

Unexpected β_2 -microglobulin sequence diversity in individual rainbow trout

(major histocompatibility complex polymorphism/evolution/teleost/salmonid/*Oncorhynchus mykiss*)

BENNY P. SHUM[†], KAORU AZUMI^{†‡}, SIWEI ZHANG[†], SUSAN R. KEHRER[§], ROBERT L. RAISON[§],
H. WILLIAM DETRICH[¶], AND PETER PARHAM^{†||}

[†]Departments of Structural Biology and Microbiology & Immunology, Stanford University, Stanford, CA 94305; [§]Immunobiology Unit, Department of Cell and Molecular Biology, University of Technology, Sydney, Broadway NSW 2007, Australia; and [¶]Department of Biology, Northeastern University, 414 Mugar Building, Boston, MA 02115

Communicated by Hugh McDevitt, Stanford University School of Medicine, Stanford, CA, November 29, 1995

ABSTRACT For mammals β_2 -microglobulin (β_2m), the light chain of major histocompatibility complex (MHC) class I molecules, is invariant (or highly conserved) and is encoded by a single gene unlinked to the MHC. We find that β_2m of a salmonid fish, the rainbow trout (*Oncorhynchus mykiss*), does not conform to the mammalian paradigm. Ten of 12 randomly selected β_2m cDNA clones from an individual fish have different nucleotide sequences. A complex restriction fragment length polymorphism pattern is observed with rainbow trout, suggesting multiple β_2m genes in the genome, in excess of the two genes expected from the ancestral salmonid tetraploidy. Additional duplication and diversification of the β_2m genes might have occurred subsequently. Variation in the β_2m cDNA sequences is mainly at sites that do not perturb the structure of the mature β_2m protein, showing that the observed diversity of the trout β_2m genes is not primarily a result of pathogen selection.

Major histocompatibility complex (MHC) class I molecules control the response of cytolytic lymphocytes to cells compromised by viral infection or malignant transformation. The class I heterodimer consists of a heavy chain associated with β_2 -microglobulin (β_2m). In mammals, the MHC harbors a multitude of class I heavy-chain genes, of which some (classical) are highly polymorphic while others (nonclassical) are conserved (reviewed in ref. 1); in contrast, β_2m is encoded by a single copy gene unlinked to the MHC (2) and it is either invariant [e.g., in human (3)] or has limited polymorphism [e.g., in mice (4)]. The discovery of a large and diversified family of MHC class I genes in mammals (reviewed in ref. 5) has stimulated investigations of their evolution and conservation by examining other vertebrate species, including fish (reviewed in ref. 6). The rainbow trout (*Oncorhynchus mykiss*) and other salmonid species share a common tetraploid ancestor (7). Although native to the drainage of the North American Pacific Coast and to the waters of the Pacific Ocean, rainbow trout have been introduced into many other countries of the world through human intervention (8). Here we describe the isolation and characterization of cDNA clones encoding β_2m and MHC class I heavy chain from *O. mykiss*.^{††} Contrasting with the mammalian paradigm, we find that the trout β_2m is encoded by a family of genes, giving an unexpected polymorphism to the cDNA expressed by an individual fish that cannot be explained by ancestral tetraploidy.

MATERIALS AND METHODS

Animals. Adult rainbow trout were purchased from Lintt's Trout Farm (Half Moon Bay, CA).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Screening of cDNA Libraries. Tissues were excised from rainbow trout and poly(A)⁺ mRNA was purified. The KA liver/spleen cDNA library and the fish J spleen cDNA library were constructed in the λ gt10 bacteriophage vector (Stratagene). For screening, replicate colony/PlaqueScreen nylon filters (DuPont/NEN) containing $\approx 5 \times 10^4$ plaque-forming units of the unamplified libraries were hybridized with the β_2m probe b2m-118 and the class I heavy-chain probe a3-200 (the KA and Ja clones) (9). These two probes were obtained by PCR amplification of spleen cDNA with the following primer sets: IP-30, 5'-GAT CTG CCA TGT GAG CAG CTT CCA CCC TCC; FD-30: 5'-AAA CTG CCA GCC CTT TTC GAA GGC CAG GTC for b2m-118; and the degenerate primer set A3-5p, 5'-TGY CWS GTG ACW GGT TTC TAC CC; A3-3p, 5'-AGR CYG STG TGW TKC ACM WGA CAG for a3-200. The fish O anterior-kidney library was constructed in the ZAP-Express vector (ref. 10; Stratagene) and screened with the a123 probe, which contains sequence corresponding to exons 2–4 of the cDNA clone Ja-1. Nucleotide sequences were determined on both DNA strands and analyzed using the Wisconsin Sequence Analysis Software Package [version 8.0; Genetics Computer Group (GCG), Madison, WI] (11).

Northern Blots. Total cellular RNA (10 μ g) or poly(A)⁺ mRNA (1 μ g) was separated by electrophoresis through 1% agarose/2.2 M formaldehyde gels (9). The labeled probes were b2m-mp, which encompasses the mature protein coding region of the trout β_2m clone Jb-1, and the class I heavy-chain probe a123, corresponding to exons 2–4 of *Onmy-UAA*01*.

Phylogenetic Analysis. The β_2m amino acid sequences were aligned with the PILEUP program of the GCG Package (11). The GenBank accession numbers of previously published sequences are as follows: human (M17987), chimpanzee (M30683), gorilla (M30684), orangutan (M30682), baboon (NCBI gi: 547299), tamarin (S71244), mouse *Mus musculus* (common allele a; X01838), mouse *Mus spretus* (L04992), rat (Y00441), guinea pig (P01886), rabbit (P01885), pig (L13854), cattle (X69084), horse (X69083), chicken (M84767), turkey (P21612), carp (L05536), zebrafish (L05383). Genetic distances were calculated from the peptide sequences according to the method of Kimura (12). A phylogenetic tree was constructed using the PHYLIP program (13) with the neighbor-joining method (14).

Abbreviations: MHC, major histocompatibility complex; β_2m , β_2 -microglobulin.

[‡]Present address: Department of Biochemistry, Faculty of Pharmaceutical Sciences, Hokkaido University, Sapporo 060, Japan.

^{||}To whom reprint requests should be addressed.

^{††}The sequences reported in this paper have been deposited in the GenBank data base [accession nos. L47354 (Jb-1), L49056 (Jb-2), L63530 (Jb-3), L63531 (Jb-4), L63532 (Jb-5), L63533 (Jb-6), L63534 (Jb-7), L63535 (Jb-8), L63536 (Jb-9), L63537 (Jb-10), L63538 (Jb-11), L63539 (Jb-12), L63541 (*Onmy-UAA*0101*), and L63542 (*Onmy-UAA*0102*).


```

I      5b-9  -----g----- 1100
      5b-11 -----g-----
III    5b-3  -----g-----
      5b-5  -----g-----
      5b-8  -----g-----
Consensus TCGAGGAAGC CCCATGCGAT ATGTATTGTG GTGAGCTGTA AATATCAACG TTCATTTTGG TTCGAAGTCA TTTATCGATA TAACATCTCA TCTTCTGTCT

I      5b-9  -----g----- 1200
      5b-11 -----g-----
III    5b-3  -----g-----
      5b-5  -----g-----
      5b-8  -----g-----
Consensus ATCTTTCGAA CATTCTCGTA AATTACCAAT TGTCAACAGT AAAATAAATG GGGGGTCTTT TTACAGGCAC ATTTTGCCAA GTAGAGAACT GTATTTTCAT

I      5b-9  -----a----- 1300
      5b-11 -----a-----
III    5b-3  -----a-----
      5b-5  -----a-----
      5b-8  -----a-----
Consensus TACTTAAAAA AAATAAGATA TTTTCTGTCT TTCGGTGTAT TATACAACAC TTTTCTGTTC AGTTGTATAC AAATAAATGG TCAACTATTA AAAATATTTA
    
```

B

Group	Position	Leader peptide							Mature β ₂ -m					
		3	9	11	12	13	14	15	16	18	16	17	19	20
I	Jb-1	S	-	-	-	-	-	-	-	-	-	-	-	-
	Jb-4	S	-	-	-	-	-	-	-	-	-	-	-	-
	Jb-9	S	-	-	-	-	-	-	-	-	-	-	-	-
	Jb-11	S	-	-	-	-	-	-	-	-	-	-	-	-
II	Jb-2	-	-	-	-	-	-	-	-	-	-	-	-	-
	Jb-10	-	-	-	-	-	-	-	-	-	-	-	-	-
III	Jb-3	-	-	-	-	-	-	-	-	-	-	-	-	-
	Jb-5	-	-	-	-	-	-	-	-	-	-	-	-	-
	Jb-7	-	-	-	-	-	-	V	-	-	-	-	-	-
	Jb-8	-	-	-	-	-	-	-	-	-	-	-	-	-
IV	Jb-6	T	A	C	V	F	L	G	F	N	Q	H	K	D
	Jb-12	T	A	C	V	F	L	G	F	N	Q	H	K	D
	Consensus	Y	V	G	L	I	Y	S	A	E	N	F	D	K

FIG. 1. Nucleotide sequences of β₂m cDNA clones and the deduced β₂m peptide sequences from the rainbow trout (*O. mykiss*). Dashes indicate identities with the consensus sequences. Stop codons are denoted by asterisks. (A) Nucleotide sequences of 12 β₂m cDNA clones isolated from the fish J spleen library. cDNA sequences are segregated into four groups by phylogenetic analysis. The poly(A) tails of individual sequences are deleted for alignment. Deduced amino acid sequence of *Onmy*-Jb-1 is shown. Proposed coding regions for the leader peptide and mature protein are indicated. (B) Positions of substitution in the deduced amino acid sequences of the 12 β₂m cDNA clones isolated from fish J.

Southern Blot. Ten-microgram samples of genomic DNA were digested to completion with restriction enzymes, and DNA fragments were separated by electrophoresis through 0.7% agarose gels in 0.5× TBE buffer (9). Hybridizations were performed with the β₂m mature protein probe b2m-mp; the MHC class I heavy-chain α₃ domain probe a3-200; or the heavy-chain extracellular domains probe a123. High-stringency hybridization was performed in a solution containing 50% formamide while low-stringency hybridization was performed in 30% formamide.

RESULTS AND DISCUSSION

Generation of Rainbow Trout DNA Probes. Based on features conserved in β₂m and MHC class I heavy-chain sequences of different species, we designed oligonucleotide primers to use in PCR amplification of rainbow trout spleen cDNA. Amplification products of the expected sizes were eluted from agarose gels, subcloned, and sequenced. Inserts having sequence similarity to β₂m and class I heavy chains of other fish (reviewed in ref. 6) were obtained and used to screen cDNA libraries.

Isolation of β₂m and MHC Class I Heavy-Chain cDNA Clones. Class I heavy-chain probes were used to screen three cDNA libraries, each made from individual rainbow trout. Two clones were obtained from each library and their nucleotide sequences showed they were either alleles or the products of alternative mRNA splicing of the same class I heavy-chain gene. The absence of polymorphism and certain conserved tyrosines in the peptide-binding site indicates this gene, designated as *Onmy*-UAA according to convention (15), is probably a nonclassical MHC class I gene (data not shown).

Clones encoding β₂m were found at high abundance (≈1% of total clones) in the spleen cDNA library made from fish J. Sequencing of three cDNA clones revealed unexpected poly-

morphism, which prompted analysis of an additional nine β₂m cDNA clones.

Sequence Diversity in Rainbow Trout β₂m cDNA. Ten distinct nucleotide sequences were obtained from the analysis of 12 cDNA clones (Fig. 1A). The cDNA clones grouped into two broad size ranges (800–950 and 1100–1250 bp) corresponding to the two major bands seen on Northern blot analysis (Fig. 2). Phylogenetic analysis divided the cDNA sequences into four groups with group IV being significantly divergent from groups I-III (Fig. 1A).

Much of the variation in trout β₂m is confined to the sequence encoding the signal peptide and the noncoding 5' and 3' flanking regions. The 3' untranslated regions are particularly variable (Fig. 1A). By comparison, the region encoding the mature protein is conserved (Fig. 1B), a distribution that argues strongly against the sequence divergence being an *in vitro* artifact. The mature proteins encoded by the group I, II, and III clones are identical, whereas the group IV clones have a cluster of 11 nucleotide substitutions that change amino acids 16, 17, 19, and 20 (Fig. 1A). When compared to β₂m from different species, the rainbow trout sequences cluster with those from other fish (Fig. 3).

Rainbow Trout β₂m Is Encoded by Multiple Genes. A likely cause of the observed sequence variation is that β₂m is encoded by multiple genes in rainbow trout. To investigate, we performed Southern blot analysis with the β₂m probe on a panel of genomic DNA preparations obtained from individual fish. Many hybridizing bands were observed with DNA digested with the restriction enzymes *Hind*III (data not shown) and *Taq* I (Fig. 4A), and the pattern was not affected by the stringency of wash conditions (Fig. 4C). Variation between fish is also apparent. The complexity of the β₂m banding patterns contrasts with the simpler pattern obtained with *Onmy*-UAA class I heavy-chain probes (Fig. 4D and E).

Contrasting with the results from trout, genomic DNA from two species of Antarctic rockcod gave cross-hybridizing bands

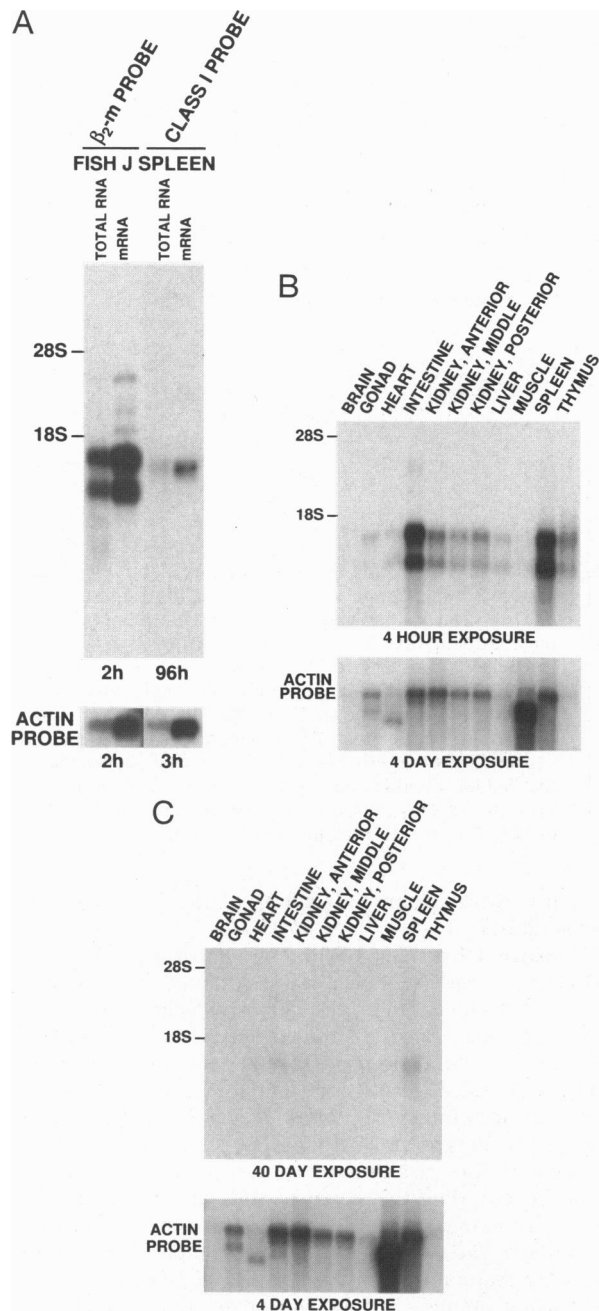


FIG. 2. Northern blot analysis of rainbow trout β_2m and *Onmy*-UAA class I heavy-chain mRNA. Exposure times for individual autoradiograms are noted. Positions of the 18S and 28S rRNA are labeled for reference. Quality of the RNA samples was examined on the same blots by hybridization with a carp β -actin probe (16). (A) Expression of β_2m and MHC class I mRNA in the spleen of rainbow trout J. mRNA samples were prepared from total RNA samples obtained from freshly isolated tissues using two rounds of oligo(dT) affinity column chromatography. (B) Expression of β_2m mRNA by various rainbow trout tissues. A single-step method was used to purify mRNA from the pooled frozen tissues of three trout (P, Q, and R) using the Oligotex direct mRNA kit of Qiagen (Chatsworth, CA). The difference in the protocols for preparing mRNA for the blot in A compared to B and C probably explains the different lengths of exposure required for the autoradiograms. (C) Expression of class I mRNA by various rainbow trout tissues. As in B, except that the hybridization was performed with the MHC class I heavy-chain probe a123.

compatible with a single-copy β_2m gene, a pattern analogous to that characteristic of mammals. Multiple β_2m bands on Southern blots are not a quirk of the Californian rainbow trout

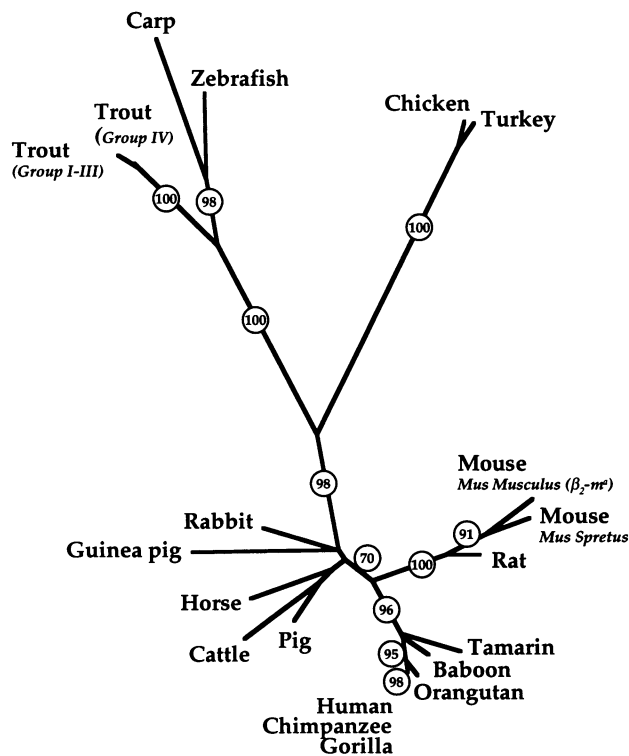


FIG. 3. Rainbow trout β_2m segregate with β_2m of other fish. A phylogenetic tree of β_2m sequences from different species was constructed by the neighbor-joining method (14). Numbers on branches refer to percentage recovery of that particular branch over 100 bootstrap replications. Nonterminal branches with a bootstrap frequency below 70% are not numbered. Method of maximum parsimony also gave a tree with similar topology, and all major branches were recovered with similar bootstrap frequencies (data not shown). Exception was the carp/zebrafish branch, which had a frequency of 69 per 100 of 1000 bootstrap replications.

population that provided the fish analyzed in Fig. 4A, as DNA from a population of Australian Steelhead trout, an anadromous form of *O. mykiss* (8), gave similar results (Fig. 4B). Multiple cross-hybridizing bands with the β_2m -mp probe were also detected in the genomic Southern analysis of another salmonid, the Atlantic salmon (*Salmo salar*; R. J. M. Stet, personal communication).

Tissue-Specific Expression of β_2m and *Onmy*-UAA mRNA. Northern blot analysis revealed a similar distribution for β_2m and *Onmy*-UAA gene expression in the tissues of rainbow trout (Fig. 2). Expression is highest in spleen and intestine and is undetectable in brain and muscle. The level of β_2m expression was ≈ 100 times higher than for the *Onmy*-UAA heavy-chain gene, consistent with the relative abundance of cDNA clones in the fish J library. This striking difference further points to *Onmy*-UAA being a low expression nonclassical class I gene and to the likely existence of yet undiscovered classical class I genes. That *Onmy*-UAA probes failed to isolate other class I heavy-chain family members from three cDNA libraries and gave simple restriction fragment length polymorphism patterns on Southern blots (Fig. 4D and E) is consistent with the finding of divergent families of class I genes in other fish species (6, 17).

Trout and other salmonids are believed to have evolved from a common tetraploid ancestor: genes are usually found in two copies and the extent to which they have diverged is variable (7). Tetraploid ancestry can therefore explain the presence of two β_2m genes in the trout genome and one might speculate that group I, II, and III cDNA clones are descendants of one ancient copy of the gene, and group IV clones are

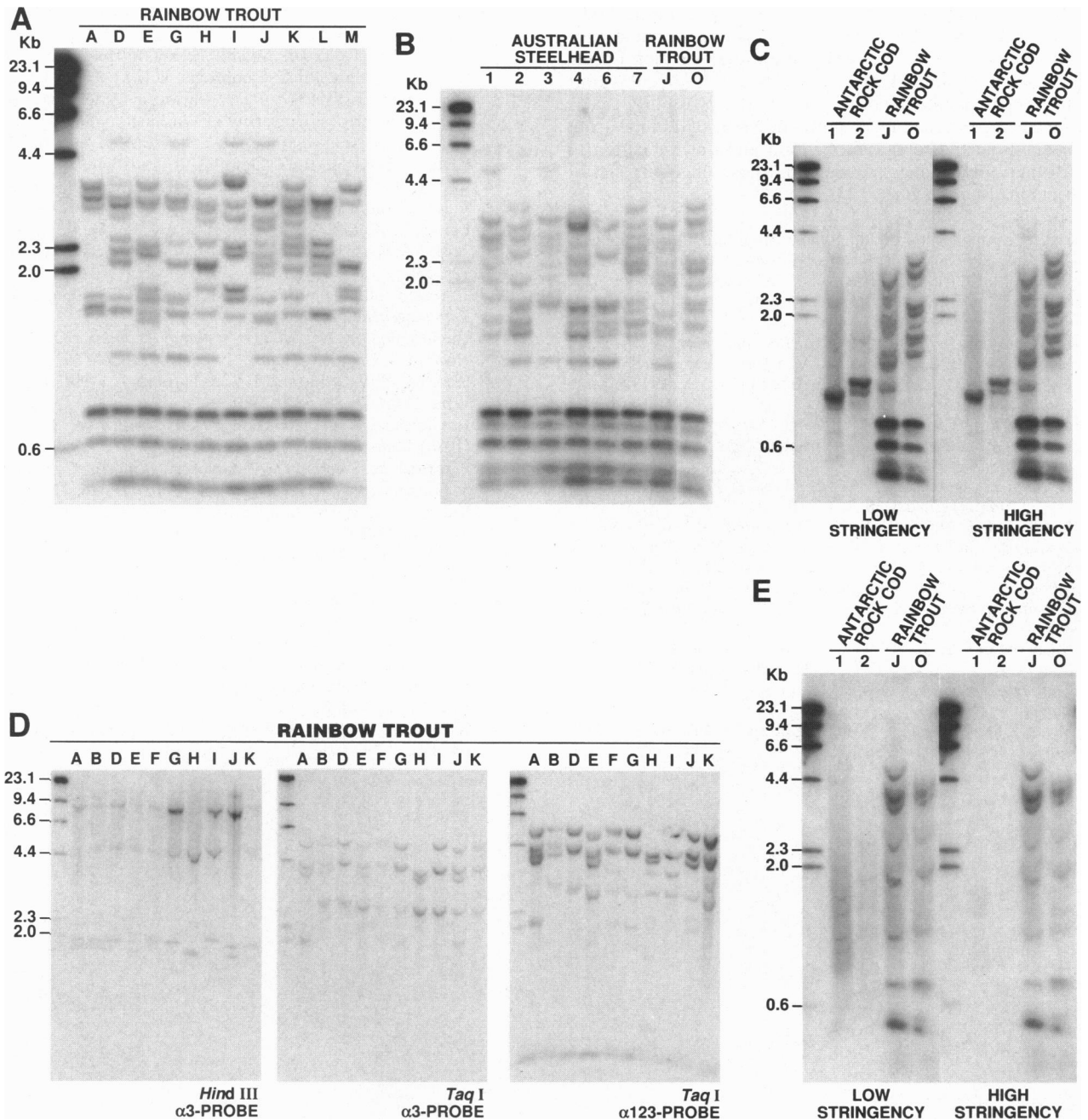


FIG. 4. Southern blot analysis of genomic DNA with probes derived from the β_2m and *Onmy-UAA*01* class I heavy-chain cDNA of rainbow trout. Sizes of *Hind*III-digested λ DNA are shown as markers. (A) Comparison of *Taq* I-digested DNA fragments from 10 individual Californian rainbow trout (lanes A–M) hybridizing under high-stringency conditions to the β_2m probe b2m-mp. (B) Comparison of *Taq* I-digested DNA fragments from seven individual Australian steelhead trout (lanes 1–7) and two Californian rainbow trout (lanes J and O) hybridizing under high-stringency conditions to the β_2m probe b2m-mp. (C) Comparison of *Taq* I-digested DNA fragments from rainbow trout J and O and from two species of Antarctic rockcod (sample 1 was from the species *Notothernia coriiceps* and sample 2 was from *Gobionotothen gibberifrons*) hybridizing with the β_2m probe b2m-mp under low stringency. Salt concentration of the final washes was $2\times$ SSPE (labeled low stringency) and $0.2\times$ SSPE (labeled high stringency). (D) Comparison of *Taq* I- and *Hind*III-digested DNA fragments from 10 individual Californian rainbow trout (lanes A–K) hybridizing with class I heavy-chain probes. Hybridizations were performed under high stringency with either the probe a3-200, which encodes only the α_3 domain of *Onmy-UAA*01*, or with the probe a123, which encodes all three extracellular domains of the class I cDNA. (E) Comparison of *Taq* I-digested DNA fragments from rainbow trout J and O and from two species of Antarctic rockcod as in C. Hybridization was performed under low stringency with the class I probe a123. Salt concentration of the final washes was $2\times$ SSPE (low stringency) and $0.2\times$ SSPE (high stringency).

descendants of the other. Tetraploidy alone, however, cannot explain the high sequence diversity of β_2m cDNA from an individual rainbow trout, indicating that subsequent duplication and divergence of β_2m genes has occurred.

Differences between trout β_2m cDNA sequences are largely at positions that do not change the primary structure of the mature protein. Therefore, selection for functional diversification of the protein is unlikely to be a major cause of the

variation, as is true for classical class I heavy-chain genes (18–20) and may be the case for the allelic differences in mouse β_2m (21–23). Conversely, purifying selection appears to have largely preserved the mature protein encoding sequence of trout β_2m genes while allowing the flanking sequences to diverge.

Similarities in the structure of β_2m and the immunoglobulin-like domains of MHC class I and II polypeptides suggest that

they share a common ancestor that was encoded in the MHC (24, 25). If that were the case, then the β_2m gene must have been translocated from the MHC to another location during evolution of the mammalian lineage (26, 27). Interpreting our results in light of this hypothesis raises the possibility that a β_2m gene may have remained within the MHC of rainbow trout and been subjected to the duplication and diversification to which other MHC genes, particularly those encoding class I heavy chains, seem prone (1, 5).

The authors thank Drs. Li Jin, Joanna Mountain, and R. Spencer Wells for discussion and assistance with phylogenetic analysis and Dr. Perry Hackett for the carp β -actin probe. This research was supported by Grant AI131168 from the National Institutes of Health.

1. Le Bouteiller, P. (1994) *CRC Crit. Rev. Immunol.* **14**, 89–129.
2. Goodfellow, P. N., Jones, E. A., Van Heyningen, V., Solomon, E., Bobrow, M., Miggiano, V. & Bodmer, W. F. (1975) *Nature (London)* **254**, 267–269.
3. Yang, S. Y. (1989) in *Immunobiology of HLA: Histocompatibility Testing 1987*, ed. Dupont, B. (Springer, New York), Vol. 1, pp. 309–331.
4. Hermel, E., Robinson, P. J., She, J.-X. & Fischer-Lindahl, K. (1993) *Immunogenetics* **38**, 106–116.
5. Kindt, T. J. & Singer, D. S. (1987) *Immunol. Res.* **6**, 57–66.
6. Dixon, B., van Erp, S. H. M., Rodrigues, P. N. S., Egberts, E. & Stet, R. J. M. (1995) *Dev. Comp. Immunol.* **19**, 109–133.
7. Allendorf, F. W. & Thorgaard, G. H. (1984) in *Evolutionary Genetics of Fishes*, ed. Turner, B. J. (Plenum, New York), pp. 1–53.
8. MacCrimmon, H. R. (1971) *J. Fish Res. Bd. Canada* **28**, 663–704.
9. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY).
10. Short, J. M., Fernandez, J. M., Sorge, J. A. & Huse, W. D. (1988) *Nucleic Acids Res.* **16**, 7583–7600.
11. Devereux, J., Haerberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
12. Kimura, M. (1983) *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, U.K.).
13. Felsenstein, J. (1993) PHYLIP, Phylogeny Inference Package (Dept. of Genetics, University of Washington, Seattle).
14. Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4**, 406–425.
15. Klein, J., Bontrop, R. E., Dawkins, R. L., Erlich, H. A., Gyllenstein, U. B., Heise, E. R., Jones, P. P., Parham, P., Wakeland, E. K. & Watkins, D. I. (1990) *Immunogenetics* **31**, 217–219.
16. Liu, Z. J., Zhu, Z. Y., Roberg, K., Faras, A., Guise, K., Kapuscinski, A. R. & Hackett, P. B. (1990) *J. DNA Sequence Map.* **1**, 125–136.
17. Takeuchi, H., Figueroa, F., O'hUigin, C. & Klein, J. (1995) *Immunogenetics* **42**, 77–84.
18. Hughes, A. L. & Nei, M. (1988) *Nature (London)* **335**, 167–170.
19. Howard, J. C. (1991) *Nature (London)* **352**, 565–567.
20. Hedrick, P. W., Whittam, T. S. & Parham, P. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 5897–5901.
21. Rammensee, H.-G., Robinson, P. J., Crisanti, A. & Bevan, M. J. (1986) *Nature (London)* **319**, 502–504.
22. Pérarnau, B., Siegrist, C.-A., Gillet, A., Vincent, C., Kimura, S. & Lemonnier, F. A. (1990) *Nature (London)* **346**, 751–754.
23. Kurtz, M. E., Martin-Morgan, D. & Graff, R. J. (1987) *J. Immunol.* **138**, 87–90.
24. Kaufman, J. F. & Strominger, J. L. (1982) *Nature (London)* **297**, 694–697.
25. Kaufman, J. F. & Strominger, J. L. (1983) *J. Immunol.* **130**, 808–817.
26. Klein, J., Ono, H., Klein, D. & O'hUigin, C. (1993) *Progress in Immunology*, eds. Gergely, J., Benczúr, M., Erdei, A., Falus, A., Füst, G., Medgyesi, G., Petrányi, G. & Rajnavölgyi, E. (Springer, New York), Vol. 8.
27. Kaufman, J., Völk, H. & Wallny, H.-J. (1995) *Immunol. Rev.* **143**, 63–88.