

The CYP1D subfamily of genes in mammals and other vertebrates

Yusuke K. Kawai · Yoshinori Ikenaka ·
Shoichi Fujita · Mayumi Ishizuka

Received: 16 January 2010 / Accepted: 4 May 2010 / Published online: 22 May 2010
© Springer Science+Business Media, LLC 2010

Abstract Members of the cytochrome P450 family 1 (CYP1s) are involved in the detoxification and bioactivation of numerous environmental pollutants and phytochemicals such as polycyclic aromatic hydrocarbons (PAHs), aromatic amines, and flavonoids. The vertebrate CYP1 gene comprises four subfamilies: CYP1A, CYP1B, CYP1C, and CYP1D. Recently, the CYP1D gene was identified in fish, and subsequently in the platypus. These findings indicate the possibility that all vertebrates have a functional CYP1D subfamily. However, there is no information on the mammalian CYP1D gene. In this study we investigated the genomic location of CYP1D genes in mammals and other vertebrates *in silico*. We also performed phylogenetic analysis and calculated the identities and similarities of CYP1D sequences. The data from synteny and phylogenetic analyses of CYP1D genes demonstrated the evolutionary history of the CYP1 gene family. The results suggested that CYP1D became a nonfunctional pseudogene in human and bovine species; however, several other mammals possess functional CYP1D genes. The promoter regions of CYP1D genes were also examined. Unlike other CYP1 isoforms, few xenobiotic responsive element (XRE)-like sequences were found upstream of the CYP1D genes. Analysis of mammalian CYP1Ds also provided new insight into the relationship between CYP1 genes and the aryl hydrocarbon receptor.

Introduction

The cytochrome P450 (CYP) superfamily consists of more than 900 gene families comprising 11,000 genes that catalyze the oxidative metabolism of various organic compounds. The roles of P450 (heme-thiolate) enzymes are divided into two functions: xenobiotic metabolism and cellular. Members of the CYP1 family have a broad affinity for polycyclic aromatic hydrocarbons (PAH), as well as aromatic amines and some endogenous substrates. CYP1 genes can be found in the deuterostome lineage. The vertebrate CYP1 gene family consists of four subfamilies: CYP1As, CYP1Bs, CYP1Cs, and CYP1Ds (Goldstone et al. 2007). Mammals and birds possess orthologs of CYP1A1 and CYP1A2, whose genes duplicated before their lineages diverged (Goldstone and Stegeman 2006). Most other vertebrates have one CYP1A gene, with the exception of *Xenopus laevis*, which possesses CYP1A6 and CYP1A7 (Fujita et al. 1999). The CYP1B subfamily is found in all vertebrates (Sutter et al. 1994). CYP1B1 is known to be expressed in the liver and other organs and plays an important role in extrahepatic metabolic processes (Nebert et al. 2004). Two orthologs of CYP1C1 and CYP1C2 were found in the fish lineage; however, the CYP1C subfamily was not found in mammals, indicating that this gene was lost in the early mammalian lineage (Godard et al. 2005). The CYP1D subfamily was recently discovered in fish (Goldstone and Stegeman 2008). Prior to this, CYP1A and CYP1B were considered the only CYP1 subfamily genes in mammalian species; however, it is now known that human CYP1A8P and fish CYP1D are orthologous (Goldstone et al. 2009). This suggests the possibility that other mammalian CYP1A8 genes are CYP1D1 orthologs. Moreover, sequence data reported from many genomes indicate that many vertebrates have a new

Y. K. Kawai · Y. Ikenaka · S. Fujita · M. Ishizuka (✉)
Laboratory of Toxicology, Department of Environmental
Veterinary Science, Graduate School of Veterinary Medicine,
Hokkaido University, N18 W9, Kita-ku, Sapporo 060-0818,
Japan
e-mail: ishizum@vetmed.hokudai.ac.jp

functional CYP1 subfamily (<http://drnelson.utmem.edu/CytochromeP450.html>).

Phylogenetic analysis showed that fish CYP1D1s are a clustered CYP1A clade. Expression of CYP1A genes is generally induced by chemicals such as PAHs via the aryl hydrocarbon receptor (AhR). Upstream of the CYP1As, there are AhR-binding sites for the xenobiotic responsive elements (XREs). CYP1B and CYP1C are also known to be induced by AhR-binding ligands (Jönsson et al. 2007). However, the tissue distribution and the expression patterns of the CYP1D genes differed from those of the other CYP1 families. In fact, reports showed that the fish CYP1D1 gene was not induced by PCB126, which is a typical inducer of CYP1 isoforms (Goldstone et al. 2009; Jönsson et al. 2009; Zanette et al. 2009). Thus, the authors suggested that the CYP1D gene regulation cascade was different from that of other CYP1 genes.

In this study we focused on investigating the mammalian CYP1D gene in silico. We showed that in addition to fish, other vertebrates, including platypus, opossum, and rhesus macaque, could express a functional (full-length) CYP1D gene. We also analyzed the evolution of the CYP1D gene in mammalian lineages.

Materials and methods

Genomic DNA sequence and synteny data

CYP1D1 genes were searched for using the National Center for Biotechnology Information (NCBI) BLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). The gene order information was retrieved from Entrez Gene, Ensemble Genome Browser (<http://www.ensembl.org/index.html>), and the University of California Santa Cruz (UCSC) Genome Browser (<http://genome.ucsc.edu/>).

Phylogenetic analysis of vertebrate CYP1 genes

The nucleotide and amino acid sequences of vertebrate CYP1s were retrieved from the GenBank (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=nucleotide>) and JGI databases (<http://www.jgi.doe.gov/>) (Table 1). The CYP1 full lengths of amino acid sequences were aligned by CLUSTAL W using Molecular Evolutionary Genetics Analysis (MEGA) (Tamura et al. 2007). The alignments were then corrected manually using the program MEGA. Phylogenetic trees for amino acid sequences were constructed by Bayesian techniques based on the MrBayes program v3.1.2 (Ronquist and Huelsenbeck 2003). We performed Metropolis–Hastings coupled Monte Carlo Markov Chain (MC³) estimates by MrBayes with uninformative prior probabilities using the JTT model of amino acid substitution and prior uniform

gamma distributions approximated with four categories (JTT + Invariant + Gamma), as indicated by the Model-Generator (Keane et al. 2006). Four incrementally heated, randomly seeded Markov chains were run for 2×10^6 generations, and topologies were sampled every 100th generation. To confirm the MC³ results, two independent, randomly seeded analyses of the data set were performed with identical results. Output of the MC³ parameter was analyzed by Tracer v1.41 (Drummond and Rambaut 2007). The MC³ burn-in values were estimated using Tracer at 500,000 generations.

Sequence analyses

Substrate recognition site (SRS) (Gotoh 1992) identification and similarities were estimated based on the BLOSUM62 matrix (<http://www.uky.edu/Classes/BIO/520/BIO520WWW/blosum62.htm>). We applied the F3 × 4 codon model and allowed for estimating synonymous and nonsynonymous substitution rates by PAML (Yang 2007).

Promoter region searching

Promoter region searching for XREs was performed by tfscan of Emboss (Rice et al. 2000) and TFSEARCH (<http://mbs.cbrc.jp/research/db/TFSEARCHJ.html>) using the Tranfac database (Heinemeyer et al. 1998). On the TFSEARCH program, the taxonomy matrix was set to “vertebrates” and the threshold was set to “75.”

Results

CYP1D genes in vertebrates

CYP1D was searched for in the nucleotide collection database (nr/nt) by BLASTN algorithms using the platypus hypothetical protein (GeneID: 100074499), which was identified as CYP1D1 on the Cytochrome P450 home page (<http://drnelson.utmem.edu/CytochromeP450.html>). Several hits were identified: opossum cyp1a1 (GeneID: 100020715), rhesus macaque cyp1A1 (GeneID: 704920), and cattle, similar to the cytochrome P450 1A1 (GeneID: 785403), which were all classified as CYP1A8X/CYP1D1 according to the Cytochrome P450 home page.

The synteny between vertebrate CYP1D loci

The zebrafish CYP1D1 gene (GeneID: 492344) was located between TMC2 and klp20 followed by ANXA on chromosome 5 (Fig. 1). The gene order of TMC–ALDH–ANXA was found in many tetrapods. The ALDH gene would become a pseudogene in fish (Cañestro et al. 2009).

Table 1 GenBank accession numbers used in phylogenetic analyses

ID	Species	Accession No.	GeneID
CYP1A			
HUMAN_1A1	<i>Homo sapiens</i>	NM_000499	1543
MACAQUE_1A1	<i>Macaca mulatta</i>	XM_001091184	678694
MOUSE_1A1	<i>Mus musculus</i>	NM_009992	13076
RAT_1A1	<i>Rattus norvegicus</i>	NM_012540	24296
GUINEA_1A1	<i>Cavia porcellus</i>	D11043	100135512
CATTLE_1A1	<i>Bos taurus</i>	XM_588298	282870
SHEEP_1A1	<i>Ovis aries</i>	NM_001129905	100170113
WHALE_1A1	<i>Balaenoptera acutorostrata</i>	AB231891	
PIG_1A1	<i>Sus scrofa</i>	NM_214412	403103
HORSE_1A1	<i>Equus caballus</i>	XM_001493909	100062285
CAT_1A1	<i>Felis catus</i>	NM_001024859	554347
MINK_1A1	<i>Neovison vison</i>	EU046493	
SEAL_1A1	<i>Phoca fasciata</i>	AB030451	
OPOSSUM_1A1	<i>Monodelphis domestica</i>	XM_001379482	100029871
PLATYPUS_1A	<i>Ornithorhynchus anatinus</i>	XM_001511750	100080942
CHICKEN_1A1	<i>Gallus gallus</i>	NM_205146	396051
HUMAN_1A2	<i>Homo sapiens</i>	NM_000761	1544
MOUSE_1A2	<i>Mus musculus</i>	NM_009993	13077
RAT_1A2	<i>Rattus norvegicus</i>	NM_012541	24297
GUINEA_1A2	<i>Cavia porcellus</i>	D50457	100135513
CATTLE_1A2	<i>Bos taurus</i>	XM_591450	503552
WHALE_1A2	<i>Balaenoptera acutorostrata</i>	AB231892	
PIG_1A2	<i>Sus scrofa</i>	NM_001159614	100152910
HORSE_1A2	<i>Equus caballus</i>	XM_001493886	100062244
CAT_1A2	<i>Felis catus</i>	NM_001048013	554345
MINK_1A2	<i>Neovison vison</i>	EU046494	
SEAL_1A2	<i>Phoca fasciata</i>	AB030452	
OPOSSUM_1A2	<i>Monodelphis domestica</i>	XM_001379504	100029908
CHICKEN_1A2	<i>Gallus gallus</i>	NM_205147	396052
XENOPUS_1A6	<i>Xenopus laevis</i>	AB022087	
XENOPUS_1A7	<i>Xenopus laevis</i>	NM_001097072	100036775
XENOPUS_1A	<i>Xenopus tropicalis</i>	NM_001097344	100037911
SALMON_1A	<i>Salmo salar</i>	NM_001123694	100136926
ZEBRAFISH_1A	<i>Danio rerio</i>	NM_131879	140634
KILLIFISH_1A	<i>Fundulus heteroclitus</i>	AF026800	
CYP1B			
HUMAN_1B1	<i>Homo sapiens</i>	NM_000104	1545
MACAQUE_1B1	<i>Macaca mulatta</i>	XM_001109361	710385
MOUSE_1B1	<i>Mus musculus</i>	NM_009994	13078
RAT_1B1	<i>Rattus norvegicus</i>	NM_012940	25426
CATTLE_1B1	<i>Bos taurus</i>	XM_588813	511470
DOG_1B1	<i>Canis lupus familiaris</i>	NM_001159684	483038
OPOSSUM_1B1	<i>Monodelphis domestica</i>	XM_001365592	100011610
CHICKEN_1B1	<i>Gallus gallus</i>	XM_419515	421466
XENOPUS_1B1	<i>Xenopus tropicalis</i>	fgenesH1_pg.C_scaffold_74000073 ^a	
KILLIFISH_1B1	<i>Fundulus heteroclitus</i>	FJ786959	
ZEBRAFISH_1B	<i>Danio rerio</i>	XM_001918464	100150054

Table 1 continued

ID	Species	Accession No.	GeneID
CYP1C			
CHICKEN_1C1	<i>Gallus gallus</i>	XM_001233594	428832
KILLIFISH_1C1a	<i>Fundulus heteroclitus</i>	DQ133570	
KILLIFISH_1C1b	<i>Fundulus heteroclitus</i>	DQ133571	
ZEBRAFISH_1C1	<i>Danio rerio</i>	NM_001020610	553637
KILLIFISH_1C2	<i>Fundulus heteroclitus</i>	FJ786960	
ZEBRAFISH_1C2	<i>Danio rerio</i>	NM_001114849	100137725
CYP1D			
HUMAN_1A8P	<i>Homo sapiens</i>		100133307
MACAQUE_1A8	<i>Macaca mulatta</i>	XM_001093299	704920
CATTLE_1A8P	<i>Bos taurus</i>	XR_028256	785403
OPOSSUM_1A8	<i>Monodelphis domestica</i>	XM_001373076	100020715
PLATYPUS_1D1	<i>Ornithorhynchus anatinus</i>	XM_001506056	100074499
XENOPUS_1A8	<i>Xenopus tropicalis</i>	NM_001126813	100145341
KILLIFISH_1D1	<i>Fundulus heteroclitus</i>	FJ786961	
ZEBRAFISH_1D1	<i>Danio rerio</i>	NM_001007310	492344
CYP2C			
RAT_2C	<i>Rattus norvegicus</i>	NM_019184	29277
HUMAN_2C9	<i>Homo sapiens</i>	NM_000771	1559

^a Xenopus: JGI database name (protein ID: 161984)

In some mammals, the CYP gene was between ALDH and ANXA (Fig. 1). In reptile, CYP gene is located between ALDH and ANXA on the anole lizard scaffold 26. The *Xenopus* CYP1D1 gene also exists on scaffold 158 and the gene order of ALDH–CYP–ANXA was found to be conserved. On rat chromosome 1 and mouse chromosome 19, the gene order showed conserved synteny. Although a CYP-like gene was not found, other genes were found between ANXA1 and ALDH of rats and mice. The online database of P450 shows that in rabbit, dog, and pig, the CYP1D1 gene became a pseudogene. From the UCSC Genome Browser, there was the CYP1D1 pseudogene between ANXA and ALDH on dog chromosome 1. There was no information about genomic location of the rabbit and pig CYP1D1 pseudogene. We found the rabbit CYP1D1 pseudogene in the region from 59,578,315 to 59,579,034 on chromosome 1, which is located between ALDH1A1 and ANXA. In pig, we found the other ANXA gene between TMC and ANXA instead of between ALDH and CYP1D1. There was a CYP1D1 pseudogene between two ANXA genes in pig. In chicken and zebra finch, ALDH and ANXA genes mapped to the Z chromosome; however, there are still unknown regions between these genes.

The length of the DNA sequence between ALDH and ANXA genes in mammals ranged from 211 to 360 kbp (commonly 250 kbp). In opossum, the length of this DNA sequence was greatest at 360 kbp. In chicken, however, this

DNA region was 83,023 bp long, which is much less than that seen in mammals.

Phylogenetic tree of CYP1 genes

From the multiple alignments of vertebrate CYP1 nucleotide sequences in the corresponding region, a phylogenetic tree of CYP1 genes was constructed by the Bayesian method using the MrBayes program (Fig. 2). Human CYP2C9 and rat CYP2C were used as outgroup species in the unrooted tree. A phylogenetic tree was constructed of two clusters, the CYP1A/1D and CYP1B/1C clades. The vertebrates' CYP1D1 genes were clustered in one clade. Although the cattle CYP1D1 gene became a pseudogene, the gene was clustered in mammalian CYP1D1 clade and reflected the mammalian phylogenetic relationship. The CYP1A genes from chicken formed one cluster and were separated from mammalian CYP1As. In mammals, opossum CYP1A1 and platypus CYP1A were also separated from the eutherian mammal CYP1A1 cluster. In the frog CYP1A clade *X. laevis*, CYP1A6 and CYP1A7 formed one clade.

Sequence analyses

Identities and similarities of vertebrate CYP1D and CYP1A substrate recognition sites (SRSs) were calculated

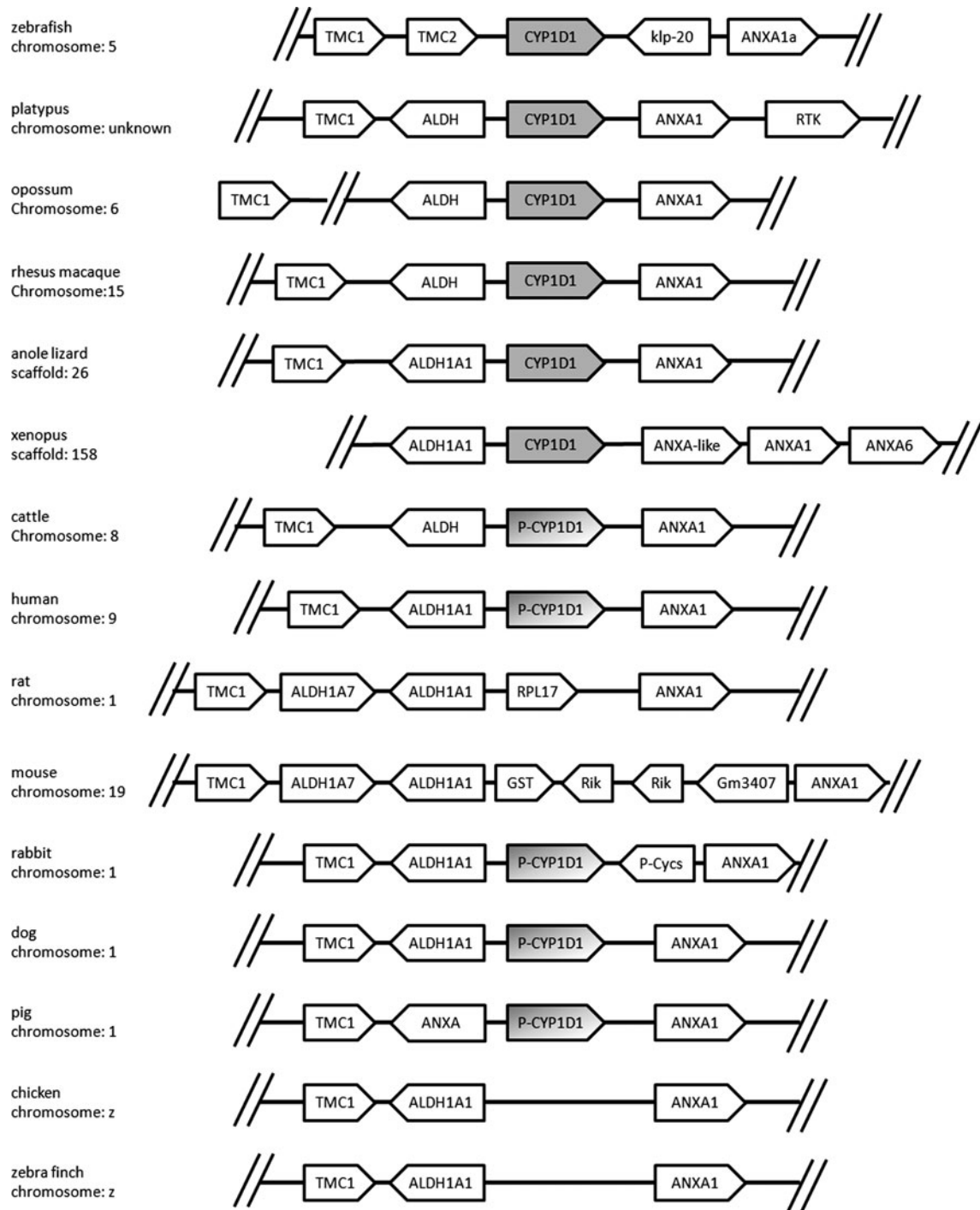
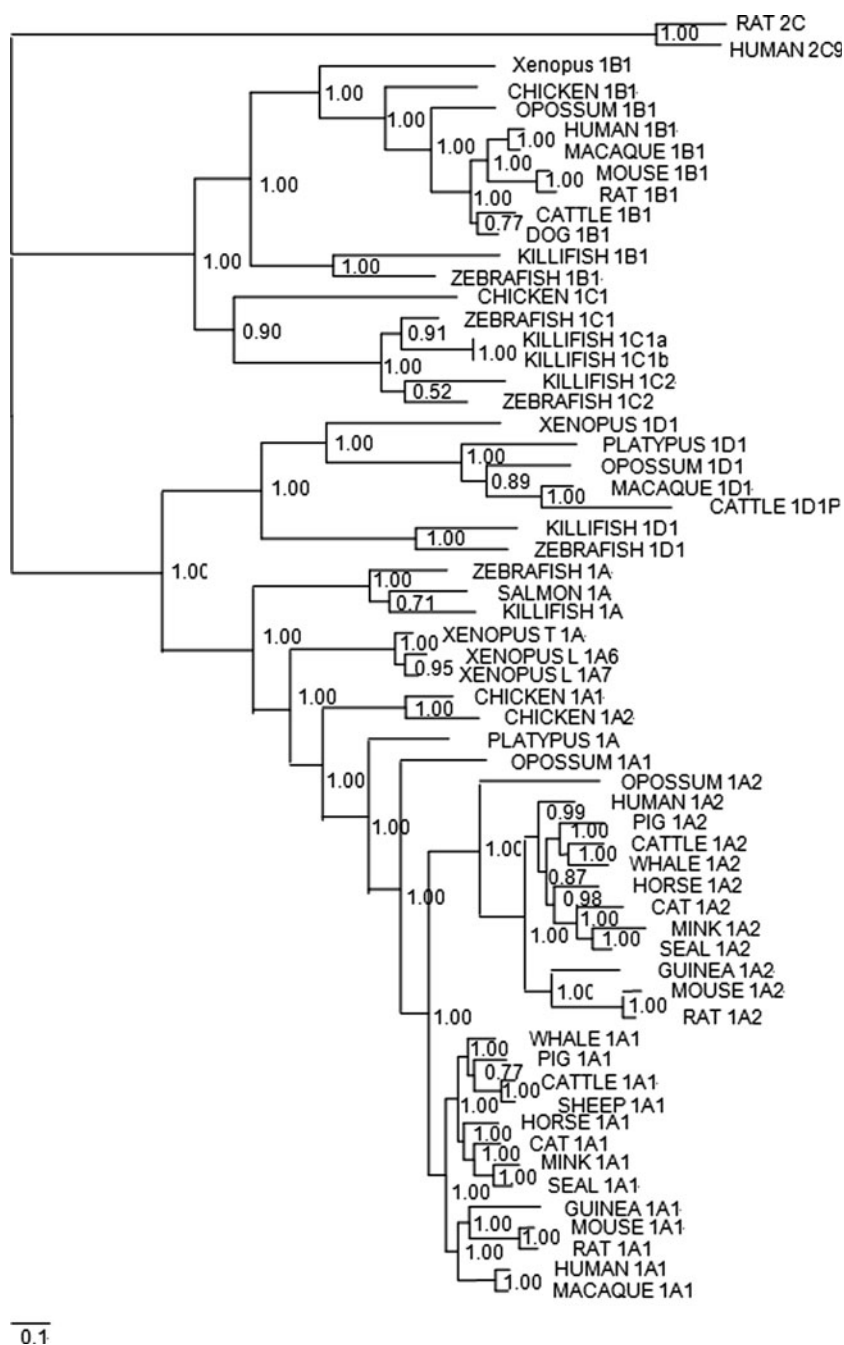


Fig. 1 Schematic representation of the gene order and orientation in the region around the cytochrome P450 gene on chromosome 5 of the zebrafish and other vertebrate chromosomes. Chromosomal DNAs were investigated to identify the CYP1D genes among vertebrates

based on the BLOSUM62 matrix. The similarity among CYP1D1SRSs showed the same scores as CYP1As (Tables 2, 3). On the other hand, more mutations occurred in the *Xenopus tropicalis* CYP1D SRSs. The dN/dS ratio

was estimated by PAML. There was no branching, suggesting that positive selection occurred in the CYP1D subfamily. In the CYP1A subfamily, CYP1A6 showed a dN/dS > 1.0.

Fig. 2 Phylogenetic analysis of mammalian and other vertebrate CYP1 genes based on the multiple alignment of the deduced amino acid sequences retrieved from the GenBank database. This is an unrooted tree for amino acid sequences constructed by the Bayesian method using the MrBayes program. The number at each branch and the length of the stem indicate posterior probabilities. Human and rat CYP2Cs were used as out-group species



Investigation of XREs

In mammalian genes, XREs were found in the upstream region of the CYP1D1 gene. In the platypus and opossum, one XRE was found in the 1-kbp region upstream of the CYP1D1 gene. Furthermore, in opossum, there were five more XRE regions in the 10-kbp upstream sequence of the CYP1D1 gene. Four XREs were identified 10 kbp upstream of this region in cattle, where the CYP1D1 mRNA was constructed from nine exons and included a

stop codon in exon 4. The cattle CYP1D1 gene was also found to be a pseudogene (Fig. 3). No XREs were found in the 1-kbp region upstream of the CYP1D1 gene in the genome of the rhesus macaque.

Discussion

In this study we investigated the CYP1D gene subfamily in vertebrates. Mammalian CYP1D genes were found to be

Table 2 Identification and similarity of CYP1A and CYP1D1 SRSs. a Identification and similarity of SRSs summary

	MACAQUE_1A1	OPOSSUM_1A1	PLATYPUS_1A
OPOSSUM_1A1	79.7 (93.7)		
PLATYPUS_1A	73.4 (87.3)	74.7 (89.9)	
XENOPUS_1A	69.6 (88.6)	64.6 (88.6)	73.4 (91.1)
	MACAQUE_1D1	OPOSSUM_1D1	PLATYPUS_1D1
OPOSSUM_1D1	79.7 (91.0)		
PLATYPUS_1D1	79.7 (87.3)	77.2 (88.6)	
XENOPUS_1D1	50.6 (75.9)	51.9 (72.2)	54.4 (77.2)

located in a region of conserved synteny. The gene order, TMC–ALDH–CYP–ANXA, was found in many vertebrates. In rat, mouse, and chicken, the ALDH and ANXA genes were arranged in tandem and the CYP gene was absent. In the mouse and rat, this region was similar in length to those of other mammals, but non-CYP genes were detected between ALDH and ANXA. This result suggested that in the rodent ancestor, the CYP1D region was lost by chromosome rearrangement. In case of pig, the gene order was not conserved. There was the other ANXA gene between ANXA and TMC instead of between ALDH and CYP1D1. This result suggested that the ANXA gene was duplicated by chromosome rearrangement and that ALDH and CYP1D1 became pseudogenes. In rabbit and dog, the CYP1D1 pseudogene was between ALDH and ANXA. This result indicated that in each lineage the CYP1D1 became a pseudogene by small-scale mutations such as point mutations, insertions, and deletions. In chicken, there were unknown regions between ALDH and ANXA. The length of this region was around 80 kbp, with about 10 kbp of unknown sequence. The CYP-like gene was not detected in chicken. CYP1D1 could not be found between ALDH and ANXA neither in zebra finch. In chicken and zebra finch, ALDH and ANXA mapped to chromosome Z, suggesting that the CYP1D1 genes of birds were located on the sex chromosome. However, in bird lineage CYP1D1 might become a pseudogene.

The phylogenetic tree of vertebrate CYP1 genes was constructed by the CYP1A/1D and CYP1B/1C clades. The CYP1D1 gene of mammals and *X. tropicalis* CYP1D formed one clade with fish CYP1D genes. The CYP1A genes from chicken formed another cluster, separated from mammalian CYP1As because of gene conversion (Goldstone and Stegeman 2006). In mammals, the opossum CYP1A1 was also separated from the mammalian CYP1A1 clade, also by gene conversion. In the platypus, only one CYP1A gene was found and this gene had also separated from other CYP1A clades. The gene was located near the Bdp–mCOX–COX sequence. This gene arrangement was different from the gene order surrounding CYP1A genes of other vertebrates. In platypus, the chromosome

rearrangement might occur and the gene order was changed. Furthermore, we constructed the other phylogenetic tree from the central region of CYP1As, which was considered a nonconverted region (data not shown). This phylogenetic tree indicated that this platypus CYP1A gene is CYP1A1, not CYP1A2. Platypus CYP1A2 might become a pseudogene. The *X. laevis* CYP1A6 and CYP1A7 genes formed one clade, which was separated from the *X. tropicalis* CYP1A gene. Because *X. laevis* is a tetraploid, this branch may not have been caused by gene conversion.

The data from synteny and phylogenetic analyses suggested that the common ancestor of vertebrates already possessed CYP1D genes. Genomic information further suggested that *X. tropicalis*, platypus, opossum, and macaque possess functional CYP1D genes, and humans and cattle have a CYP1D pseudogene. In the case of rat and mouse, CYP1D1 was not found in the normal CYP1D region, which indicated that these CYP1D1s had become pseudogenes. According to mammalian phylogeny, human, cattle, and rodent CYP1D1s became pseudogenes independently (Fig. 4). In addition, other mammalian orders may have functional CYP1D genes.

To investigate CYP1D conservation, we calculated the similarity between the CYP1D1 SRSs. The similarities present in CYP1D1 SRSs were also conserved in CYP1As. This result suggested that mammalian CYP1D genes play an important role in the enzymatic function of CYP1As. However, in *X. tropicalis*, SRS similarity among CYP1Ds was lower than among CYP1As. The high rate of amino acid mutations in the *X. tropicalis* CYP1D protein sequence indicated two possible reasons. First, in *X. tropicalis*, the restraints on sequence conservation were decreased because the role of this gene became less important over time. Second, in *X. tropicalis*, amino acid substitution frequently occurred, allowing the organism to adapt to changing environments. Thus, the dN/dS ratios of CYP1As/1Ds SRSs were calculated and used to estimate the selection pressure. We found no branching, indicating positive selection. However, when estimating dN/dS ratios for the aligned regions of the CYP1A and CYP1D genes, the dN/dS ratio of *X. tropicalis* CYP1As was greater than

Table 3 Identification and similarity of CYP1A and CYP1D1 SRSs. Value of each SRS

	MACAQUE_1A1	OPOSSUM_1A1	PLATYPUS_1A
OPOSSUM_1A1			
SRS1	72 (92)		
SRS2	75 (75)		
SRS3	75 (87.5)		
SRS4	89.5 (100)		
SRS5	90.9 (100)		
SRS6	75 (100)		
PLATYPUS_1A			
SRS1	56 (76)	76 (80)	
SRS2	62.5 (87.5)	75 (87.5)	
SRS3	75 (87.5)	62.5 (100)	
SRS4	94.7 (100)	84.2 (100)	
SRS5	90.9 (100)	90.9 (100)	
SRS6	62.5 (75)	37.5 (75)	
XENOPUS_1A			
SRS1	56 (80)	60 (84)	68 (96)
SRS2	62.5 (87.5)	50 (87.5)	62.5 (100)
SRS3	62.5 (100)	37.5 (87.5)	62.5 (87.5)
SRS4	84.2 (94.7)	78.9 (94.7)	89.5 (94.7)
SRS5	81.8 (90.9)	81.8 (90.9)	81.8 (90.9)
SRS6	75 (87.5)	62.5 (87.5)	62.5 (62.5)
	MACAQUE_1D1	OPOSSUM1D1	PLATYPUS_1D1
OPOSSUM_1D1			
SRS1	88 (96)		
SRS2	50 (87.5)		
SRS3	75 (75)		
SRS4	84.2 (94.7)		
SRS5	72.7 (90)		
SRS6	87.5 (87.5)		
PLATYPUS_1D1			
SRS1	84 (88)	80 (92)	
SRS2	75 (87.5)	50 (75)	
SRS3	62.5 (75)	62.5 (75)	
SRS4	84.2 (94.7)	78.9 (89.5)	
SRS5	81.8 (90.9)	90.9 (100)	
SRS6	75 (75)	87.5 (87.5)	
XENOPUS_1D1			
SRS1	48 (68)	52 (72)	52 (76)
SRS2	37.5 (62.5)	25 (37.5)	50 (62.5)
SRS3	12.5 (62.5)	12.5 (50)	0 (62.5)
SRS4	68.4 (100)	68.4 (94.7)	78.9 (94.7)
SRS5	63.6 (90.9)	63.6 (90.9)	54.5 (90.9)
SRS6	50 (50)	62.5 (62.5)	62.5 (75)

The value in *parenthesis* is similarity of amino acid sequences

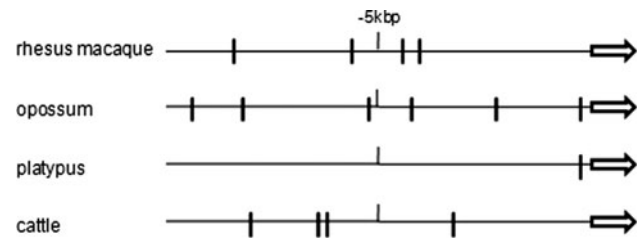


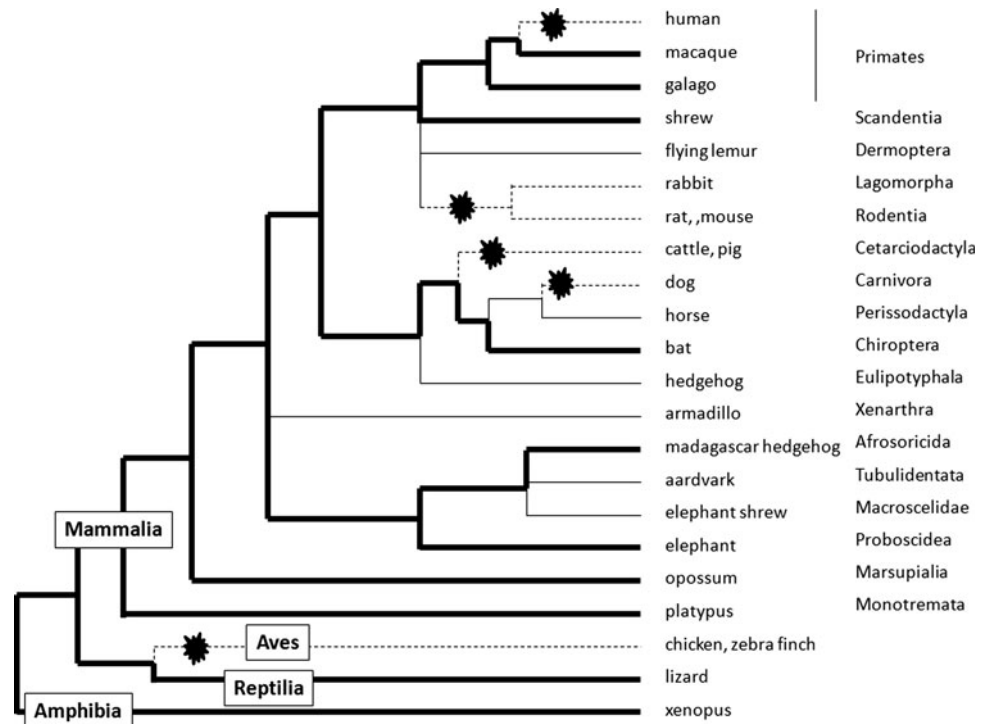
Fig. 3 Xenobiotic responsive element (XRE) sequences in the upper regions of the predicted CYP1D genes. The upper regions of the CYP1D gene-coding regions were investigated. A few XRE sites were identified in the upper region of CYP1D1, unlike other CYP1 genes

1.0 and positive selection was detected. Considering this result, the many mutations in the *X. tropicalis* CYP1D1 gene sequence suggest that positive selection has occurred.

In general, CYP1A genes are induced by AhR ligands, and CYP1D was also expected to be regulated by AhR due to the similarity to CYP1A. However, in fish, previous reports indicated that there were few XRE regions upstream of CYP1D1 and that CYP1D1 was not induced by AhR ligands (Goldstone et al. 2009; Jönsson et al. 2009; Zanette et al. 2009). There are markedly fewer XRE regions upstream of mammalian CYP1D1 genes than upstream of mammalian CYP1A1 genes. This result suggested that the mammalian CYP1D1 regulation mechanism is different from that of CYP1As and could be induced via other signal cascades, as is the case for the fish CYP1D1 gene. CYP1A, as well as CYP1B1 and CYP1C1 genes, is known to be induced via AhR. This result indicated that the ancestral CYP1 genes were regulated by AhR and that following divergence from CYP1A, CYP1D became independent of AhR regulation. In this study we focused on the XRE, but also considered other regulatory elements such as XRE II (Sogawa et al. 2004). A specific element was not found upstream of the mammalian CYP1D1 gene. Further study is needed to identify the transcriptional factor that regulates mainly mammalian CYP1D1.

CYP1 genes induced via AhR play an important role in the xenobiotic metabolism of PAH and food components such as carotenoids and flavonoids. Understanding the evolution of the relationship between mammalian CYP1s and AhR is important in predicting and evaluating the ability of animals to adapt to the risk of exogenous chemicals. It is still unknown why AhR has the ability to induce CYP1s. CYP1D1 is a unique gene because it is not induced via AhR. Characterizing the CYP1D subfamily of genes will help us understand CYP1 regulation and evolution, and further studies focusing on CYP1D are required.

Fig. 4 Analysis of CYP1D1 evolution. The *bold line* indicates functional CYP1D1 gene evolution. The *black points* and *broken line* indicate that CYP1D1 became a pseudogene. *Normal lines* are still unknown



Acknowledgments This study was supported in part by a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan which was awarded to M. Ishizuka (No. 19671001).

References

- Cañestro C, Catchen JM, Rodríguez-Marí A, Yokoi H, Postlethwait JH (2009) Consequences of lineage-specific gene loss on functional evolution of surviving paralogs: ALDH1A and retinoic acid signaling in vertebrate genomes. *PLoS Genet* 5:e1000496
- Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214
- Fujita Y, Ohi H, Murayama N, Saguchi KI, Higuchi S (1999) Molecular cloning and sequence analysis of cDNAs coding for 3-methylcholanthrene-inducible cytochromes P450 in *Xenopus laevis* liver. *Arch Biochem Biophys* 371(1):24–28
- Godard CA, Goldstone JV, Said MR, Dickerson RL, Woodin BR et al (2005) The new vertebrate CYP1C family: Cloning of new subfamily members and phylogenetic analysis. *Biochem Biophys Res Commun* 331(4):1016–1024
- Goldstone HM, Stegeman JJ (2006) A revised evolutionary history of the CYP1A subfamily: gene duplication, gene conversion, and positive selection. *J Mol Evol* 62:708–717
- Goldstone JV, Stegeman JJ (2008) Gene structure of the novel cytochrome P4501D1 genes in stickleback (*Gasterosteus aculeatus*) and medaka (*Oryzias latipes*). *Mar Environ Res* 66(1):19–20
- Goldstone JV, Goldstone HMH, Morrison AM, Tarrant A, Kern SE et al (2007) Cytochrome P450 1 genes in early deuterostomes (tunicates and sea urchins) and vertebrates (chicken and frog): origin and diversification of the CYP1 gene family. *Mol Evol* 24(12):2619–2631
- Goldstone JV, Jönsson ME, Behrendt L, Woodin BR, Jenny MJ et al (2009) Cytochrome P450 1D1: a novel CYP1A-related gene that is not transcriptionally activated by PCB126 or TCDD. *Arch Biochem Biophys* 482(1–2):7–16
- Gotoh O (1992) Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *J Biol Chem* 267:83–90
- Heinemeyer T, Wingender E, Reuter I, Hermjakob H, Kel AE et al (1998) Databases on transcriptional regulation: TRANSFAC, TRRD, and COMPEL. *Nucleic Acids Res* 26:364–370
- Jönsson ME, Orrego R, Woodin BR, Goldstone JV, Stegeman JJ (2007) Basal and 3,3',4,4',5-pentachlorobiphenyl-induced expression of cytochrome P450 1A, 1B and 1C genes in zebrafish. *Toxicol Appl Pharmacol* 221(1):29–41
- Jönsson ME, Franks DG, Woodin BR, Jenny MJ, Garrick RA et al (2009) The tryptophan photoproduct 6-formylindolo[3,2-b]carbazole (FICZ) binds multiple AHRs and induces multiple CYP1 genes via AHR2 in zebrafish. *Chem Biol Interact* 181(3):447–454
- Keane TM, Creevey CJ, Pentony MM, Naughton TJ, McInerney JO (2006) Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol Biol* 6:29
- Nebert DW, Dalton TP, Okey AB, Gonzalez FJ (2004) Role of aryl hydrocarbon receptor-mediated induction of the CYP1 enzymes in environmental toxicity and cancer. *J Biol Chem* 279(23):23847–23850
- Rice P, Longden I, Bleasby A (2000) EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet* 16(6):276–277
- Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19(12):1572–1574
- Sogawa K, Numayama-Tsuruta K, Takahashi T, Matsushita N, Miura C et al (2004) A novel induction mechanism of the rat CYP1A2

- gene mediated by Ah receptor-Arnt heterodimer. *Biochem Biophys Res Commun* 318:746–755
- Sutter TR, Tang YM, Hayes CL, Wo YY, Jabs EW et al (1994) 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* 269(18):13092–13099
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24(8):1596–1599
- Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24(8):1586–1591
- Zanette J, Jenny MJ, Goldstone JV, Woodin BR, Watka LA et al (2009) New cytochrome P450 1B1, 1C2, and 1D1 genes in killifish *Fundulus heteroclitus*: basal expression and response of five killifish CYP1s to the AHR agonist PCB126. *Aquat Toxicol* 93(4):234–243