

Hagfish and lancelet fibrillar collagens reveal that type II collagen-based cartilage evolved in stem vertebrates

GuangJun Zhang* and Martin J. Cohn**

*Department of Zoology and †Department of Anatomy and Cell Biology, University of Florida, P.O. Box 118525, Gainesville, FL 32611

Edited by Kathryn V. Anderson, Sloan-Kettering Institute, New York, NY, and approved September 19, 2006 (received for review July 5, 2006)

The origin of vertebrates was defined by evolution of a skeleton; however, little is known about the developmental mechanisms responsible for this landmark evolutionary innovation. In jawed vertebrates, cartilage matrix consists predominantly of type II collagen (Col2 α 1), whereas that of jawless fishes has long been thought to be noncollagenous. We recently showed that Col2 α 1 is present in lamprey cartilage, indicating that type II collagen-based cartilage evolved earlier than previously recognized. Here, we investigate the origin of vertebrate cartilage, and we report that hagfishes, the sister group to lampreys, also have Col2 α 1-based cartilage, suggesting its presence in the common ancestor of crown-group vertebrates. We go on to show that lancelets, a sister group to vertebrates, possess an ancestral clade A fibrillar collagen (ColA) gene that is expressed in the notochord. Together, these results suggest that duplication and diversification of ColA genes at the chordate-vertebrate transition may underlie the evolutionary origin of vertebrate skeletal tissues.

development | gene duplication | skeleton evolution | chordate | notochord

The phylogenetic relationships of the vertebrates were established largely based on anatomical characters, particularly those of the skeleton. The skeletons of jawed vertebrates (gnathostomes) are composed of cartilage and bone, which contain high levels of type II and type I collagen (Col2 α 1 and Col1 α 1/2 protein, respectively). By contrast, the cartilaginous skeletons of lampreys and hagfishes, the only extant jawless fishes (agnathans), have been reported to be noncollagenous and to contain instead the elastin-like proteins lamprin and myxinin (1). This difference in the extracellular matrices of vertebrate skeletons led to the idea that type II collagen became the major structural component of gnathostome cartilage after the divergence of these two lineages (1). However, this view was challenged by our recent report that lampreys have two Col2 α 1 orthologs and that both genes are expressed during chondrogenesis. Adult lamprey cartilage was also shown to be rich in Col2 α 1 protein (2). Furthermore, we identified a lamprey ortholog of Sox9, a direct transcriptional regulator of Col2 α 1 in gnathostomes, and we showed that it is coexpressed with Col2 α 1 (2). This discovery revealed that the genetic pathway for vertebrate chondrogenesis predated the divergence of lampreys and gnathostomes, although precisely how early this character arose is unknown. Here, we investigate the evolutionary origin of Col2 α 1-based cartilage by expanding our analysis to the most inclusive clade of vertebrates, which includes the hagfishes, and to a sister group to the vertebrates, the lancelets (3, 4). We report that hagfishes also have Col2 α 1-based cartilage, suggesting that this type of cartilage was present in common ancestor of all crown vertebrates. Our analysis of lancelets revealed the presence of an ancestral clade A fibrillar collagen (ColA) gene that is expressed in the notochord. Thus, during the chordate-vertebrate transition, an ancestral ColA gene underwent duplication and diversification, and this process may underlie the evolutionary origin of vertebrate skeletal tissues.

Results

Identification of Col2 α 1 in Hagfishes. We searched for an expressed Col2 α 1 ortholog in hagfish using degenerate RT-PCR, and we recovered a 2,214-bp cDNA fragment with a sequence that corresponds to the region between the major triple helix and the C-propeptide domains of gnathostome Col2 α 1. The deduced amino acid sequence of the hagfish clone was 76% identical to mouse Col2 α 1, and comparison with the two lamprey Col2 α 1 orthologs showed 80% identity to lamprey Col2 α 1a and 77% identity to lamprey Col2 α 1b. We next conducted molecular phylogenetic analyses using Bayesian phylogenetics, minimum evolution, and maximum likelihood methods. All three analyses placed the hagfish sequence within the vertebrate Col2 α 1 clade, supporting the designation of this gene as hagfish Col2 α 1 (Fig. 1 and Figs. 5 and 6, which are published as supporting information on the PNAS web site). Interestingly, each tree further positioned hagfish Col2 α 1 as a sister to lamprey Col2 α 1a, with lamprey Col2 α 1b falling out as the sister branch (Fig. 1, 5, and 6). The topology of these trees suggests that Col2 α 1 was present in the common ancestor of agnathans and gnathostomes and that an additional duplication gave rise to Col2 α 1a and Col2 α 1b in the cyclostome (lampreys plus hagfishes) lineage.

Col2 α 1 Localizes to Hagfish "Soft" Cartilage. We next asked whether Col2 α 1 is expressed in hagfish cartilage. Little is known about hagfish embryonic development because of difficulties in obtaining fertilized embryos (5–7). *Eptatretus* embryos were last collected in 1930 (7), and only three embryos of *Myxine* have ever been found (5, 6). The unavailability of embryos therefore precludes analysis of hagfish Col2 α 1 expression during development; however, Col2 α 1 protein is known to be detectable in adult cartilage of gnathostomes and lampreys (2, 8). We therefore investigated whether Col2 α 1 is present in hagfish cartilage by using a human antibody against the highly specific N-terminal region of Col2 α 1 (2, 9–11). Immunohistochemical analysis revealed the presence of Col2 α 1 in the extracellular matrix of several cartilage elements in the hagfish head and tail (Fig. 2). Col2 α 1 protein was also detected in the notochord (data not shown), which is known to contain collagen fibers (12). Although a number of cartilage elements in hagfish were rich in Col2 α 1 protein (Fig. 2A' and B'), others showed a mosaic distribution (Fig. 2C') or lacked Col2 α 1 altogether (Fig. 2D'). Hagfishes have been described as having two types of cartilage: "soft" cartilage, which contains large hypertrophic chondrocytes that stain with

Author contributions: G.Z. and M.J.C. designed research; G.Z. performed research; G.Z. and M.J.C. analyzed data; and G.Z. and M.J.C. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS direct submission.

Abbreviations: Col1 α 1, type I collagen; Col2 α 1, type II collagen;

Data deposition: The sequences reported in this paper have been deposited in the GenBank database [accession nos. DQ647925 (lancelet *AmphiColA*) and DQ647926 (hagfish Col2 α 1)].

†To whom correspondence should be addressed. E-mail: cohn@zoo.ufl.edu.

© 2006 by The National Academy of Sciences of the USA

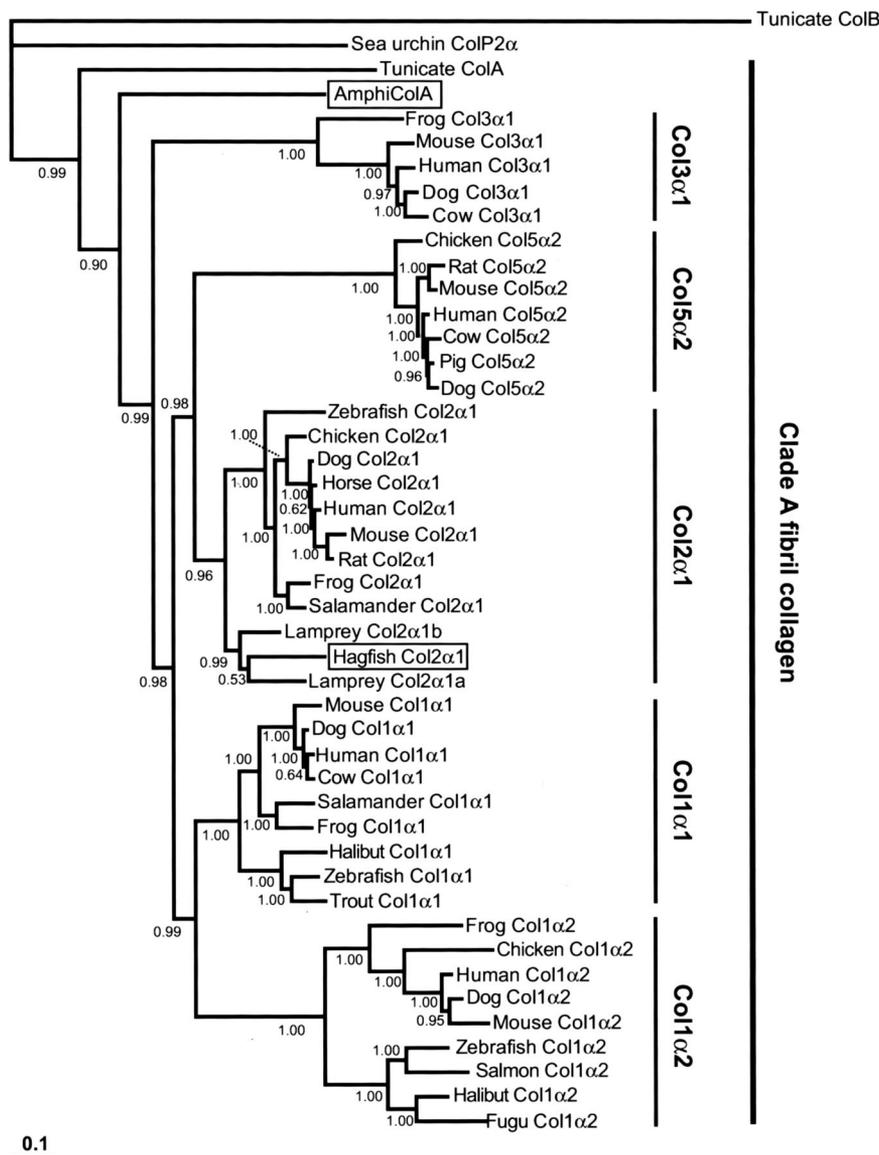


Fig. 1. Extended majority-rule consensus tree for the Bayesian phylogenetic analysis of clade A fibrillar collagen proteins. Numbers at each node indicate posterior probability (pp) values based on one million replicates. Branch lengths are proportional to means of the pp densities for their expected replacements per site. The tree is rooted by tunicate (*C. intestinalis*) clade B fibrillar collagen (ColB) and sea urchin fibrillar collagen (ColP2 α). Hagfish Col2 α 1 (boxed) is grouped with lamprey Col2 α 1a and Col2 α 1b with a pp of 0.99. This cyclostome Col2 α 1 clade joins to the base of the gnathostome Col2 α 1 clade with a pp of 0.96. Lancelet clade A fibrillar collagen (AmphiColA; boxed) is joined to the vertebrate clade A collagen family with a pp of 0.99. Minimum evolution and maximum likelihood methods confirm these positions for hagfish Col2 α 1 and AmphiColA (see also Figs. 5 and 6).

hematoxylin (blue) and are surrounded by a thin extracellular matrix; and “hard” cartilage, which contains smaller chondrocytes that are surrounded by an abundance of extracellular matrix (13, 14). Cartilage elements that were positive for Col2 α 1 had the cellular characteristics of soft cartilage (Fig. 2 *A, A', B,* and *B'*), whereas cartilages that lacked Col2 α 1 exhibited the features of hard cartilage (Fig. 2 *D* and *D'*). Elements that showed a mosaic distribution of Col2 α 1 also showed a mosaicism of hard and soft features, and Col2 α 1 was restricted to the soft regions of these structures (Fig. 2 *C* and *C'*). Hagfish soft cartilage has been proposed to be structurally similar to lamprey cartilage (14), and our finding that each is composed of Col2 α 1 protein supports this idea. The presence of Col2 α 1 in the cartilages of hagfishes, lampreys (2), and gnathostomes (15) strongly suggests that their last common ancestor had a Col2 α 1-based endoskeleton.

Duplication and Divergence of Clade A Fibrillar Collagen Genes Occurred in Stem Vertebrates. Col2 α 1 belongs to the clade A fibrillar collagen family, which includes collagen types I, II, III, and V α 2 (16). To investigate the relationship between evolution of the clade A collagens and the origin of the vertebrate skeleton, we extended our analysis to lancelets, a sister group to the vertebrates (4). Comparative studies of a multitude of genes in lancelets and vertebrates show that the lancelet lineage diverged before the duplication events that occurred in the vertebrate genome (17). We screened for lancelet fibrillar collagen cDNAs using degenerate RT-PCR, and we isolated a 2,196-bp clone with a deduced amino acid sequence that is 54% identical to mouse Col2 α 1 and 53% identical to mouse Col1 α 1 and Col3 α 1. To determine the relationship of the lancelet clone to the vertebrate fibrillar collagen proteins, we carried out molecular phylogenetic analyses using C-propeptide of clade A, B, and C fibrillar

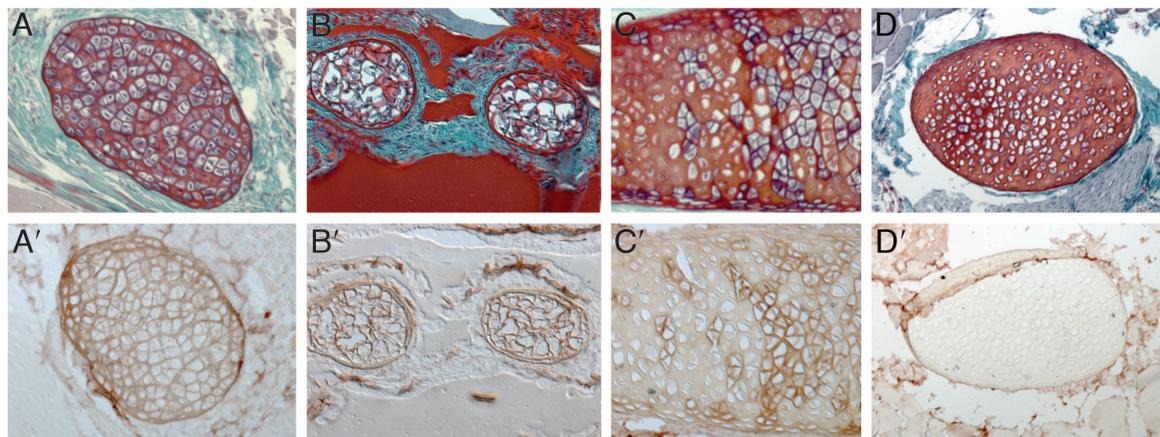


Fig. 2. Col2 α 1 in cranial and tail fin cartilages of Atlantic hagfish. Transverse sections through adult hagfish stained with hematoxylin/eosin and fast green (A–D) or with Col2 α 1 antibody (A'–D') are shown. (A and A') Soft cranial cartilage stained with hematoxylin contains large hypertrophic chondrocytes surrounded by thin layer extracellular matrix (A), and it is positive for Col2 α 1 (A'). (B and B') Section through caudal fin shows cartilaginous fin rays with soft-cartilage characteristics (B) that are rich in Col2 α 1 (B'). (C and C') Cranial cartilage element exhibiting a mosaic distribution of soft (blue-stained chondrocytes) and hard (red-stained chondrocytes) cartilage (C). Note that Col2 α 1 protein is restricted to the soft cartilage region (C'). (D and D') Hard cranial cartilage stained with eosin but not hematoxylin contains small chondrocytes surrounded by a thick layer of extracellular matrix (D), and it is negative for Col2 α 1 (D').

collagens. Bayesian phylogenetics, minimum evolution, and maximum likelihood methods each supported its designation as a clade A fibrillar collagen (*AmphiColA*; Fig. 7, which is published as supporting information on the PNAS web site). We then used the C-propeptide and triple-helix domains to refine further the position of *AmphiColA* within the chordate clade A family, and all three methods placed it as the sister clade to the vertebrate (including cyclostome) clade A collagens (Fig. 1). These results suggest that duplication of the ancestral *ColA*, the precursor of the clade A fibrillar collagen multigene family, occurred in the vertebrate lineage after the divergence of lancelets.

***AmphiColA* Is Expressed in the Lancelet Notochord and Neural Tube.**

To gain insight into the expression pattern of the ancestral *ColA* gene, we investigated the developmental expression of *AmphiColA* by RNA *in situ* hybridization on lancelet embryos. During neurulation, *AmphiColA* was expressed in the notochord (Fig. 3A). At 30 h postfertilization, *AmphiColA* remained in the notochord, and a new domain of expression was detected in the neural tube (Fig. 3B). By 36 h, *AmphiColA* expression was being down-regulated in the middle third of the neural tube and notochord, but it remained strong anteriorly and posteriorly (Fig. 3C). In 5-day-old larvae, *AmphiColA* was detected in the

tail bud and in the anterior region of the neural tube, up to the base of the cerebral vesicle (Fig. 3D). Expression of *AmphiColA* in the lancelet notochord and neural tube is strikingly similar to that of clade A fibrillar collagens in gnathostomes and lampreys. We did not detect *AmphiColA* in the embryonic and larval pharynx by *in situ* hybridization, although Rychel *et al.* (18) showed that the gill bars of adult cephalochordates can be stained with a chicken Col2 α 1 antibody. Our phylogenetic analyses indicate that *AmphiColA* is not a strict ortholog of Col2 α 1 but rather is a sister to the entire vertebrate *ColA* family, and whether the Col2 α 1 antibody can recognize *AmphiColA* protein is unclear. Nonetheless, our results cannot exclude the possibility that *AmphiColA* is activated in the pharynx during metamorphosis, when the gill region undergoes extensive remodeling.

Discussion

Col2 α 1-Based Cartilage Is a Shared Character of Crown-Group Vertebrates. Cartilage based on type II collagen has long been considered a unifying character of gnathostomes that separates them from lampreys and hagfishes (1, 19). Recently, we reported that lampreys possess two Col2 α 1 orthologs and that adult lamprey cartilage is composed of Col2 α 1 protein (2). Here, we have gone on to show that hagfishes, the sister group to lampreys, also possess a Col2 α 1 ortholog, and we demonstrate that Col2 α 1 protein is localized to their soft cartilage. Taken together, these results suggest that the common ancestor of all crown-group (the living jawed and jawless) vertebrates had Col2 α 1-based cartilage. The presence of an undifferentiated clade A fibrillar collagen in lancelets and tunicates (20) suggests that the expansion of the *ColA* gene family occurred in stem vertebrates after the divergence of lancelets and tunicates. Thus, Col2 α 1-based cartilage is a synapomorphy of all crown-group vertebrates.

Our finding that hagfish hard cartilage lacks Col2 α 1 indicates that hagfishes also possess some non-Col2 α 1-based cartilage. This finding highlights a relationship between the profile of collagen expression and the structure of skeletal tissues in hagfishes. In gnathostomes, cartilage matrix is composed predominantly of Col2 α 1, whereas bone matrix is mostly Coll1 α 1. During endochondral ossification, which involves the transition from cartilage to bone, Col2 α 1 is replaced by Coll1 α 1 in the skeletal matrix. Our observation that a single skeletal element in the hagfish can have a mosaic structure (both soft and hard

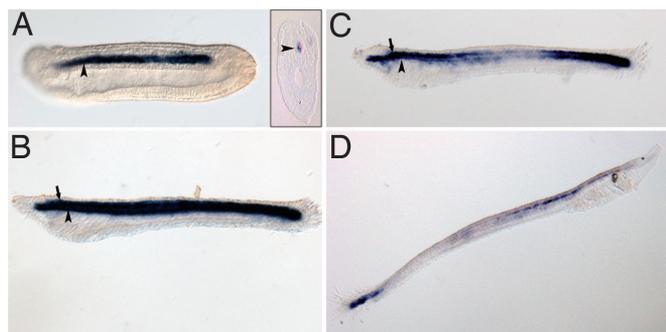


Fig. 3. Expression of *AmphiColA* during lancelet development. Whole-mount *in situ* hybridizations of lancelet embryos with an *AmphiColA* antisense riboprobe are shown. Anterior is to the left in A–C and to the right in D. Stages shown are 18 h (A), 30 h (B), 36 h (C), and 5 days (D) postfertilization. (A Inset) Transverse section of 18-h embryo. The arrowhead indicates notochord, and the arrows point to the neural tube.

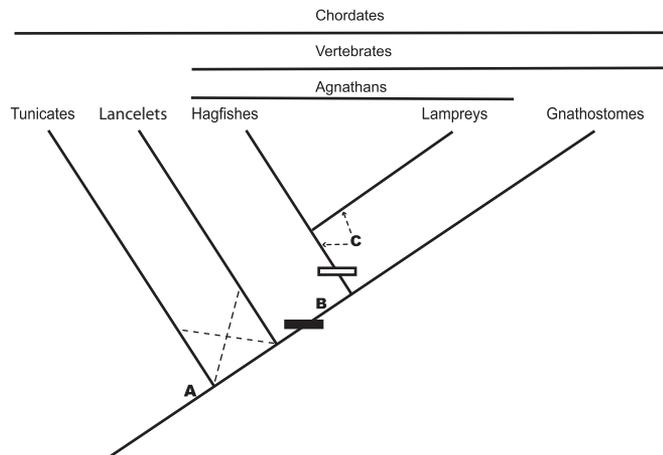


Fig. 4. Origin of type II collagen-based cartilage. Phylogenetic relationships of tunicates, lancelets, hagfishes, lampreys, and gnathostomes (37, 38) are shown. Dashed lines indicate alternative branch positions (27). Letters point out major steps in *Col2α1*-based cartilage evolution. The solid rectangle indicates the origin of the *Col2α1* gene by *ColA* gene duplication(s). The hollow rectangle indicates the origin of *Col2α1a* and *Col2α1b* by *Col2α1* gene duplication. (A) Ancestral clade A collagen gene expression in notochord of stem chordates. (B) Origin of *Col2α1*-based cartilage in stem vertebrates. (C) Subfunctionalization of *Col2α1*, in which the expression domain of the ancestral gene is partitioned between the duplicate genes (*Col2α1a* and *Col2α1b*); arrows indicate possible positions of this event.

cartilage) that corresponds to mosaic distribution of *Col2α1* raises the question of whether one type of cartilage may develop from the other, perhaps by altering the proportion of *Col2α1* relative to other types of collagen or noncollagen matrix proteins. Alternatively, soft and hard cartilage elements may arise from distinct chondrocyte lineages (*Col2α1*-positive and -negative), with mosaic cartilages having a mixed lineage. Further characterization of hagfish hard cartilage will clarify whether it is composed of other types of collagen or whether this subset of the skeleton is noncollagenous.

Did Vertebrate Chondrocytes Evolve from the Notochord? It has been suggested that the notochord may represent a primitive form of cartilage, based on their many shared structural, cellular, and molecular properties, and that vertebrate chondrocytes may have evolved from notochordal cells (21, 22). In gnathostomes, the notochord and/or notochordal sheath expresses most of the vertebrate clade A fibrillar collagen genes (15, 23–25). Our finding that *AmphiColA* is expressed in the notochord and notochordal sheath of *Branchiostoma floridae*, taken together with the recently reported data on *Ciona intestinalis* *CiFColl1* and *Branchiostoma belcheri* *BbFColl1* (both clade A fibrillar collagens), supports the idea that an ancestral *ColA* gene was expressed in the notochord of stem-group chordates (14, 20, 26). We suggest that duplication and divergence of the clade A collagen genes in stem-group vertebrates may have facilitated the evolutionary diversification of chondrocytes and notochordal cells. This hypothesis deals specifically with the origin of vertebrate chondrocytes, and it is important to note that cartilage is also found in several invertebrates, including cephalopods, snails, and horseshoe crabs (1, 21). Future work on the molecular basis of invertebrate chondrogenesis should reveal whether fibrillar collagens also were used in these independent evolutionary events.

Clade A Fibrillar Collagen Duplication Facilitated Evolution of the Vertebrate Skeleton. The identification of an undifferentiated, ancestral *ColA* gene in lancelets indicates that the duplication event that gave rise to *Col2α1* occurred after the divergence of lancelets and vertebrates. We have focused our analysis on

hagfishes and lancelets; however, it has recently been suggested that tunicates are the closest sister group to vertebrates (27). As noted above, tunicates also have an undifferentiated *ColA* gene (20), indicating that the duplication also postdates their divergence from the lineage leading to vertebrates. Taken together, the data show that the origin of *Col2α1* from the ancestral *ColA* gene must have occurred in stem vertebrates. We propose that this duplication event may have been critical for the origin of the vertebrate skeleton (Fig. 4).

Our data fit with the hypothesis that at least one round of genome duplication occurred between the origin of chordates and the origin of vertebrates (28, 29). The expansion of the clade A fibrillar collagen gene family, particularly the origin of *Col2α1* and *Colla1*, may account for the unique skeletal matrices of vertebrate cartilage and bone. Although cartilage has evolved multiple times in metazoa, it is unclear from the fossil record whether cartilage or bone evolved first in the vertebrates (30). Our results raise the possibility that *Colla1* and *Col2α1* arose from the same duplication event, and thus the major matrix components of bone and cartilage may have evolved at the same time in vertebrates. Finally, it is noteworthy that clade A fibrillar collagens are physically linked with *Hox* gene clusters in vertebrates and in echinoderms (31). Duplication of the clade A fibrillar collagen genes therefore may have coincided with the *Hox* cluster duplications (32, 33). Coordinated expansion of the *ColA* and *Hox* gene families may have facilitated the diversification of vertebrate connective tissue types and provided a mechanism for differentially patterning them.

Materials and Methods

Animals. Adult lancelets (*B. floridae*) were collected in Tampa Bay, FL. Spawning was induced in the laboratory according to published methods (34). Atlantic hagfishes (*Myxine glutinosa*) were purchased from Ward's Natural Science (Rochester, NY).

Gene Cloning. RNA was extracted by using TRIzol Reagent (Invitrogen, Carlsbad, CA). cDNA was made by reverse transcription reactions by using a SuperScript II first-strand cDNA synthesis kit (Invitrogen). Fibrillar collagen genes were amplified with degenerate primers designed by using the CODEHOP program (35). PCRs were carried out by using BD Advantage 2 PCR enzyme system (Clontech, Mountain View, CA) from cDNA. In 50- μ l reactions containing 1- μ l forward (5'-GGCCCTCCCGCCTGCARGGNATGCC-3') and 1- μ l reverse (5'-GGGGCCGATGTCCACGCCRAAYTCYTG-3') primers (20 pmol/ μ l), 5 μ l of 10 \times buffer, 1 μ l of cDNA template, 1 μ l of dNTP at 10 mM (each), 1 μ l of *Taq* polymerase mixture, and 40 μ l of double-distilled water. Reactions were amplified as follows: 94°C for 1 min followed by 35 cycles of 94°C for 45 s, 65°C for 45 s, 68°C for 3 min, and a final 10-min elongation at 68°C. PCR products were purified by using QIAGEN II gel extraction kit (Qiagen, Valencia, CA) then cloned into pCRII-TOPO (Invitrogen) for sequencing.

Sequence Analysis and Molecular Phylogenetics. Inferred protein sequences for hagfish and amphioxus cDNAs were initially assigned to the clade A fibrillar collagen families on the basis of BLAST searches and conserved domains. These preliminary assignments were followed by estimates of their amino acid identities and phylogenetic relationships. Multiple sequence alignments for available fibrillar A, B, and C collagen proteins, including the new hagfish and amphioxus sequences, were generated with ClustalX. Phylogenetic analyses of these multiple protein alignments were conducted with Bayesian phylogenetics, minimum evolution, and maximum likelihood methods, as described previously (2). The following amino acid sequences were retrieved from GenBank for inclusion in our phylogenetic analyses: mouse *Col2α1*, B41182; rat *Col2α1*, NP_037061; dog *Col2α1*, NP_001006952; chick *Col2α1*,

