

# Evolutionary analyses of hedgehog and *Hoxd-10* genes in fish species closely related to the zebrafish

(developmental genes/Sonic hedgehog/functional domains/homeobox/paralogues)

RAFAEL ZARDOYA, EHAB ABOUHEIF, AND AXEL MEYER\*

Department of Ecology and Evolution and Program in Genetics, State University of New York, Stony Brook, NY 11794-5245

Communicated by Ernst Mayr, Museum of Comparative Zoology, Cambridge, MA, August 22, 1996 (received for review April 19, 1996)

**ABSTRACT** The study of development has relied primarily on the isolation of mutations in genes with specific functions in development and on the comparison of their expression patterns in normal and mutant phenotypes. Comparative evolutionary analyses can complement these approaches. Phylogenetic analyses of Sonic hedgehog (*Shh*) and *Hoxd-10* genes from 18 cyprinid fish species closely related to the zebrafish provide novel insights into the functional constraints acting on *Shh*. Our results confirm and extend those gained from expression and crystalline structure analyses of this gene. Unexpectedly, exon 1 of *Shh* is found to be almost invariant even in third codon positions among these morphologically divergent species suggesting that this exon encodes for a functionally important domain of the hedgehog protein. This is surprising because the main functional domain of *Shh* had been thought to be that encoded by exon 2. Comparisons of *Shh* and *Hoxd-10* gene sequences and of resulting gene trees document higher evolutionary constraints on the former than on the latter. This might be indicative of more general evolutionary patterns in networks of developmental regulatory genes interacting in a hierarchical fashion. The presence of four members of the hedgehog gene family in cyprinid fishes was documented and their homologies to known hedgehog genes in other vertebrates were established.

Many vertebrate homologues of *Drosophila* developmental regulatory genes have recently been identified (1–4). These discoveries provide intriguing evidence that comparable ontogenetic processes, even in species from different phyla, can be regulated by homologous and evolutionarily conserved signaling factors. Despite their high degree of sequence conservation and their similarity of interactions in developmental networks, the expression of these homologous developmental control genes results in drastically a divergent Baupläne like those of insects and vertebrates. The homology and often extensive degree of sequence similarity illustrates the somewhat paradoxical contradiction between biological diversification among different animal phyla and the phylogenetic conservation of some developmental genes and their interactions.

Originally identified by Nüsslein-Volhard and Wieschaus (5) in *Drosophila*, the hedgehog (*hh*) gene has emerged as one of the most interesting and important developmental genes characterized so far. In *Drosophila*, *hh* has been reported to regulate embryonic segmentation and patterning (6), whereas in vertebrates one member of the *hh* gene family, the protein encoded by the Sonic hedgehog (*Shh*) gene, controls several developmental processes including dorsal–ventral patterning of the neural tube (7), left–right distinction (8, 9), and limb bud morphogenesis (10). One of the most interesting long-range signaling activities of the Shh protein is the regulation of limb bud patterning in vertebrate embryos (1, 3, 11–13). Recent evidence indicates that the effect of *Shh* signal on mesenchy-

mal limb bud proliferation is indirect and is mediated through a wide variety of secondary transductional factors (10). Notably, the expression of posterior members of the HoxD gene cluster (i.e., *Hoxd-10* to *Hoxd-13*) is associated with *Shh* gene expression, in the early developmental stages of the mesoderm primordium of fish and tetrapods (14, 15). Furthermore, differential late distal expression patterns of *Shh* and *Hoxd-10* to *Hoxd-13* genes in fish and tetrapod limb buds is responsible for the development of fins and limbs (15). Both hh and Shh proteins are expressed as precursors that undergo self-cleaving and secretion events (16, 17). After autoprocessing, N-terminal and C-terminal fragments are generated and locally released. The N-terminal processed form corresponds to exons 1 and 2 and seems to be implicated in short- as well as long-range signaling activities, whereas the C-terminal portion of the protein is encoded by exon 3 and has the autoproteolytic activity (18).

The recently determined crystal structure of the N-terminal signaling domain of Shh revealed the presence of an unsuspected zinc-coordinated catalytic site (19). The discovery of this catalytic site, which had not been predicted based on the amino acid sequences, emphasizes that Shh might have features and functions that still await discovery. Herein, we use an alternative approach to improve our understanding of both the evolution of ontogenetic processes and the developmental genes that regulate them (20). This method involves the study of the evolutionary history of developmental genes to characterize the evolutionary and presumably functionally permissible variation within them and to discover conserved functional domains.

## MATERIALS AND METHODS

**Samples and DNA Extraction.** Total DNA was extracted as described (21) from muscle of individuals of 18 cyprinid species related to the zebrafish (*Danio rerio*): *Danio frankei*, *Danio kerri*, *Danio pulcher*, *Danio* sp. aff. *albolineatus*, *Danio* sp. aff. *tweediei*, *Devario devario*, *Devario* cf. *aequipinnatus*, *Devario pathirana*, *Devario malabaricus*, *Rasbora heteromorphia*, *Rasbora elegans*, *Rasbora paviei*, *Amblypharyngodon chulabhornae*, *Pseudorasbora* cf. *parva*, *Tanichthys albonubes*, *Puntius tetrazona*, *Puntius conchoni*, and *Carassius auratus*.

**PCR Amplification, Cloning, and DNA Sequencing.** Four sets of PCR primers were designed to consistently amplify *Shh* exon 1 (Shh1-F, 5'-CTGGCCTGTGGTCCCTGGCAGAGG-3'; Shh1-R, 5'-CTGAGTCATGAGCCGGTCC GCTCC-3'), *Shh* exon2 (Shh2-F, 5'-GACAAGCTGAACGCCTGGCC-ATCTC-3'; Shh2-R, 5'-CTTTGGACTCGTAATAGACCC-

Abbreviations: MP, maximum parsimony; NJ, neighbor joining; ML, maximum likelihood.

Data deposition: The sequences reported in this paper have been deposited in the GenBank data base (accession nos. U51339–U51423 and U68236–U68241).

\*To whom reprint requests should be addressed. e-mail: ameyer@life.bio.sunysb.edu.

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

AGT-3'), *Hoxd-10* exon 1 (*Hoxd10.1-F*, 5'-ATGTCCTTTC-CCAACAGCTCTCC-3'; *Hoxd 10.1-R*, 5'-TTTGCCTTCTC-TGTGTGGCAATT-3'), and *Hoxd-10* exon 2 (*Hoxd10.2-F*, 5'-GCAGAATCTAAAAACGACACACC-3'; *Hoxd10.2-R*, 5'-CTAGGTTTTTGGATTTGCACTTGT-3'). Amplification by the polymerase chain reaction (PCR) was conducted with 30 PCR cycles (denaturing at 94°C for 60 s, annealing at 50–56°C for 60 s, and extending at 72°C for 60 s). An aliquot of the PCR product was cloned with the pGEM-T vector (Promega) following the manufacturer's instructions. Positive clones were sequenced in both directions with an Applied Biosystems 373A Stretch DNA sequencer using the Taq Dye Deoxy Terminator Cycle Sequencing kit (Applied Biosystems) and M13 universal (-40) and reverse sequencing primers following the manufacturer's instructions. Sequencing errors resulting in frame shifts in the originally reported *Hoxd-10* zebrafish sequence (15) were discovered shortening the actual size of the gene.

**Sequence Analyses.** A multiple alignment was performed using CLUSTAL W (22). DNA sequence data were analyzed (23) by maximum parsimony (MP) method (23) (Heuristic search, TBR, MULPARS, 10 random stepwise addition) with PAUP version 3.1.1 (23), and by neighbor joining (NJ) (24) (based on Kimura distance matrixes; jumble option in effect) and maximum likelihood (ML) (25) (Ti:Tv of 2:1; one category of substitution rates; input order randomized 10 times) methods using PHYLIP version 3.5 (25). The robustness of the inferred trees was assessed by the bootstrap method (26) as implemented in PAUP and PHYLIP with 500 pseudoreplications. Establishment of orthologous relationships among the different members of the vertebrate hedgehog gene family by the NJ

method (24) was conducted with PAUP version 4.0d47 (27) using the midpoint rooting option and 1000 bootstrap pseudoreplications.

**RESULTS AND DISCUSSION**

***Shh* Gene Is Phylogenetically More Constrained Than *Hoxd-10*.** We amplified via the PCR, exons 1 and 2 of *Shh* and the two exons of the *Hoxd-10* gene in 18 cyprinid fish species that are closely related to the zebrafish (*Danio rerio*), an important model system in vertebrate developmental biology (28). Interestingly, for both sets of genes, most of the nucleotide variation at this level of evolutionary comparison is limited to synonymous substitutions (replacement/silent substitutions ratios: *Shh* exon 1 = 1:5; *Shh* exon 2 = 2:5; *Hoxd-10* exon 1 = 1:1; *Hoxd-10* exon 2 = 1:3). When all these gene sequences were combined, all three commonly used methods of phylogeny reconstruction arrived at identical hypotheses about the evolutionary relationships among these genes and the species that contain them (Fig. 1A). These results are in almost complete agreement with those based on morphological (29) and mitochondrial DNA characters (20, 30).

In the zebrafish and its relatives, *Shh* and *Hoxd-10* genes were found to contain different levels of DNA variation (proportion of variable sites: *Shh* = 12.6%; *Hoxd-10* = 20.1%) and phylogenetic information (Fig. 1B and C). The level of mutation variation of *Shh* exons 1 and 2 is lower than that of *Hoxd-10* exons 1 and 2 (Fig. 1B and C) indicating that this upstream control gene is more stringently constrained in its molecular evolution than is the downstream effector gene. Moreover, the observed permissible occurrence of insertion

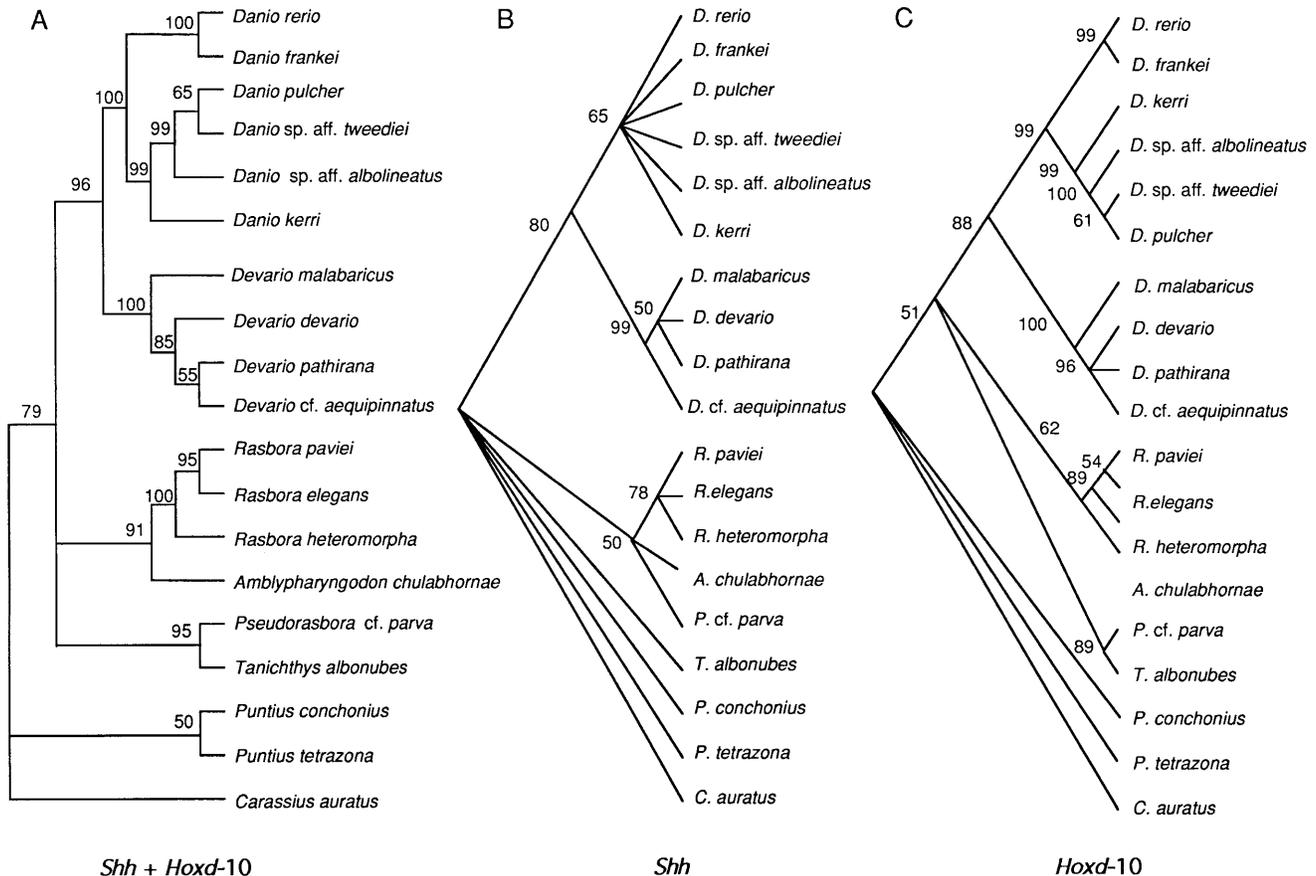


FIG. 1. Majority-rule bootstrap MP tree (500 replications; 10 random stepwise addition of taxa, TBR branch swapping, MULPARS) (23) based on (A) the combined data set of *Shh* gene exons 1 and 2 and *Hoxd-10* exons 1 and 2 (about 1300 bp per species); (B) *Shh* exons 1 and 2 combined; and (C) *Hoxd-10* exons 1 and 2 combined. Nodes with bootstrap values below 50% were forced to collapse and yield polytomies. NJ and ML methods arrived at identical topologies to that shown for MP. Goldfish (*Carassius auratus*) was used as an outgroup for all trees.

and deletion events in vertebrate *Hoxd-10* (31, 32) but not in *Shh* (1–3, 33) reinforces the notion of an extremely high degree of sequence conservation in *Shh* genes. It will be interesting to investigate whether these divergent patterns of evolutionary constraints on upstream and downstream regulator genes are more widespread than previously recognized. These different levels of evolutionary and functional constraint may be due to the effect of the hierarchical position of these genes in the nexus of interactions that control developmental events. Alternatively, these divergent patterns may be associated with the biochemical properties that are demanded by the secreted and long-range (*Shh*) or intracellular and short-range (*Hoxd-10*) signaling character of these proteins.

***Shh* Gene Exon 1 Is Evolutionarily Conserved in Vertebrates.** An important and unanticipated outcome of our evolutionary analyses of the *Shh* genes is that exon 1 is at least as conserved as exon 2 (proportion of variable sites: *Shh* exon 1 = 9.5%; *Shh* exon 2 = 16.0%; see Fig. 2); the latter is thought to encode the major functional domain of this gene (3, 19). The extremely high evolutionary conservation of *Shh* exon 1 that extends even to third codon positions among the relatives of the zebrafish suggests that there are strong selective constraints against changes in this region and might imply a functional role for the DNA sequence itself. One, among other, potential explanations for this strict conservation might be that fidelity of codon recognition by the most abundant tRNA types is required since developmental genes are often

expressed at high levels during short periods of time in early development (34).

Krauss *et al.* (3) compared the predicted amino acid sequences of zebrafish *Shh* and *Drosophila hh* and identified a highly conserved domain in exon 2 (77% identity) that was proposed to be essential for the signaling activity of the protein. We find that the exon 2 domain shares 90% amino acid identity among vertebrates. The same comparisons for exon 1 show that zebrafish *Shh* and *Drosophila hh* exon 1 domains share only 55% amino acid identity, whereas zebrafish *Shh* and its vertebrate counterparts {mouse (1), chicken (2), frog (33), rat (35), and human [V. Marigo, D. J. Roberts, S. M. K. Lee, O. Tsukurov, T. Levi, J. M. Gastier, D. J. Epstein, D. J. Gilbert, G. G. Martin, N. G. Copeland, C. E. Seidman, N. A. Jenkins, J. G. Seidman, A. P. McMahon, and C. Tabin (1995) GenBank accession nos. L38517 and L38518.]} share 97% amino acid identity, an even higher level than that of exon 2. These comparisons suggest that the functional domain of exon 1 might be a novel acquisition of vertebrates. This hypothesis can be tested by determining and comparing the rates of evolution of both exons in invertebrate phyla.

**Vertebrate Hedgehog Gene Family.** Two paralogues (Fig. 3) of *Shh* exon 2 were identified by sequencing up to 30 clones for each of the zebrafish-related species. For example, the cloning and sequencing of 30 clones of the *Danio rerio* PCR product yielded 11 *Shh*, 16 Indian *hh* (*Ihh*) and 2 Desert *hh* (*Dhh*) clones [following the nomenclature established by Echelard *et*

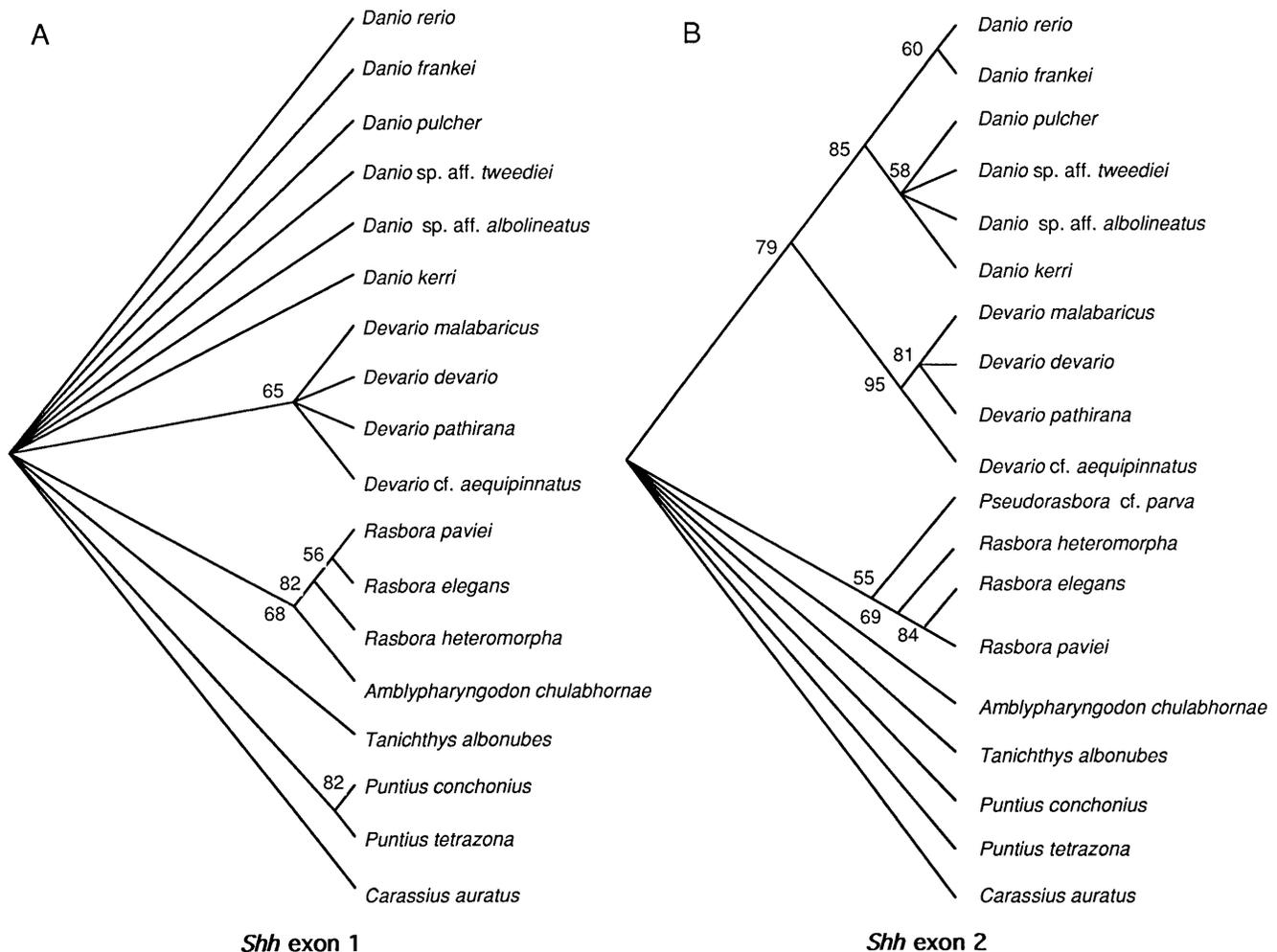


FIG. 2. Fifty percent majority-rule NJ bootstrap tree (500 replications) (24) and groupings compatible with it, based on *Shh* exon 1 (A) and *Shh* exon 2 (B). Nodes with bootstrap values below 50% were forced to collapse. *Shh* exon 1 was found to have less phylogenetic resolution than *Shh* exon 2, reflecting its extremely low rate of evolution. Hence, it is likely that both exon 1 and exon 2 domains are functionally important. Goldfish (*Carassius auratus*) was used as an outgroup for all trees.

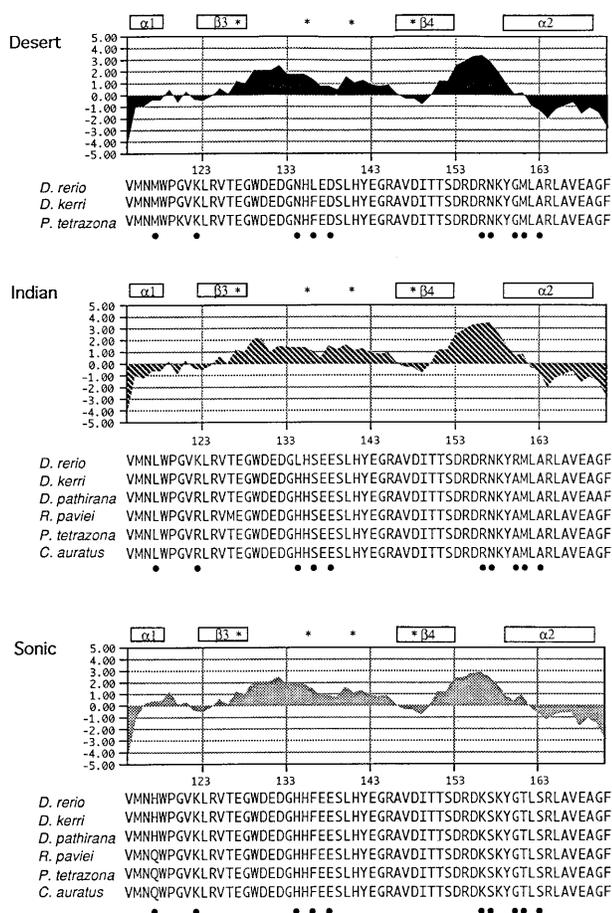


FIG. 3. Hydrophatic profiles of cyprinid hedgehog exon 2 paralogues. PCR primers designed to amplify *Shh* exon 2 within cyprinids, also amplified two other exon 2 paralogues [*Ihh* and *Dhh* according to the nomenclature of Echelard *et al.* (2)]. The hydrophobicity of the deduced amino acid sequences for all exon 2 paralogues was calculated by the Kyte–Doolittle method (36) using the MACVECTOR program (IBI). Secondary structures (19) assigned to *Shh* exon 2 are shown. Coordinating and noncoordinating residues likely to be important for catalysis are indicated by an asterisk. Positions that characterize each of the paralogues are indicated by a dot.

*al.*(2)]. These two paralogues are the same as those described by Krauss *et al.* (3) from zebrafish. A fourth member of the hedgehog family, tigg-y-winkle hedgehog (*twhh*), has recently been found in zebrafish (37); however, with this set of primers we were unable to detect it in any of the species assayed. Nevertheless, we were able to amplify *twhh* in one representative species from each genus (with the exception of *Amblypharyngodon* and *Rashora*) with specifically designed primers that discriminate this paralogue from others (*twhh*-2F, 5'-CGCTGTAAGGACAAGTTA-3' and *twhh*-2R, 5'-TGCAAGCCTGGATAGCA-3'). In some relatives of the zebrafish, two alleles were found suggesting that this developmental gene is likely to contain variation at the population level, as has been recently demonstrated for the *Drosophila Ubx* gene (38). Comparisons of the deduced amino acid sequences for each of the paralogues (Fig. 3) revealed that changes among paralogues are localized at the edge of  $\alpha$ -helix 1 and  $\beta$ -sheet 3, in the core of  $\alpha$ -helix 2, and in the loops connecting those structures (19). The putative residues involved in the zinc coordination site (19) are highly conserved among all three paralogues. The exon 2 domain for each of the paralogues is highly hydrophilic as shown by the deduced hydrophatic profiles (36) (Fig. 3). Interestingly, differences at the amino acid level among paralogues do not affect the hydrophatic character and the putative tertiary structure of the domain (Fig. 3). This

might imply that the other hedgehog family members may have similar regulatory functions to those of *Shh*; however, this hypothesis remains to be tested by comparing expression patterns as well as phenotypes induced by ectopic expression assays. There is increasing evidence that the selective maintenance of functionally redundant paralogous copies of *Shh* may be explained by the tissue-specific expression of particular paralogues—i.e., *Shh* is mainly expressed in notochord, floor plate, and limb (e.g., refs. 2 and 3), whereas *twhh* is detected in floor plate (37), *Ihh* is expressed in gut and cartilage (39, 40), and *Dhh* is expressed in Sertoli cell precursors (41).

The existence of several members of the hedgehog gene family has also been reported for chicken (1, 40), mouse (2), frog (33), and human [V. Marigo *et al.* (1995) GenBank accession nos. L38517 and L38518]. The phylogenetic analysis of the different cognate members of the hedgehog gene family from vertebrates allowed us to establish likely homology relationships among members of this gene family (Fig. 4). The complete amino acid sequences (exons 1–3) of the known vertebrate hedgehog genes were subjected to all three commonly used methods of phylogenetic inference (MP, NJ, and ML). Ambiguous alignments (mainly in exon 3) were excluded from the analyses. By using our results, the members of the vertebrate hedgehog gene family can be classified into three major orthology groups, *Shh*, *Ihh*, and *Dhh* (Fig. 4). Previous studies have suggested homologous relationships for some of the hedgehog gene family based on simple comparisons of percentage identity (e.g., refs. 2, 3, 33, and 37). However, these reported relationships should be interpreted with caution since both convergent and divergent autopolymorphisms (i.e., unique evolutionary changes) can influence the degree of similarity and can potentially lead to false assignments of homology based on this criterion. For example, residue 122 in zebrafish *Dhh* (lysine) (Fig. 3) might have been considered constant in all three paralogues increasing the percentage of identity of this copy to the other two zebrafish hedgehog gene family members. However, the amino acid sequences of other cyprinid fish Indian hedgehog copies found this residue to be a convergent homoplasious change in *Dhh* and *Shh* copies. Therefore, statements about homology among members of a gene family must be based on gene-tree phylogenetic analyses (Fig. 4). Kumar *et al.* (43) also recently analyzed the evolution of a smaller set of vertebrate hedgehog genes and report finding similar relationships to ours. Correct homology assignment is crucial for the comparative study of function and inferences about evolutionary shifts in function; i.e., changes in spatial or temporal expression patterns among paralogues of a multigene family. Establishing orthology on the basis of similar expression patterns in time or space has led to the unfortunate concept of “functional homology” (44, 45). This is not a valid criterion to establish true orthologous relationships since paralogous members of a gene family might serve similar functions in different species or, alternatively the functional roles of orthologous genes can change through evolution. Case in point is the Echidna hedgehog (*Ehh*) gene that was recently discovered in zebrafish (42). The expression of this gene is necessary for the induction of muscle pioneers in zebrafish (42). However, although this gene performs novel functions in development, it does not belong to an entirely new group of vertebrate hedgehog genes, as had been suggested (42), but rather appears to be the *Ihh* orthologue in zebrafish (Fig. 4). Therefore, similarity or dissimilarity in function are not a reliable indicators of homology, a relationship that is due to shared evolutionary history and can only be discovered through evolutionary analyses.

As far as is known, there is no evidence for the existence of more than one copy of *hh* in invertebrates (46, 47). According to our analyses (Fig. 4), the three major members of the hedgehog gene family were likely generated by two gene

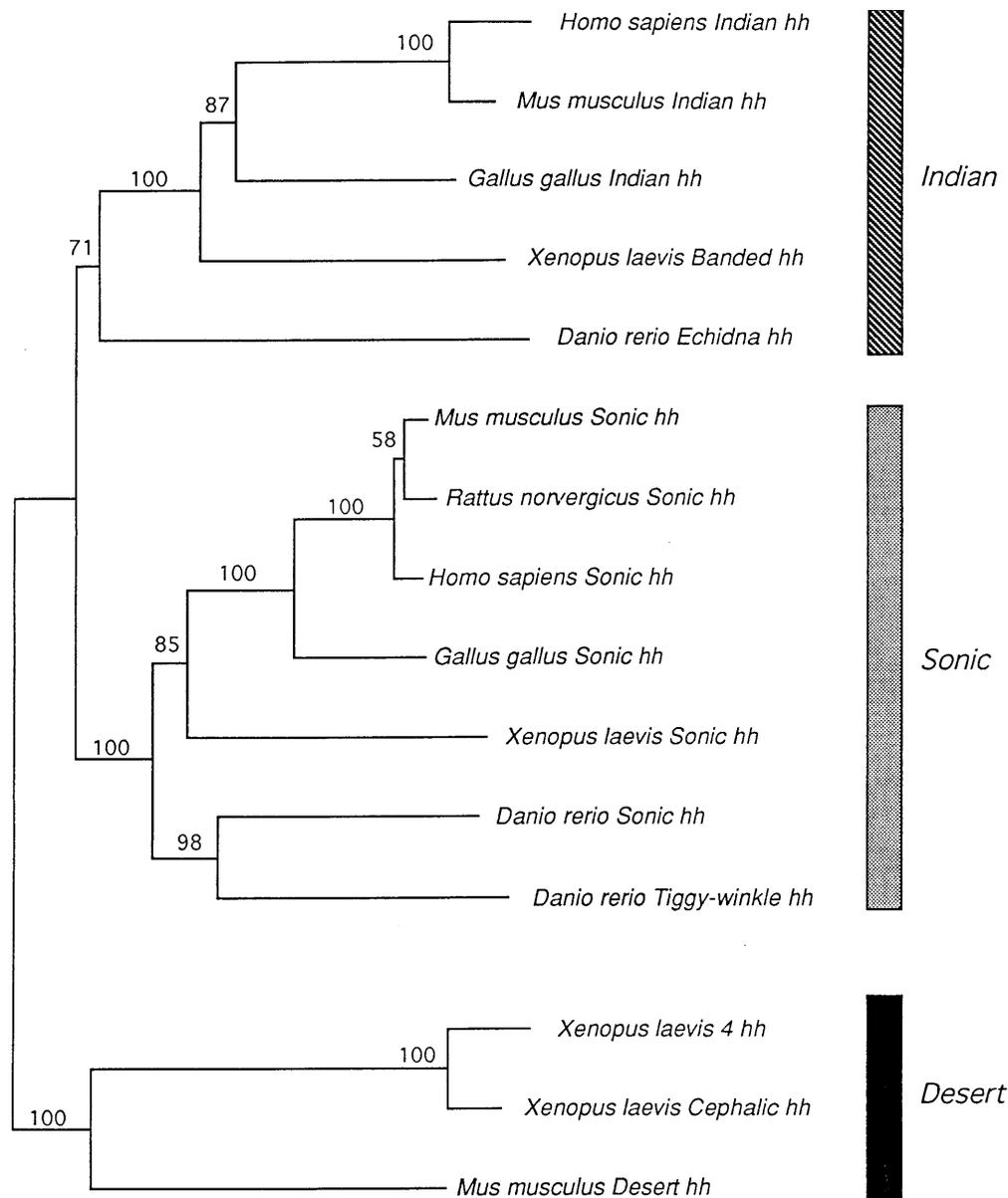


FIG. 4. Homologous relationships among members of the vertebrate hedgehog gene family. Majority-rule bootstrap NJ tree (1000 replications). Numbers above branches indicate bootstrap values. Midpoint rooting option of PAUP Version 4.0d47 (27) was used to root the NJ tree. Branch lengths are proportional to distances. According to this NJ tree, sequences were classified into three groups: *Dhh*, *Ihh*, and *Shh* as shown by three distinct shaded bars. The existence of these three orthology clades is corroborated by MP and ML methods (data not shown). Nomenclature of the paralogues follows that of the original studies [refs. 2, 3, 33, 37, 40, and 42; and V. Marigo *et al.* (1995) GenBank accession nos. L38517 and L38518].

duplication events in which *Shh* and *Ihh* arose more recently, whereas the *Dhh* paralogue is the most basal member of this gene family. There appear to have been recent duplication events of the *Shh* gene in zebrafish that resulted in a new member of the hedgehog gene family, *twhh* (37), and of the *Dhh* gene in frog that resulted in a new member, *4hh*. It has been suggested that the establishment of the vertebrate hedgehog gene family clearly preceded the evolution of vertebrates and possibly chordates (43). Future studies might want to focus on the precise phylogenetic timing of hedgehog gene duplication events to elucidate whether these events are correlated with increasing morphological complexity during evolution such those that have been suggested at a macroevolutionary scale for homeobox genes in Hox-gene clusters (48, 49).

P. W. H. Holland, S. V. Edwards, M. Peng-Pilou, D. Lans, D. Wake, and D. Weisblat provided insightful comments on the manuscript. P. J. Perl assisted in the DNA cloning and sequencing. D. Swofford granted

permission to publish results based on the test version of his PAUP\* program. R.Z. is a postdoctoral fellow of the Ministerio de Educacion y Ciencia of Spain, and E.A. is supported by the Fonds pour la Formation de Chercheurs et l'Aide à la Recherche of the government of Quebec. Partial financial support came from the National Science Foundation and the Max-Planck-Society, Germany, to A.M. This publication was prepared during A.M.'s tenure as John Simon Guggenheim Fellow and a Miller Visiting Research Professor at the University of California, Berkeley. The kind hospitality of the Departments of Integrative Biology and Molecular and Cell Biology and the Miller Institut at Berkeley is gratefully acknowledged.

1. Riddle, R. D., Johnson, R. L., Laufer, E. & Tabin, C. (1993) *Cell* **75**, 1401-1416.
2. Echelard, Y., Epstein, D. J., St-Jacques, B., Shen, L., Mohler, J., McMahon, J. A. & McMahon, A. P. (1993) *Cell* **75**, 1417-1430.
3. Krauss, S., Concordet, J. P. & Ingham, P. W. (1993) *Cell* **75**, 1431-1444.
4. Quiring, R., Walldorf, U., Kloter, U. & Gehring, W. J. (1994) *Science* **265**, 785-789.

5. Nüsslein-Volhard, C. & Wieschaus, E. (1980) *Nature (London)* **287**, 795–801.
6. Tabata, T. & Kornberg, T. B. (1994) *Cell* **76**, 89–102.
7. Ericson, J., Muhr, J., Placzek, M., Lints, T., Jessell, T. M. & Edlund, T. (1995) *Cell* **81**, 747–756.
8. Yost, H. J. (1995) *Cell* **82**, 689–692.
9. Levin, M., Johnson, R. L., Stern, C. D., Kuehn, M. & Tabin, C. (1995) *Cell* **82**, 803–814.
10. Tabin, C. (1995) *Cell* **80**, 671–674.
11. Duboule, D. (1994) *Science* **266**, 575–576.
12. Fietz, M. J., Concordet, J. P., Barbosa, R., Johnson, R., Krauss, S., McMahon, A. P., Tabin, C. & Ingham, P. W. (1994) *Development (Cambridge, U.K.)* **120**, Suppl. 43–51.
13. Smith, J. C. (1994) *Cell* **76**, 193–196.
14. Morgan, B. A. & Tabin, C. (1994) *Development (Cambridge, U.K.)* **120**, Suppl. 181–186.
15. Sordino, P., van der Hoeven, F. & Duboule, D. (1995) *Nature (London)* **375**, 678–681.
16. Lee, J. J., Ekker, S. C., von Kessler, D. P., Porter, J. A., Sun, B. I. & Beachy, P. A. (1994) *Science* **266**, 1528–1537.
17. Roelink, H., Porter, J. A., Chiang, C., Tanabe, Y., Chang, D. T., Beachy, P. A. & Jessell, T. M. (1995) *Cell* **81**, 445–455.
18. Johnson, R. L. & Tabin, C. (1995) *Cell* **81**, 313–316.
19. Tanaka Hall, T. M., Porter, J. A., Beachy, P. A. & Leahy, D. J. (1995) *Nature (London)* **378**, 212–216.
20. Meyer, A., Ritchie, P. A. & Witte, K. E. (1995) *Philos. Trans. R. Soc. London* **349**, 103–111.
21. Kocher, T. D., Thomas, W. K., Meyer, A., Edwards, S. V., Paabo, S., Villablanca, F. X. & Wilson, A. C. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 6196–6200.
22. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 673–680.
23. Swofford, D. L. (1993) PAUP: Phylogenetic Analysis Using Parsimony (Illinois Natural History Survey, Champaign, IL), Version 3.1.1.
24. Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4**, 406–425.
25. Felsenstein, J. (1989) *Cladistics* **5**, 164–166.
26. Felsenstein, J. (1985) *Evolution* **39**, 783–791.
27. Swofford, D. L. (1996) PAUP Star: Phylogenetic Analysis Using Parsimony (Sinauer, Sunderland, MA), Version 4.0d47.
28. Mullins, M. C., Hammerschmidt, M., Haffter, P. & Nüsslein-Volhard, C. (1994) *Curr. Biol.* **4**, 189–202.
29. Chu, X. I. (1981) *Zool. Res.* **2**, 145–156.
30. Meyer, A., Biermann, C. H. & Orti, G. (1993) *Proc. R. Soc. London B* **252**, 231–236.
31. Simon, H. G. & Tabin, C. J. (1993) *Development (Cambridge, U.K.)* **117**, 1397–1407.
32. Garcia-Fernandez, J. & Holland, P. H. (1994) *Nature (London)* **370**, 563–566.
33. Ekker, S. C., McGrew, L. L., Lai, C. H., Lee, J. J., von Kessler, D. P., Moon, R. T. & Beachy, P. A. (1995) *Development (Cambridge, U.K.)* **121**, 2337–2347.
34. Akashi, H. (1994) *Genetics* **136**, 927–935.
35. Roelink, H., Augsburg, A., Heemskerk, J., Korzh, V., Norlin, S., Ruiz i Altaba, A., Tanabe, Y., Placzek, M., Edlund, T., Jessell, T. M. & Dodd, J. (1994) *Cell* **76**, 761–775.
36. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132.
37. Ekker, S. C., Ungar, A. R., Greenstein, P., von Kessler, D. P., Porter, J. A., Moon, R. T. & Beachy, P. A. (1995) *Curr. Biol.* **5**, 944–955.
38. Gibson, G. & Hogness, D. S. (1996) *Science* **271**, 200–202.
39. Bitgood, M. J. & McMahon, A. P. (1995) *Dev. Biol.* **172**, 126–138.
40. Vortkamp, A., Kaechoong, L., Lanske, B., Segre, G. V., Kronenberg, H. M. & Tabin, C. J. (1996) *Science* **273**, 613–622.
41. Bitgood, M. J., Shen, L. & McMahon, A. P. (1996) *Curr. Biol.* **6**, 298–304.
42. Currie, P. D. & Ingham, P. W. (1996) *Nature (London)* **382**, 452–455.
43. Kumar, S., Balczarek, K. A. & Zhi-Chun, L. (1996) *Genetics* **142**, 965–972.
44. Dickinson, W. J. (1995) *Trends Genet.* **11**, 119–122.
45. Bolker, J. A. & Raff, R. A. (1996) *BioEssays* **18**, 489–494.
46. Mohler, J. & Vani, K. (1992) *Development (Cambridge, U.K.)* **115**, 957–971.
47. Chang, D. T., Lopez, A., von Kessler, D. P., Chiang, C., Simandl, B. K., Zhao, R., Seldin, M. F., Fallon, J. F. & Beachy, P. A. (1994) *Development (Cambridge, U.K.)* **120**, 3339–3353.
48. Holland, P. W. H. (1994) *BioEssays* **14**, 267–273.
49. Carroll, S. B. (1995) *Nature (London)* **376**, 479–485.