)	*intron* I (210)	*exon* II (258)	*intron* II (209)	*exon* III (273)	*intron* III (1032)			*exon* IV (285)	*intron* IV (90)	*exon* V (108)	*intron* V (120)	*exon* VI (20)	*3′* downstream of stop codon			
		s	d	i		s	s	s	s	s	s	d	i	s	s	s	s	d	s	d	i	
chinook 1	*Onts-UAA*0101*	1	0	0	(89)	2	0	0	0	0	4	3	0	0	0	0	0	0	0	0	0	(63)
chinook 2		1	0	0	(89)	2	0	0	0	0	4	3	0	1	0	0	0	0	0	0	0	(123)
coho 1	*Onki-UAA*0101*	3	6	4	(227)	0	3	5	2	4	36	50	5	3	1	1	0	7	13	42	4	(153)
coho 2		3	9	4	(227)	0	2	5	2	5	40	51	2	3	0	1	0	7	13	42	4	(153)

Fig. 8 Comparison of *UAA* genes of chinook and coho salmon with *Onmy-UAA*0101* of rainbow trout J. For each part of gene, the numbers of nucleotide differences with the *Onmy-UAA*0101* sequence from trout J are given. Nucleotide substitutions are denoted by *s*, deletion by *d*, and insertion by *i*. In *parentheses* is the length in nucleotides of the corresponding region in *Onmy-UAA*0101* used for comparison

unique characteristics, further investigation was focused on *Onmy-UAA*01*.

Within rainbow trout populations we defined two allotypes of *Onmy*-UAA*01, which differ by a single non-conservative amino acid substitution, at position 223 of the α_3 domain. Of interest is that this position corresponds to residue 227 of the mammalian class I molecule, an important component of the CD8-binding site (Potter et al. 1989; Salter et al. 1990). Both residues seen at position 223 are found in UAA*01 allotypes of other species of *Oncorhynchus*, suggesting that this polymorphism predates the divergence of these species. Diversity in the non-coding regions of the *Onmy-UAA*01* gene is greater than that in the coding region, suggesting that natural selection operates to conserve the *Onmy-UAA*01* sequence. This inference is further supported by the comparison of five salmonid species, which also reveals a relative conservation of the coding region and diversification of non-coding sequences. However, we failed to reveal the presence of the *UAA*01* gene in non-salmonid species, indicating either that the gene is restricted in its species distribution or that it is sufficiently diverged in other species to escape detection by the methods we used.

The relative conservation of the *UAA*01* gene and its oligomorphism is like that of the non-classical class I genes of mammals. Also shared with certain non-classical class I genes is the relatively low level of expression of *Onmy-UAA*01* as compared with *Onmy-UBA*01*. In summary, the data we obtained on *UAA*01* point to it being a non-classical gene of the salmonids. By analogy with the non-classical class I genes of mammals, we predict that *UAA*01* has a specialized function which contributes to the immune response.

Acknowledgments We thank Dr. William Cox from the California Department of Fish and Game (Rancho Cordova, Calif.), for providing various salmon samples; Dr. Robert Raison of the University of Technology (Sydney, Australia), for the Australian steelhead and lungfish samples; and Dr. Perry Hackett of the University of Minnesota (St. Paul), for the carp β-actin probe. This research was supported by grant AI 31168 to P.P. from the National Institutes of Health and a grant to R.P.H. from the Whirling Disease Foundation. K.E.M. was the recipient of a fellowship from the Natural Sciences and Engineering Research Council of Canada, and B.D. was a recipient of a fellowship from Medical Research Council of Canada. The work reported here was performed in compliance with all applicable regulations.

References

De Guerra A, Charlemagne J (1997) Genomic organization of the TcR beta-chain diversity (Dβ) and joining (Jβ) segments in the rainbow trout: presence of many repeated sequences. Mol Immunol 34:653–662

Devlin RH (1993) Sequence of sockeye salmon type 1 and 2 growth hormone genes and the relationship of rainbow trout with Atlantic and Pacific salmon. Can J Fish Aquat Sci 50:1738–1748

Dixon B, Shum B, Adams EJ, Magor KE, Hedrick RP, Adkison M, Muir DG, Parham P (1998) CK-1, a putative chemokine of rainbow trout (*Oncorhynchus mykiss*). Immunol Rev 166:341–348

Goodier JL, Davidson WS (1994) Tc1 transposon-like sequences are widely distributed in salmonids. J Mol Biol 241:26–34

Grimholt U, Hordvik I, Fosse VM, Olsaker I, Endresen C, Lie Ø (1993) Molecular cloning of major histocompatibility complex class I cDNAs from Atlantic salmon (*Salmo salar*). Immunogenetics 37:469–473

Hanfman DT (1993) The status and potential of aquaculture in the United States: an overview and bibliography. National Agricultural Library, Maryland

Hansen JD, Strassburger P, Du Pasquier L (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

Hashimoto K, Nakanishi T, Kurosawa Y (1990) Isolation of carp genes encoding major histocompatibility complex antigens. Proc Natl Acad Sci USA 87:6863–6867

Hedrick SM (1992) Dawn of the hunt for nonclassical MHC function. Cell 70:177–180

Hoffman GL (1970) Intercontinental and transcontinental dissemination and transfaunation of fish parasites with emphasis on whirling disease (*Myxosoma cerebralis*). In: Snieszko SF (ed) A symposium on diseases of fishes and shellfishes. American Fisheries Society, pp 69–81

Kaufman J, Salomonsen J, Flajnik M (1994) Evolutionary conservation of MHC class I and class II molecules – different yet the same. Semin Immunol 6:411–424

Klein J, Bontrop RE, Dawkins RL, Erlich HA, Gyllensten UB, Heise ER, Jones PP, Parham P, Wakeland EK, Watkins DI (1990) Nomenclature for the major histocompatibility complexes of different species: a proposal. Immunogenetics 31:217–219

Klein J, O'hUigin C (1994) The conundrum of nonclassical major histocompatibility complex genes. Proc Natl Acad Sci USA 91:6251–6252

Lawlor DA, Zemmour J, Ennis PD, Parham P (1990) Evolution of class-I MHC genes and proteins: from natural selection to thymic selection. Annu Rev Immunol 8:23–63

MacCrimmon H (1971) World distribution of rainbow trout (*Salmo gairdneri*). J Fish Res Bd Canada 28:663–704
Madden DR (1995) The three-dimensional structure of peptide-MHC complexes. Annu Rev Immunol 13:587–622
Marcadet A, P. OC, Cohen D (1989) Standardized Southern Blot Workshop Technique. In: Duport B (ed) Immunobiology of HLA. Springer, New York, pp 553–560
Mason PM, Parham P (1998) HLA class I region sequences, 1998. Tissue Antigens 51:417–466
Pamer E, Cresswell P (1998) Mechanisms of MHC class I-restricted antigen processing. Annu Rev Immunol 16:323–358
Parham P (1994) The rise and fall of great class I genes. Semin Immunol 6:373–382
Parham P, Adams EJ, Arnett KL (1995) The origins of HLA-A,B,C polymorphism. Immunol Rev 143:141–180
Potera C (1997) Fishing for answers to whirling disease. Science 278:225–226
Potter TA, Rajan TV, Dick RFd, Bluestone JA (1989) Substitution at residue 227 of H-2 class I molecules abrogates recognition by CD8-dependent, but not CD8-independent, cytotoxic T lymphocytes. Nature 337:73–75
Pullen JK, Horton RM, Cai ZL, Pease LR (1992) Structural diversity of the classical H-2 genes: K, D, and L. J Immunol 148:953–967
Rosenzweig B, Liao LW, Hirsh D (1983) Target sequences for the *C. elegans* transposable element Tc1. Nucleic Acids Res 11:7137–7140
Saitou N and Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425
Salter RD, Benjamin RJ, Wesley PK, Buxton SE, Garrett TP, Clayberger C, Krensky AM, Norment AM, Littman DR, Parham P (1990) A binding site for the T-cell co-receptor CD8 on the alpha 3 domain of HLA-A2. Nature 345:41–46
Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
Sato A, Klein D, Sultmann H, Figueroa F, O'hUigin C, Klein J (1997) Class I *Mhc* genes of cichlid fishes: identification, expression, and polymorphism. Immunogenetics 46:63–72
Shum BP, Azumi K, Zhang S, Kehrer SR, Raison RL, Detrich HW, Parham P (1996) Unexpected beta2-microglobulin sequence diversity in individual rainbow trout. Proc Natl Acad Sci U S A 93:2779–2784
Stearley RF, Smith GR (1993) Phylogeny of the Pacific trouts and salmons (*Oncorhynchus*) and genera of family Salmonidae. Transact Am Fish Soc 122:1–33
Stickney RR (1996) Aquaculture in the United States: a historical survey. John Wiley & Sons, New York
Tiersch TR, Chandler RW, Wachtel SS, Elias S (1989) Reference standards for flow cytometry and application in comparative studies of nuclear DNA content. Cytometry 10:706–710
Van Erp SH, Dixon B, Figueroa F, Egberts E, Stet RJ (1996) Identification and characterization of a new major histocompatibility complex class I gene in carp (*Cyprinus carpio L.*). Immunogenetics 44:49–61