Cytochrome P450 1 Genes in Early Deuterostomes (Tunicates and Sea Urchins) and Vertebrates (Chicken and Frog): Origin and Diversification of the CYP1 Gene Family

Jared V. Goldstone, Heather M. H. Goldstone,† Ann M. Morrison,* Ann Tarrant,* Suzanne E. Kern,* Bruce R. Woodin,* and John J. Stegeman**

*Department of Biology, Woods Hole Oceanographic Institution; and †Josephine Bay Paul Center, Marine Biological Laboratory, Woods Hole

Cytochrome P450 family 1 (CYP1) proteins are important in a large number of toxicological processes. *CYP1A* and *CYP1B* genes are well known in mammals, but the evolutionary history of the *CYP1* family as a whole is obscure; that history may provide insight into endogenous functions of CYP1 enzymes. Here, we identify *CYP1*-like genes in early deuterostomes (tunicates and echinoderms), and several new *CYP1* genes in vertebrates (chicken, *Gallus gallus* and frog, *Xenopus tropicalis*). Profile hidden Markov models (HMMs) generated from vertebrate CYP1A and CYP1B protein sequences were used to identify 5 potential *CYP1* homologs in the tunicate *Ciona intestinalis* genome. The *C. intestinalis* genes were cloned and sequenced, confirming the predicted sequences. Orthologs of 4 of these genes were found in the *Ciona savignyi* genome. Bayesian phylogenetic analyses group the tunicate genes in the *CYP1* family, provisionally in 2 new subfamilies, *CYP1E* and *CYP1F*, which fall in the CYP1A and CYP1B/1C clades. Bayesian and maximum likelihood analyses predict functional divergence between the tunicate and vertebrate CYP1s, and regions within CYP substrate recognition sites were found to differ significantly in position-specific substitution rates between tunicates and vertebrates. Subsequently, 10 *CYP1*-like genes were found in the echinoderm *Strongylocentrotus purpuratus* (sea urchin) genome. Several of the tunicate and echinoderm *CYP1*-like genes are expressed during development. Canonical xenobiotic response elements are present in the upstream genomic sequences of most tunicate and sea urchin *CYP1s*, and both groups are predicted to possess an aryl hydrocarbon receptor (AHR), suggesting possible regulatory linkage of AHR and these *CYPs*. The *CYP1* family has undergone multiple rounds of gene duplication followed by functional divergence, with at least one gene lost in mammals. This study provides new insight into the origin and evolution of CYP1 genes.

Introduction

Cytochromes P450 (CYPs) comprise a large and ancient superfamily of genes encoding heme-thiolate monooxygenase enzymes, which function in a great array of biological processes in plants, animals, and microbes. Substrate specificity of individual CYP enzymes ranges from the highly specific (e.g., biosynthetic enzymes) to the exceptionally diverse (the xenobiotic-oxidizing enzymes). Cytochrome P450 family 1 (CYP1) enzymes are of broad biomedical interest for their roles in toxicological and physiological processes (Ioannides and Lewis 2004; Nebert and Dalton 2006). Collectively, vertebrate CYP1 enzymes catalyze the oxidation of many xenobiotics including environmental chemicals and many drugs. Metabolism by CYP1s can result in detoxification but also bioactivation (e.g., of procarcinogens benzo[a]pyrene and aflatoxin B_1) (Conney 1982; Crespi et al. 1990; Shimada and Guengerich 2006). CYP1s also oxidize a variety of endogenous substrates, including uroporphyrin (Lambrecht et al. 1992), estradiol (Spink et al. 1992), retinoids (Raner et al. 1996), and arachidonic acid, resulting in formation of eicosanoid regulatory molecules (Nebert and Russell 2002).

Vertebrate *CYP1* genes occur in 2 major subclades, the *CYP1As* and the *CYP1B*/*1Cs*. *CYP1As* occur in all vertebrate groups examined. Mammals have 2 *CYP1A* paralogs, *CYP1A1* and *CYP1A2*. The avian genes *CYP1A4* and *CYP1A5* recently were shown to be orthologs of mammalian *CYP1A1* and *CYP1A2*, respectively, a phylogenetic relationship that had been obscured by gene conversion (Goldstone and Stegeman 2006). The frog *Xenopus laevis* also has 2 closely related *CYP1As* (Fujita et al. 1999), a duplication possibly reflecting tetraploidy in *X. laevis*. Most fish have one *CYP1A* gene, although there are multiple *CYP1As* in some lines that have recently undergone tetraplodization (Gooneratne et al. 1997). Vertebrate CYP1A enzymes generally are inducible, via the ligand-activated aryl hydrocarbon receptor (AHR) (Hahn et al. 1998).

The other *CYP1* subclade consists of the *CYP1B* and *CYP1C* subfamilies. Mammals and fish possess a single *CYP1B1* (Sutter et al. 1994; Leaver and George 2000). The *CYP1C* subfamily was identified recently in fishes and is paralogous to the *CYP1Bs* (Godard et al. 2005; Itakura et al. 2005). To date, no *CYP1C* has been found in any mammalian genome, suggesting that this subfamily was lost during mammalian evolution (Godard et al. 2005).

Elucidating the evolutionary history of the *CYP1* family may provide insight into the origins of physiological and toxicological functions of CYP1 enzymes. Some aspects of CYP1 evolution likely are driven by xenobiotic exposure, and evolutionary processes forming this family may be inferred by analyzing this CYP family in detail. To date, *CYP1* genes have been identified only in vertebrates and do not occur in protostomes. In this study, we address the emergence of the *CYP1* gene family in prevertebrate deuterostomes.

We sought *CYP1* genes in genomes of the ascidian tunicates *Ciona intestinalis* and *Ciona savignyi* and an echinoderm, the purple sea urchin *Strongylocentrotus purpuratus* (Dehal et al. 2002; Vinson et al. 2005; Sodergren et al. 2006). The tunicate lineage is believed to be the most basal among the chordates, diverging prior to the cephalochordates and the vertebrates, and the echinoderms are perhaps the earliest diverging

Key words: cytochrome P450, CYP1A, CYP1B, CYP1C, substrate recognition site, functional selection.

E-mail: jgoldstone@whoi.edu.

Mol. Biol. Evol. 24(12):2619–2631. 2007
doi:10.1093/molbev/msm200
Advance Access publication Month 00, 2007

deuterostomes. Our results show the presence of both *CYP1A*-like and *CYP1B*/*1C*-like genes in the tunicates and a suite of *CYP1*-like genes in the sea urchin. The presence of multiple *CYP1* genes in early deuterostomes raises questions regarding the functional significance of *CYP1* gene diversity as well as the nomenclature for evolutionarily distant *CYP* lineages. Although the emphasis is on *CYP1s* in prevertebrate deuterostomes, we also consider *CYP1* occurrence and loss in the vertebrates, and the results enhance our understanding of the history of the *CYP1* subfamilies in vertebrates.

Materials and Methods

Profile HMMs

The *C. intestinalis* genome (release 1.0) (Dehal et al. 2002) predicted protein database was searched using hidden Markov models (HMMs) of CYP1s constructed using Hmmer 2.2g and Hmmer 2.3 (Eddy 1998). The HMM was constructed with 28 CYP1As, 4 CYP1Bs, and 2 CYP1Cs, including human, mouse, rat, and fish sequences. Both global and local multidomain HMMs were constructed and used to search the predicted protein database. Predicted proteins were aligned with known CYP1s using ClustalW (v1.82; EMBL) and GCG (v. 10.3; Accelrys, San Diego, CA). Examination of the *C. intestinalis* genome assembly using Blast searching confirmed the CYP1 protein sequence predictions. CYP1 sequences were likewise obtained from the respective genomes of *C. savignyi* (ascidian, v1.0), *Takifugu rubripes* (torafugu, v.3), *Tetraodon nigroviridis* (freshwater pufferfish, v3.0), *Danio rerio* (zebrafish, Zv5), *Gallus gallus* (chicken, v1.0), *Xenopus tropicalis* (clawed frog, v2), *S. purpuratus* (purple sea urchin, v2.1), and *Monodelphis domestica* (gray short-tailed opossum, v1.0) by a combination of HMM and Blast searches. Genome sequences were obtained from GenBank, Ensembl, and the Joint Genome Institute. Some preliminary gene predictions were performed using Genewise (Birney et al. 2004) and Genscan (Burge and Karlin 1998).

Phylogenetic Analyses

Multisequence alignments of protein translations were generated using ClustalX with the Gonnet series of protein weight matrices. The alignments were corrected by hand as necessary using GCG and MacClade. Alignments were subjected to 10 rounds of randomization and manual masking prior to phylogenetic analyses. Trees were rooted using known CYP2 sequences (see supplementary table S3 in Supplementary Material online for accession numbers).

Phylogenetic relationships were investigated using Bayesian techniques as implemented in the computer program MrBayes (v 3.1.1; Ronquist and Huelsenbeck 2003). MrBayes estimates posterior probabilities using Metropolis–Hastings coupled Monte Carlo Markov chains (MC^3). We performed MC^3 estimates with uninformative prior probabilities using the model of Whelan and Goldman (2001; WAG) of amino acid substitution and prior uniform gamma distributions approximated with 4 categories (WAG + Invariant + Gamma), as indicated by analysis with ProtTest (Abascal et al. 2005). Four incrementally heated, randomly seeded Markov chains were run for 10^7 generations, and topologies were sampled every 100th generation. Analysis of the MC^3 parameter output using Bayesian Output Analysis (BOA; v1.71; Smith 2003) indicated that this degree of sampling was sufficient to avoid significant sampling autocorrelation. In order to confirm the MC^3 results, 4 independent, randomly seeded analyses of the data set were performed with identical results. The MC^3 burn-in values were calculated using BOA and conservatively set at 200,000 generations based on convergence statistics (Raftery and Lewis 1992). Posterior probabilities of topologies and clades were estimated from the sampled topologies after removal of the initial MC^3 burn-in. Bayes factors are defined as the ratio of the posterior to the prior odds for the 2 hypotheses in question (Kass and Raftery 1995; Huelsenbeck and Imennov 2002; Suchard et al. 2005). In testing of the monophyly of certain clades within the same tree, the model prior odds are the same, and thus, the Bayes factor is computed as the ratio of the frequencies of the 2 hypotheses in the filtered MC^3 run, corrected for the prior number of possible trees. Following Suchard et al. (2005), we considered the cluster of taxa for which we are testing the hypothesis of monophyly to be rooted within the overall unrooted phylogenetic tree.

Sequence Analysis

Prediction of both overall and site-specific rates of evolutionary divergence of amino acid sequences was performed using DIVERGE (v1.04, Gu and Vander Velden 2002). Masked regions were removed from the alignment prior to the DIVERGE analysis. The input tree used to assign clade groupings was the consensus tree determined in the Bayesian phylogenetic analysis with several polytomies altered to conform to the tree displaying the highest Bayesian posterior probability. Analysis of the site-specific rates of amino acid substitution was also performed using the likelihood method of Knudsen and Miyamoto (2001). Pairwise relative rates tests to examine relative rates of substitution were performed using the program HYPHY (Pond et al. 2005).

Upstream flanking regions up to 2 kb in length, adjacent (5′) to the predicted translational start sites of each gene were searched for known transcription factor recognition sequences housed in the TRANSFAC sites database. Specific pattern searches for degenerate versions of the consensus xenobiotic receptor element (XRE) were performed using GCG.

Protein Structure Calculations

Secondary structure predictions for the major indels in 2 pairs of *Ciona* sequences (*C. intestinalis* CYP1F4 and its *C. savignyi* ortholog; *C. intestinalis* CYP1F1 and its *C. savignyi* ortholog) were done with PredictProtein (Rost and Liu 2003), JPRED (Cuff and Barton 1999), and PHD (Rost 1996). Three-dimensional structure prediction was done with SWISS-MODEL (Peitsch 1995; Guex and Peitsch 1997; Schwede et al. 2003) after alignment of the respective Ciona predicted amino acid sequence to CYP2C5 (PDB: 1DT6) using ClustalX.

Cloning of *Ciona CYP1* Gene cDNAs

Adult *C. intestinalis* were collected from floating docks in Eel Pond (Woods Hole, MA). *Ciona intestinalis* individuals were separated from other tunicates and fouling organisms using a razor blade and maintained in aquaria with flowing seawater at ambient temperatures. Whole adult *C. intestinalis* were frozen in liquid nitrogen and pulverized using a mortar and pestle. Genomic DNA was extracted from frozen pulverized tissue using NucleoSpin columns (BD Biosciences, San Jose, CA). Total RNA was extracted from pulverized tissue using RNA STAT-60 (Tel-Test, Inc, Friendswood, TX). cDNA was synthesized from total RNA using Powerscript reverse transcriptase (BD Biosciences) with random hexamers or oligo dT.

Specific oligonucleotide primers (Sigma Genosys) were designed for each predicted *CYP1*-like gene based on the genome assembly. Polymerase chain reactions (PCRs) were performed using the Advantage 2 polymerase kit (BD Biosciences); 5% dimethyl sulfoxide was added to all PCRs. Full length CI0100138492 (CYP1F3, GenBank: EU139258), CI0100131189 (CYP1E1, GenBank: EU139 256), CI0100143263 (CYP1F1, GenBank: EU139257), and a 370-bp fragment of CI0100136792 (CYP1F2, GenBank: EU155006) were amplified from cDNA derived from total mRNA, and a 284-bp fragment of CI0100132188 (CYP1F4) was amplified from genomic DNA using PCR primers and conditions listed in supplementary table S1 (Supplementary Material online). All PCR products were subcloned into the pGEM-T Easy vector (Promega, Madison, WI). Both strands from multiple clones of each PCR product were sequenced. DNA sequences were analyzed, assembled, and translated using GCG and Sequencher (Gene Codes Corporation, Ann Arbor, MI) sequence analysis software.

Developmental Expression

Spawning of C. *intestinalis* was initiated by light induction (Cirino et al. 2002). Floating glass and plastic petri dishes served as substrates for the settling larvae. Upon hatching, larvae swim upward to settle and metamorphose. This behavior was used to separate the larvae from the embryos because development was asynchronous. Water was decanted into 15-ml centrifuge tubes and centrifuged for 5 min to pellet the larvae, which were then frozen in liquid nitrogen. After 20 days, juvenile *Ciona* were scraped from glass surfaces using a razor blade, centrifuged to allow aspiration of excess water, and frozen in liquid nitrogen

Poly(A)$^+$ RNA (1 μg) was reverse transcribed with random primers (Gene-Amp RNA–PCR kit, PerkinElmer), and an equal aliquot of cDNA was used in each of 3 PCRs with AmpliTaq Gold DNA polymerase (PerkinElmer, Waltham, MA). Specific primers were designed for *C.intestinalis* β-actin and for each of the *C. intestinalis CYP1* genes to be used in semiquantitative PCR. The linear range for the PCR was determined by varying the number of cycles from 20 to 35 with 3-cycle increments and using 2, 4, 6, and 8 μl of template cDNA (data not shown). Subsequent reactions used 5 μl of template cDNA and 28 cycles. The cycling conditions were 95 °C/10 min (94 °C/15 s, 60 °C/30 s) for 28 cycles and 72 °C/7 min. Under these conditions, the amount of PCR products amplified from cDNA was linearly related to cycle number and amount of template. Ten-microliter aliquots of each reaction (volume verified to be in the linear range for imaging) were subjected to agarose gel electrophoresis and subsequent ethidium bromide staining. The integrated density of each amplified fragment was determined from the digital image. The intensity of each *CYP1* gene fragment was normalized to β-actin intensity.

Results

Identification and Cloning of Tunicate *CYP1* Genes

The database of *C. intestinalis* predicted peptide sequences (v2.0) was searched for CYP1s using CYP1A and CYP1B profile HMMs, constructed with sequences from *Homo sapiens*, *Mus musculus*, *G. gallus*, *D. rerio*, *Stenotomus chrysops*, *X. laevis*, and *Pleuronectes platessa* (plaice). Gene predictions for potential *CYP1s* were refined by comparing known vertebrate *CYP1* coding sequences directly with those identified in the *C. intestinalis* genome. Distance-based hierarchical clustering of the top 10 hits from each HMM search showed that the majority of the matches clustered with *CYP2s*; those sequences are not considered here. Predicted protein sequences for the 5 remaining genes, CI0100131189, CI0100143263, CI0100138492, CI0100136792, and CI0100132188 were subjected to Bayesian phylogenetic analysis (see below) and all 5 were found to cluster with the CYP1s. The new *Ciona* genes have been provisionally named as *CYP1E1* (CI0100131189) and *CYP1F1–CYP1F4* (CI0100143263, CI0100136792, CI0100138492, and CI0100132188, respectively).

cDNAs corresponding to predicted *C. intestinalis* genes *CYP1E1*, *CYP1F1*, *CYP1F2*, and *CYP1F3* were obtained via reverse transcriptase (RT)-PCR and sequenced to verify predicted coding sequences and intron–exon boundaries. *CYP1F4* could not be obtained by RT-PCR and was cloned from a genomic PCR product. Based on unambiguously aligned positions, the cloned *C. intestinalis* deduced amino acid sequences share 29.4 ± 4.0% identity with each other and 33.5 ± 3.2% identity (mean ± standard deviation) with known vertebrate CYP1s (see table 1). These identities rise as high as 41.2% (mean 39.8%) between CYP1F1 and fish CYP1B1 amino acid sequences (Supplementary table S7, Supplementary Material online).

The 5 CYP1-like sequences in *C. intestinalis* were used to search the *C. savignyi* genome using Blast. Four predicted CYP1 coding sequences were assembled manually from the Blast results. These 4 *C. savignyi* predicted proteins exhibit 58–81% amino acid identity with their corresponding *C. intestinalis* homologs. For our purposes, these sequences are termed *C. savignyi CYP1E1*, *CYP1F1–CYP1F3*, respectively. No homolog of *CYP1F4* could be identified in the current assembly of the *C. savignyi* genome.

In Vivo Expression of *Ciona CYP1* Genes

Semiquantitative RT-PCR confirmed expression of *CYP1E1*, *CYP1F1*, and *CYP1F2* in various life stages of *C. intestinalis*. *CYP1E1* and *CYP1F1* were strongly

Table 1
Mean Percent Amino Acid Identity ± Standard Deviation (on a masked unambiguously aligned basis) between Selected *CYP1* Genes, Including the Cloned Tunicate *Ciona intestinalis* CYP1E and CYP1F and Predicted Sea Urchin *Strongylocentrotus purpuratus CYP1*-like Genes

	CYP2 (%)	CYP1A (%)	CYP1B (%)	CYP1C (%)	Tunicate 1E (%)	Tunicate 1F (%)	Sea Urchin (%)
CYP2	46.0 ± 4.1	32.6 ± 1.6	32.7 ± 1.5	34.1 ± 1.5	25.9 ± 1.3	27.0 ± 2.0	31.5 ± 1.3
CYP1A		71.0 ± 11.2	45.4 ± 2.3	45.4 ± 1.5	31.6 ± 1.1	34.1 ± 3.0	38.6 ± 3.2
CYP1B			67.8 ± 11.0	56.8 ± 2.7	30.3 ± 0.9	35.0 ± 3.8	37.6 ± 2.6
CYP1C				82.3 ± 3.6	28.7 ± 0.7	34.1 ± 3.9	37.0 ± 3.0
Tunicate CYP1E					60.4 ± 0	25.5 ± 1.3	27.6 ± 1.6
Tunicate CYP1F						37.2 ± 14.2	30.4 ± 2.8
Sea urchin							57.8 ± 19.2

expressed in larvae (at 18 h after fertilization), in 20-day-old juveniles, and in adults (fig. 1), whereas gene *CYP1F2* was expressed weakly in all 3 stages, but more strongly in adult tissues, compared with the earlier stages. Searching the Ghost EST database (Satou et al. 2005) also showed that *CYP1E1* is expressed in blood cells, gonads, and digestive glands. In addition, the Ghost EST database showed that *CYP1F3*, which we had not examined, is very strongly expressed (5,059 of 23,897 total ESTs) in stage 1 juveniles. In contrast to the other genes, we could not find any good evidence for expression of *CYP1F4* in the Ghost database, suggesting that this gene is expressed at very low levels, if at all (see also supplementary table S4, Supplementary Material online). The lack of an obvious *C. savignyi* ortholog and of clear expression data for *CYP1F4* suggests that it could be nonfunctional, which would have to be confirmed.

CYP1 Genes in *S. purpuratus*

With evidence for *CYP1* genes in tunicates, we extended the search for CYP1-related sequences to the earlier diverging *Echinodermata* (represented by *S. purpuratus*; Sodergren et al. 2006). Ten putative CYP1 homologs were identified in the genome of *S. purpuratus* using a combination of Blast and profile HMM searches (Goldstone et al. 2006). Genscan and FGENESH+ were used to refine these gene predictions. *S. purpuratus* CYP1-like sequences exhibit amino acid identities ranging from 29.3 % to 45.3% (low: *G. gallus* CYP1B, high: *Anguilla anguilla* CYP1A; average of 38.3 ± 3.2%) with various vertebrate CYP1s (see also table 1; supplementary table S7, Supplementary Material online).

The majority of these sea urchin genes are single-exon genes. However, 2 (SPU_019883 and SPU_017582) are multi-exon genes with 9 exons each. Several of the single-exon predicted sea urchin genes are syntenic: SPU_010719, SPU_010720, and SPU_010721 are located together on a scaffold, as are SPU_007404 and SPU_07406 (Goldstone et al. 2006). Based on sea urchin microarray data, we previously reported that the *CYP1*-like genes SPU_007404, SPU_07406, SPU_010720, SPU_019883, and SPU_017582 are expressed during development (Goldstone et al. 2006). Searches of EST libraries support these results (data not shown). EST data indicate that SPU_07406 appears to be expressed in primary mesenchyme cells, and SPU_019883 and SPU_017582 are expressed throughout sea urchin development. Expression of the other sea urchin *CYP1*-like genes is unknown.

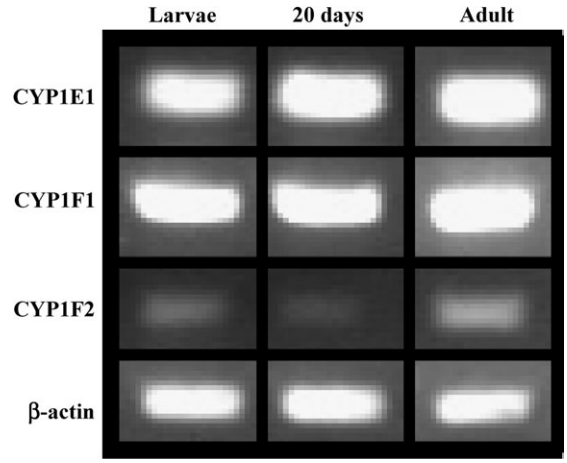

FIG. 1.—Semiquantitiative developmental expression of CYP1 genes in *Ciona intestinalis*. Shown are images of ethidium bromide–stained agarose gels of RT-PCR products for *CYP1E1*, *CYP1F1*, and *CYP1F2*, as well as β-actin in 18-h postfertilization swimming larvae, 20-day-old juveniles, and adult *C. intestinalis*. PCRs were performed under conditions in which formation of product was linearly related to cycle number.

New Avian and Amphibian *CYP1* Genes

Using Blast and Genewise searches, several new *CYP1* genes were identified in the genomes of the chicken *G. gallus* and the frog *X. tropicalis*. In both species, there was one sequence identified that resembled the recently described fish *CYP1Cs* and one that appears to be a CYP1B1 ortholog. EST evidence indicates that these predicted avian and amphibian *CYP1Bs* and *CYP1C*s are expressed at least at the mRNA level (see supplementary table 2, Supplementary Material online). In addition to the *CYP1B* and *CYP1C* genes, a single *CYP1A* gene was identified in the genome of *X. tropicalis*.

Bayesian Inference of Phylogeny

The newly predicted CYP1-like sequences from sea urchin, tunicates, and vertebrates were aligned with all available complete or nearly complete vertebrate CYP1 peptide sequences, totaling 110 at the time of analysis. Five

CYP2 sequences from vertebrates provided an outgroup to the CYP1s. (A complete list of gene names and GenBank accession numbers is provided in supplementary table S3 (Supplementary Material online.) We used the protein substitution matrices of WAG (Whelan and Goldman 2001) in phylogenetic analyses because fish and mammalian CYP1B1s were determined to have amino acid compositions significantly different from the commonly used JTT model (Jones et al. 1992).

As suggested above, these Bayesian phylogenetic analyses show that the newly identified *G. gallus* and *X. tropicalis* CYP1s share specific orthologous relationships with vertebrate CYP1 subfamilies (fig. 2; supplementary fig. S1, Supplementary Material online). Note that the *X. tropicalis* and *X. laevis* CYP1As cluster with the fish CYP1As rather than with the avian or mammalian CYP1As.

Both Bayesian and maximum likelihood phylogenetic analyses show that the predicted tunicate sequences fall within the CYP1 family (fig. 2). *Ciona intestinalis CYP1E1* and its *C. savignyi* ortholog fall within the *CYP1A* clade, whereas the remaining 7 tunicate genes are in the *CYP1B*/*C* clade. When tested against alternative topologies using Bayes factor tests, these groupings received decisive support (Bayes factors were 71.6 and 66.2) for the inclusion of tunicate genes in the *CYP1B*/*1C* clade and the *CYP1A* clade, respectively. Notably, the new tunicate CYP1-like sequences share a lower absolute identity with vertebrate CYP1s than do the *Strongylocentrotus* genes (table 1), yet the topology recovered by our phylogenetic analyses reflects the known species phylogeny, placing the echinoderm genes more distant to the vertebrate genes. Thus, the new *Ciona* genes have been assigned to new subfamilies within the CYP1 family and the sea urchin genes remain "*CYP1*-like." Formal nomenclature of the sea urchin genes will require analyses with additional taxa.

Although the *Ciona* proteins clearly are related to the 2 major CYP1 subclades, they exhibited unusually long branch lengths. We performed pairwise relative rates tests to examine whether the long branches were the result of significantly increased rates of substitution in the tunicate CYP1 sequences, compared with CYP1s in other species. Maximum likelihood estimates of all possible 3-taxa trees with a fixed outgroup (*D. rerio* CYP2K; 4467 comparisons) were generated using the program HYPHY (Pond et al. 2005). We compared the trees obtained with fixed branch lengths (equal substitution rates) with those with unconstrained branch lengths using a likelihood ratio test (LRT) to determine the significance of the increased substitution rates (fig. 3). Most rate comparisons (62%) involving *Ciona* showed substitution rates significantly greater than those of the CYP1 protein data set as a whole ($P < 0.01$). In sharp contrast, only 1% of pairwise comparisons of CYP1s of any other species (including <0.1% of comparisons of the sea urchin genes) showed significantly elevated substitution rates relative to the CYP1 data set as a whole ($P < 0.01$).

Functional Constraint and Divergence

To begin to address possible functional evolution of CYP1 enzymes, we performed an analysis of the amino acid alignment in the context of the hypothesized phylogenetic tree using DIVERGE (Gu and Vander Velden 2002). DIVERGE detects differing site-specific rates of amino acid substitution following gene duplication events, which imply altered functional constraints, by comparing site-specific evolutionary rates in amino acid sequences among subclades within a phylogenetic tree (Gaucher et al. 2002). DIVERGE analyses were based on pairwise comparison among 6 CYP1 subclades: fish CYP1As, mammalian CYP1A1s, mammalian CYP1A2s, the CYP1B/1C subclade, the *Ciona* CYP1F sequences, and the outgroup CYP2s (table 2). The coefficient of evolutionary functional divergence (θ) between the tunicate CYP1F sequences and the fish CYP1A clade is very large but between tunicate CYP1Fs and vertebrate CYP1Bs or CYP1Cs is low (table 2), suggesting that some functional conservation among the CYP1Bs and CYP1Cs occurs in the CYP1Fs as well. The divergence coefficient between the fish CYP1A and mammalian CYP1A1 and CYP1A2 subclades is surprisingly large in this analysis.

The *Ciona* CYP1A-like sequences (CYP1E1) were subjected to a maximum likelihood site-specific divergence analysis, which requires fewer sequences than DIVERGE to calculate site-specific functional divergence (Knudsen and Miyamoto 2001). Site-specific analysis of θ revealed a nonrandom distribution of increased and decreased substitution rates along the alignment between the tunicate CYP1E1 and selected vertebrate CYP1As, including residues within the putative substrate recognition sites (SRSs) (Gotoh 1992). In particular, sites within SRS 1, 4, and 5 exhibit significantly decreased substitution rates (i.e., show conservation) between *Ciona* and vertebrate CYP1A subclades and also between mammalian CYP1A1 and fish CYP1A proteins. However, there are residue differences in the SRSs between the CYP1Es and the CYP1As that could affect substrate specificity, based on significant differences in amino acid properties at the sites. Notable substitutions within SRS4 include a valine for a phenylalanine (V334 in tunicate CYP1E1, F319 in human CYP1A1) and a methionine for a valine (M337/V322, respectively) within the I-helix. Some of these sites have been examined using site-specific mutation and homology modeling (Liu et al. 2003, 2004; Prasad et al. 2007).

In contrast, both SRS 3 and 6 exhibit significantly higher substitution rates in the *Ciona* CYP1A-like sequences relative to vertebrate CYP1As, suggesting increased functional divergence. There are substitutions at aligned positions (notably *Ciona* L233, vertebrate F224) within a region of SRS2 that appears to be significant for differences in substrate binding between mammalian CYP1A1 and fish CYP1A (Prasad et al. 2007). A complete list of CYP1A residues identified as having significantly altered rates of evolutionary divergence is in the supplementary table S5 (Supplementary Material online).

Despite the evidence for increased substitution rates in most of the SRSs, both Bayesian and maximum likelihood analyses of the SRS regions alone place the tunicate genes within the CYP1 family, in accordance with the phylogeny estimated using the entire data set (supplementary fig. S2, Supplementary Material online).

FIG. 2.—Bayesian phylogenetic tree of CYP1 family protein sequences using several vertebrate CYP2 sequences as an outgroup. The tunicate *CYP1* genes are divided between the 2 known *CYP1* subclades. Values at branch points are the posterior probabilities of the branch observed at 10^7 generations of MC^3. Bifurcations with 100% support are indicated with an asterisk. Sequences discussed in the text are in bold.

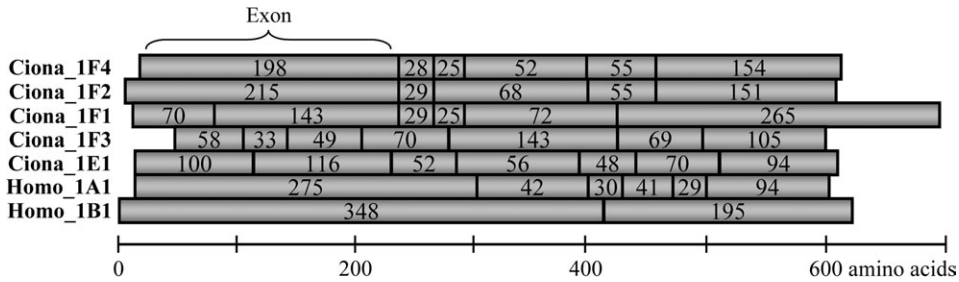

FIG. 3.—Gene structure of Ciona *CYP1* genes compared with human CYP1A1 and CYP1B1. The numbers in gray boxes indicate number of amino acids in individual exons. No exon–exon boundaries are shared between mammalian CYP1A1 genes and *Ciona* CYP1E genes or between mammalian CYP1B1 genes and *Ciona* CYP1Fs. Vertebrate CYP1Cs are single-exon genes (data not shown).

Tunicate and Echinoderm *CYP1* Protein and Gene Structure

In addition to a high level of divergence at the primary sequence level, several tunicate CYP1s contain insertions that could alter secondary and tertiary structure in ways affecting function. CYP1F4 includes a 53 amino acid insertion at positions 347–404, between the H and I helices at a surface-exposed turn. This insertion was found in the predicted and the cloned *C. intestinalis* CYP1F3 and in the predicted *C. savignyi* ortholog. This insert showed no sequence similarity to any known gene. Secondary structure prediction algorithms JPRED and PHD calculate a coil for this region of CYP1F3, although 3-dimensional structure modeling using SWISS-MODEL/Gromos96 suggests that it may contain up to 3 short helices and a sheet (data not shown). Additional data are required to accurately determine the orientation of this region relative to the surrounding helices.

CYP1F1 in *C. intestinalis* and *C. savignyi* contains an extension at the C-terminal end of the deduced protein sequence. This was confirmed in the sequenced cDNA. The 607 amino acid deduced protein sequence is approximately 75 residues longer than the average CYP1 protein (~530 aa). Blast searching with this C-terminal extension produced no significant matches in GenBank. Structure predictions indicate that this additional portion of the protein contains significant secondary structure (data not shown). Three secondary structure algorithms (PHD, PROF, and JPRED) predict β sheets between positions 528–531, 539–541, and 571–573, and a 10-residue α-helix from 596 to 606.

The general gene structure of *Ciona CYP1s* is different from other *CYP1s*. All known vertebrate *CYP1A*s have 7 exons generally with homologous exon boundaries, whereas *CYP1B* and *CYP1C* genes have 2 and 1 coding exons, respectively. The *Ciona CYP1* genes share several exon boundaries in common with one another but share only one exon–exon boundary with human *CYP1A* genes (fig. 4). Among the *Ciona CYP1* genes, the ones that share the largest number of homologous exon boundaries (4) are those that are phylogenetically most closely related (*CYP1F2* and *CYP1F4*) with progressively fewer shared boundaries apparent as phylogenetic distances increase. Neither of the 2 multi-exon sea urchin genes share exon boundaries with the vertebrate or tunicate *CYP1* genes.

Examination of the genomic arrangement of the tunicate *CYP1* genes disclosed no shared synteny with vertebrate *CYP1* genes. Interestingly, the *C. intestinalis* AHR homolog is located immediately adjacent to the *CYP1A*-like *CYP1E1* on chromosome 12. However, we did not find any shared synteny between the Ciona *AHR* and any vertebrate *AHR*.

Promoter Analysis

To explore whether these tunicate or echinoderm *CYP1*-like genes might have regulatory controls like the vertebrate *CYP1s*, we searched upstream promoter regions for consensus AHR binding motifs, TNGCGTG, known as the XRE (Sun et al. 2004). A large number of XREs were identified within 2 kb of the predicted translation start site of *C. intestinalis CYP1F1* (table 3). *Ciona intestinalis CYP1F3* and its *C. savignyi* ortholog exhibit 2 clusters of XREs similarly situated 3 and 7 kb upstream of the translation start site. Such clusters of XREs are present upstream of known AHR-inducible mammalian *CYP1A1* genes (Sun et al. 2004). The *C. intestinalis* gene *CYP1E1*, which is

Table 2
Coefficient of Functional Divergence (θ ± standard error, upper right diagonal) and LRT Values for Significance as Computed with DIVERGE

	CYP1A_fish	CYP1A1	CYP1A2	CYP1B	CYP1C	Tunicate 1B/1C
CYP1A_fish		0.60 ± 0.18	0.57 ± 0.10	0.69 ± 0.08	0.50 ± 0.06	0.88 ± 0.10
CYP1A1	48.1*		0.18 ± 0.14	0.41 ± 0.16	0.56 ± 0.11	0.14 ± 0.23
CYP1A2	32.6*	1.6		0.37 ± 0.14	0.46 ± 0.06	0.05 ± 0.21
CYP1B	75.4*	12.9*	6.8**		0.15 ± 0.10	0.13 ± 0.18
CYP1C	62.5*	28.7*	15.0*	2.3		0.28 ± 0.15
Tunicate	81.3*	0.4	0.1	0.5	3.6	

NOTE.—Significant at $*P < 0.01$ (χ^2) and $**P < 0.05$.

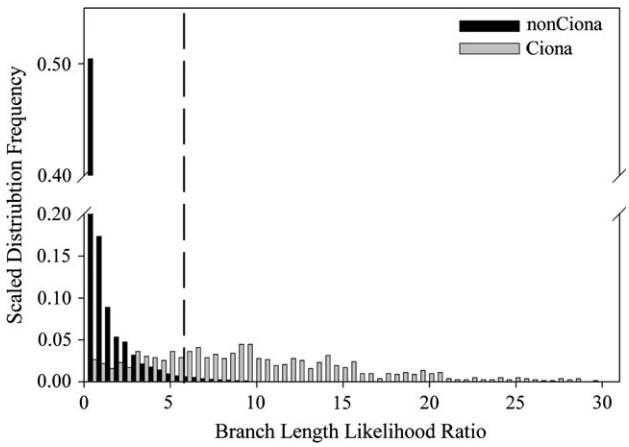

FIG. 4.—Increased substitution rates in the *Ciona CYP1* genes relative to all other *CYP1*-like genes. Maximum likelihood estimates of all possible 3-taxa trees with a fixed outgroup comparing the trees with fixed branch lengths (equal substitution rates) with those with unconstrained branch lengths using a LRT to determine the significance of the increased substitution rates. Likelihood ratios for *Ciona CYP1* genes versus all other *CYP1* genes are in grey, all other *CYP1* intercomparisons are in black. The dashed line indicates the $P < 0.01$ value relative to a χ^2 distribution with 1 degree of freedom.

more closely related to *CYP1A*, has 6 XREs located within 10 kb of the translational start site. As with *Ciona*, the sea urchin CYP1-like genes also had variable numbers of XREs in the upstream regions (table 3). Without detailed functional assays, it is not possible to predict whether XREs in these promoter regions are functional. It is known that not all XREs in CYP1 promoter regions are functional (Tsuchiya et al. 2003; ZeRuth and Pollenz 2005).

Discussion

In this study we used HMM searching and Bayesian phylogenetic analysis to examine genomes of early deuterostomes, including tunicates and echinoderms, for *CYP1*-like genes. We also examined selected vertebrate groups for additional *CYP1* sequences. The *CYPs* we identified in 2 tunicate genomes are phylogenetically *CYP1s*, with some that are *CYP1A*-like and others that are *CYP1B*/*CYP1C*-like. The genes found in the *S. purpuratus* genome, although not falling within specific *CYP1A* or *CYP1B*/*1C* clades, are nevertheless distinctly *CYP1*-like, ostensibly earlier diverging representatives of the line leading to the vertebrate *CYP1s*. The newly identified *CYP1s* in the chicken and frog include genes orthologous to the recently described *CYP1Cs*, extending this gene line forward to tetrapods.

Deuterostome Origin of *CYP1* Gene Family

Phylogenetic analysis clearly supports the assignment of the *Ciona CYP* genes we identified to the *CYP1* family. Based on these analyses, the tunicate gene provisionally named *CYP1E1* is CYP1A-like and the *CYP1F* genes are CYP1B/1C-like, forming a monophyletic gene family and extending both CYP1 subclades back in evolutionary time to the early chordate lineages. The tunicate and vertebrate *CYP1* genes appear to have evolved from a common ancestor present before the split of tunicates and vertebrates.

On average, *Ciona* CYP1s share just 26.1–41.2% (mean 34%) amino acid identity with known vertebrate CYP1s, in most cases lower than the nominal 40% cutoff for membership in a given CYP family (Nelson et al. 1993). This raises a question about whether these tunicate genes might be placed in a new CYP family. The tunicate CYP1s also exhibit features that contribute to classification difficulties, namely long unique insertions and poorly conserved exon boundaries. Although these features suggests that the tunicate genes could be placed in a new CYP family according to the current nomenclature guidelines, our phylogenetic analyses nevertheless support the assignment of the *Ciona* CYP genes to the CYP1 family. Interestingly, we found a higher shared sequence identity (average 38%) between echinoderm and vertebrate CYP1 sequences than between tunicate and vertebrate sequences, yet the phylogenetic analyses place the echinoderm CYP1s more basal in the tree than the tunicate genes. This suggests that nomenclature guidelines based on percent identity may need adjustments to accommodate evolutionary distances such as those represented here and the definition of a CYP family as monophyletic, as noted previously (Degtyarenko and Archakov 1993). There is need for a nomenclature with principles (Thornton and DeSalle 2000) that accommodate evolutionary distances greater than evident among vertebrates. Although molecular phylogeny clearly places the tunicate genes in the CYP1 family, whether the echinoderm genes are properly termed CYP1s requires resolution. Regardless, the echinoderm genes and the tunicate genes apparently are descended from an earlier *CYP1* antecedent.

The degree of divergence of the *Ciona* CYP1 sequences and, thus, long branch lengths in phylogenetic reconstructions may be explained by an overall accelerated rate of tunicate evolution. Elevated substitution rates are apparent not only in the *Ciona* CYP1s (fig. 3) but also in tunicate mitochondrial genes and in several nuclear gene families (Yokobori et al. 1999; Swalla 2001; Yokobori et al. 2003). This accelerated rate of substitution is corroborated by the average 4.6% substitution rate between the 2 haplotypes of *C. savignyi*, although there is extreme heterogeneity of these haplotype differences over the genome (Vinson et al. 2005).

Elevated substitution rates and divergent gene structure observed in *Ciona CYP1s* may not be a feature common to all *CYP* families in *Ciona*. The *CYP3* genes in *C. intestinalis* and *C. savignyi* share 34–44% amino acid identity and a marked similarity of gene structure with the vertebrate *CYP3s* (Verslyke et al. 2006). The greater divergence between the tunicate and vertebrate CYP1 sequences suggests that the CYP1s may be less functionally constrained than the CYP3s.

New Vertebrate *CYP1* Genes

We identified several new *CYP1* genes in the genomes of the frog, *X. tropicalis*, and the chicken, *G. gallus*. A single *CYP1A* gene was detected in the genome of *X. tropicalis*, which is a diploid frog (Hughes and Hughes 1993). In

Table 3
Number of Consensus AHR Binding Motifs or XREs

		Number of Consensus XREs within[a]				
Gene	Exon	1 kb	2 kb	5 kb	10 kb	Notes
Ciona intestinalis						
CYP1E1 (CI131189)	7	1	1	2	6	No clusters
CYP1F1 (CI143263)	6	2	3	4	6	
CYP1F2 (CI136792)	5	0	0	1	1	
CYP1F4 (CI132188)	6	0	—	—	—	<1.3 kb upstream sequence available
CYP1F3 (CI138492)	7	0	0	5	16	Clusters at −3 kb and −7 kb
Ciona savignyi						
CYP1E1	7	0	0	1	1	
CYP1F1	6	0	2	2	3	
CYP1F2	6	0	1	3	7	No clusters
CYP1F3	7	0	1	11	11	Cluster at −4 kb
Homo sapiens						
CYP1A1	7	3	5	11	15	Cluster at −1 kb
CYP1A2	7	1	1	5	6	
CYP1B1	3	4	6	8	9	
Danio rerio						
CYP1A	7	1	1	7	12	
CYP1B1	2	0	1	2	2	5′ untranslated region length unknown
CYP1C1	1	1	1	2	3	CYP1C1 and CYP1C2
CYP1C2	1	0	0	2	6	Intergenic region is 4 kb
Strongylocentrotus purpuratus						
SPU_003038	1	0	0	—	—	~1.95 kb upstream sequence available
SPU_014843	1	0	0	—	—	~2.2 kb upstream sequence available
SPU_010719	1	0	1	2	3	
SPU_010720	1	0	0	1	3	
SPU_010721	1	0	0	1	2	
SPU_007404	1	3	7	8	10	
SPU_007406	1	6	8	12	18	
SPU_006989	1	0	1	2	2	
SPU_017582	9[b]	0	0	1	2	
SPU_019883	9	0	1	2	5	

[a] Genomic sequences were searched with degenerate consensus XRE sequences (KNGCGTG). Distances for the *Ciona* genes are upstream from the translation start site, whereas the distances for the vertebrate genes are from the mRNA transcription start site.
[b] Incomplete assembly prevents the full prediction of exon 7.

contrast, *X. laevis* has 2 *CYP1A* genes (*CYP1A6* and *CYP1A7*) (Fujita et al. 1999), a duplication possibly resulting from tetraploidy in *X. laevis*. Polyploidy occurs frequently in the genus *Xenopus* ranging up to the dodecaploid *Xenopus ruwenzoriensis* (108 chromosomes) (Fischberg and Kobel 1978), and it will be interesting to determine how many *CYP1A* genes occur in these amphibians. We also identified one predicted *CYP1B1* gene in both *X. tropicalis* and *G. gallus*. The presence of a *CYP1B1* in these groups is not unexpected, given that *CYP1B1* occurs in both fish and mammals (Sutter et al. 1994; Leaver and George 2000) and CYP1B1 phylogenetic placement is consistent with species relationships among tetrapods.

In addition, in both the *X. tropicalis* and the *G. gallus* genomes, we identified one sequence orthologous to the fish CYP1C genes, expressed at the transcript level. We have been unable to identify a *CYP1C* gene in mammalian genomes. The presence of expressed *CYP1Cs* in an amphibian and a bird supports the idea that the *CYP1C* subfamily has been lost in mammals (Godard et al. 2005).

Timing of Steps in *CYP1* Divergence

The new genes reported here suggest a revised evolutionary history of the CYP1 family (fig. 5). Based on our results, it appears that 2 primary clades—CYP1A/1Es and CYP1B/1C/1Fs—resulted from an early duplication in a chordate ancestor prior to the divergence of tunicates from the lineage leading to vertebrates (~590 MYA) but after the divergence of echinoderms (~ 620 MYA) (Ayala and Rzhetsky 1998). Subsequently, multiple independent gene duplication events led to expansions in both *CYP1* clades. However, what happened between the divergence of the tunicates (580 MYA) and before the origins of the bony fish (450 ± 35 MYA) is unclear.

The presence of only one *CYP1A* gene in the diploid frog enables the gene duplication leading to *CYP1A1* and *CYP1A2* to be dated more narrowly to 310–360 MYA, between the amniote–amphibian and mammal–avian divergences (Benton 1990; Hedges et al. 1996; Kumar and Hedges 1998). The avian *CYP1A4* and *CYP1A5* and the mammalian *CYP1A1* and *CYP1A2* paralog pairs are the result of the same tandem-inverted gene duplication event (Goldstone and Stegeman 2006).

In addition to the main gene duplications giving rise to the various subfamilies, lineage-specific gene duplication (symparalogy) has occurred. This duplication has given rise to at least 2 *CYP1C* paralogs in fish, multiple *CYP1F* genes in *Ciona*, and multiple *CYP1*-like genes in *S. purpuratus*. The duplicate *CYP1C* genes in fish are probably not due to the ancient whole-genome duplication in ray-finned fishes (Van de Peer et al. 2003), as they are immediately adjacent

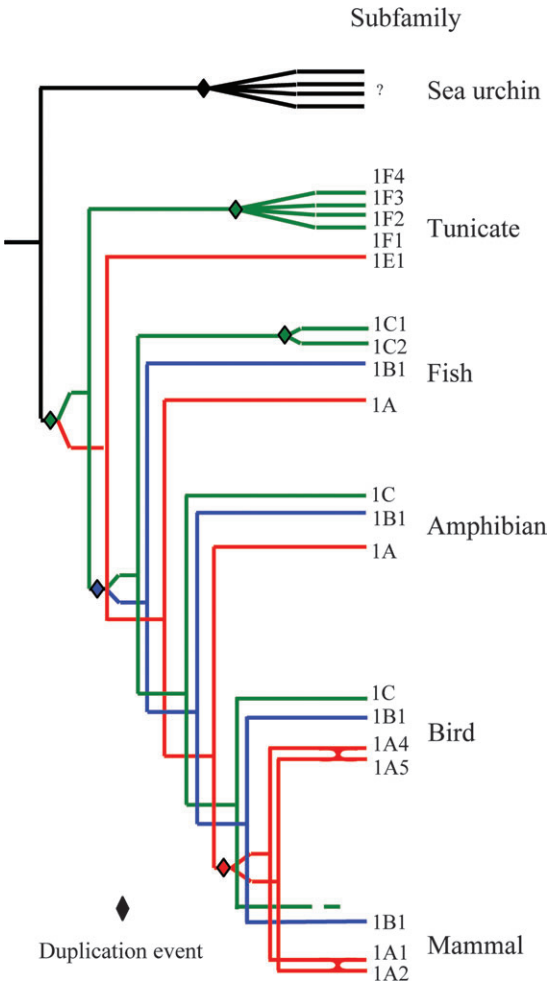

FIG. 5.—Schematic representation of the gene duplication and speciation leading to the observed distribution of *CYP1* genes. Two primary clades—CYP1A/1Es and CYP1B/1C/1Fs—resulted from an early divergence in an ancestral "protochordate" prior to the divergence of tunicates from chordates. Subsequently, multiple independent gene duplication events led to expansions of both *CYP1* subfamilies. The CYP1A/1B split appears to have occurred between the time that echinoderms and tunicates diverged from the chordate lineage. The CYP1C subfamily originated after the divergence of the tunicates and before the origins of the bony fish. The gene duplication leading to CYP1A1 and CYP1A2 occurred between the mammal–amphibian and mammal–avian divergences as the duplicated chicken and mammalian CYP1As are the result of the same tandem-inverted gene duplication event obscured by gene conversion (indicated by an x in the figure; Goldstone and Stegeman 2006).

single-exon genes that are likely the result of a reinsertion of a processed mRNA followed by tandem gene duplication. There are duplicate *CYP1A* (Gooneratne et al. 1997) and *CYP1B* (El-kady et al. 2004) genes in some fish species, possibly arising from tetraploidization in those lineages (e.g., salmonids [Johnson et al. 1987]). Determining the lineage specificity of these gene duplications will require significantly greater taxonomic sampling.

Structural and Functional Divergence

Comparing site-specific evolutionary rates in amino acid sequences among subclades suggests differences in functional constraints between the CYP1As and CYP1B/1Cs. Low overall coefficients of functional divergence (θ) (Gaucher et al. 2002) in the tunicate CYP1Fs relative to other members of the CYP1B/1C/1F clade suggest that all of these genes share similar, although unknown, functional constraints. There were high values of θ between vertebrate CYP1As and most CYP1B/1C/1Fs, suggesting that these 2 CYP1 subclades have significantly different evolutionary constraints on function. However, the substrate specificity is not known for the new CYP1s, or for any CYP1C, precluding correlation of these results with catalytic activities.

Overall CYP structure generally is highly conserved, particularly within families (Werck-Reichhart and Feyereisen 2000), and altered functional constraints are most likely to have effects if they occur within a SRS (Gotoh 1992) or otherwise within the active site. Changes in substrate specificity therefore may be determined by a relatively small number of amino acid residues. Our phylogenetic analysis of the SRS regions supports the similarity of the tunicate SRSs with those of other CYP1s (supplementary fig. S2, Supplementary Material online). In agreement with this phylogeny, analysis of site-specific functional constraints showed significantly decreased substitution rates between the tunicate CYP1A-like genes (CYP1E1) and the vertebrate CYP1As, especially within SRSs 1, 4, and 5. Homology modeling and docking results indicate that catalytic differences between mammalian CYP1A1s and fish CYP1As result from substitutions within SRS 2, 4, and 5 (Prasad et al. 2007). The high *Ciona* vertebrate pairwise similarities at sites within the SRSs 1, 4, and 5, and the relatively slow substitution rates found in these regions suggest that the CYP1E1 may share some substrate specificity with vertebrate CYP1As. Clustering of sites with increased substitution rates in functionally relevant regions of the proteins suggests the action of positive selection on these sites rather than an effect of the generally elevated substitution rates in tunicates. These possibilities are under investigation.

Two of the tunicate CYP1Fs had large insertions that are expected to add helix and sheet structures exposed on the surfaces of the proteins (in the case of CYP1F3, near the purported substrate entry site). These insertions could be important to the function of these enzymes. It will be interesting to assess substrate binding by these unusual CYPs, computationally and directly.

Establishing the factors involved in regulation of the early *CYP1s* also may give clues to biological roles. We observed that several of the *CYP1s* (*CYP1E1* and *CYP1F1–CYP1F3*) are expressed during development in *C. intestinalis*. CYP1 family genes are either basally expressed or inducible by xenobiotics throughout the development in many tissues in zebrafish and mice (Jonsson, Orrego, et al. 2007) and mice (Choudhary et al. 2003) and play roles in tissue patterning and humans (Choudhary et al. 2006).

Expression of *CYP1A* genes in vertebrates is regulated largely by the AHR (Hahn et al. 1998); vertebrate AHR also regulates other genes, including other CYP1s and some CYP2s (Rivera et al. 2002; Arpiainen et al. 2005; Jonsson, Jenny, et al. 2007). *Ciona intestinalis* possesses an AHR homolog, which is located immediately adjacent to *CYP1A*-like *CYP1E1* (CI0100131189) on chromosome 12. This

synteny suggests a possible functional linkage of the 2 genes. Although no data are available on AHR involvement in gene regulation in *Ciona*, numerous AHR binding sites (XREs) are present in the upstream region of the *CYP1E1* as well as in other *Ciona CYP1s*. In vitro studies suggest that the *C. intestinalis* AHR homolog does not bind the vertebrate AHR ligand 2,3,7,8 tetrachlorodibenzo-*p*-dioxin (Hahn ME, personal communication), a characteristic shared by the AHR homolog in several invertebrates, including Drosophila, the clam *Mya arenaria*, and the nematode *Caenorhabditis elegans* (Butler et al. 2001; McMillan and Bradfield 2007). However, this observation presumably reflects a phylogenetic difference in AHR ligand binding and does not preclude involvement of AHR in *CYP* regulation or activation of the *Ciona* AHR by other ligands. Two *AHR*-like genes occur also in the *S. purpuratus* genome (Goldstone et al. 2006), and multiple XREs are present in the upstream regions of many of the *S. purpuratus CYP1*-like genes, including 8 in the 2,000-bp region immediately upstream of both SPU_007404 and SPU_007406. It will be important to determine ligand-binding properties of the echinoderm and tunicate AHRs and whether any of these XREs are functional. This should help to determine when the AHRs and the CYP1s became functionally linked, a key question regarding *CYP* and *AHR* evolution.

The long divergence times and variable functional constraints manifested by the CYP1s present a challenge for accurate phylogenetic analysis. When faced with these challenges, rigorous phylogenetic analyses, such as Bayesian and likelihood techniques, give better estimates of evolutionary relationships than similarity or distance-based approaches (Thornton and DeSalle 2000). Increased taxonomic sampling also is known to reduce phylogenetic error (Pollock et al. 2002; Zwickl and Hillis 2002).

The *CYP1* lineage clearly predates the emergence of vertebrates, with the *CYP1A*/*1E* and the *CYP1B*/*1C*/*1F* clades established in the tunicates. Thus, the new genes analyzed here present a more comprehensive picture of *CYP1* phylogeny, with an evolutionary history that spans some 600 million years. The stability of our provisional nomenclature for the tunicate genes (*CYP1E1* and *CYP1F1–1F4*) will depend on the resolution of the entire *CYP1* phylogeny, including earlier events in *CYP1* evolution. The *CYP1s* are in Clan 2, together with the *CYP2s*, *CYP17s*, and *CYP21s*, and may have evolved from one of these families (Nelson 1998). Interestingly, the echinoderm genes we describe appear intermediate between the *CYP1s* and *CYP2s* in molecular phylogeny. These new genes provide a foundation for studies of CYP1 functional evolution in the deuterostomes. Analysis of prebilaterian genomes and of *CYP1s* from taxa between the tunicates and vertebrates should elucidate the evolutionary antecedent of the *CYP1* family and the fate of the descendant gene lines in deuterostomes as well as protostomes.

Supplementary Material

Supplementary figures 1 and 2 and tables 1–7 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We gratefully acknowledge Dr David R. Nelson for discussions regarding nomenclature, Dr Andrew McArthur for his assistance with phylogenetic analyses, and Dr Mark E. Hahn for comments on the text. This work was supported by National Institutes of Health grants 5P42ES007381 (the Boston University Superfund Basic Research Program), 1R01ES015912 (J.J.S.), and F32ES012794 (J.V.G.). Computational support was provided by the W. M. Keck Ecological and Evolutionary Genetics Facility at the Marine Biological Laboratory.

Literature Cited

Abascal F, Zardoya R, Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics. 21: 2104–2105.

Arpiainen S, Raffalli-Mathieu F, Lang MA, Pelkonen O, Hakkola J. 2005. Regulation of the Cyp2a5 gene involves an aryl hydrocarbon receptor-dependent pathway. Mol Pharmacol. 67:1325–1333.

Ayala FJ, Rzhetsky A. 1998. Origin of the metazoan phyla: molecular clocks confirm paleontological estimates. Proc Natl Acad Sci USA. 95:606–611.

Benton MJ. 1990. Phylogeny of the major tetrapod groups: morphological data and divergence dates. J Mol Evol. 30:409–424.

Birney E, Clamp M, Durbin R. 2004. GeneWise and Genomewise. Genome Res. 14:988–995.

Burge CB, Karlin S. 1998. Finding the genes in genomic DNA. Curr Opin Struct Biol. 8:346–354.

Butler RA, Kelley ML, Powell WH, Hahn ME, Van Beneden RJ. 2001. An aryl hydrocarbon receptor (AHR) homologue from the soft-shell clam, Mya arenaria: evidence that invertebrate AHR homologues lack 2,3,7,8-tetrachlorodibenzo-p-dioxin and beta-naphthoflavone binding. Gene. 278:223–234.

Choudhary D, Jansson I, Sarfarazi M, Schenkman JB. 2006. Physiological significance and expression of P450s in the developing eye. Drug Metab Rev. 38:337–352.

Choudhary D, Jansson I, Schenkman JB, Sarfarazi M, Stoilov I. 2003. Comparative expression profiling of 40 mouse cytochrome P450 genes in embryonic and adult tissues. Arch Biochem Biophys. 414:91–100.

Cirino P, Toscano A, Caramiello D, Macina A, Miraglia V, Monte A. 2002. Laboratory culture of the ascidian *Ciona intestinalis* (L.): a model system for molecular developmental biology research. Mar Mod Elec Rec. [serial online]; Available from http://www.mbl.edu/BiologicalBulletin/MMER/cirino/CirTit.html. Accessed 15 Sept 2007.

Conney AH. 1982. Induction of microsomal enzymes by foreign chemicals and carcinogenesis by polycyclic aromatic hydrocarbons: G.H.A. Clowes Memorial Lecture. Cancer Res. 42:4875–4917.

Crespi CL, Steimel DT, Aoyama T, Gelboin HV, Gonzalez FJ. 1990. Stable expression of human cytochrome P450IA2 cDNA in a human lymphoblastoid cell line: role of the enzyme in the metabolic activation of aflatoxin B1. Mol Carcinog. 3:5–8.

Cuff JA, Barton GJ. 1999. Evaluation and improvement of multiple sequence methods for protein secondary structure prediction. Proteins. 34:508–519.

Degtyarenko KN, Archakov AI. 1993. Molecular evolution of P450 superfamily and P450-containing monooxygenase systems. FEBS Lett. 332:1–8.

Dehal P, Satou Y, Campbell RK, et al. (87 co-authors). 2002. The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. Science. 298:2157–2167.
Eddy SR. 1998. Profile hidden Markov models. Bioinformatics. 14:755–763.
El-kady MA, Mitsuo R, Kaminishi Y, Itakura T. 2004. Isolation of cDNA of novel cytochrome P450 1B gene, CYP1B2, from Carp (*Cyprinus carpio*) and its induced expression in gills. Environ Sci. 11:345–354.
Fischberg M, Kobel HR. 1978. Two new polyploid *Xenopus* species from western Uganda. Experientia. 34:1012–1014.
Fujita Y, Ohi H, Murayama N, Saguchi K, Higuchi S. 1999. Molecular cloning and sequence analysis of cDNAs coding for 3-methylcholanthrene-inducible cytochromes P450 in *Xenopus laevis* liver. Arch Biochem Biophys. 371:24–28.
Gaucher EA, Gu X, Miyamoto MM, Benner SA. 2002. Predicting functional divergence in protein evolution by site-specific rate shifts. Trends Biochem Sci. 27:315–321.
Godard CA, Goldstone JV, Said MR, Dickerson RL, Woodin BR, Stegeman JJ. 2005. The new vertebrate CYP1C family: cloning of new subfamily members and phylogenetic analysis. Biochem Biophys Res Commun. 331:1016–1024.
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, Hamdoun A, Cole BJ, Howard-Ashby M, Nebert DW, Scally M, Dean M, Epel D, Hahn ME, Stegeman JJ. 2006. The chemical defensome: environmental sensing and response genes in the *Strongylocentrotus purpuratus* genome. Dev Biol. 300:366–384.
Gooneratne R, Miranda CL, Henderson MC, Buhler DR. 1997. Beta-naphthoflavone induced CYP1A1 and 1A3 proteins in the liver of rainbow trout (*Oncorhynchus mykiss*). Xenobiotica. 27:175–187.
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.
Gu X, Vander Velden K. 2002. DIVERGE: phylogeny-based analysis for functional-structural divergence of a protein family. Bioinformatics. 18:500–501.
Guex N, Peitsch MC. 1997. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modelling. Electrophoresis. 18:2714–2723.
Hahn ME, Woodin BR, Stegeman JJ, Tillitt DE. 1998. Aryl hydrocarbon receptor function in early vertebrates: inducibility of cytochrome P450 1A in agnathan and elasmobranch fish. Comp Biochem Physiol C Pharmacol Toxicol Endocrinol. 120:67–75.
Hedges SB, Parker PH, Sibley CG, Kumar S. 1996. Continental breakup and the ordinal diversification of birds and mammals. Nature. 381:226–229.
Huelsenbeck JP, Imennov NS. 2002. Geographic origin of human mitochondrial DNA: accommodating phylogenetic uncertainty and model comparison. Syst Biol. 51:155–165.
Hughes MK, Hughes AL. 1993. Evolution of duplicate genes in a tetraploid animal, *Xenopus laevis*. Mol Biol Evol. 10: 1360–1369.
Ioannides C, Lewis DF. 2004. Cytochromes P450 in the bioactivation of chemicals. Curr Top Med Chem. 4:1767–1788.
Itakura T, El-kady MAH, Mitsuo R, Kaminishi Y. 2005. Complementary DNA cloning and constitutive expression of cytochrome P450 1C1 in the gills of carp (*Cyprinus carpio*). Environ Sci. 12:111–120.
Johnson KR, Wright JE Jr, May B. 1987. Linkage relationships reflecting ancestral tetraploidy in salmonid fish. Genetics. 116:579–591.
Jones DTB, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci. 8:275–282.
Jonsson ME, Jenny MJ, Woodin BR, Hahn ME, Stegeman JJ. 2007. Role of AHR2 in the expression of novel cytochrome P450 1 family genes, cell cycle genes, and morphological defects in developing zebrafish exposed to 3,3′,4,4′,5-pentachlorobiphenyl or 2,3,7,8-tetrachlorodibenzo-p-dioxin. Toxicol Sci. 100:180–193.
Jonsson 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. 221:29–41.
Kass RE, Raftery AE. 1995. Bayes factors. J Am Stat Assoc. 90:773–795.
Knudsen B, Miyamoto MM. 2001. A likelihood ratio test for evolutionary rate shifts and functional divergence among proteins. Proc Natl Acad Sci USA. 98:14512–14517.
Kumar S, Hedges SB. 1998. A molecular timescale for vertebrate evolution. Nature. 392:917–920.
Lambrecht RW, Sinclair PR, Gorman N, Sinclair JF. 1992. Uroporphyrinogen oxidation catalyzed by reconstituted cytochrome P450IA2. Arch Biochem Biophys. 294:504–510.
Leaver MJ, George SG. 2000. A cytochrome P4501B gene from a fish, *Pleuronectes platessa*. Gene. 256:83–91.
Rost B, Liu J. 2003. The PredictProtein server. Nucleic Acids Res. 31:3300–3304.
Liu J, Ericksen SS, Besspiata D, Fisher CW, Szklarz GD. 2003. Characterization of substrate binding to cytochrome P450 1A1 using molecular modeling and kinetic analyses: case of residue 382. Drug Metab Dispos. 31:412–420.
Liu J, Ericksen SS, Sivaneri M, Besspiata D, Fisher CW, Szklarz GD. 2004. The effect of reciprocal active site mutations in human cytochromes P450 1A1 and 1A2 on alkoxyresorufin metabolism. Arch Biochem Biophys. 424:33–43.
McMillan BJ, Bradfield CA. 2007. The aryl hydrocarbon receptor sans xenobiotics: endogenous function in genetic model systems. Mol Pharmacol. 72:487–498.
Nebert DW, Dalton TP. 2006. The role of cytochrome P450 enzymes in endogenous signalling pathways and environmental carcinogenesis. Nat Rev Cancer. 6:947–960.
Nebert DW, Russell DW. 2002. Clinical importance of the cytochromes P450. Lancet. 360:1155–1162.
Nelson DR. 1998. Metazoan cytochrome P450 evolution. Comp Biochem Physiol C Pharmacol Toxicol Endocrinol. 121:15–22.
Nelson DR, Kamataki T, Waxman DJ, et al. 1993. The P450 superfamily: update on new sequences, gene mapping, accession numbers, early trivial names of enzymes, and nomenclature. DNA Cell Biol. 12:1–51.
Peitsch MC. 1995. Protein modeling by E-mail. Biotechnology. 13:658–660.
Pollock DD, Zwickl DJ, McGuire JA, Hillis DM. 2002. Increased taxon sampling is advantageous for phylogenetic inference. Syst Biol. 51:664–671.
Pond SL, Frost SD, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics. 21:676–679.
Prasad J, Goldstone JV, Camacho C, Stegeman JJ, Vajda S. 2007. Ensemble modeling of substrate binding to cytochromes P450: analysis of catalytic differences between CYP1A orthologues. Biochemistry. 46:2640–2654.
Raftery A, Lewis S. 1992. How many iterations in the Gibbs sampler? In: Bernardo J, Berger J, Dawid A, Smith A, editors. Bayesian statistics 4. Oxford: Oxford University Press. p. 763–774.
Raner GM, Vaz AD, Coon MJ. 1996. 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. 49:515–522.
Rivera SP, Saarikoski ST, Hankinson O. 2002. Identification of a novel dioxin-inducible cytochrome P450. Mol Pharmacol. 61:255–259.
Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19:1572–1574.
Rost B. 1996. PHD: predicting one-dimensional protein structure by profile based neural networks. Meth Enzymol. 266:525–539.
Satou Y, Kawashima T, Shoguchi E, Nakayama A, Satoh N. 2005. An integrated database of the ascidian, *Ciona intestinalis*: towards functional genomics. Zool Sci. 22:837–843.
Schwede T, Kopp J, Guex N, Peitsch MC. 2003. SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res. 31:3381–3385.
Shimada T, Guengerich FP. 2006. Inhibition of human cytochrome P450 1A1-, 1A2-, and 1B1-mediated activation of procarcinogens to genotoxic metabolites by polycyclic aromatic hydrocarbons. Chem Res Toxicol. 19:288–294.
Smith BJ. 2003. Bayesian output analysis program (BOA).
Sodergren E, Weinstock GM, Davidson EH et al. (228 co-authors). 2006. The genome of the sea urchin *Strongylocentrotus purpuratus*. Science. 314:941–952.
Spink DC, Eugster HP, Lincoln DW 2nd, Schuetz JD, Schuetz EG, Johnson JA, Kaminsky LS, Gierthy JF. 1992. 17 beta-estradiol hydroxylation catalyzed by human cytochrome P450 1A1: a comparison of the activities induced by 2,3,7,8-tetrachlorodibenzo-p-dioxin in MCF-7 cells with those from heterologous expression of the cDNA. Arch Biochem Biophys. 293:342–348.
Suchard MA, Weiss RE, Sinsheimer JS. 2005. Models for estimating Bayes factors with applications to phylogeny and tests of monophyly. Biometrics. 61:665–673.
Sun YV, Boverhof DR, Burgoon LD, Fielden MR, Zacharewski TR. 2004. Comparative analysis of dioxin response elements in human, mouse and rat genomic sequences. Nucleic Acids Res. 32:4512–4523.
Sutter TR, Tang YM, Hayes CL, Wo YY, Jabs EW, Li X, Yin H, Cody CW, Greenlee WF. 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. 269:13092–13099.
Swalla BJ. 2001. Phylogeny of the urochordates: implications for chordate evolution. In: Sawada H, Yokosawa H, Lambert C, editors. The biology of ascidians. Tokyo: Springer Verlag. p. 219–224.
Thornton JW, DeSalle R. 2000. Gene family evolution and homology: genomics meets phylogenetics. Annu Rev Genomics Hum Genet. 1:41–73.
Tsuchiya Y, Nakajima M, Yokoi T. 2003. Critical enhancer region to which AhR/ARNT and Sp1 bind in the human CYP1B1 gene. J Biochem (Tokyo). 133:583–592.
Van de Peer Y, Taylor JS, Meyer A. 2003. Are all fishes ancient polyploids? J Struct Funct Genomics. 3:65–73.
Verslyke T, Goldstone JV, Stegeman JJ. 2006. Isolation and phylogeny of novel tunicate clan 3 P450 genes. Mol Phylogenet Evol. 40:760–771.
Vinson JP, Jaffe DB, O'Neill K, et al. (13 co-authors). 2005. Assembly of polymorphic genomes: algorithms and application to *Ciona savignyi*. Genome Res. 15:1127–1135.
Werck-Reichhart D, Feyereisen R. 2000. Cytochromes P450: a success story. Genome Biol. 1:REVIEWS3003
Whelan S, Goldman N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol. 18:691–699.
Yokobori S, Ueda T, Feldmaier-Fuchs G, Paabo S, Ueshima R, Kondow A, Nishikawa K, Watanabe K. 1999. Complete DNA sequence of the mitochondrial genome of the ascidian *Halocynthia roretzi* (Chordata, Urochordata). Genetics. 153:1851–1862.
Yokobori S, Watanabe Y, Oshima T. 2003. Mitochondrial genome of *Ciona savignyi* (Urochordata, Ascidiacea, Enterogona): comparison of gene arrangement and tRNA genes with *Halocynthia roretzi* mitochondrial genome. J Mol Evol. 57:574–587.
ZeRuth G, Pollenz R. 2005. Isolation and characterization of a dioxin-inducible CYP1A1 promoter/enhancer region from zebrafish (*Danio rerio*). Zebrafish. 2:197–210.
Zwickl DJ, Hillis DM. 2002. Increased taxon sampling greatly reduces phylogenetic error. Syst Biol. 51:588–598.

Billie Swalla, Associate Editor

Accepted September 5, 2007