Evolution of the *hedgehog* Gene Family

Sudhir Kumar,* Kristi A. Balczarek*,† and Zhi-Chun Lai*,†

**Institute of Molecular Evolutionary Genetics and Department of Biology, †Department of Biochemistry and Molecular Biology, The Pennsylvania State University, PA 16802*

Manuscript received August 30, 1995
Accepted for publication November 21, 1995

ABSTRACT

Effective intercellular communication is an important feature in the development of multicellular organisms. Secreted *hedgehog* (*hh*) protein is essential for both long- and short-range cellular signaling required for body pattern formation in animals. In a molecular evolutionary study, we find that the vertebrate homologs of the Drosophila *hh* gene arose by two gene duplications: the first gave rise to *Desert hh*, whereas the second produced the *Indian* and *Sonic hh* genes. Both duplications occurred before the emergence of vertebrates and probably before the evolution of chordates. The amino-terminal fragment of the *hh* precursor, crucial in long- and short-range intercellular communication, evolves two to four times slower than the carboxyl-terminal fragment in both Drosophila *hh* and its vertebrate homologues, suggesting conservation of mechanism of *hh* action in animals. A majority of amino acid substitutions in the amino- and carboxyl-terminal fragments are conservative, but the carboxyl-terminal domain has undergone extensive insertion-deletion events while maintaining its autocleavage protease activity. Our results point to similarity of evolutionary constraints among sites of Drosophila and vertebrate *hh* homologs and suggest some future directions for understanding the role of *hh* genes in the evolution of developmental complexity in animals.

MORPHOGENESIS and pattern formation require intercellular communication for which extracellular signaling molecules are essential. The *hedgehog* (*hh*) gene was first identified as an embryonic segment polarity gene in Drosophila (NÜSSLEIN-VOLHARD and WIESCHAUS 1980). It encodes a secreted protein product that provides neighboring cells with positional information in embryos and imaginal discs (reviewed in INGHAM 1994; HEBERLEIN and MOSES 1995; PERRIMON 1995). Homologs of the Drosophila *hh* gene have been identified in other invertebrates (*e.g.*, sea urchin, leech, and beetle) and in vertebrates (*e.g.*, zebrafish, chicken, and mammals). One of the vertebrate *hh* protein products, *Sonic hh* (*Shh*), has been found to play crucial roles in the development of neural tube, somites, and limbs (ECHELARD *et al.* 1993; KRAUSS *et al.* 1993; RIDDLE *et al.* 1993; CHANG *et al.* 1994; ROELINK *et al.* 1994; SMITH 1994).

The secreted *hh* gene product has dual functions, one being a short-range contact-dependent inducer for maintaining gene expression in neighboring cells, and the other being a long-range contact-independent signaling activity (see JOHNSON and TABIN 1995). A protease cleavage site in the *hh* precursor protein has been identified and the resulting amino-terminal fragment has been shown to be active in both short- and long-range signaling in Drosophila, while the carboxyl-terminal peptide is required for autocleavage and may be involved in regulating the range of action of the amino terminal peptide (FIETZ *et al.* 1994; LEE *et al.* 1994; PORTER *et al.* 1995). Similarly, vertebrate *hh* precursor proteins are processed to generate two distinct domains (BUMCROT *et al.* 1995; PORTER *et al.* 1995). In one example, the mouse amino-terminal domain of *hh* protein, but not the carboxyl domain, has been shown to induce distinct ventral cell types in a cultured CNS system (MARTI *et al.* 1995; ROELINK *et al.* 1995).

The *hh* action is coordinated with other signaling molecules that include members of *Wnt* and *TGF-β* super gene families. *Shh* expression in the vertebrate limb bud has been shown to be induced by *Wnt-7a* and *FGF-4* expressed in neighboring tissues (PARR and MCMAHON 1995; YANG and NISWANDER 1995). It is possible that bone morphogenetic protein 2 (BMP2), a member of *TGF-β* gene family, also participates with *Shh* in controlling vertebrate limb development (FRANCIS *et al.* 1994). In Drosophila, *hh* activity is critical for the expression of *dpp* (*decapentaplegic*) and *wg* (*wingless*) genes. For instance, *dpp* expression near the morphogenetic furrow in eye imaginal discs is abolished in *hh* mutants, resulting in failure of eye formation (HEBERLEIN *et al.* 1993; MA *et al.* 1993). Thus, the coordinated action of *hh* seems essential in many aspects of animal development, such as embryonic segmentation, limb outgrowth, and retinal patterning.

Even though the knowledge of *hh* signaling has been rapidly accumulating, many questions remain unanswered. For instance, it is not clear how widely the *hh* gene and its mechanisms of action have been conserved in evolution, and how two other vertebrate *hh* genes

Corresponding author: Zhi-Chun Lai, 208 Mueller Laboratory, The Pennsylvania State University, University Park, PA 16802.
E-mail: zcl1@psuvm.psu.edu

(*Indian* and *Desert*) are evolutionarily and functionally related to *Sonic hh*. By using a molecular evolutionary approach, we have reconstructed the evolutionary history, estimated the rates of amino acid substitution at individual sites, and inferred the relative frequency of different amino acid substitutions to understand the forces that have shaped the evolution of the *hh* gene family. We have also estimated the approximate times of major evolutionary events that led to different vertebrate homologs of the fly *hh* gene in an attempt to assess the ubiquity of these genes in the animal kingdom. The results presented may be useful in devising future molecular genetic experiments to study the function(s) of *hh* gene products and to probe for the presence of *hh*-like genes in unexplored invertebrate lineages.

MATERIAL AND METHODS

Amino acid sequence data: We obtained all published nucleotide and amino acid sequences of Drosophila *hh* and their homologs (GenBank accession numbers in the parentheses). The data set contained *Sonic hh* (*Shh*) sequences of a human (*Homo sapiens*, L38518; MARIGO *et al.* 1995), a mouse (*Mus musculus*, X76290; ECHELARD *et al.* 1993), a chicken (*Gallus gallus*, L28099; RIDDLE *et al.* 1993), an African clawed frog (*Xenopus laevis*, L39213; STOLOW and SHI 1995), and a zebrafish (*Brachydanio rerio*, Z35669; FIETZ *et al.* 1994); *Vhh-1* sequences of a rat (*Rattus norvegicus*, L27340; ROELINK *et al.* 1994), a frog (*X. laevis*, L35248; RUIZ *et al.* 1995), and a zebrafish (*B. rerio*, L27585; ROELINK *et al.* 1994); *Indian hh* (*Ihh*) sequences of a human (*H. sapiens*, L38517; MARIGO *et al.* 1995) and a mouse (*M. musculus*, X76291; ECHELARD *et al.* 1993); *Desert hh* (*Dhh*) sequence of a mouse (*M. musculus*, X76292; ECHELARD *et al.* 1993); and *Drosophila melanogaster hh* (L02793, LEE *et al.* 1992; Z11840, S50757, S50758, MOHLER and VANI 1992; S66384, TABATA and KORNBERG 1994; L05404, L05405, TASHIRO *et al.* 1993) and *D. hydei hh* sequences (CHANG *et al.* 1994).

Before the sequence alignment, we noticed that the zebrafish *Shh* and *Vhh-1* were identical in their amino acid and nucleotide sequences. Thus, only one of them was included in the phylogenetic analysis. The published amino acid sequences of *Shh* and *Vhh-1* genes of *X. laevis* differ. However, an alignment of their nucleotide sequences revealed that the observed differences may be due to multiple shifts of the reading frame in the reported *Shh* sequence and that the *X. laevis Shh* sequence may contain a stop codon. The *D. melanogaster hh* sequence reported by TASHIRO *et al.* (1993) differs from others' in the signal peptide region. This difference can also be explained by shifts in the reading frame. Thus, these sequences were excluded from the evolutionary analysis.

The final data set contained 11 amino acid sequences that were first aligned using the default options of CLUSTAL V program (HIGGINS *et al.* 1992) and were further aligned by eye following the alignments presented in CHANG *et al.* (1994) and FIETZ *et al.* (1994). This alignment is given in Figure 1. The autoproteolytic action cleaves between residues 272 and 273 in the consensus sequence (PORTER *et al.* 1995). We refer to the fragment from residue 96 to 272 in the consensus sequence as the N domain and the fragment from the cleavage site to the end as the C domain (273–526).

Evolutionary analysis: The amino acid sequences were primarily used for the evolutionary analysis because the species analyzed are distantly related. The evolutionary divergences (number of amino acid replacements per site) were estimated by the Poisson correction distance (*d*-distance) to account for multiple substitutions at the same site. For a pair of amino acid sequences, the *d*-distance is related to the proportion of different amino acids (p) by the following equation (NEI 1987, pp. 41)

$$d = -\log_e(1 - p) \quad (1)$$

The neighbor-joining (NJ) method (SAITOU and NEI 1987) based on the *d*-distance was used to infer phylogenetic relationships and Drosophila *hh* sequences were used as outgroups to establish the root of the tree. The reliability of the inferred relationships was evaluated by the bootstrap test with 2000 replications and by the interior branch length test (FELSENSTEIN 1985; RZHETSKY and NEI 1992) (see Figure 2 legend). The MEGA program package was employed for these computations (KUMAR *et al.* 1993). Phylogenetic trees were also reconstructed by the maximum parsimony methods implemented in the PAUP program (SWOFFORD 1993).

To examine the differences in the rates of amino acid replacement among sites (and, thus, differences in evolutionary constraints), we generated the variability profiles for *Shh* and other *hh* genes. These profiles were generated by first estimating the rate of evolution at each amino acid site in the alignment by the maximum-likelihood approach of YANG (1994) and YANG and WANG (1995). (For directly comparing different variability profiles, evolutionary rates were scaled such that the mean of rates across sites was one.) Then the moving-average evolutionary rates in segments of length 5 were estimated. The moving-average rate in the sth segment (r_s) was computed by the formula: r_s Equation $(r_{i-2} + 2r_{i-1} + 4r_i + 2r_{i+1} + r_{i+2})/10$, where r_i is the rate at the ith site. The segments were offset from each other by one amino acid and the average rates were plotted at the center of each segment. Evolutionary rates could not be estimated for sites with alignment gaps or missing data, and whenever a segment contained such sites the moving average was adjusted by taking average over sites in the segment for which rates could be estimated.

ZUCKERKANDL and PAULING (1965) showed that the number of amino acid substitutions accumulate at a steady rate with time in different evolutionary lineages. This clock-like behavior of amino acid sequence evolution has been observed for a variety of genes (see NEI 1987, pp. 47–50). Whenever the molecular clock ticks at a constant rate, the evolutionary divergence (d) is directly proportional to the time (T), *i.e.*,

$$T = d/(2r), \quad (2)$$

where r is the rate of evolution per site per million years; a factor of 2 appears because the total time separating the two groups is two times the age of their common ancestor.

For the *hh* data, we estimated r by using a widely accepted date of 350 mya for the divergence of amphibian and mammalian lineages (BENTON 1990; AHLBERG and MILNER 1994). Given this time of divergence and the average *d*-distance (d_{avg}) between these two groups, we obtain r by the equation

$$r = d_{avg}/(2 \times 350), \quad (3)$$

Substituting r into equation (2), we get

$$T = 350\ (d\ /\ d_{avg}). \quad (4)$$

A program for calculating T and a rough estimate of its standard error is available from S.K. upon request.

RESULTS

Evolutionary relationships of genes of the *hedgehog* family: Figure 2 shows that the vertebrate homologs of the Drosophila *hh* gene fall into three groups: *Sonic,*

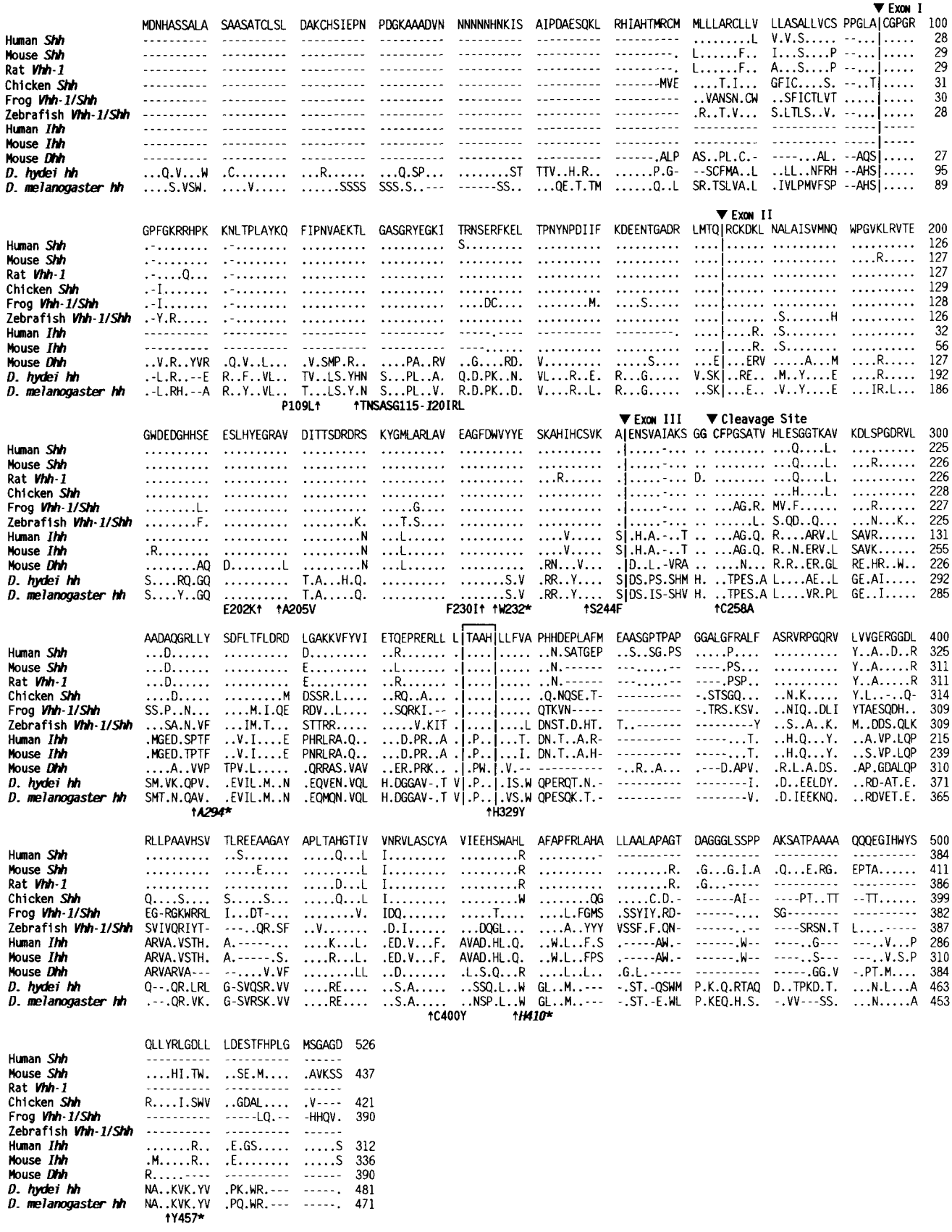

FIGURE 1.—An alignment of amino acid sequences of the *hedgehog* genes. Residues identical to the consensus sequence are shown with a dot (·) and dashes (-) represent alignment gaps or missing data. Landmarks above the consensus sequence show exon boundaries and the site where autocleavage occurs. The amino terminal fragment (N domain) is from site 9 to 272 and the carboxyl terminal fragment is from site 273 to 526 in the consensus sequence. Primary sequence lesions in the open reading frame from PORTER *et al.* (1995; Table 1) are shown with bold letters beneath the *D. melanogaster* sequence, and the stop codon mutations reported in LEE *et al.* (1994) are shown with bold-italics. Regions C1 and C2 in the C domain are from 273 to 348 and from 418 to 456, respectively.

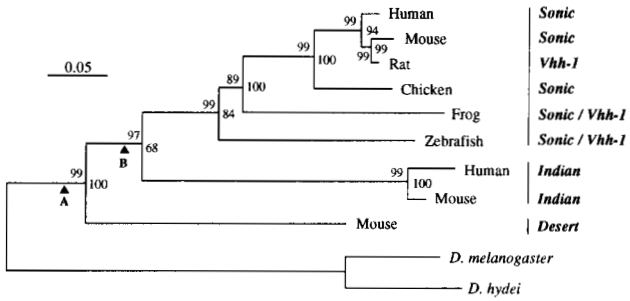

FIGURE 2.—Evolutionary relationships of genes of the *hedgehog* family. The NJ method based on *d*-distance with the complete-deletion option in MEGA (KUMAR *et al.* 1993) program was used. Only N and C domains sequences (96–526) were used. Results from the interior branch length test (confidence probability that a branch length is significantly different from 0) are shown above the corresponding branches, and the numbers to the right of each fork represent the percentage of the bootstrap replicates that support the branch. The scale bar has the units of number of amino acid replacements per site. In the maximum parsimony branch-and-bound analysis, a single most parsimonious tree was recovered, which was identical in topology to the NJ tree.

Indian, and *Desert*. *Shh* and *Vhh-1* form a cohesive group, and the vertebrate relationships within this group are identical to what is known from the fossil record and other molecular evidence. The two *Ihh* sequences are more closely related to *Shh*/*Vhh-1* than to either mouse *Dhh* or Drosophila *hh* sequences. The inferred phylogenetic relationships among different *Drosophila hedgehog* homologues are statistically supported in the interior-branch test at 95% confidence level. However, some of these groupings are only moderately supported in the bootstrap analysis because the bootstrap is known to be a conservative test (*e.g.*, SITNIKOVA *et al.* 1995). Phylogenetic analyses using data from the N and C domains independently also indicated that all *Shh* and *Vhh-1* genes form a monophyletic group and that *Dhh* is a sister group to the cluster of *Shh* and *Ihh* genes.

The close relationship of the mouse *Shh* and the rat *Vhh-1* and the conspicuous absence of the human and mouse *Vhh-1* sequences in the literature may indicate that *Shh* and *Vhh-1* are not distinct genes in mammals. As mentioned earlier, reported *Shh* and *Vhh-1* sequences of *X. laevis* are almost identical. If these genes are indeed distinct in *X. laevis*, it is likely that they have arisen from a recent genome duplication which resulted in the tetraploidization of the *X. laevis* genome (*e.g.*, BISBEE *et al.* 1977). In this case, one of the copies may have been inactivated (or diversified in function) because of a nonsense mutation in the position 313. Thus, we refer to *Shh* and *Vhh-1* genes as *Shh*.

The evolutionary relationships of the *hh* family genes indicate that the three vertebrate homologs of Drosophila *hh* arose by two gene duplications: the first duplication event gave rise to *Desert hh* and the second produced *Indian* and *Sonic hh*. Because both of the *hh* gene duplications occurred before the emergence of vertebrates (Figure 2), we expect to find the three *hh* genes in all vertebrates as in mouse. KRAUSS *et al.* (1993) amplified three different fragments from zebrafish genomic DNA. One of these was shown to correspond to *Shh* and the other two were named hh[a] and hh[b]. In our phylogenetic analyses, zebrafish hhb appeared to be more similar to the human and mouse *Ihh* than to others; and, hha appears to be a homologue of mouse *Dhh*. However, these results are not supported with high statistical confidence that may be due to their short sequence length. RIDDLE *et al.* (1993) sequenced *Shh* in chicken and detected two other "unique" bands with *hh*-specific probes, suggesting that chicken may have three different *hh* genes. CHANG *et al.* (1994) reported detecting five clones each for frog (*X. laevis*) and zebrafish (*B. rerio*), but did not indicate how many were distinct at the nucleotide sequence level.

After this manuscript was submitted, EKKER *et al.* (1995) reported the isolation of four *X. laevis hh* genes: *sonic hh* (*X-shh*), *banded hh* (*X-bhh*), *cephalic hh* (*X-chh*), and *hh4* (*X-hh4*). Based on sequence similarity, they suggested that *X-shh* and *X-bhh* are closely related to the *Sonic* and *Indian hh*, respectively, and that *X-chh* and *X-hh4* are most similar to the mouse *Desert hh*. Our phylogenetic analyses confirmed their suggestions (results not shown). In agreement with our initial speculation that the recent genome duplication in *X. laevis* may be responsible for the presence of two homologs of vertebrate *Sonic hh* in *X. laevis*, EKKER *et al.* suggested that the *X-chh* and *X-hh4* have been produced by a recent genome duplication in *X. laevis*. In our phylogenetic analyses, *X-chh* and *X-hh4* were closer to each other than either is to *Shh* or *Ihh* of other vertebrates.

If *Desert*, *Indian*, and *Sonic hh* are indeed specific to vertebrates, we expect that diverse coelomic invertebrates including proterostomes (*e.g.*, arthropods and annelids) and deuterostomes (*e.g.*, echinoderms) will carry only one *hh* gene. Genomic library screenings of Drosophila (an arthropod), leech (an annelid), and sea urchin (an echinoderm) have suggested the presence of only one *hh* gene (*e.g.*, KRAUSS *et al.* 1993; CHANG *et al.* 1994).

Variability in amino- and carboxyl-terminal domains: The alignment given in Figure 1 clearly shows that the amino acid sequences of the N domain are highly conserved. This is reflected in large differences in the rates of amino acid substitution between the N and C domains (Figure 3). The signal peptide is highly variable in its amino terminal, but is well conserved near the start of the first exon. High amino acid sequence similarity (low evolutionary rates) are also observed near other exon boundaries. In *Shh*, the average rate of amino acid substitution in the N domain is fourfold lower than that in the C-domain (Figure 3A). An analysis of *Shh*, *Dhh*, and Drosophila *hh* shows a 2.5 times lower rate of evolution in the N domain than in the C domain (Figure 3B). A nice correlation exists between the variability profiles given in Figure 3, A and B, which suggests that the evolutionary constraints on different

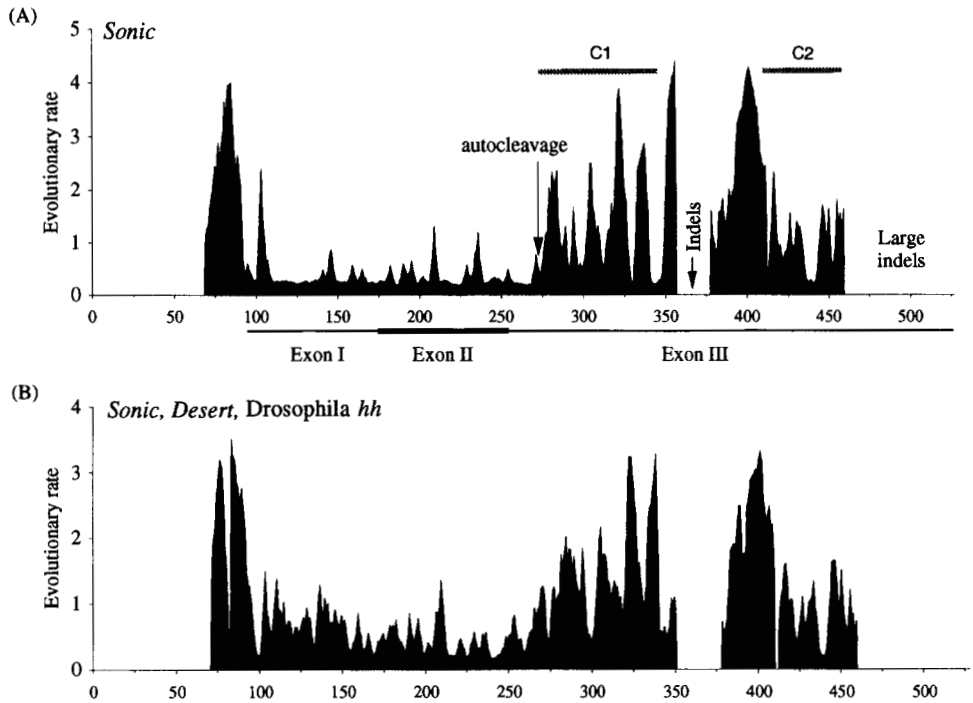

FIGURE 3.—The number of amino acid replacements per site (A) in only *Shh* genes and (B) in all *hh* genes except *Ihh*, which was excluded because the sequence of the first exon was not available at the time of the analysis. The rate of amino acid substitution at each site were determined by the maximum-likelihood approach and the moving average in segments of length 5 were plotted at the center of each segment (see text for details). Sites are numbered following the consensus sequence in Figure 1. Regions with extensive alignment gaps or missing data are marked as *indel* regions. Various landmarks shown in Figure 1 are also marked.

regions of the *hh* gene have not varied considerably in the long evolutionary history of vertebrates (>500 mya). Experimental evidence for the similarity of developmental functions of *Shh* in vertebrates and the Drosophila *hh* gene is now accumulating (KRAUSS *et al.* 1993; INGHAM 1994, SMITH 1994; JOHNSON and TABIN 1995).

DISCUSSION

Age of *hedgehog* gene duplications: By using the *hh* protein sequence data, approximate times of the two gene duplications that can be estimated provided that the evolutionary rates in the lineages leading from ancestral nodes A and B have remained the same (Figure 3). In statistical tests, existence of the molecular clock was not rejected at 5% significance level at these nodes (TAKEZAKI *et al.* 1995). That is, even though the rates of amino acid replacements vary *among sites* because of functional constraints, the rates of evolution have remained constant *among lineages* leading to the extant sequences of *Shh, Ihh,* and *Dhh* genes. Thus, the times of two gene duplications (T_A and T_B) can be estimated. The average *d*-distance between the amphibian *X. laevis* and mammals is 0.251. Thus, the rate of evolution is 3.59×10^{-4} substitutions per site per million year (0.251/700). The average *d*-distance between the mouse and human *Ihh* and all vertebrate *Shh* is 0.404. Thus, *Indian* and *Sonic hh* diverged 563 mya. Similarly, the average *d*-distance between the mouse *Dhh* and the *Ihh* and *Shh* is 0.475, which translates to 662 mya (Figure 4).

Clearly, both *hh* gene duplications preceded the evolution of vertebrates (Figure 2) and possibly occurred before the emergence of chordates (570 mya) (HARLAND *et al.* 1990; BENTON 1993). If the gene duplication that produced *Indian* and *Sonic hh* occurred after the split of vertebrates from other chordates, we expect to find *Desert* and *Sonic*-like *hh* genes in the primitive chordate lineages (tunicates and cephalochordates). The likelihood of finding *Sonic hh* in all chordates is rather high because the notochord is a unique feature of chordates and the expression of *Shh* in the notochord appears to be critical for inducing neural tube formation at least in vertebrates (MARTI *et al.* 1995). Of all complete or partial *hh* sequences available, sea urchin (an echinoderm) is most closely related to vertebrates and appears to carry only one *hh* gene (CHANG *et al.* 1994). The evolutionary relationships of this partial sea urchin *hh* gene were uncertain in our evolutionary anal-

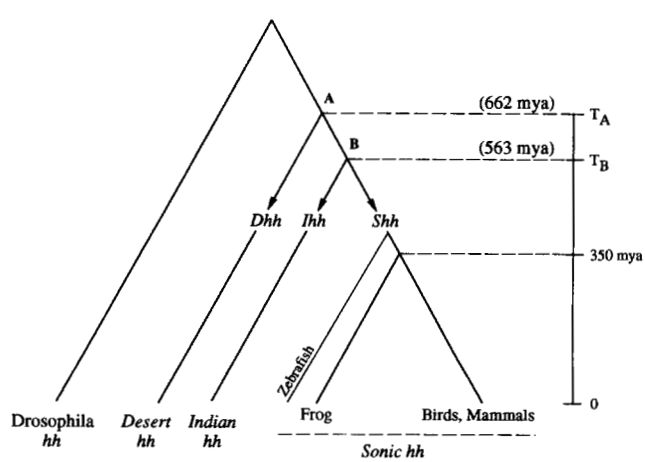

FIGURE 4.—Model for the evolution of the *hedgehog* gene family. The approximate times of the gene duplication events at nodes A and B in Figure 2 were computed by equation (4). The standard errors of the estimates of T_A and T_B are of the order of 100 mya.

yses; however, it appear to be closely related to mouse *Dhh* and Drosophila *hh* genes. The knowledge of the number of *hh* genes and their primary sequences in early diverging chordate lineages and closely related deuterostomes may provide new insights into the role of *hh* genes in the evolution of the chordates and vertebrates.

The divergence of triploblasts was rapid and the relative branching order of major triploblastic lineages has been difficult to establish (CONWAY MORRIS 1993). However, it is clear that deuterostomes (*e.g.*, chordates and echinoderms) and proterostomes (*e.g.*, arthropods and annelids) constitute two major lineages of coelomate triploblasts that diverged earlier than 600 mya (CONWAY MORRIS 1993). The time for the first *hh* gene duplication provides a lower bound of 662 mya on the age of *hh* gene family. This lower bound suggests that the *hh* gene most likely evolved prior to the proterostome-deuterostome split. However, CHANG *et al.* (1994) failed to detect any *hh* homologs in a pseudocoelomate triploblast (*Caenorhabditis elegans*). Evolutionary analyses of the 18S small subunit ribosomal RNA and RNA polymerase II have suggested that the nematodes (*e.g.*, *C. elegans*) may constitute one of the basal triploblastic lineages (SIDOW and THOMAS 1994; WINNEPENNINCKX *et al.* 1995). Thus, it is possible that the *hh* gene family evolved after the divergence of nematodes from other triploblasts. However, the *hh* gene may be present in nematodes or may have been lost independently, and the origin of *hedgehog* gene family may have predated the evolution of all triploblasts. By using the estimate of r previously obtained and assuming that the amino acid sequences of *hh* genes have evolved in a clock-like fashion throughout the evolutionary history of *hh* genes, we obtain a tentative estimate of 900 mya for the age of the common ancestor of Drosophila and vertebrate *hh* genes. This estimate is similar to that obtained by NEWFELD and GELBART (1995) in the comparison of intercellular signaling molecules of the *TGF-β* super-family from human and Drosophila, which suggests that the *hh* gene may have evolved at the same time as some other intercellular signaling molecules. However, this estimate of time for the origin of *hh* genes is not reliable because the constancy of molecular clock could not be tested at the root of the tree in Figure 2 due to the absence of an outgroup sequence.

Conserved features of the *hedgehog* proteins: The *hh* protein is characterized by the presence of a signal sequence near the amino-terminal (except in the Drosophila *hh* protein, which has an internal signal sequence) and a highly conserved autocleavage site that splits the *hh* protein into the N and C domains. The evolutionary conservation of the primary sequence of the N domain, including the carboxyl end of the signal peptide, correlates well with the functional significance of this domain in the *hh* signaling activity (Figure 3). The zebrafish and mouse *Shh* genes were shown to function in a similar manner as fruit fly *hh* to induce gene

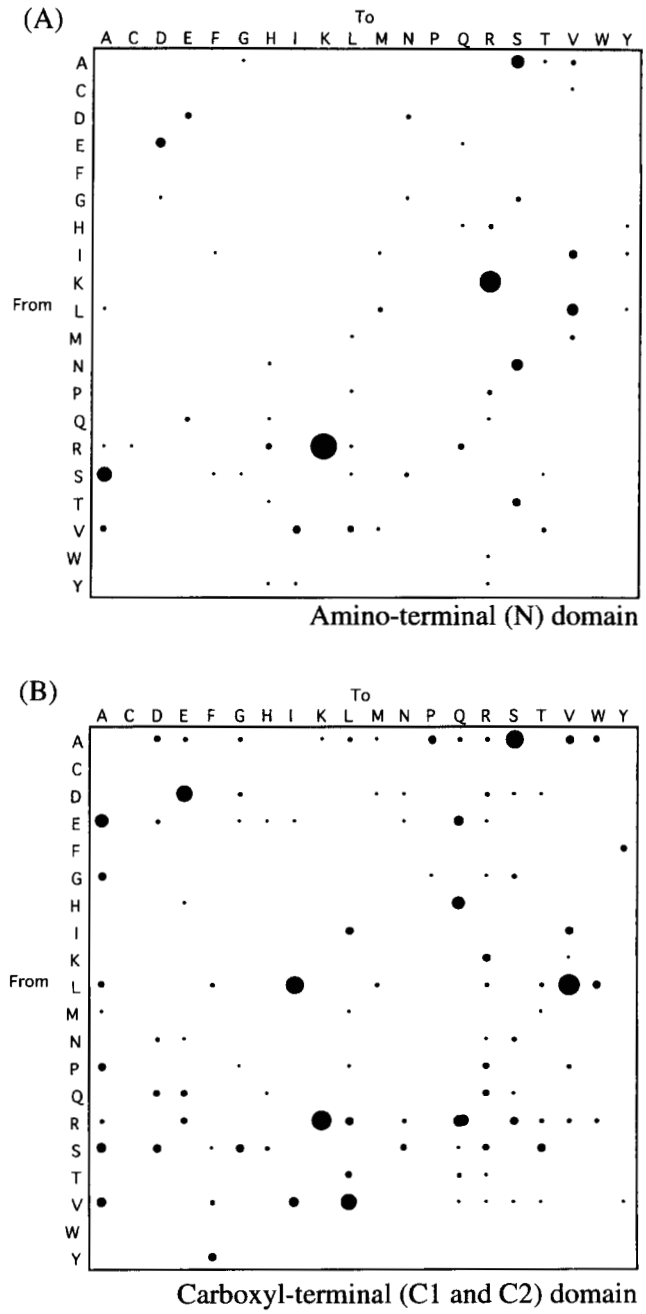

FIGURE 5.—The average frequency of different amino acid replacements for (A) the N domain (96–272 in consensus) and (B) the C1 and C2 domains (273–348, 418–456). Ancestral states were inferred by using the MacClade program (MADDISON and MADDISON 1992). The diameter of the largest circle in A is proportional to nine changes, and other circles in A and B are scaled in diameter proportionally.

expression in Drosophila embryos (KRAUSS *et al.* 1993; CHANG *et al.* 1994), suggesting that not only the *hh* activity, but also the *hh* signaling pathway has been conserved in evolution. Furthermore, various primary sequence lesions identified in the open reading frames of *hh* mutant alleles are mapped to amino acid positions that have remained invariant in the *hh* gene family for ≥500 million years. This points to the similarity of functional constraints on these sites (and corresponding

domains) in the Drosophila *hh* and its vertebrate homologs (Figures 1 and 3).

In the C domain, several regions have undergone extensive insertion-deletions in the course of evolution, yet it maintains its autocleavage activity. These insertion-deletion events have not only occurred among different *hh* genes, but also in *Shh* genes in different vertebrate lineages. Interspersed among these regions are two fragments that appear to have evolved with lower rates of amino acid replacement and contain only a few alignment gaps: C1: 273–348 and C2: 418–456 (Figures 1 and 3). It is possible that the autocleavage activity of the C domain has been conserved through the C1 and C2 regions. Indeed, replacement of the Cys residue with Ala at the start of the C1-domain (site 273 in the consensus) prevents proteolysis and diminishes *hh* activity and a short sequence (342–345) with limited similarity to the catalytic site of serine proteases is crucial for autocleavage activity (LEE *et al.* 1994; PORTER *et al.* 1995).

Because the N domain and the C1- and C2-domains are evolutionarily conserved and crucial for cellular signaling activities, we examined the relative frequency of the amino acid replacements that have been permitted in these domains during the course of evolutionary change. Figure 5 shows the average frequency of amino acid changes between different residues in the N domain and in the C1 and C2 domains. Arg ↔ Lys and Ala ↔ Ser changes occur most frequently in the N domain, whereas Leu ↔ Val, Leu ↔ Ile, and Asp ↔ Glu changes are abundant in the C1 and C2 domains in addition to Arg ↔ Lys and Ala ↔ Ser changes. Of these, only Ala ↔ Ser changes are radical amino acid replacements. The primary sequence mutations in the N and C domains that disrupt the function of the *hh* genes in the laboratory studies appear to occur with rather low frequency naturally (Figure 1), and, thus, rarely tolerated in the evolution of *hh* genes. It would be interesting to examine correlations between the inferred frequency of amino acid replacements (Figure 5) and those observed in the mutagenesis experiments. However, amino acid mutations that do not alter the phenotype of the organism in the laboratory are generally not reported. If such data were available, we could study the evolutionary and functional relationships of the inferred and laboratory mutagenesis amino acid replacements at different sites.

Although structurally similar, it is not clear how *Dhh* and *Ihh* may be functionally related to *Shh*. Based on their structural conservation, it is likely that they act in a similar fashion as *Shh*. *Ihh* and *Dhh* proteins are expected to be processed like *Shh*, including signal sequence and autoproteolytic cleavages. The structure of a putative receptor that is recognized by the N domain may also exhibit conserved features. However, *Ihh* and *Dhh* may show distinct expression profiles that may reflect their distinct roles in development. After the submission of this manuscript, LAI *et al.* (1995) have shown that the N domain in *X-shh* is critical in cellular signaling. Furthermore, EKKER *et al.* (1995) have reported distinct patterns of expressions of *X-shh* (*Shh* homologue), *X-bhh* (*Ihh* homologue) and *X-chh* (*Dhh* homologue), which provide support for some of our speculations.

We thank E. SIEGFRIED, J. ZHANG, C. KAPPEN, and M. NEI for critically reading an earlier draft of this paper and for helpful discussions. This work was supported in part by grants from the National Science Foundation to Z.-C.L. and from the National Science Foundation and National Institute of Health to Dr. M. Nei.

LITERATURE CITED

AHLBERG, P. E., and A. R. MILNER, 1994 The origin and early diversification of tetrapods. Nature **368:** 507–514.

BENTON, M. J., 1990 Phylogeny of the major tetrapod groups: morphological data and divergence dates. J. Mol. Evol. **30:** 409–424.

BENTON, M. J., 1993 *The Fossil Record 2.* Chapman and Hall, New York.

BISBEE, C., M. BAKER, A. C. WILSON, I. HADJI-AZIMI and M. FISCHBERG, 1977 Albumin phylogeny for clawed frogs (*Xenopus*). Science **195:** 785–787.

BUMCROT, D., R. TAKADA, and A. MCMAHON, 1995 Proteolytic processing yields two secreted forms of Sonic hedgehog. Mol. Cell. Biol. **15:** 2294–2303.

CHANG, D., A. LOPEZ, D. VON KESSLER, C. CHIANG, B. SIMANDL *et al.*, 1994 Products, genetic linkage and limb patterning activity of a murine hedgehog gene. Development **120:** 3339–3353.

CONWAY MORRIS, S., 1993 The fossil record and the early evolution of the Metazoa. Nature **361:** 219–225.

ECHELARD, Y., D. EPSTEIN, B. ST-JACQUES, L. SHEN, J. MOHLER *et al.*, 1993 Sonic Hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity. Cell **75:** 1417–1430.

EKKER, S. C., L. L. MCGREW, C.-J. LAI, J. J. LEE D. P. VON KESSLER *et al.*, 1995 Distinct expression and shared activities of members of the *hedgehog* gene family of *Xenopus laevis.* Development **121:** 2337–2347.

FELSENSTEIN, J., 1985 Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:** 783–791.

FIETZ, M., J. CONCORDET, R. BARBOSA, R. JOHNSON, S. KRAUSS *et al.*, 1994 The hedgehog gene family in Drosophila and vertebrate development. Development Supplement 43-51.

FRANCIS, P. H., M. K. RICHARDSON, P. M. BOCKELL and C. TICKLE, 1994 Bone morphogenetic proteins and a signaling pathway that controls patterning in the development of chick limb. Development **120:** 209–281.

HARLAND, W. B., R. L. ARMSTRONG, A. V. COX, L. E. CRAIG, A. G. SMITH *et al.*, 1990 *A Geologic Time Scale 1989.* Cambridge, New York.

HEBERLEIN, U., and K. MOSES, 1995 Mechanisms of Drosophila retinal morphogenesis: the virtues of being progressive. Cell **81:** 987–990.

HEBERLEIN, U., T. WOLFF and G. RUBIN, 1993 The TGF-β homolog and the segment polarity gene hedgehog are required for propagation of a morphogenic wave in the Drosophila retina. Cell **75:** 913–926.

HIGGINS, D. G., A. J. BLEASBY and R. FUCHS 1992 CLUSTALV: improved software for multiple sequence alignment. Comput Appl. Biosci. **8:** 189–191.

INGHAM, P. W., 1994 Hedgehog points the way. Curr. Biol. **4:** 347–350.

JOHNSON, R. L., and C. TABIN, 1995 The long and short of hedgehog signaling. Cell **81:** 313–316.

KRAUSS, S., J. P. CONCORDET and P. W. INGHAM, 1993 A functionally conserved homolog of the Drosophila segment polarity gene hh is expressed in tissues with polarizing activity in zebrafish embryos. Cell **75:** 1431–1444.

KUMAR, S., K. TAMURA, and M. NEI, 1993 *MEGA: Molecular Evolutionary Genetics Analysis. Version 1.02.* The Pennsylvania State University, University Park, PA.

LAI, C.-J., S. C. EKKER, P. A. BEACHY and R. T. MOON, 1995 Patterning of the neural ectoderm of Xenopus laevis by the amino-terminal product of *hedgehog* autoproteolytic cleavage. Development **121:** 2349–2360.

LEE, J., D. VON KESSLER, S. PARKS and P. BEACHY, 1992 Secretion and

localized transcription suggest a role in positional signaling for products of the segmentation gene hedgehog. Cell **71:** 33–50.

LEE, J., S. EKKER, D. VON KESSLER, J. PORTER, B. SUN *et al.*, 1994 Autoproteolysis in hedgehog protein biogenesis. Science **266:** 1528–1537.

MA, C., Y. ZHOU, P. BEACHY and K. MOSES, 1993 The segment polarity gene hedgehog is required for progression of the morphogenic furrow in the developing Drosophila eye. Cell **75:** 927–938.

MADDISON, W. and D. MADDISON, 1992 *MacClade: Analysis of Phylogeny and Character Evolution, Version 3.* Sinauer Associates, Sunderland, MA.

MARIGO, V., D.J. ROBERTS, S. LEE, O. TSUKUROV, T. LEVI *et al.*, 1995 Cloning, expression and chromosomal location of SHH and IHH, two human homologs of the Drosophila segment polarity gene hedgehog. Genomics **28:** 44–51.

MARTI, E., D. BUMCROT, R. TAKADA and A. MCMAHON, 1995 Requirement of 19K form of Sonic hedgehog for induction of distinct ventral cell types in CNS explants. Nature **375:** 322–325.

MOHLER, J., and K. VANI, 1992 Molecular organization and embryonic expression of the *hedgehog* gene involved in cell-cell communication in segmental patterning of Drosophila. Development **115:** 957–971.

NEI, M., 1987 *Molecular Evolutionary Genetics.* Columbia University Press, New York.

NEWFELD S. J, and W. M. GELBART, 1995 Identification of two Drosophila TGF-β family members in the grasshopper *Schistocerca americana.* J. Mol. Evol. **41:** 155–160.

NÜSSLEIN-VOLHARD, C., and E. WIESCHAUS, 1980 Mutations affecting segment number and polarity in Drosophila. Nature **287:** 795–801.

PARR, B. A., and A. P. MCMAHON, 1995 Dorsalizing signal Wnt-7a required for normal polarity of D-V and A-P axes of mouse limb. Nature **374:** 350–353.

PERRIMON, N., 1995 Hedgehog and beyond. Cell **80:** 517–520.

PORTER, J., D. VON KESSLER, S. EKKER, K. YOUNG, J. LEE *et al.*, 1995 The product of hedgehog autoproteolytic cleavage active in local and long-range signaling. Nature **374:** 363–366.

RIDDLE, R., R. JOHNSON, E. LAUFER and C. TABIN, 1993 Sonic hedgehog mediates the polarizing activity of the ZPA. Cell **75:** 1401–1416.

ROELINK, H., A. AUGSBURGER, J. HEEMSKERK, V. KORZH, S. NORLIN *et al.*, 1994 Floor plate and motor neuron induction by vhh-1, a vertebrate homolog of hedgehog expressed by the notochord. Cell **76:** 761–775.

ROELINK, H., J. A. PORTER, C. CHIANG, Y. TANABE, D. T. CHANG *et al.*, 1995 Floor plate and motor neuron induction by different concentrations of the amino-terminal cleavage product of Sonic hedgehog autoproteolysis. Cell **81:** 445–455.

RUIZ I ALBATA, A., H. ROELINK and T. JESSELL, 1995 Restrictions to floor plate induction by vertebrate homolog and winged helix genes in the neural tube of frog embryos. Mol. Cell. Neurosci. **6:** 106–121.

RZHETSKY, A., and M. NEI, 1992 A simple method for estimating and testing minimum-evolution trees. Mol. Biol. Evol. **9:** 945–967.

SAITOU, N., and M. NEI, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406–425.

SIDOW, A., and THOMAS, W. K., 1994 A molecular evolutionary framework for eukaryotic model organisms. Curr. Biol. **4:** 596–603.

SITNOKOVA, T. L., A. RZHETSKY, and M. NEI., 1995 Interior-branch and bootstrap tests of phylogenetic trees. Mol. Biol. Evol. **12:** 319–333.

SMITH, J. C., 1994 Hedgehog, the floor plate, and the zone of polarizing activity. Cell **76:** 193–196.

STOLOW, M. A., and Y. B. SHI, 1995 Xenopus sonic hedgehog as a potential morphogen during embryogenesis and thyroid hormone-dependent metamorphosis. Nucleic Acids Res. **23:** 2555-2562.

SWOFFORD, D. L., 1993 *PAUP: Phylogenetic Analysis Using Parsimony. Version 3.1.1.* University of Illinois, Champaign.

TABATA, T., and T. KORNBERG, 1994 hedgehog is a signaling protein with a key role in patterning Drosophila imaginal disks. Cell **76:** 89–102.

TAKEZAKI, N., A. RZHETSKY and M. NEI, 1995 Phylogenetic test of the molecular clock and linearized trees. Mol. Biol. Evol. **12:** 823–833.

TASHIRO, S., T. MICHIUE, S. I. HIGASHIJIMA, S. ZENNO, S. ISHIMARU *et al.*, 1993 Structure and expression of hedgehog, a Drosophila segment-polarity gene required for cell-cell communication. Gene **124:** 183–189.

WINNEPENNINCKX, B., T. BACKELJAU, L. MACKEY, J. BROOKS, R. DE-WACHTER *et al.*, 1995 18S rRNA data indicate that aschelminthes are polyphyletic in origin and consist of at least three distinct clades. Mol. Biol. Evol. **12:** 1132–1137.

YANG, Z. 1994 Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J. Mol. Evol. **39:** 306–314.

YANG, Y., and L. NISWANDER, 1995 Interaction between signaling molecules *Wnt-7a* and SHH during vertebrate limb development: dorsal signals regulate anterioposterior patterning. Cell **80:** 939–947.

YANG, Z., and T. WANG, 1995 Mixed model analysis of DNA sequence evolution. Biometrics **51:** 552–561.

ZUCKERKANDL, E., and L. PAULING, 1965 Evolutionary divergence and convergence in proteins, pp. 97–166 in *Evolving Genes and Proteins,* edited by V. BRYSON and H. J. VOGEL. Academic Press, New York.

Communicating editor: N. TAKAHATA