ELSEVIER

Comparative Biochemistry and Physiology Part C 121 (1998) 231–240

CBP

Molecular cloning of CYP1A from the estuarine fish *Fundulus heteroclitus* and phylogenetic analysis of CYP1 genes: update with new sequences[1]

Hilary G. Morrison [a], E. Jennifer Weil [a], Sibel I. Karchner [b], Mitchell L. Sogin [a], John J. Stegeman [b,*]

[a] *Marine Biological Laboratory, Woods Hole, MA 02543, USA*
[b] *Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA*

Received 5 March 1998; received in revised form 14 April 1998; accepted 1 May 1998

Abstract

Since we published a phylogenetic analysis of the CYP1A subfamily in 1995, several additional full-length sequences have been reported, including three members of an entirely new subfamily, CYP1B. Two avian sequences were recently published, so that CYP1A sequence data are now available from three of the five major vertebrate lineages. The two new branches that have been added to the CYP1 family tree significantly add to our understanding of P450 evolution. The inclusion of the CYP1Bs to the phylogenetic analysis allows us to root inferred trees. Addition of the avian CYP1As indicates that the CYP1A1/CYP1A2 duplication present in the mammalian lineage may have occurred after the divergence of birds and mammals. The number of fish species from which full-length coding regions of CYP1A genes have been sequenced has increased from four (trout, plaice, toadfish, and scup) to nine. These include CYP1A sequences from tomcod, butterflyfish, sea bream, sea bass, and the full-length sequence of CYP1A from the killifish *Fundulus heteroclitus* that is reported here. Phylogenetic analyses incorporating the new fish CYP1A sequences support our original conclusion that the fish CYP1As are monophyletic and indicate that the genes are evolving at very different rates in different species.

Keywords: CYP1A subfamily; CYP1B; Vertebrate lineages; Mammalian lineage; Divergence; Phylogenetic analyses

1. Introduction

Members of the cytochrome P450 (CYP) gene family 1 catalyze the activation of numerous hydrocarbon carcinogens and are induced by polynuclear and planar halogenated aromatic hydrocarbons and various natural compounds [34]. Apart from their clear importance as critical xenobiotic-metabolizing P450 enzymes, members of the *CYP1* gene family are an intriguing group in which to study gene evolution. Evidence indicates that this gene family is present in all vertebrate lineages [34], yet its endogenous functions remain to be discovered, even though some endogenous substrates are known [28]. Multiple gene duplication events have occurred in the CYP1 family. Correlating these duplications with branch points or key events in vertebrate evolution can suggest functions for the diverging genes.

The publication of the first non-mammalian CYP1A sequence in 1988 [12] marked the beginning of a vertebrate CYP1 database, to which many non-mammalian gene sequences have been added in recent years. Each new sequence has clarified relationships among the family members. This led to predictions about gene orthologues in different vertebrate lineages and to cor-

* Corresponding author. Tel.: +1 508 2892320; fax: +1 508 4572169; e-mail: jstegeman@whoi.edu

[1] This article was invited by Guest Editors Dr John J. Stegeman and Dr David R. Livingstone to be part of a special issue of CBP on cytochrome P450 (Comp. Biochem. Physiol. 121 C, pages 1–412, 1998).

0742-8413/98/$ - see front matter

relations of gene duplication with the emergence of new enzymatic functions. The more recent publication of two full-length avian *CYP1A* genes [9] adds a third vertebrate lineage to the *CYP1A* gene tree. While their addition answers some long-standing questions, it also raises some interesting new ones and highlights the need for investigating these genes in particular vertebrate groups, such as primitive fishes, amphibians, reptiles, and early diverging mammals. There are continuing questions regarding the identity, hence the classification, of *CYP1A* genes in non-mammals.

CYP1 genes in non-mammalian taxa may be particularly important in evaluating the ecological or physiological pressures contributing to diversification. The estuarine killifish, *Fundulus heteroclitus*, is a model system for studying rapid gene evolution. It is widely distributed in estuarine waters of the Atlantic coast. However, populations are non-migratory, with a limited home range. They represent discrete populations with different levels of exposure to environmental contaminants that can drive gene evolution.

2. Methods

2.1. Isolation of killifish CYP1A gene

The *CYP1A* gene from killifish, *Fundulus heteroclitus*, was obtained from a liver cDNA library prepared in Lambda Zap XR (Stratagene, La Jolla CA). Plaques were plated on XL1 Blue MRF′ cells grown on NZ-CYM agar at a density of approximately 10000 plaques per 150 mm plate. Plaques were lifted onto nitrocellulose membranes and the phage was eluted from the membrane in SM medium. Ten μl of each eluate were used as a template in a polymerase chain reaction using degenerate primers to conserved domains of the *CYP1A* genes. The primer sequences were 5′-CTG-CAG-GAT-CCY-GTG-GTK-GTK-CTG-AGY-GG-3′ (forward) and 5′-AAT-CGA-ATT-CAG-CAR-GAT-GGC-CAR-GAA-GAG-RAA-3′ (reverse). The reaction mixture included a buffer modified for long PCR [6], 1.75 units of rTth polymerase and 0.02 units of Vent polymerase, 1 ng of each primer, 200 μM dNTPs, and 1.15 mM Mg(OAc)2. Cycling conditions were 30 cycles at 94°C for 30 s, 45°C for 1 min, and 65°C for 12 min. The PCR products were electrophoresed in agarose gel and blotted to NYTRAN membrane. The blot was probed with a radiolabeled scup (*Stenotomus chrysops*) *CYP1A* gene [23] to determine which eluate contained a positive phage clone. This eluate was plated at low density, plaques were lifted onto NYTRAN membrane, and hybridized to the scup CYP1A probe. Hybridization conditions were 30% formamide, 0.1% SDS, 6 × SSC at 42°C; filters were washed at room temperature in 2 × SSC, 0.1% SDS. A single positive clone was isolated after two additional rounds of screening, and the clone was excised as a pBluescript plasmid following Stratagene protocols. Initial sequencing demonstrated that the clone contained two inserts ligated together at an *Eco*RI site within the CYP1A coding region, so the *CYP1A* gene was subcloned by *Eco*RI digestion. Double-stranded sequence was determined using a LI-COR 4000L automated sequencer with end-labeled M13 primers and CYP1A-specific primers (Fig. 1).

2.2. Sequence alignment and phylogenetic analyses

Additional *CYP1* genes were located through http://drnelson.utmem.homepage.html/ (D. Nelson's web server) and through NCBI BLAST search programs (www.ncbi.nlm.nih.gov) (Table 1). New mammalian CYP1As include CYP1A1 from sheep and CYP1A2 from guinea pig, as well as duplicate, partial, or allelic forms of previously described mammalian 1As. Two variant forms of trout 1A1 were recently reported in GenBank, which bring the number of trout 1A sequences to five. The killifish CYP1A and other recently reported full-length CYP1A and CYP1B sequences were added to our previous data set of mammalian and fish *CYP1A* genes and aligned using Clustal [13] and the PILEUP multiple sequence alignment program (Wisconsin Package version 9.0, Genetics Computer Group (GCG), Madison, WI). Unambiguously aligned positions were used to infer *CYP1* gene trees by both distance-matrix and parsimony methods, as previously described [23].

3. Results and discussion

3.1. Killifish CYP1A

The *Fundulus heteroclitus* CYP1A cDNA is 2537 bp in length and encodes a predicted protein 521 amino acids in length, comparable to the length of other fish CYP1As. The 5′ noncoding region is 138 bp long, starting at the *Eco*RI cloning site. The 3′ noncoding region is 833 bp and includes a polyadenylated region of 16 nucleotides. The inferred amino acid sequence of the *Fundulus* gene has a putative heme binding cysteine at position 463, in a typical P450 signature sequence of FGLGRRRCIG. The *Fundulus* gene also has a sequence 308-SDEKIVGIVNDLFGAGFDT-326 that is very highly conserved among teleost CYP1As. This region contains distinct residues that are conserved also with mammalian CYP1A1s, with mammalian CYP1A2s, or among fish CYP1As only, as described before [23]. The coding sequence and translated protein were readily aligned to the known fish CYP1As. The coding sequence shares 74–79% nucleotide similarity

```
-138                                                        gaattcggcacgagggac
-120  acctctgcaaaacacatttttttttctgttgttggacacgcatctctggaattagagtgtt
 -60  cgtcttctttttttatcattcagctaaaggttgagcagagaacagagaaaagttgtcatc

       M--A--L--M--I--L--P--F--I--G--A--L--S--V--S--E--G--L--I--A-
   1   ATGGCATTAATGATACTGCCATTCATTGGAGCACTCTCGGTGTCTGAGGGTTTGATAGCC
       -L--V--T--V--C--L--V--Y--L--T--L--K--H--F--R--R--E--I--P--E-
  61   TTGGTGACGGTGTGCTTGGTCTACCTGACCCTTAAGCATTTCCGCAGAGAGATCCCAGAG
       -G--L--R--R--L--P--G--P--T--P--L--P--I--I--G--N--F--L--E--L-
 121   GGGCTACGTCGACTCCCCGGCCCGACGCCACTTCCTATCATTGGGAATTTTCTGGAGTTG
       -G--S--K--P--Y--L--S--L--T--E--M--S--K--R--F--G--D--V--F--Q-
 181   GGAAGCAAGCCTTATCTAAGTCTTACTGAAATGAGCAAGCGCTTCGGCGATGTGTTCCAG
       -I--Q--L--G--M--R--P--V--V--I--L--S--G--Y--E--T--V--K--Q--A-
 241   ATTCAGCTCGGCATGCGCCCCGTGGTCATTCTGAGTGGTTATGAAACCGTTAAACAAGCT
       -L--T--K--Q--G--D--D--F--A--G--R--P--D--L--Y--S--F--R--F--I-
 301   CTCACCAAACAGGGGACGACTTTGCCGGCAGGCCAGACCTGTACAGCTTCCGCTTCATC
       -N--D--G--K--S--L--A--F--S--T--D--K--A--G--V--W--R--A--R--R-
 361   AATGATGGAAAGAGTCTGGCCTTCAGCACGGACAAAGCTGGCGTTTGGAGGCTCGCAGA
       -K--L--A--Y--S--A--L--R--S--F--S--S--L--E--G--K--L--P--E--Y-
 421   AAGCTTGCCTACAGTGCACTGCGTTCTTTTCCTCCCTGGAGGAAAGCTCCCAGAGTAC
       -S--C--V--L--E--E--H--I--C--K--E--T--E--H--L--I--K--E--L--H-
 481   TCCTGTGTGCTGGAGGAGCACATCTGCAAAGAGACTGAGCATCTGATCAAGGAACTCCAT
       -N--V--M--T--A--E--G--K--F--D--P--F--R--Y--I--V--V--S--V--A-
 541   AATGTCATGACAGCGGAAGGCAAATTTGACCCTTTCCGCTACATAGTTGTTTCTGTTGCC
       -N--V--I--C--G--M--C--F--G--R--R--Y--D--H--H--N--Q--E--L--L-
 601   AATGTGATCTGTGGCATGTGCTTTGGCCGACGTTATGACCATCATAACCAGGAGCTGCTG
       -S--L--V--N--L--A--E--D--F--V--Q--V--T--G--S--G--N--P--A--D-
 661   AGCTTGGTAAACCTCGCCGAAGATTTTGTCCAGGTGACAGGCAGCGGCAACCCAGCAGAT
       -F--I--P--A--L--Q--F--L--P--N--K--S--M--K--K--F--V--N--L--N-
 721   TTCATCCCTGCTCTGCAGTTTCTGCCCAACAAGTCAATGAAGAAGTTTGTCAACCTCAAC
       -N--R--F--N--N--F--V--Q--K--I--V--S--E--H--Y--S--T--F--D--K-
 781   AACCGCTTCAACAACTTTGTTCAGAAGATCGTCAGTGAGCACTACTCCACCTTTGACAAG
       -D--N--I--R--D--I--T--D--S--L--I--D--H--C--E--D--R--K--L--D-
 841   GACAACATCCGTGACATCACAGACTCCTTAATAGATCACTGCGAGGACCGGAAGCTCGAT
       -E--N--S--N--I--Q--M--S--D--E--K--I--V--G--I--V--N--D--L--F-
 901   GAGAACTCCAACATCCAGATGTCAGATGAGAAAATTGTTGGCATTGTCAACGACCTCTTC
       -G--A--G--F--D--T--I--S--T--A--L--S--W--A--V--M--Y--L--V--A-
 961   GGAGCTGGTTTCGACACCATCTCTACCGCTTTGTCATGGGCAGTGATGTACCTTGTGGCT
       -Y--P--E--V--E--E--R--L--Y--E--E--I--K--E--K--V--G--L--D--R-
1021   TATCCAGAGGTTGAGGAGAGGCTTTATGAAGAAATCAAGGAGAAAGTCGGTCTGGATCGT
       -T--P--V--M--S--D--R--S--N--L--P--L--L--E--S--F--I--L--E--L-
1081   ACTCCTGTTATGTCTGACAGAAGCAACTTGCCTTTACTTGAGTCTTTCATTCTGGAGCTC
       -F--R--H--S--S--Y--L--P--F--T--I--P--H--C--S--T--K--D--T--S-
1141   TTTCGTCATTCTTCATACCTGCCTTTCACAATCCCACACTGCTCTACAAAAGATACATCT
       -L--N--G--Y--F--I--P--K--D--T--C--V--F--V--N--Q--W--Q--I--N-
1201   CTGAACGGCTACTTCATTCCCAAAGACACCTGCGTGTTTGTCAACCAGTGGCAGATAAAC
       -H--D--P--E--L--W--K--D--P--S--M--F--I--P--D--R--F--L--S--A-
1261   CACGACCCAGAGCTCTGGAAAGACCCGTCTATGTTCATCCCAGACCGCTTCCTCAGCGCT
       -D--G--T--E--V--N--K--Q--E--G--E--K--V--L--I--F--G--L--G--R-
1321   GACGGCACAGAGGTAAACAAGCAAGAGGGAGAGAAGGTGCTTATTTTCGGCTTGGGAGA
       -R--R--C--I--G--E--V--I--A--R--N--E--V--F--L--F--L--A--I--I-
1381   CGGCGGTGCATCGGTGAGGTTATCGCACGAAACGAAGTCTTCCTCTTCCTGGCAATCATC
       -I--Q--K--L--H--F--Y--K--L--P--G--E--P--V--D--M--T--P--E--Y-
1441   ATCCAGAAACTGCACTTTTACAAGTTGCCCCGAGAGCCCGTGGACATGACCCCGGAGTAT
       -G--L--T--M--K--H--K--R--C--Y--L--G--V--A--M--R--A--K--D--V-
1501   GGCCTCACGATGAAGCACAAACGCTGTTACCTGGGAGTCGCAATGAGAGCTAAGGACGTG
       -Q--*
1561   CAGTGAagctctgcgttatttagaatgtaagactttgaaggtggcctatgttgactgtga
1621   cactgtaaattaagagttacattaagtttgagtgaaataagtatatttctaagaatggag
1681   gacactggatgaccttaatttatagagctaacgacattgaggcaaatccaagaaatttgc
1741   ttgatgcagactgcaagacatgtctacatttaggttcattgctttgtagtttgaggctct
1801   gcaaagtaacattcctacacgttgtctctgcttctaaatgaagaacaaaaaggtggatgc
1861   ttcgctgtcgtgttgtcctgtaatctaaaagcaatccatgcagaagtgactaagcagaca
1921   agagtttttggattgagagtaaatttttataggcattaccatgaagggcactgcaatgtgt
1981   gcattgtaacagaaaactctggagcaaacagatattacagtatgacaccattcacgtttg
2041   taagctaattatttttcatttcattatactgtaaatgccaatcctgaagctatatttgta
2101   tccctaaatgtgattatttgtgtatgtgtgtctgtattttcttataggtcatttttttta
2161   tgatccaagcattttgtttatgtttgaggtgaactacttatctgtgccaaaaaatgattt
2221   atcagggcagatcaaatgattctggaaaaattatcacaatgctgtacaattttatatatc
2281   ttatgtatcattcattgtaaaagaaaaaaaatgtcagggtttgagttattttatgaacag
2341   aaataaaattgaatgctctacggataaaaaaaaaaaaaaaactcgctcgtgccgaattc
```

Fig. 1. *Fundulus heteroclitus* cytochrome P450 1A sequences. The killifish CYP1A cDNA is 2537 bp long and encodes a predicted protein of 521 amino acids. The clone was obtained from a liver cDNA library.

Table 1
Sequences used to infer CYP1 gene trees

Species	Label used	Reference	Acc. no.
Homo sapiens	Human 1B1	[37] Sutter et al., 1994	U03688
	Human 1A1	[14] Jaiswal et al., 1985	K03191
	Human 1A2	[15] Jaiswal et al., 1987	M55053
Macaca irus	Monkey 1A1	[26] Ohmachi et al., 1993	D17575
Oryctolagus cuniculus	Rabbit 1A1	[16] Kagawa et al., 1987	D00212
	Rabbit 1A2	[16] Kagawa et al., 1987	D00213
Mus musculus	Mouse 1B1	[31] Savas et al., 1994	X78445
	Mouse 1A1	[19] Kimura et al., 1987	Y00071
	Mouse 1A2	[18] Kimura et al., 1984	X00479
Rattus norvegicus	Rat 1B1	[1] Battacharyya et al., 1995	X83867
	Rat 1A1	[32] Sogawa et al., 1984	K02246
	Rat 1A2	[17] Kawajiri et al 1984	K02422
Mesocricetus auratus	Hamster 1A1	[30] Sagami et al., 1991	D12977
	Hamster 1A2	[20] Lai and Chiang 1990	M34446
Cavia cobaya	Guinea pig 1A1	[25] Ohgiya et al., 1993	D11043
Cavia porcellus	Guinea pig 1A2	[4] Black 1995	U23501
Ovis sp.	Sheep 1A1	[11] Hazinski et al., 1995	S79795
Gallus gallus	Chicken 1A4	[9] Gilday et al., 1996	X99453
	Chicken 1A5	[9] Gilday et al., 1996	X99454
Oncorhynchus mykiss	Trout 1A1V1	[5] Berndtson and Chen 1994	S69278
	Trout 1A1V2	[2] Bailey et al., 1997	U62796
	Trout 1A1V3	[2] Bailey et al., 1997	U62797
	Trout 1A3	[5] Berndtson and Chen 1994	S69277
Pleuronectes platessa	Plaice 1A	[21] Leaver et al., 1993	X73631
Stenotomus chrysops	Scup 1A	[23] Morrison et al., 1995	U14162
Opsanus tau	Toadfish 1A	[23] Morrison et al., 1995	U14161
Chaetodon capistratus	Butterflyfish 1A	[38] Vrolijk et al., 1995	U19855
Pagrus major	Seabream 1A	[22] Mizukami et al., 1994	—
Microgadus tomcod	Tomcod 1A	[29] Roy et al., 1995	L41886
Dicentrarchus labrax	Seabass 1A	[36] Stien et al., 1996	U78316
Fundulus heteroclitus	Killifish 1A	Morrison et al., 1998 (this paper)	AF026800

and 65–80% amino acid identity with other fish CYP1As. This cDNA has been used to recover a *Fundulus* CYP1A clone from a genomic DNA library. The sequence data will be used to design primers to amplify the CYP1A coding region from *Fundulus* from discrete populations to examine amino acid variations that may be correlated with environmental exposures.

3.2. Sequence conservation of CYP1 genes and proteins

In contrast to the *Fundulus CYP1A* gene, which showed close relationship to other CYP1As, the inclusion of the *CYP1B1* genes and some other fish CYP1As resulted in fewer unambiguously aligned nucleotide positions. To compare aligned sequences, 1446 nucleotide positions encoding 482 amino acids were used, rather than the 1515 nucleotides and 505 amino acids of the original analysis [23]. Separate analyses were done of the fish *CYP1A* genes using 1515 nucleotide positions. The CYP1A genes of tomcod and seabream were particularly difficult to align, and comparison of their nucleotide sequences versus amino acid sequences showed specific areas where mutations leading to amino acid deletions and frameshifts appear to have occurred.

The seabream sequence contains two sites where a three-amino acid gap is required to maintain the alignment (residues 173–177 and 300–302). The first gap is located within α-helix D and the second is just upstream of α-helix I [7]. Two single amino acid deletions occurred in the tomcod sequence, at residues 40 and 106. In the region spanning residues 57–74, three nucleotide gaps were introduced that resulted in a single amino acid deletion at residue 69. Tomcod CYP1A is quite different from other CYP1As in this region (which includes part of α-helix A) at both the nucleotide and protein levels. The most critical alignment decisions occurred in residues 133–182 of the tomcod CYP1A, which includes α-helices C and D. A dinucleotide gap was introduced at the beginning of this region and a single nucleotide gap was placed at the end, to align the nucleotide sequence maximally. This resulted in nucleotide sequence identity of 73% to this region in the other fish CYP1As, while the amino acid identity is approximately 4%. It will be interesting to determine whether these seabream and tomcod CYP1A sequences code for functional proteins.

The sheep CYP1A1, as reported in GenBank, encodes protein of only 397 amino acids. The truncation results

Table 2
Average sequence identity among CYP1 gene groups

	Mammalian 1B1 ($n=3$) (%)	Mammalian 1A1 ($n=8$) (%)	Mammalian 1A2 ($n=6$) (%)	Chicken 1A ($n=2$) (%)	Fish 1A ($n=12$) (%)
Mammalian 1B1	86	50	51	50	50
Mammalian 1A1	—	83	75	67	62
Mammalian 1A2	—	—	82	67	61
Chicken 1A	—	—	—	84	63
Fish 1A	—	—	—	—	80

from a single base insertion at nucleotide 997, creating an in-frame stop codon. Preliminary phylogenetic analyses on the gene confirmed its identity as a CYP1A1.

Table 2 shows the average sequence identity shared within and between the CYP1 subgroups: CYP1B1s, CYP1A1s, CYP1A2s, avian CYP1As and fish CYP1As. Generally, genes within a subgroup share at least 80% sequence identity, while identity shared between the different CYP1As ranges from 61 to 75%. The shared sequence identity between the CYP1As and CYP1B1s is only about 50%, yet the CYP1A and CYP1B subfamilies share several highly conserved domains and were not difficult to align. Table 3 shows the shared nucleotide and amino acid identities between selected CYP1 genes.

3.3. *CYP1A gene trees*

It is apparent from these analyses that separate gene duplication events occurred in the mammalian CYP1As, in the avian CYP1As, and in the trout CYP1As (Fig. 2). Gene trees based on the aligned amino acids rather than nucleotides were poorly resolved (data not shown). As discussed previously [5,23], the duplication event in trout is a much more recent event than the mammalian CYP1A1/CYP1A2 duplication. The three trout 1A1 sequences appear to form a monophyletic group distinct from trout 1A3. However, since the sequences share over 97% nucleotide identity, classification of these additional forms as gene duplications or allelic variants requires analysis of their intronic and flanking regions. Additional avian sequences will be needed to determine when the CYP1A4/CYP1A5 duplication occurred in the avian lineage. The coding sequence identity of the two chicken genes is only 80%, compared to $>96\%$ for the two trout genes, indicating that duplication occurred earlier in the avian lineage than in the fish lineage. We predict that orthologues of the multiple trout CYP1As will be specific to salmonids. In contrast, we expect that other avian families will have both AHH (1A4) and AA (1A5) forms of CYP1A.

Have gene duplication events comparable to the mammalian CYP1A1/CYP1A2 or the avian AHH (CYP1A4)/AA (CYP1A5) divergences occurred in earlier-branching lineages? Recent evidence suggests that there are two CYP1A proteins in reptiles [8,39]. Genes for these enzymes must be identified and sequenced in order to determine whether they are orthologous to the mammalian CYP1A1 and CYP1A2 or to the avian forms, or neither. It also remains to be seen whether the two chicken CYP1As occur in other avian species. Although the distance between the chicken sequences suggests an early divergence, these genes could have resulted from a duplication sufficiently recent so that orthologues would not occur in all avian families, comparable to the trout CYP1A1/1A3 split. These questions can only be addressed through careful phylogenetic analyses of full-length genes from diverse vertebrate taxa.

Separate phylogenetic analyses done of the fish CYP1A genes included a greater number of aligned sites (Fig. 3). Human CYP1A1, CYP1B1, and chicken CYP1A4 were added as outgroups. The parsimony and distance methods generated congruent trees. All branches were resolved in this distance tree (Fig. 3b) in contrast to the distance tree of all taxa (Fig. 2b). Tomcod CYP1A is the most deep-branching of the fish CYP1As, even with the optimized nucleotide alignment; the basis for this is not known. The next long branch includes the trout CYP1A1s and CYP1A3. A relatively long branch in the distance tree (Fig. 3a) and bootstrap value of 98 in the parsimony tree (Fig. 3b) strongly support the grouping of the remaining *CYP1A* genes of killifish, plaice, toadfish, scup, seabream, sea bass, and butterflyfish. Within this group, the toadfish CYP1A is the most deep-branching followed by the CYP1As of killifish and plaice. Seabass and butterfly fish CYP1As are grouped, as are those of scup and seabream. The grouping of CYP1A from scup and seabream, which occur in the family Sparidae, is very strongly supported in both the distance and parsimony analyses. The long branch length of killifish CYP1A suggests that this gene is evolving rapidly relative to the other fish genes; in contrast, scup CYP1A appears to be evolving slowly as previously noted [23].

3.4. *Current nomenclature*

CYP1 gene nomenclature remains problematic. The first trout *CYP1A* gene reported by Heilmann et al. [12] was called CYP1A1. Two trout genes reported subsequently by Berndtson and Chen [5] were named

Table 3
Nucleotide and amino acid identity among selected CYP1 genes

	Human 1B1 (%)	Mouse 1B1 (%)	Human 1A1 (%)	Mouse 1A1 (%)	Human 1A2 (%)	Mouse 1A2 (%)	Chicken 1A4 (%)	Chicken 1A5 (%)	Trout 1A1 (%)	Scup 1A (%)	Tomcod 1A (%)
Human 1B1	—	84	40	42	39	39	37	38	39	39	36
Mouse 1B1	83	—	40	41	40	38	39	39	39	39	35
Human 1A1	51	50	—	81	75	70	61	64	57	58	52
Mouse 1A1	51	49	83	—	71	74	63	66	58	59	50
Human 1A2	52	51	81	76	—	75	58	65	56	57	50
Mouse 1A2	51	50	75	77	81	—	56	60	53	54	46
Chicken 1A4	51	50	67	66	66	64	—	80	59	58	50
Chicken 1A5	51	49	67	67	70	67	84	—	59	60	52
Trout 1A1	51	48	63	62	64	61	64	64	—	83	71
Scup 1A	51	49	63	63	65	63	64	64	79	—	69
Tomcod 1A1	52	50	65	62	65	61	66	65	79	78	—

Above diagonal, amino acid identity; below diagonal, nucleotide identity.

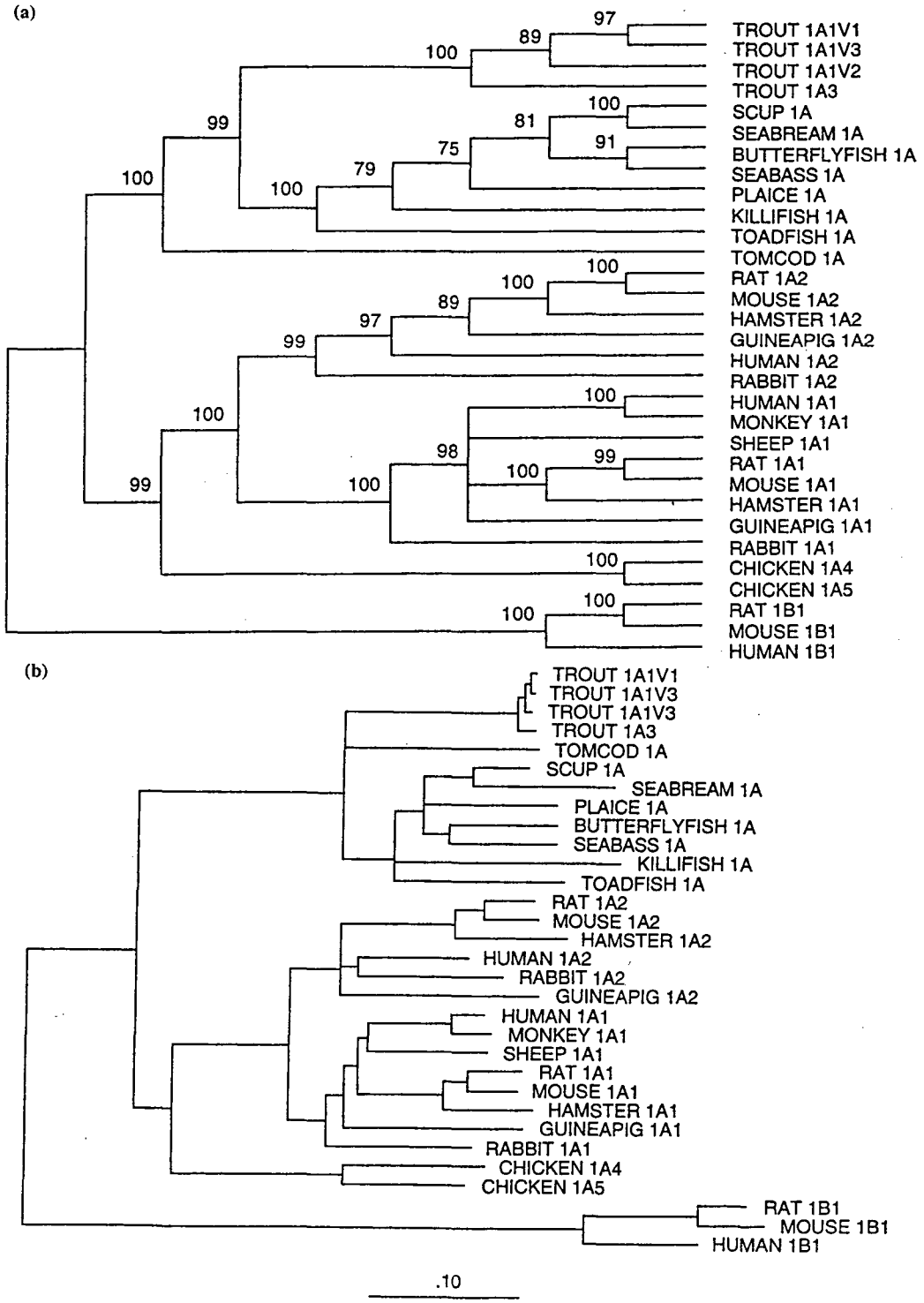

Fig. 2. Phylogenetic analysis of CYP1A genes. The tree inferred by the maximum parsimony method is shown in (a). Numbers above branches are bootstrap values based on 100 samplings and are a measure of relative confidence in an assembly. The tree inferred by the distance method is shown in (b). Distances between genes are represented by the sum of their horizontal separation. The scale bar under the tree indicates the distance corresponding to 10 differences in 100 positions. A total of 1446 residues in the alignment were used in the analysis.

CYP1A1 and CYP1A2 by the authors. 'CYP1A1' has since been renamed CYP1A3, and 'CYP1A2' has been renamed CYP1A1, based on sequence identities of 96.7 and 97.9%, respectively, with the original trout CYP1A1. However, the Heilmann gene sequence apparently represents a chimeric sequence with the 5′ portion derived from CYP1A3 and the 3′ derived from CYP1A1 (http://drnelson.utmem.homepage.html/). The

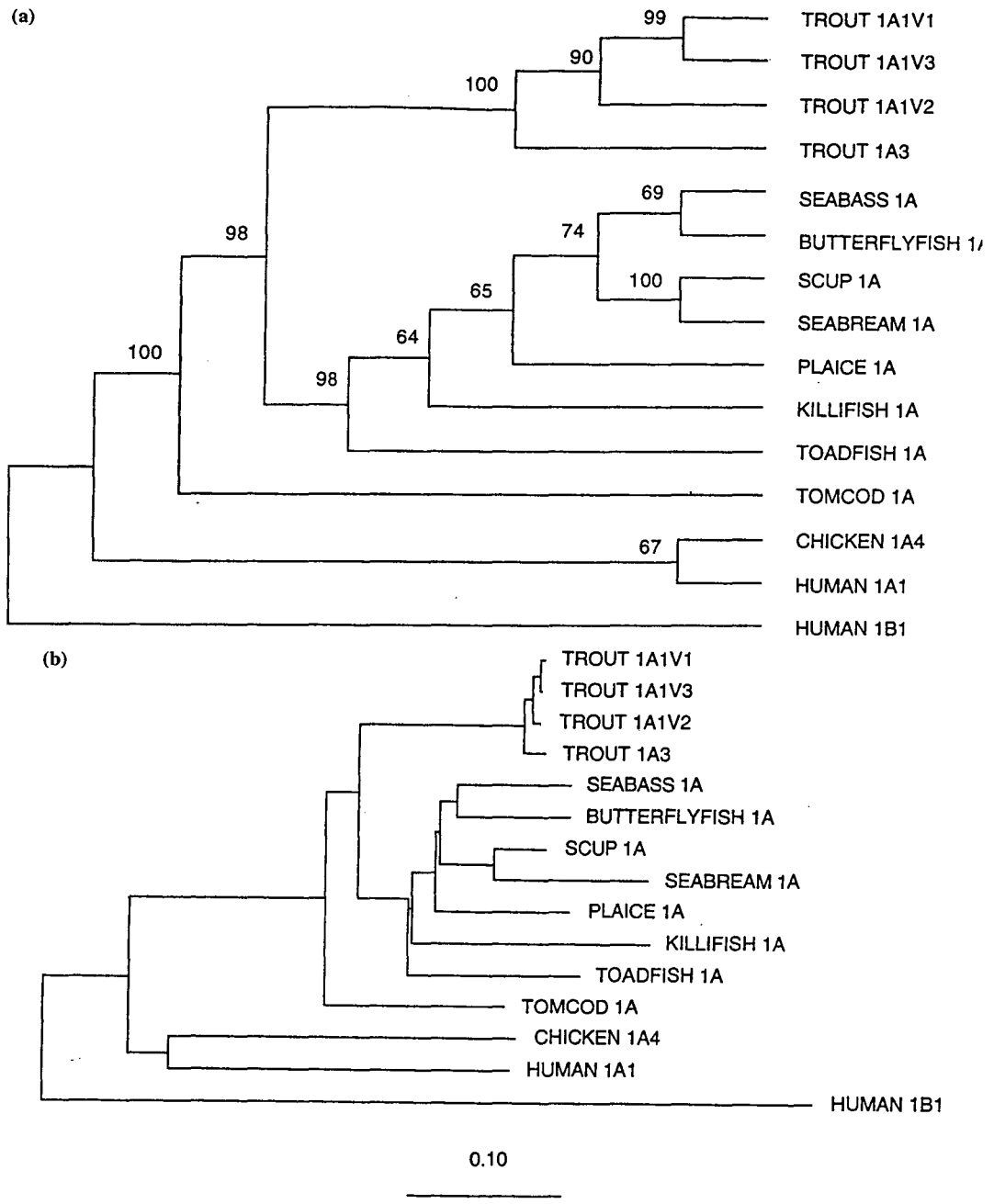

Fig. 3. Phylogenetic analysis of teleost CYP1A genes. Trees inferred by parsimony (a) and distance methods (b) for 12 fish CYP1A genes. A total of 1515 residues in the alignment were used.

decision to call one form CYP1A1 is thus somewhat arbitrary until more is known about the function of each. In fact, the *CYP1A3* gene contains two xenobiotic-responsive elements (XRE) within the 5′ flanking region upstream of the start site, while the *CYP1A1* gene does not have XRE in this region [5]. Thus, the gene originally called CYP1A1 by the authors, now called CYP1A3, may indeed be functionally more CYP1A1-like, as suggested by Berndtsen and Chen [5]. Regardless of the naming of these trout genes, their recent divergence precludes the possibility that one of the trout genes is orthologous to mammalian CYP1A1 and the other separately orthologous to CYP1A2.

As suggested before [33], multiple CYP1As, whether representing different loci or alleles, can be anticipated in many groups of fish, a suggestion confirmed by the multiple genes in trout and the possibility of multiple genes in tilapia (K. Chan, personal communication). The variants of trout CYP1A (Table 1) and genetic polymorphism in the 3′ untranslated region of the tomcod CYP1A, indicate the probability of allelic variation, further complicating the nomenclature issues. As

discussed above, it will not be possible to determine whether such variants represent gene duplications or alleles without data on the gene loci. Moreover, some variants may encode non-functional proteins. In the interval between the report of a gene sequence and analysis of its function, classification should be considered tentative. Notably, however, all of the teleost CYP1As have residues conserved only with mammalian CYP1A1s and others conserved only with CYP1A2s, consistent with the teleost CYP1As being descended from a single gene, representing the gene that diverged into mammalian CYP1A1 and CYP1A2. Functional analysis seems to support such a 'hybrid' character of teleost CYP1As [35], although on the basis of sequence, function and regulation, all of the fish CYP1As appear to have CYP1A1-like character rather than CYP1A2-like character.

The avian *CYP1A* genes were originally designated AHH and AA by the authors based on, respectively, aryl hydrocarbon hydroxylase and arachidonic acid epoxygenation activities [9]. Sequence analysis, however, indicates that the genes are equally distant from the mammalian CYP1A1 and CYP1A2 lines, so that neither one can be identified as more CYP1A1-like on the basis of sequence alone. The authors suggest naming them CYP1A4 and CYP1A5. We agree with this suggestion, even though the CYP1A4 appears to be functionally more like CYP1A1 than is CYP1A5. Whether it would be preferable to continue numbering the non-mammalian CYP1A genes sequentially as they are reported is an issue for discussion.

CYP nomenclature based on sequence alone is becoming increasingly complex. It is now clear that members of the major xenobiotic metabolizing CYP gene families identified in mammals occur as well in fish and other vertebrates. However, although CYP sequences in different species may be classified to the same CYP subfamily, functional counterparts may not be known. In *CYP* gene family 2, where diversification is extreme, identifying relationships at the level of subfamily is difficult between fish and mammals. Even in the CYP1A gene subfamily, in which there are few genes in most species, confusion is growing. Including data on properties additional to sequence is essential if we are to identify orthologous relationships that are important to identify fundamental roles of CYP genes. In the recent classification update [24], the designation of fish *CYP1A* genes as *CYP1A1* was based on the greater identity between the fish genes and mammalian CYP1A1s, the greater similarity in catalytic function and sites of expression, and the fact that there was but one gene evident in most species studied. This implies that CYP1A1 is the more ancient lineage, with conserved functions. The degree to which the CYP1A lineage is conserved in the Agnatha (hagfish and lamprey), where there appear to be differences in the CYP1A-like proteins [10], or in the invertebrates, may point to those functions. However, with the evidence that there are multiple CYP1A genes in some fish, in the absence of evidence of function it is misleading to identify only one of those genes in each species as CYP1A1; the implied orthology to other CYP1A1 genes may not be correct. At present the designation as CYP1A is used for the single genes in most species. Where there are multiple genes, designations must be different. Clearly, rules of nomenclature, much like the genes themselves, will continue to evolve. A system of 'tentative' designations, with final names based on a combination of functional and evolutionary data, might be workable.

Finally, with the cloning of a *CYP1A* gene from *Fundulus heteroclitus*, and from many other fish species, there can be no question that *CYP1A* is a constituent gene in teleost fishes generally; in addition to the sequences here, we are aware of partial sequences from several more fish species. It is still relevant, however, to determine whether diversification of this gene occurs in many teleost groups. Moreover, the function of these proteins in non-mammalian vertebrates still is not well known. Establishing the identity of CYP1As, and whether there are multiple CYP1As expressed in many species, is of interest not just phylogenetically, but will also shade our efforts to understand the susceptibility of organisms to toxicity of foreign chemicals. In this regard, the *Fundulus CYP1A* gene reported here may be important in furthering an understanding of the development of resistance to toxic effects of Ah receptor agonists in this species [3,27].

Acknowledgements

This research was supported in part by grants from the NIH Superfund Basic Research Program NIH P42-07381, by the US EPA Grant R823890 and by a Mellon Award at the WHOI (JJS); by NIH ES0558 (HGM), NIH GM32964 (MLS) and by a fellowship from the Beth Israel Deaconess Medical Center (EJW). We thank Dr Mark Hahn for helpful comments on the manuscript. Contribution No. 9692 from the Woods Hole Oceanographic Institution.

References

[1] Battacharyya KK, Brake PB, Eltom SE, Otto SA, Jefcoate CR. Identification of a rat adrenal cytochrome P450 active in polycyclic hydrocarbon metabolism as rat CYP1B1. Demonstration of a unique tissue-specific pattern of hormonal and aryl hydrocarbon receptor-linked regulation. J Biol Chem 1995;270:11595–602.

[2] Bailey G, You L, Harttig U. Cloning, sequencing and functional expression of two trout CYP1A cDNAs in yeast. Genbank U62796 and U62797.

[3] Bello S, Stegeman JJ, Hahn ME. Lack of cytochrome P450 1A inducibility in Fundulus heteroclitus from New Bedford Harbor: A case of dioxin resistance? Society of Environmental Toxicology

and Chemistry 17th Annual Meeting, Washington, DC (Abstract # P0820); 1996:272.
[4] Black VH, Wang A, Henry M, Shaw P. Molecular cloning of cDNA for guinea pig CYP1A2 comparison with guinea pig CYP1A1. Arch Biochem Biophys 1997;344:11–7.
[5] Berndtson AK, Chen TT. Two unique CYP1 genes are expressed in response to 3-methylcholanthrene treatment in rainbow trout. Arch Biochem Biophys 1994;310:187–95.
[6] Cheng S, Fockler C, Barnes WM, Higuchi R. Effective amplification of long targets from cloned inserts and human genomic DNA. Proc Natl Acad Sci USA 1994;91:5695–9.
[7] Edwards RJ, Murray BP, Boobis AR, Davies DS. Identification and Location of α-helices in mammalian cytochromes P450. Biochemistry 1989;28:3762–70.
[8] Ertl RP, Stegeman JJ, Winston GW. Induction time-course of cytochrome P450 by phenobarbital and 3-methylcholanthrene in liver microsomes of *Alligator mississippiensis*. Biochem Pharmacol 1998;55:1513–21.
[9] Gilday D, Gannon M, Yutzey K, Bader D, Rifkind AB. Molecular cloning and expression of two novel avian cytochrome P450 1A enzymes induced by 2,3,7,8-tetrachlorodibenzo-p-dioxin. J Biol Chem 1996;271:33054–9.
[10] Hahn ME, Woodin BR, Stegeman JJ, Tillitt DE. Aryl hydrocarbon receptor function in early vertebrates: Inducibility of cytochrome P4501A in agnathan and elasmobranch fish. Comp Biochem Physiol 1998;120C:67–75.
[11] Hazinski TA, Noisin E, Hamon I, DeMatteo A. Sheep lung cytochrome P4501A1 (CYP1A1): cDNA cloning and transcriptional regulation by oxygen tension. J Clin Invest 1995;96:2083–9.
[12] Heilmann LJ, Sheen Y-Y, Bigelow SW, Nebert DW. Trout P450IA1: cDNA and deduced protein sequence, expression in liver, and evolutionary significance. DNA 1988;7:379–87.
[13] Higgins DG, Sharp PM. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene 1988;73:237–44.
[14] Jaiswal AK, Gonzalez FJ, Nebert DW. Human dioxin-inducible cytochrome P1-450: complementary DNA and amino acid sequence. Science 1985;228:80–3.
[15] Jaiswal AK, Nebert DW, McBride OW, Gonzalez FJ. Human P(3)450: cDNA and complete protein sequence, repetitive Alu sequences in the 3′ nontranslated region, and localization of gene to chromosome 15. J Exp Pathol 1987;3:1–17.
[16] Kagawa N, Mihara K, Saito R. Structural analysis of cloned cDNAs for polycyclic hydrocarbon-inducible forms of rabbit liver microsomal cytochrome P-450. J Biochem 1987;101:1471–9.
[17] Kawajiri K, Gotoh O, Sogawa K, Tagashira Y, Muramatsu M, Fujii-Kuriyama Y. Coding nucleotide sequence of 3-methylcholanthrene-inducible cytochrome P-450d cDNA from rat liver. Proc Natl Acad Sci USA 1984;81:1649–53.
[18] Kimura S, Gonzalez FJ, Nebert DW. Mouse cytochrome P3-450: complete cDNA and amino acid sequence. Nucleic Acids Res 1984;12:2917–28.
[19] Kimura S, Smith HH, Hankinson O, Nebert DW. Analysis of two benzo[*a*]pyrene-resistant mutants of the mouse hepatoma Hepa-1 P(1)450 gene via cDNA expression in yeast. EMBO J 1987;6:1929–33.
[20] Lai TS, Chiang JY. Cloning and characterization of two major 3-methylcholanthrene inducible hamster liver cytochrome P450s. Arch Biochem Biophys 1990;283:429–39.
[21] Leaver MJ, Pirrit L, George SG. Cytochrome P450 1A1 cDNA from plaice (*Pleuronectes platessa*) and induction of P450 1A1 mRNA in various tissues by 3-methylcholanthrene and isosafrole. Mol Mar Biol Biotechnol 1993;2:338–45.
[22] Mizukami Y, Okauchi M, Arizono K, Ariyoshi T, Kito H. Isolation and sequence of cDNA encoding a 3-methylcholanthrene-inducible cytochrome P450 from wild red sea bream, *Pagrus major*. Mar Biol 1994;120:343–9.
[23] Morrison HG, Oleksiak MF, Cornell NW, Sogin ML, Stegeman JJ. Identification of cytochrome P-450 1A (CYP1A) genes from two teleost fish, toadfish (*Opsanus tau*) and scup (*Stenotomus chrysops*), and phylogenetic analysis of CYP1A genes. Biochem J 1995;308:97–104.
[24] Nelson DR, Koymans L, Kamataki T, Stegeman JJ, Feyereisen R, Waxman DJ, Waterman MR, Gotoh O, Coon MJ, Estabrook RW, Gunsalus IC, Nebert DW. P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. Pharmacogenetics 1996;6:1–42.
[25] Ohgiya S, Ishizaki D, Shinriki N. Molecular cloning of guinea pig CYP1A1: complete primary structure and fast mobility of expressed protein on electrophoresis. Biochim Biophys Acta 1993;1216:237–44.
[26] Ohmachi T, Sagami I, Kikuchi H, Fujii H, Suzaki Y, Fujiwara T, Watanabe M. Molecular cloning and sequence analysis of cDNA encoding a crab-eating monkey (*Macaca irus*) cytochrome P-450. Unpublished. Genbank Accession D17575, 1993.
[27] Prince R, Cooper KR. Comparisons of the the effects of 2,3,7,8-tetrachlorodibenzo-p-dioxin on chemically impacted and nonimpacted subpopulations of *Fundulus heteroclitus*: I. TCDD toxicity. Environ Toxicol Chem 1995;14:579–87.
[28] Raner GM, Vaz AD, Coon MJ. Metabolism of all-*trans*, 9-*cis*, and 13-*cis* isomers of retinal by purified isozymes of microsomal cytochrome P450 and mechanism-based inhibition of retinoid oxidation by citral. Mol Pharmacol 1996;9:515–22.
[29] Roy NK, Konkle BA, Kreamer G-L, Grunwald C, Wirgin II. Characterization and prevalence of a polymorphism in the 3′ untranslated region of cytochrome P4501A1 in cancer-prone Atlantic tomcod. Arch Biochem Biophys 1995;322:204–13.
[30] Sagami I, Ohmachi T, Fujii H, Kikuchi H, Watanabe M. Hamster cytochrome P-450 IA gene family, P-450IA1 and P-450IA2 in lung and liver: cDNA cloning and sequence analysis. J Biochem 1991;110:641–7.
[31] Savas U, Bhattacharyya KK, Christou M, Alexander DL, Jefcoate CR. Mouse cytochrome P-450EF, representative of a new 1B subfamily of cytochrome P-450s. Cloning, sequence determination, and tissue expression. J Biol Chem 1994;269:14905–11.
[32] Sogawa K, Gotoh O, Kawajiri K, Fujii-Kuriyama Y. Distinct organization of methylcholanthrene- and phenobarbital-inducible cytochrome P-450 genes in the rat. Proc Natl Acad Sci USA 1984;81:5066–70.
[33] Stegeman JJ. Cytochrome P450 forms in fish: Catalytic, immunological and sequence similarities. Xenobiotica 1989;19:1093–110.
[34] Stegeman JJ, Hahn ME. Biochemistry and molecular biology of monooxygenases: Current perspectives on multiplicity, functions and regulation of cytochrome P450 in aquatic species. In: Ostrander D, Malins, D, editors. Biochemical, Cellular and Molecular Biological Perspectives in Aquatic Toxicology. Lewis Press, 1994:87–206.
[35] Stegeman JJ, Soballa V, Morrison HG, Jacob J, Doehmer J. Catalytic properties of scup cytochrome P450 1A expressed in Chinese hamster V79 cells. Mar Environ Res 1998 (in press).
[36] Stien X, Amichot M, Berge J-B, Lafaurie M. Molecular cloning of a CYP1A cDNA from the teleost fish *Dicentrarchus labrax*. Unpublished. Genbank Accession U78316.
[37] Sutter TR, Tang YM, Hayes CL, Wo YY, Jabs EW, Li X, Yin H, Cody CW, Greenlee WF. Complete cDNA sequence of a human dioxin-inducible mRNA identifies a new gene subfamily of cytochrome P450 that maps to chromosome 2. J Biol Chem 1994;269:13092–9.
[38] Vrolijk NH, Lin C, Chen TT. Characterization and expression of a CYP1A gene from the tropical teleost, *Chaetodon capistratus*. Unpublished. Genbank Accession U19855.
[39] Yawetz A, Woodin BR, Stegeman JJ. Cytochromes P450 (CYP) in liver of the turtle *Chrysemys picta picta* and the induction and partial purification of CYP1A-like proteins. Biochim Biophys Acta 1998;1381:12–26.