ABSTRACT The aryl hydrocarbon receptor (AHR) is a ligand-activated transcription factor through which halogenated aromatic hydrocarbons such as 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) cause altered gene expression and toxicity. The AHR belongs to the basic helix–loop–helix/Per-Arnt-Sim (bHLH-PAS) family of transcriptional regulatory proteins, whose members play key roles in development, circadian rhythmicity, and environmental homeostasis; however, the normal cellular function of the AHR is not yet known. As part of a phylogenetic approach to understanding the function and evolutionary origin of the AHR, we sequenced the PAS homology domain of AHRs from several species of early vertebrates and performed phylogenetic analyses of these AHR amino acid sequences in relation to mammalian AHRs and 24 other members of the PAS family. AHR sequences were identified in a teleost (the killifish Fundulus heteroclitus), two elasmobranch species (the skate Raja erinacea and the dogfish Mustelus canis), and a jawless fish (the lamprey Petromyzon marinus). Two putative AHR genes, designated AHR1 and AHR2, were found both in Fundulus and Mustelus. Phylogenetic analyses indicate that the AHR2 genes in these two species are orthologous, suggesting that an AHR gene duplication occurred early in vertebrate evolution and that multiple AHR genes may be present in other vertebrates. Database searches and phylogenetic analyses identified four putative PAS proteins in the nematode Caenorhabditis elegans, including possible AHR and ARNT homologs. Phylogenetic analysis of the PAS gene family reveals distinct clades containing both invertebrate and vertebrate PAS family members; the latter include paralogous sequences that we propose have arisen by gene duplication early in vertebrate evolution. Overall, our analyses indicate that the AHR is a phylogenetically ancient protein present in all living vertebrate groups (with a possible invertebrate homolog), thus providing an evolutionary perspective to the study of dioxin toxicity and AHR function.

Halogenated aromatic hydrocarbons such as 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) are potent modulators of cellular growth and differentiation and thus are highly toxic to vertebrate animals (1). These effects are mediated by the aryl hydrocarbon receptor (Ah receptor, AHR, or “dioxin receptor”), a ligand-activated transcription factor that acts in concert with the Ah receptor nuclear translocator [ARNT (2)] to alter the expression of target genes, such as cytochrome P450 1A1 (1, 3). The AHR and ARNT belong to the Per-ARNT-Sim (PAS) family of transcriptional regulatory proteins (3, 4), whose members play key roles in development (5), adaptation to hypoxia (6, 7), control of circadian rhythmicity (8, 9), and phototransduction (8, 10, 11). The physiological function of the AHR is not yet known, but an important role in the developing liver and immune system has been suggested by the phenotypes of mice bearing a targeted disruption of the AHR locus (12, 13).

The AHR has been studied almost exclusively in mammals, in which a single gene has been identified (14, 15). The mammalian AHR contains basic helix–loop–helix (bHLH) and PAS homology domains that define the PAS family. The bHLH domain contains basic and HLH motifs involved in protein–DNA and protein–protein interactions, respectively. The PAS domain forms a secondary dimerization surface for heteromeric interactions between AHR and ARNT, as well as among other bHLH-PAS proteins (16, 17). It includes two imperfect repeats of 51 amino acids [PAS-A and PAS-B (18)] separated by an intervening sequence of approximately 110 amino acids. Importantly, the distal portion of this region (PAS-B) is part of the ligand-binding domain of the AHR (15, 19–22).

In contrast to the extensive literature on the mammalian AHR, knowledge of the AHR in other vertebrate and invertebrate animals is limited (23–25). The objective of the present work was to investigate the evolutionary history of the AHR and its relationship to other members of the PAS family. Our approach was to sequence the AHR PAS domains from early chordates and to assess their relationships by phylogenetic inference, a powerful tool for understanding the evolution and interrelationships of multigene families (26). We focused on early chordates because previous results had suggested the first appearance of an AHR protein in cartilaginous fish (24). The PAS domain was chosen because it is a well-conserved and functionally important region of the mammalian AHR and other members of the PAS family (8, 15, 16, 27); except for the bHLH domain (28), other regions of PAS proteins are not highly conserved (29) and, therefore, are less suitable for phylogenetic analysis.

The results of these studies show that the AHR is a phylogenetically ancient protein that exists in bony and cartilaginous fish, as well as lamprey, the most “primitive” (i.e., early diverging) living vertebrate. We also report a second AHR in two species of gnathostome (jawed) fish and provide evidence that an AHR gene duplication occurred early in vertebrate evolution. Possible invertebrate AHR and ARNT homologs are also described. We discuss these results in relation to the diversification of the PAS family.

METHODS

Animals and RNA Isolation. Killifish (Fundulus heteroclitus), smooth dogfish (Mustelus canis), little skate (Raja erina-
parsimony (PAUP 3.1 (33)). Alignment positions with gaps were using the Neighbor-Joining (NJ) algorithm (32) and maximum domain only) were used to construct phylogenetic trees by poly(A) RNA (RNA STAT-60; Tel-Test, Friendswood, TX) and FastTrack kit (InVitrogen) or by sequential isolation of total and hagfish), from the anterior region of the lamprey, or from prepared from frozen liver powders (killifish, skate, dogfish, and hagfish), from the anterior region of the lamprey, or from the total visceral organs of amphioxus, either directly using a FastTrack kit (Invitrogen) or by sequential isolation of total RNA (RNA STAT-60; Tel-Test, Friendswood, TX) and poly(A) mRNA [mini-oligo(dT)-cellulose spin column kit; 5 Prime → 3 Prime].

RT–PCR, Cloning, and Sequencing. Degenerate inosine-containing oligonucleotides AHR-A1 and AHR-B1 were designed as described (25, 30). Reverse transcription coupled-PCR (RT–PCR) was performed by using the Gene-Amp RNA-PCR kit (Perkin–Elmer) and a GeneAmp 2400 thermocycler. Reverse transcription was primed with random hexamers; for the PCR, AHR-A1 and AHR-B1 were used at 1 µM. PCR conditions were optimized for each species. For hagfish, lamprey, and amphioxus, MgCl₂ concentration was 3.0 mM rather than 2.0 mM. PCR cycles were as follows: 105 sec at 95°C, 35 cycles of 95°C for 15 sec and 50°C for 30 sec, followed by 7 min at 72°C. PCR products were analyzed by Southern blotting using oligonucleotide J2u (5'-GGCTAYCAGTTYATYCATGC-3'), targeted to the conserved sequence GYQFIHA (corresponding to amino acids 315–321 of the mouse AHR). Hybridizing bands were cloned into pCNTR (5 Prime → 3 Prime) or p17BlueR (Novagen) and sequenced in both directions by using SequiTherm and SequiTherm Excel long-read cycle sequencing kits (Epicentre Technologies, Madison, WI) and an automated DNA sequencer (LI-COR). Three to seven clones were sequenced for each PCR fragment. The sequence of AHR1 from Fundulus was obtained from genomic DNA clones and confirmed by RT–PCR (S.I.K. and M.E.H., unpublished results).

Sequence Analysis. The sequences of RT–PCR products were assembled and translated. Multiple alignment of the deduced amino acid sequences was performed by using CLUSTALW version 1.6 (31). The aligned amino acid sequences (PAS domain only) were used to construct phylogenetic trees by using the Neighbor-Joining (NJ) algorithm (32) and maximum parsimony [PAUP 3.1 (33)]. Alignment positions with gaps were excluded. Bootstrap analysis (34) was performed to assess relative confidence in the topologies obtained.

RESULTS

AHRs in Early Vertebrates. To examine the evolutionary history of the AHR, degenerate PCR primers (25, 30) were used to amplify cDNA sequences from the cephalochordate amphioxus and representative species of early chordates, including the teleost Fundulus, two cartilaginous fish (smooth dogfish and little skate), and two jawless fish (sea lamprey and Atlantic hagfish). Lamprey are the most ancient living vertebrates, and hagfish are considered invertebrate chordates (35–36).] Products of the predicted size (~700 bp) that hybridized to an AHR-specific probe were obtained from Fundulus, dogfish, skate, and lamprey (data not shown). The nucleotide and deduced amino acid sequences of these RT–PCR products were most closely related to PAS domains of mammalian AHRs (60–76% amino acid identity; Table 1).

Initially, a single AHR sequence was obtained from Fundulus, as reported (25). Subsequently, a second Fundulus AHR sequence was obtained by screening a genomic DNA library; its expression was confirmed by RT–PCR. Similarly, RT–PCR using dogfish RNA revealed two different AHR-like sequences, which were confirmed by sequencing multiple clones from two independent RT–PCRs. We have designated the two putative AHRs in each of these species as AHR1 and AHR2; in each case AHR1 shares greater sequence identity with mammalian AHR sequences than does AHR2 (Table 1). Both Fundulus AHR1 and AHR2 possess bHLH motifs that are closely related to those of mammalian AHRs [AHR1, 83% amino acid identity; AHR2, 73% amino acid identity; both exhibit 100% identity of amino acids critical for DNA binding (37)]. The designation of these fish sequences as AHRs is based on the high bHLH and PAS sequence identities in comparisons with mammalian AHRs and on the phylogenetic analyses described below. The full-length sequence and other properties of both Fundulus AHRs will be described in detail elsewhere (S.I.K. and M.E.H., unpublished results).

Herein, it is important to note that the sequence difference between the two apparently paralogous AHRs within each species is as great or greater than the interspecies differences (Table 1), suggesting an ancient duplication.

Alignment of the PAS domain sequences of all vertebrate AHRs reveals several conserved regions within and between the PAS-A and PAS-B boxes (Fig. 1). Overall, 82 residues (41%) are conserved in the PAS domains of all of these vertebrate AHRs. These include 20 amino acids in PAS-A, 23 in PAS-B, and 35 in the region between the two PAS boxes. At several positions, characteristic amino acids distinguish the fish and mammalian AHRs. Of the fish sequences, dogfish AHR1 appears to be most closely related to the mammalian AHRs (Fig. 1 and Table 1). The skate AHR is the most divergent, both overall and at certain residues conserved in all of the other sequences. For example, the sequence RCLLDNSSGFL is identical in all AHRs except skate, where five differences occur. At five positions the two AHR2 sequences share unique residues that are not present in any of the other AHRs.

Molecular Phylogeny of AHR Genes. Phylogenetic analyses were used to assess the relationship of the two putative AHR2 forms to each other and to the other AHR sequences. In both NJ (Fig. 2A) and maximum parsimony analyses (Fig. 2B), the AHR2 forms from Fundulus and dogfish form a monophyletic group. Bootstrap analysis provides strong (96%) support for the AHR2 cluster and, thus, for an orthologous relationship between the Fundulus and dogfish AHR2 forms. The relationship of the skate and lamprey AHRs to the other fish AHR sequences is not resolved in these rooted trees. In a further

<table>
<thead>
<tr>
<th></th>
<th>Fundulus AHR1</th>
<th>Fundulus AHR2</th>
<th>Dogfish AHR1</th>
<th>Dogfish AHR2</th>
<th>Skate AHR</th>
<th>Lamprey AHR</th>
<th>C. elegans C41G7.5</th>
<th>Human AHR</th>
<th>Mouse AHR b-1</th>
<th>Rat AHR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fundulus AHR1</td>
<td>100</td>
<td>63</td>
<td>71</td>
<td>69</td>
<td>61</td>
<td>70</td>
<td>36</td>
<td>70</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>Fundulus AHR2</td>
<td>100</td>
<td>66</td>
<td>68</td>
<td>35</td>
<td>57</td>
<td>59</td>
<td>33</td>
<td>62</td>
<td>65</td>
<td>64</td>
</tr>
<tr>
<td>Dogfish AHR1</td>
<td>100</td>
<td>70</td>
<td>62</td>
<td>70</td>
<td>35</td>
<td>76</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>Dogfish AHR2</td>
<td>100</td>
<td>63</td>
<td>67</td>
<td>35</td>
<td>68</td>
<td>68</td>
<td>67</td>
<td>67</td>
<td>67</td>
<td></td>
</tr>
<tr>
<td>Skate AHR</td>
<td>100</td>
<td>66</td>
<td>30</td>
<td>61</td>
<td>61</td>
<td>61</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>Lamprey AHR</td>
<td>100</td>
<td>35</td>
<td>68</td>
<td>68</td>
<td>68</td>
<td>67</td>
<td></td>
<td>67</td>
<td>67</td>
<td></td>
</tr>
<tr>
<td>P. aeruginosa C41G7.5</td>
<td>100</td>
<td>35</td>
<td>34</td>
<td>34</td>
<td>34</td>
<td>34</td>
<td></td>
<td>34</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>Human AHR</td>
<td>100</td>
<td>87</td>
<td>88</td>
<td>88</td>
<td></td>
<td></td>
<td></td>
<td>88</td>
<td>88</td>
<td></td>
</tr>
<tr>
<td>Mouse AHR b-1</td>
<td>100</td>
<td>95</td>
<td>100</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Rat AHR</td>
<td>100</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>
the complete sequences of these and additional of the same gene [mosaic evolution (38)], phylogenetic analysis of change are known to differ between functionally distinct regions protein–protein interactions. Because rates of evolutionary conserved based on its important roles in ligand binding and

AHR2 clusters are each monophyletic requires 202 steps. In these analyses, the vertebrate AHRs, including skate and lamprey AHR) and PAS proteins, we conducted phylogenetic analyses of all AHR relationship of these new AHR sequences to sequences of other PAS domain sequences and representative sequences of the 24 PAS proteins, we conducted phylogenetic analyses of all AHR genes will be necessary to establish their relationships with greater certainty. Nevertheless, the present results provide support for orthology of the AHR2 genes in Fundulus and dogfish.

**Phylogenetic Analysis of the PAS Superfamily.** To examine the relationship of these new AHR sequences to sequences of other PAS proteins, we conducted phylogenetic analyses of all AHR PAS domain sequences and representative sequences of the 24 other PAS proteins from vertebrate and invertebrate animals, plants, and bacteria, as identified from the literature and through searches of the nonredundant GenBank protein sequence database (Table 2). In these analyses, the vertebrate AHRs, including all AHR1 and AHR2 genes, form a distinct clade that is strongly supported by bootstrap analysis (NJ tree, 100%; MP tree, 85%; Fig. 3). This strong clustering is the basis for our designation of all the fish sequences reported herein as AHRs.

Interestingly, a search of the GenBank protein database for sequences related to AHR PAS domains revealed that the CEC41G7 locus in the C. elegans genome (39) encodes a predicted protein (C41G7.5) that, although clearly distinct from vertebrate AHRs, is more closely related to them than to any other known member of the PAS gene family (bootstrap values of $\geq 96\%$ using either criterion; Fig. 3). Of the 82 amino acid residues that are conserved in all of the vertebrate AHR PAS domains, 40 are also conserved in this nematode protein (Fig. 1). Overall, this region of C41G7.5 shares 30–36\% identity with the PAS domains of the AHR sequences (Table 1). C41G7.5 also possesses a bHLH domain that shares 57\% amino acid identity with that of mammalian AHRs (data not shown). In addition to C41G7.5, BLAST searches revealed three other putative PAS family proteins in C. elegans, including a possible ARNT homolog (C25A1.11) that also possesses a
The AHR Is an Ancient Protein. Identification of AHR cDNA sequences in living representatives of early vertebrates (jawless, cartilaginous, and bony fish) provides evidence that the AHR is an ancient protein that existed early in vertebrate evolution, at least 450–510 million years ago. Its conservation in all vertebrate groups suggests that it serves an important function, as suggested also by recent findings of liver and immune system dysfunction after targeted disruption of an AHR gene in mice (12, 13). Although originally of interest because of its role in dioxin toxicity, the AHR likely has a more fundamental significance with regard to gene regulation, development, or other aspects of cellular homeostasis.

The identification of AHR cDNA sequences in the dogfish Mustelus confirms our previous report of an AHR protein in this species (24). In pairwise comparisons (Table 1) and phylogenetic analyses (Fig. 2), dogfish AHR1 consistently appears as the fish most divergent with our earlier ligand-binding results (24) and with the lack of concordance of gene and species phylogenies has been seen of the vertebrate AHR sequences. The AHR phylogeny does not match accepted phylogenetic relationships of these species, suggesting unequal rates of change in some lineages. A similar lack of concordance of gene and species phylogenies has been seen with the LDH-A genes of mammals, Fundulus, and another species of dogfish (40, 41).

In previous studies, we failed to detect AHR proteins by photoaffinity labeling of hepatic cytosol from adult lamprey (24). Similarly, induction of CYP1A in response to planar aromatic hydrocarbons—the “classical” AHR-dependent response—is not apparent in adult lamprey (60). The lamprey AHR sequence reported herein was obtained from the anterior section of larvae (amnioseres), suggesting that expression of the AHR may be regulated developmentally or in a cell- or tissue-specific manner in this species. Our inability to identify an AHR in adult hagfish liver in the present study is consistent with our earlier ligand-binding results (24) and with the lack of CYP1A inducibility in adult animals (42, 60). However, in light of the lamprey results, a similar AHR-related gene may yet be found in hagfish or in other invertebrate chordates.

The presence of a gene in the nematode C. elegans that bears strong similarity to vertebrate AHRs is intriguing. Because this sequence (C41G7.5) contains both bHLH and PAS domains, it appears to represent a structural homolog of the vertebrate AHR—the first such invertebrate sequence identified. Interestingly, closer examination of this sequence reveals that the
PAS-B box, which is part of the ligand-binding domain of the mammalian AHR (15, 19–22), is poorly conserved in the C. elegans C41G7.5 sequence. Thus, although the PAS-A box of C41G7.5 shares 43–50% amino acid identity with the homologous region of the vertebrate AHRs, the PAS-B box is only 25–29% identical to those of vertebrate AHRs (Fig. 1). The bHLH domain of C41G7.5 is more highly conserved with respect to the bHLH regions of vertebrate AHRs, including conservation of amino acids that have been shown (37) to be critical for DNA binding of the murine AHR (data not shown). These observations suggest that the C. elegans protein may participate in protein–protein and protein–DNA interactions that are qualitatively like those of the vertebrate AHRs but that its ligand-binding properties could be substantially different. Thus, this apparent AHR homolog in C. elegans and the possible ARNT homolog C25A1.11 may provide a system with which to examine possible ancestral functions of the AHR, especially those that may be ligand-independent.

PAS Family Gene Duplications. The presence of a duplicated AHR gene in cartilaginous and bony fish and the degree of difference between the paralogous forms are consistent with a duplication event occurring early in vertebrate evolution. Phylogenetic analysis using two different methods support the orthology of Fundulus AHR2 and dogfish AHR2, suggesting that this duplication occurred prior to the divergence of bony and cartilaginous fish. The two AHR genes may have arisen as an isolated gene duplication or, alternatively, as a result of the genome duplications that are thought to have occurred early in chordate evolution (43). Such duplications have contributed to the diversification of Hox gene clusters (44, 45) and other gene families (46–49). Because the complexity of such gene families is similar in fish and mammals (50), multiple AHRs may also occur in other vertebrates. A recent report of a second AHR in mice (51) is consistent with this hypothesis.

The existence of a second AHR is reminiscent of the two forms of other PAS proteins (ARNT, Sim) recently described in mammals (52, 53). We suggest that these and several other pairs of PAS proteins are paralogs, i.e., homologous by gene duplication (54). Thus, the following pairs of proteins share extensive amino acid identity (64–90%) in the PAS domain and cluster together in both NJ and MP trees: AHR1 ∼ AHR2, ARNT1 ∼ ARNT2, SIM1 ∼ SIM2, HIF-1α ∼ MOP2, SRC-1 ∼ TIF2, and CLOCK ∼ MOP4. Duplication of these PAS genes may have occurred at about the same time as the proposed AHR gene duplication (i.e., near the origin of the gnathostomes), consistent with the genome duplication scenario. Thus, the PAS gene family—like other gene families (47, 49)—appears to contain sets of related genes (paralog groups), which might exhibit some degree of functional redundancy (28, 55). Such redundancy has been suggested to occur within the ARNT1 ∼ ARNT2 pair (56), possibly in conjunction with the hypoxia-responsive paralogs HIF-1α and MOP2/EPAS1/HLF (6, 7, 57).

Molecular Evolution of the PAS Gene Family. Recent findings suggest that the PAS domain had its origin in early photoreceptor proteins, the descendants of which exist in modern bacteria, fungi, and plants (8, 10, 11, 27). Some of these proteins may have subsequently become involved in regulation of circadian rhythms (8, 9, 27). In animals, PAS domain-containing proteins and their functions have diversified fur-
ther, evolving roles in development and the response to environmental variables, including oxygen tension (hypoxia) and small ligands (dioxin). In the phylogenetic analysis reported herein, we identify several clusters of metazoan PAS proteins, including invertebrate orthologs of vertebrate PAS proteins, that suggest evolutionary and possibly functional relationships. Bradfield and coworkers (29) recently presented a phylogenetic analysis of 16 PAS members. Our analysis of 26 PAS proteins confirms some, but not all, of their groupings and reveals additional relationships. Our trees are consistent with an initial diversification of the PAS family in invertebrates, followed by extensive gene duplication and further diversification in early vertebrates. Because of the rapid pace at which new PAS family members are being discovered (e.g., refs. 8, 9, 29, 57, and 58), a definitive description of evolutionary relationships within this family must await a more complete cataloguing of its members and will require continuing phylogenetic analyses.

Conclusions. The vertebrate AHR plays a critical role in susceptibility to dioxin toxicity (59), but its conservation in all vertebrate groups suggests that it has a more fundamental role in cellular physiology. The existence of a second AHR-like gene in fish and mammals raises questions concerning the functions and possible interactions of these two genes. Understanding the phylogenetic relationships among these AHR genes and other members of the PAS family may provide an evolutionary context within which to interpret the functions of these proteins in gene regulation, development, environmental homeostasis, and toxicity.

We thank Drs. J. Stegeman, W. Powell, and H. Morrison for comments on the manuscript. This work was supported in part by the National Institute of Environmental Health Sciences (Grants R29 ES06272, F32 ES05644, and P42 ES07381), the Donaldson Charitable Foundation, and a grant from the Illinois Natural History Survey (Illinois Natural History Survey, Champaign, IL). We thank Drs. J. Stegeman, W. Powell, and H. Morrison for comments on the manuscript. This work was supported in part by the National Institute of Environmental Health Sciences (Grants R29 ES06272, F32 ES05644, and P42 ES07381), the Donaldson Charitable Foundation, and a grant from the Illinois Natural History Survey (Illinois Natural History Survey, Champaign, IL).