

Demosponge and Sea Anemone Fibrillar Collagen Diversity Reveals the Early Emergence of A/C Clades and the Maintenance of the Modular Structure of Type V/XI Collagens from Sponge to Human^{*[S]}

Received for publication, June 16, 2008, and in revised form, July 29, 2008 Published, JBC Papers in Press, August 11, 2008, DOI 10.1074/jbc.M804573200

Jean-Yves Exposito^{†1}, Claire Larroux[§], Caroline Cluzel[‡], Ulrich Valcourt[‡], Claire Lethias[‡], and Bernard M. Degnan[§]

From the [‡]Institut de Biologie et Chimie des Protéines (IBCP), 7 passage du Vercors, Lyon, F69367, France, CNRS, UMR 5086; Université de Lyon, France, Université Lyon 1, France, IFR 128, Lyon, France, and the [§]School of Integrative Biology, University of Queensland, Brisbane, Queensland 4072, Australia

Collagens are often considered a metazoan hallmark, with the fibril-forming fibrillar collagens present from sponges to human. From evolutionary studies, three fibrillar collagen clades (named A, B, and C) have been defined and shown to be present in mammals, whereas the emergence of the A and B clades predates the protostome/deuterostome split. Moreover, several C clade fibrillar collagen chains are present in some invertebrate deuterostome genomes but not in protostomes whose genomes have been sequenced. The newly sequenced genomes of the choanoflagellate *Monosiga brevicollis*, the demosponge *Amphimedon queenslandica*, and the cnidarians *Hydra magnipapillata* (*Hydra*) and *Nematostella vectensis* (sea anemone) allow us to have a better understanding of the origin and evolution of fibrillar collagens. Analysis of these genomes suggests that an ancestral fibrillar collagen gene arose at the dawn of the Metazoa, before the divergence of sponge and eumetazoan lineages. The duplication events leading to the formation of the three fibrillar collagen clades (A, B, and C) occurred before the eumetazoan radiation. Interestingly, only the B clade fibrillar collagens preserved their characteristic modular structure from sponge to human. This observation is compatible with the suggested primordial function of type V/XI fibrillar collagens in the initiation of the formation of the collagen fibrils.

Collagen is often defined as one of the specific components of metazoan extracellular matrices, although collagenous Gly-Xaa-Yaa repeated sequences have been reported in some bacteria, viruses, and fungi (1–4). In humans, 29 types of collagen have been described that can be divided into several families according to their primary structures and supramolecular organization (5, 6). Among these metazoan families, only two of them, the fibrillar and the basement membrane type IV collagens, have been described in the earliest branching multicellular animals, sponges and cnidarians (7–10).

Like all collagens, the fibrillar molecules are made of three α chains, which can either be identical or result from a combination of two or three genetically distinct α chains. Each α chain consists of a major uninterrupted triple helical or collagenous domain made up of \sim 338 Gly-Xaa-Yaa triplets, which is flanked by two noncollagenous regions, the N- and C-propeptides. During the maturation of procollagens into collagen molecules, the two propeptides are generally cleaved by specific proteinases yielding processed molecules consisting of an \sim 300-nm-long rod-like structure, representing the triple helix, flanked by short noncollagenous segments, the N- and C-telopeptides.

Once processed, these fibrillar collagen molecules are involved in the formation of the well known cross-striated fibrils. In mammals, the fibrillar collagens involved in the formation of cross-striated fibrils are types I–III, V, and XI. These fibrils are usually heterotypic structures, consisting of one quantitatively minor (V or XI) and one or two quantitatively major (I–III) types of fibrillar collagen. Fibrils present in cartilage, which are constructed with type II and XI collagens, can be distinguished from those located in noncartilaginous tissues, which include type I, III, and V collagens. More recently, it has been shown that a newly characterized fibrillar collagen, type XXVII, is involved in the formation of thin nonstriated fibrils (11, 12). As is the case for type XXIV collagen, the major triple helix of type XXVII is slightly shorter than other fibrillar collagen chains and has two glycine substitutions and one Gly-Xaa-Yaa imperfection (13–15).

From evolutionary studies, mammalian fibrillar collagen chains have been divided into three clades: the A clade including types I–III and the pro α 2(V) chains; the B clade comprising pro α 1(V), pro α 3(V), pro α 1(XI), and pro α 2(XI) chains; and the C type XXIV and XXVII collagen chains (14, 16). The α chains of the A clade possess a VWC² module (absent in the pro α 2(I) chain) in their N-propeptide, in addition to a minor triple helix. The other two fibrillar collagen clades possess a TSPN module in their N-propeptide in addition to or in the absence of a minor

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

[S] The on-line version of this article (available at <http://www.jbc.org>) contains supplemental Figs. S1–S4.

¹ To whom correspondence should be addressed. Tel.: 33-4-72-72-26-77; Fax: 33-4-72-72-26-04; E-mail: jy.exposito@ibcp.fr.

² The abbreviations used are: VWC, von Willebrand factor-type C; TSPN, thrombospondin N-terminal-like domain; EST, expressed sequence tag; contig, group of overlapping clones; COLF1, Fibrillar collagens C-terminal domain; WAP, whey acidic protein; RT, reverse transcriptase; NJ, neighbor-joining; ML, maximum likelihood.

triple helix for B and C clade members, respectively. These three fibrillar collagen clades defined in vertebrates have been found in deuterostome invertebrates, whereas only A and B clade members have been characterized in protostomes (17–19). More recently, several Hydra fibrillar collagen chains have been characterized, which from phylogenetic studies appear to be most closely related to the A clade members (9), although none of these collagens possess a VWC module in their N-propeptide. Moreover, Blast analyses of sponge or Hydra ESTs suggest the presence of B clade collagens in these taxa (9, 20). Fibrillar collagen chains of undetermined clade have also been characterized in the freshwater demosponge *Ephydatia mülleri* (21) and in the marine demosponge *Suberites domuncula* (22).

In this report, we have studied the evolution of fibrillar collagens by taking advantage of the publicly available genome data from the sponge *Amphimedon queenslandica* (formerly classified as *Reniera* sp.) and the sea anemone *Nematostella vectensis*. We demonstrate that the formation of an ancestral B clade fibrillar collagen chain predated the eumetazoan radiation. Moreover, demosponge and cnidarian data suggest that although the emergence of the three fibrillar collagen clades occurred early in evolution, only the B clade preserves its characteristic modular structure in modern metazoans, from sponges to humans.

EXPERIMENTAL PROCEDURES

Data Base Searching—Sequences from metazoan fibrillar collagen chains used in this study were obtained either from the European Bioinformatics Institute, from previous work on mosquito, honeybee, and ascidian collagens (17), or from blast searches in sponge and cnidarian genomes. To identify fibrillar collagens in early branching metazoans, different approaches were used. For the demosponge *A. queenslandica*, simple reads from whole genome shotgun sequencing available in Trace Archive data bases at National Center for Biotechnology Information were mined with Blast using sequences encoding C-propeptides and triple helix from different sources. Each read isolated by this approach and its mate were used to construct contigs by repeating several cycles of blast analysis and contig assembly. Thus, this strategy permitted us to assemble five genomic contigs and to characterize seven distinct genes encoding fibrillar collagen chains (termed Amq1 α to Amq7 α). Each contig sequence was checked by compiling all of the overlapping whole genome shotgun reads, the coverage being greater than 4. Gene structures were (i) predicted by using the *ab initio* program GENSCAN (23) at the Massachusetts Institute of Technology and (ii) completed by careful examination of the open reading frame and by matching these assemblies with existing ESTs. Several ESTs covering part of the coding region were obtained for Amq1 α (CAYH1651 and CAYI6544), Amq2 α (CABF4997, CAYH5770, CAYI7575, and CAYI5900), Amq3 α (CABF3931, CAYH3995, CAYH3529, CAYH6239, CAYH4772, and CABF22710), Amq4 α (CAYI5983, CAYH5842, and CAYI1569), Amq5 α (CAYH3025, CAYI6261, CAYI7211, and CAYI7901), Amq6 α (CABF3357, CAYH6535, and CAYI9396), and Amq7 α

TABLE 1
Metazoan fibrillar collagens

Asterisk indicates that a score is less significant than the required threshold (*E* value > 1).

Common name	Species name	Accession number or sequence identity	Abbreviation	Smart TSPN E-Value
Freshwater sponge	<i>E. mülleri</i>	P18856, Q06452	Emu1 α	
Sea sponge	<i>A. queenslandica</i>	WGS, ESTs ^a	Amq1 α	
		WGS, ESTs ^a	Amq2 α	
		WGS, ESTs ^a	Amq3 α	
		WGS, ESTs ^a	Amq4 α	
		WGS, ESTs ^a	Amq5 α	2.78e-01
		WGS, ESTs ^a	Amq6 α	1.15e+00*
		WGS, ESTs ^a	Amq7 α	6.37e+01*
Hydra	<i>H. Magnipapillata</i>	WGS, ESTs ^a	HCol1	
		WGS, ESTs ^a	HCol2	
		WGS, ESTs ^a	HCol3	
		WGS, ESTs ^a	HCol5	
		WGS, ESTs ^a	Nve1 α	
Sea anemone	<i>N. vectensis</i>	ID # 25350	Nve1 α	
		ID # 6496	Nve2 α	
		ID # 21796	Nve3 α	
		ID # 36007	Nve4 α	
		ID # 17317	Nve5 α	
		ID # 26644	Nve6 α	
		ID # 166	Nve7 α	1.24e-23
		ID # 40962	Nve8 α	1.14e+02*
Mosquito	<i>Anopheles gambiae</i>	ENSANGG00000016690	Agao1	
		ENSANGG00000018512	Agao2	4.68e-19
Honeybee	<i>Apis mellifera</i>	AADG02012353	Ame1 α	
		AADG02005865	Ame2 α	1.11e-19
Abalone	<i>Haliotis discus</i>	O97405	Hdcol1 α	
		O97406	Hdcol2 α	
Sea urchin	<i>S. purpuratus</i>	Q26634	Spu1 α	
		Q26639	Spu2 α	
Ascidian	<i>Paracentrotus lividus</i> <i>Ciona intestinalis</i>	XP_001192332	Spu6 α	1.87e-09
		CAE53096	Plf5 α	
		ci0100150759	Cin759	
		ci0100154301	Cin301	8.34e-16
		ci0100131606	Cin606	1.28e+03*
Human	<i>Homo sapiens</i>	ci0100144916	Cin916	
		P02452	Hsac1(I)	
		P08123	Hsac2(I)	
		Q14027	Hsac1(II)	
		P02461	Hsac1(III)	
		P20908	Hsac1(V)	6.53e-82
		P05997	Hsac2(V)	
		P25940	Hsac3(V)	1.96e-21
		P12107	Hsac1(XI)	1.15e-68
		P13942	Hsac2(XI)	7.68e-79
		Q7Z5L5	Hsac1(XXIV)	1.02e-05
Q8IZC6	Hsac1(XXVII)	1.46e-05		

^a See Experimental Procedures.

(CAYI8810 and CAYI1553). *A. queenslandica* ESTs can be downloaded from the Ensembl Trace Server.

For the cnidarian *Hydra magnipapillata*, the same approach was used to identify orthologous genes of *Hydra vulgaris*, *Hcol1*, *Hcol2*, *Hcol3*, and *Hcol5* (9). Several ESTs covering part of the coding region were obtained for *H. magnipapillata* *Hcol1* (DN602947, DN253482, DT620184, D612454, DN603678, CCF658308, DR435484, and DT613951), *Hcol2* (DN243434, CX832112, CX835856, DN812915, DT616768, DT609283, and DT606480), *Hcol3* (DN636048 and DN811440), and *Hcol5* (CX770252, DN603429, DN815968, DT616007, DN810685, and CF675546).

For the recently completed genomes of the cnidarian *N. vectensis* and the sea urchin *Strongylocentrotus purpuratus*, Blast analysis was carried out using *Nematostella* (evodevo.bu.edu/stellabase/) and sea urchin servers. This allowed us to identify eight fibrillar collagen genes in *Nematostella* (called *Nve1 α* to *Nve8 α*) and the *Strongylocentrotus* ortholog of the *Paracentrotus lividus* COLP6 α gene (17). ESTs covering part of the coding region were obtained for *Nve1 α* (CAGN4739 and CAGN8651), *Nve3 α* (CAGN3324), *Nve4 α* (CAGN8932), *Nve5 α* (CAGN9683), *Nve7 α* (DV089101), and *Nve8 α* (CAGF11976).

The accession numbers, species abbreviations, and sources are compiled in Table 1. The modular architecture of the fibrillar

Early Evolution of Fibrillar Collagen Genes

lar collagen chains has been analyzed using the Smart server (24). The nucleotide sequences of the sponge and Hydra contigs assembled in this work are available upon request. Sequences from the choanoflagellate *Monosiga brevicollis* genome (25) encoding triple helical or C-propeptide (COLF1) modules were obtained using the advanced search tool of the Joint Genome Institute server. Blast analysis failed to recover additional sequences encoding TSPN or COLF1 modules.

Amphimedon Molecular Techniques—*A. queenslandica* adult sponges, larvae, embryos, and postlarvae were procured as previously described (26, 27). After storage in RNA later (Sigma), total RNA was extracted using an RNeasy extraction kit (Qiagen) as per the manufacturer's instruction. RT-PCR was successfully conducted on total RNA as previously described (28) using the following primer pairs: Amq1 α .1 GTCTTCAATTCAGTTCCTAGTTG + Amq1 α .2 GTCCAGACGTTCCTTTGTC, Amq3 α .1 GTGTTTCTACTGCAGGCT + Amq3 α .2 GTCAACTCCATCTTTACCATC, Amq5 α .1 TGCATTAGGTGTAGCAGCCAC + Amq5 α .2 CTTGGTTCAGCCAGACTGAG, Amq6 α .3 CTTTGAGACCCTGTCATTTAGAC + Amq6 α .2 GTGCAGCGTTACCAGT-GTC, and Amq7 α .3 CTCCTCTGATGGACGTAATGAG + Amq7 α .2 CTTTAGCTCCATCTGTACCAG. RT-PCR sequences are available (GenBank™/European Bioinformatics Institute Data Bank accession numbers FM165586, FM165587, FM165588, FM165589, and FM165590).

Alignment and Evolutionary Analysis—The major triple helix of metazoan fibrillar collagen chains were first aligned using ClustalW (29) with the PAM alignment matrix at the European Bioinformatics Institute. The resulting initial alignments were manually improved using the SeaView alignment editor (30) and were scanned using the RASCAL component (31) of PipeAlign, a toolkit for protein family analysis. Neighbor-joining (NJ), maximum likelihood (ML), and Bayesian phylogenetic analyses were performed on the validated multiple alignments. NJ trees were determined using the Phylo win program (30). NJ bootstrap support was based on 1000 replicates using SEQBOOT and CONSENSE (majority rule extended) of the PHYLIP package (32) to generate data replicates and consensus trees, respectively. The PHYML v2.4.4 algorithm (33) was applied for the ML analyses, under the WAG amino acid substitution model and with 100 bootstrapped data sets using the PhyML Online server. Bayesian analysis was carried out as implemented in MrBayes 3.1.2 (34, 35) with a mixed amino acid model incorporating invariant sites and gamma parameter. Analyses in MrBayes were conducted using default settings, with two parallel runs of one million generations each, using four chains and a sample frequency of 100. Likelihoods of Bayesian analyses converged after the initial 21,000 generations. We therefore discarded the initial 210 trees from each of the two parallel runs and computed posterior probabilities from the remaining trees. The standard deviation of split frequencies after 1 million generations was below 0.01.

The illustrations were drawn using the TreeView program (36) and then annotated using Adobe Illustrator. In the absence of a good outgroup, some trees were rooted using the midpoint rooting method and the retree editor from the PHYLIP package

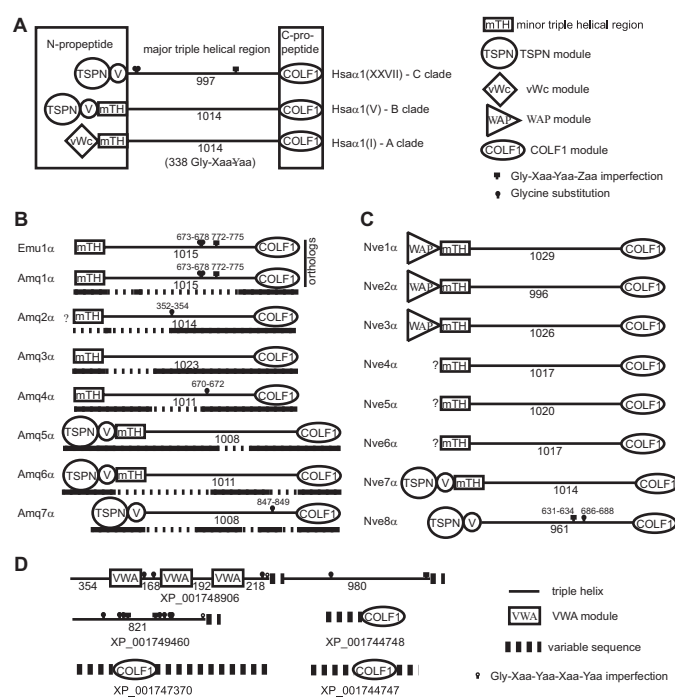


FIGURE 1. Modular structure of sponge and sea anemone fibrillar collagens and of choanoflagellate collagen-like proteins. A–C, the general modular architecture of human members of the three fibrillar collagen clades (A), sponge (B), and sea anemone (C) α chains is depicted. For each α chain, the Arabic numbers under their domain architecture indicate the lengths in amino acids of the major triple helices. For the sponge fibrillar collagens chains, the Arabic numbers above the glycine substitutions and Gly-Xaa-Yaa-Zaa imperfections represent their precise location according to their numbering in the major triple helix. Under the *A. queenslandica* α chains, the solid lines represent regions confirmed by EST and RT-PCR sequences, and the dashed lines indicate that these sequences were deduced from genomic analysis. Amq, marine demosponge *A. queenslandica*; Emu, freshwater demosponge *E. mülleri*; Nve, starlet sea anemone *N. vectensis*; ?, undetermined sequence. D, putative proteins including triple helical or COLF1 (C-propeptide) modules present in the choanoflagellate *M. brevicollis*. The accession numbers of the predicted *M. brevicollis* proteins are indicated below their modular representation. The Arabic numbers under the triple helix indicate their lengths in amino acids. In this figure, the different modules are not represented to scale.

(32). Midpoint rooting places the root at the middle of the longest path between the two most distantly related taxa.

RESULTS

Demosponge and Cnidarian Fibrillar Collagen Diversity—Aiming to understand the early evolution of fibrillar collagens, we searched the genomes and ESTs of one demosponge and two cnidarians for genes possessing triple helix and C-propeptide sequences similar to Hydra and human fibrillar collagens. This survey led to the characterization of seven fibrillar collagen α chains in the demosponge *A. queenslandica*, named Amq1 α to Amq7 α , and eight fibrillar collagen α chains in the sea anemone *N. vectensis*, named Nve1 α to Nve8 α . For *H. magnipapillata*, we could detect only the complete major triple helical sequences for the α chains orthologous to those previously characterized in *H. vulgaris* (Hcol1, Hcol2, Hcol3, and Hcol5) (9). The sponge and cnidarian α chains have the general modular structure of eumetazoan fibrillar collagens, i.e. a major triple helix flanked by the N- and C-propeptides (Fig. 1). Parts of the primary structure of the sponge fibrillar collagens (mostly the N terminus) were deduced from reading frame predictions

(see "Experimental Procedures" and Fig. 1B). For this reason, the mRNA sequences encoding N-propeptide domain of sponge chains Amq1 α , Amq3 α , Amq5 α , Amq6 α and Amq7 α were amplified by RT-PCR to confirm our predictions and to analyze the expression of the related collagen genes throughout the life cycle of *A. queenslandica* (Fig. 2). Most of these genes, specifically Amq1 α , Amq3 α , Amq5 α , and Amq7 α , are expressed throughout the sponge life cycle. Among these genes, Amq1 α and Amq3 α have higher expression levels during metamorphosis than other stages, whereas Amq7 α has lower expression levels (barely detectable) during this period. In contrast, Amq5 α seems to be more highly expressed during embryogenesis and in the larva, with levels dropping off during metamorphosis and in adults. The last gene, Amq6 α , does not appear to be expressed in sponge adults.

Taking into account the cnidarian and sponge fibrillar collagens for which the N-propeptide sequence is available, three groups of α chains can be distinguished by their modular orga-

nization. The first group includes fibrillar collagen chains that have an N-propeptide restricted to a minor triple helix (Amq1 α , Amq3 α , Amq4 α , and Hydra Hcol1), a situation observed in some A clade members such as mammalian pro α 2(I) and sea urchin 1 α chains (37). Moreover, Amq1 α appears to be the ortholog of the *E. mülleri* Emu1 α chain with two glycine substitutions and one Gly-Xaa-Yaa-Zaa imperfection in identical positions (Fig. 1B).

The second group comprises cnidarian α chains possessing WAP modules in their N-propeptide (Nve1 α , Nve2 α , Nve3 α , Hcol2, and Hcol3; Fig. 1C and Ref. 9). It should be noted that despite the high level of amino acid identity between fibrillar collagens of the two Hydra species (close to 100%), two differences were observed for two collagen genes. First, the major triple helix of *H. magnipapillata* Hcol3 is comprised of 1023 residues instead of 969 amino acids for the *H. vulgaris* Hcol3 chain. In *H. magnipapillata*, the additional 54 residues are encoded by a 162-bp exon (supplemental Fig. S1A), suggesting that the discrepancy between these two species results from (i) an alternative splicing event in *H. vulgaris*, (ii) the loss of this exon in *H. vulgaris*, or (iii) a cloning artifact in the cDNA characterized in the *H. vulgaris* study. Second, the *H. magnipapillata* Hcol2 fibrillar collagen chain possesses one less Gly-Xaa-Yaa triplet in the main collagenous domain compared with the *H. vulgaris* chain. Hence, one of the two GEQ triplets (31st or 32nd of the major triple helix) present in the *H. vulgaris* Hcol2 chain is absent from *H. magnipapillata* (supplemental Fig. S1B).

In the third group, which includes sponge Amq5 α , Amq6 α , and Amq7 α and sea anemone Nve7 α and Nve8 α collagen chains (Fig. 1), we noted the presence of a large noncollagenous domain in their N-propeptide. From SMART analyses, this noncollagenous domain clearly corresponds to a TSPN module for Amq5 α and Nve7 α . As shown in Table 1, the TSPN module of sea anemone Nve7 α and sponge Amq5 α chains have *E* value scores comparable with those obtained with bilaterian B clade and C clade collagen TSPN modules, respectively. From multiple alignments (Fig. 3 and Table 2), the TSPN domain of Amq5 α and Nve7 α share on average 18 and 30% of identity with comparable modules of human B clade members, respectively. For the other sponge (Amq6 α and Amq7 α) and sea anemone (Nve8 α) chains in this group, the large noncollagenous domain might also correspond to a TSPN module (Table 1), although the *E* values for the TSPN modules are not significant in SMART.

The peculiar location of this noncollagenous domain in their N-propeptide strongly supports that these sequences are TSPN modules. This result is in agreement with poor identity observed for this module between B and C clade members or diploblast α chains and B/C clade fibrillar collagens (Table 2 and supplemental Fig. S2, for multiple alignment analysis). Based on the modular structure of their N-propeptide, Amq5 α , Amq6 α , and Nve7 α are B clade

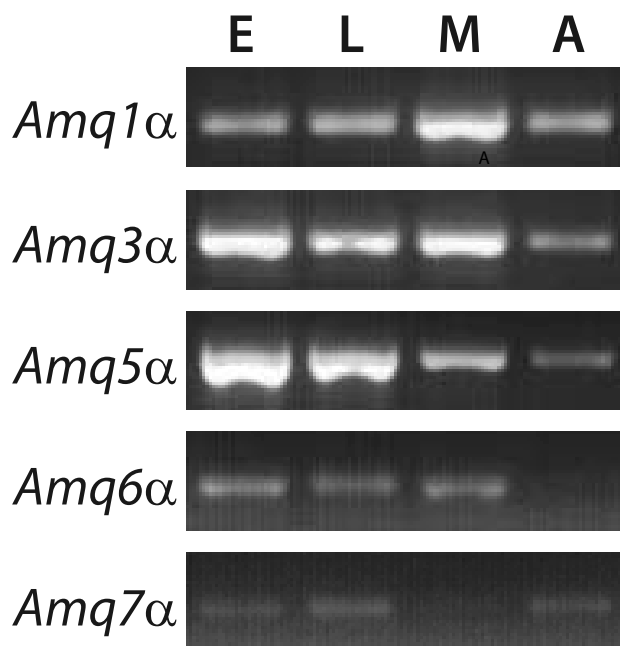


FIGURE 2. Gene-specific RT-PCR of 5 collagen genes during the *Amphimedon* life cycle. Columns correspond to the different stages. E, embryogenesis; L, larva; M, metamorphosis; A, adult. The embryogenesis, larva, and metamorphosis stages correspond to a pool of embryos, larvae, and postlarvae of different ages, respectively.



FIGURE 3. Representative multiple alignments of the TSPN module from sponge, starlet sea anemone, and human fibrillar collagen chains. The alignments were generated with ClustalW, and gray boxes denote sequence identity (Threshold = 0.75). See Table 1 for abbreviations.

Early Evolution of Fibrillar Collagen Genes

members (they contain a TSPN module and a minor triple helix), whereas Amq7 α and Nve8 α are related to C clade chains (they contain a TSPN module but not a minor triple helix). However, the latter two chains differ from human C clade collagens in the length of their major triple helix (Fig. 1): longer for Amq7 α (1008 residues) and significantly shorter for Nve8 α (961 residues). From the data available to date, we could not investigate the N-propeptide structure of three *N. vectensis* chains, Nve4 α to Nve6 α , and of the sponge Amq2 α fibrillar collagen chain.

Collagen-like Proteins and COL1 Module in the Choanoflagellate *M. brevicollis*—Choanoflagellates are considered the closest known relatives of metazoans. In the recent work describing the genome of *M. brevicollis*, the authors indicated the presence of sequence encoding triple helical domains and COL1, a module always retrieved to date at the C terminus of metazoan fibrillar collagen chains (25). One of the two *M. brevicollis* putative collagenous proteins has several triple helical domains interspersed with VWA modules (Fig. 1D). This modular organization is reminiscent of the Hydra collagen Hcol6 (9), although the distribution of VWA and triple helical domains are not identical. Three other *M. brevicollis* proteins contain a COL1 module. Sequence alignments of *M. brevicol-*

lis COL1 modules with comparable domains of some demosponge and cnidarian fibrillar collagens are represented in Fig. 4. Among the eight specific cysteine residues of metazoan COL1 module, the three choanoflagellate sequences lack cysteine residues 5 and 8, which are responsible for an intrachain disulfide bond in fibrillar α chains (38). These sequences also lack cysteine residues 2 and 3. Interestingly, either cysteine 2 or cysteine 3 is absent in some fibrillar collagen chains (for example, the vertebrate pro α 2(I) chain), whereas both of these cysteine residues are absent in sponge Amq3 α and Amq4 α and in the worm *Arenicola marina* FAm1 α chain (16).

Genomic Context of Genes Encoding Demosponge and Cnidarian Fibrillar Collagen Chains—Two of the *N. vectensis* chains (Nve2 α and Nve8 α) have major triple helices shorter than A and B clade chains (Fig. 1). As shown in supplemental Fig. S3, the loss of exons might explain the shortening of their major triple helix. With the exception of these two genes, all of the sponge and cnidarian fibrillar collagen genes have intron/exon structures in their regions encoding the major triple helix that match the suggested organization of an ancestral fibrillar collagen gene (39). Moreover, in the sponge *A. queenslandica*, two sets of paralogous collagen genes are arranged in tandem (Amq1 α -Amq5 α and Amq2 α -Amq4 α) in a head-to-head fashion. The base pair distance between the two ATG start codons of the gene pairs Amq1 α -Amq5 α and Amq2 α -Amq4 α is 932 and less of 3400, respectively. In the sea anemone, two fibrillar collagen genes (Nve6 α -Nve7 α) are also arranged in tandem, but in a tail-to-tail configuration, with 2699 bp separating their TGA stop codons.

Presence of B and C Clade Fibrillar Collagen Chains in Demosponges and Cnidarians—Comparison of modular organization of metazoan fibrillar collagen chains provides evidence for the presence of A clade and B/C clade-related chains in early

TABLE 2
Percentage of identity between the TSPN module of metazoan fibrillar collagens

	Amq5 α	Amq6 α	Amq7 α	Hsaa1(V)	Hsaa3(V)	Hsaa1(XI)	Hsaa1(XXIV)	Hsaa1(XXVII)	Nve7 α
Amq5 α	100	14.66	7.30	16.58	18.43	18.13	15.08	15.30	15.03
Amq6 α		100	12.04	12.31	13.92	12.44	19.90	13.27	13.85
Amq7 α			100	13.54	14.05	14.87	16.46	17.42	16.06
Hsaa1(V)				100	38.02	73.04	19.79	21.28	32.16
Hsaa3(V)					100	40.10	19.13	27.72	29.90
Hsaa1(XI)						100	20.31	21.24	29.65
Hsaa1(XXIV)							100	32.58	18.13
Hsaa1(XXVII)								100	19.17
Nve7 α									100

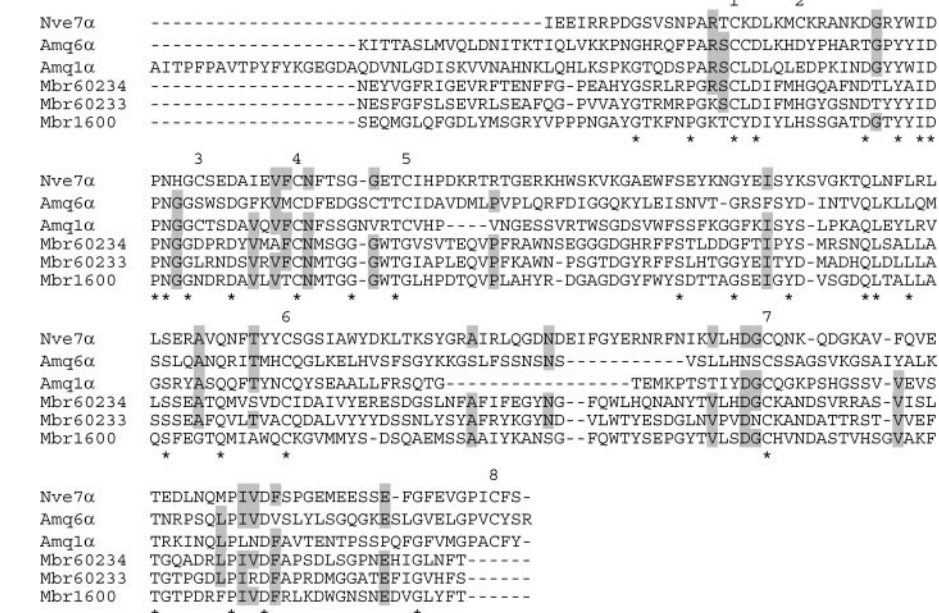


FIGURE 4. Representative multiple alignments of the COL1 module from choanoflagellate, sponge, and sea anemone proteins. The alignments were generated with ClustalW. The gray boxes and asterisks denote sequence identity (Threshold = 0.66) and perfectly conserved amino acids between the proteins analyzed, respectively. Numbering of the cysteine residues is indicated above the multiple alignments. See Table 1 and Fig. 1D for abbreviations.

branching metazoans. We further explored the affinities of sponge and cnidarian collagen chains with phylogenetic analyses. Because of the N-propeptide modular organization diversity and the C-propeptide variability, we only used the major triple helical sequences in these analyses. Multiple alignments of bilaterian major triple helix confirm the specific pattern of exon lengths of fibrillar collagen genes (supplemental Fig. S3 and Ref. 17). However, this is not always the case when multiple alignments are performed using all metazoan α chains (data not shown), except when sponge Amq1 α , Emu1 α , Amq7 α , sea anemone Nve8 α , and ascidian Cin906 α chains are removed. With the assumption that conservation of exon/intron organization of metazoan fibrillar collagen genes in the region encoding the major triple helix reflects conservation of related

Early Evolution of Fibrillar Collagen Genes

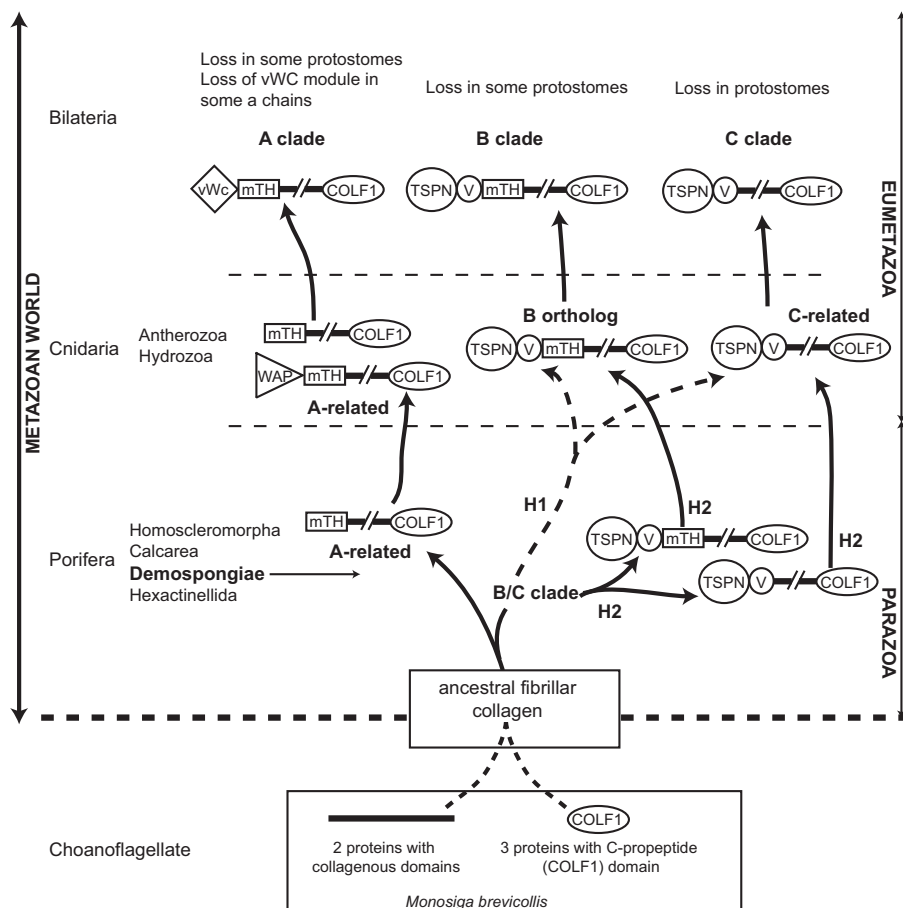


FIGURE 6. Early evolution of the fibrillar collagen family. In this model, we propose that an ancestral fibrillar collagen gene arose in the lineage leading up to the Metazoa. Indeed, the choanoflagellate *M. brevicollis* possesses genes encoding either triple helical or COLF1 module, whereas in demosponges several fibrillar collagen genes are present. From modular structure comparison and phylogenetic analysis, we propose that the split into the three fibrillar collagen clades occurred before the Parazoa/Eumetazoa split. Regarding the A clade, demosponges and cnidarians have several fibrillar collagen chains related to this clade. The acquisition of a VWC module in the N-propeptide domain of an A clade-related chain occurred at the base of the Bilateria, possibly evolving from a WAP module present in cnidarians. Some protostomes have lost gene(s) encoding A clade collagens. Regarding the B/C clade, two different scenarios (H1 and H2) are proposed. In hypothesis H1, demosponges possess fibrillar collagen chains descended from an ancestral B/C clade fibrillar collagen chain containing a TSPN module in its N-propeptide. In this scenario, the emergence of B and C clades occurred between Parazoa-Eumetazoa and Cnidaria-Bilateria splits. In hypothesis H2, the divergence of B and C clades from an ancestral B/C clade α chain predated metazoan cladogenesis. Whatever the scenario, the B clade seems to correspond to the most conserved fibrillar collagen family, whether it is at the modular level (from sponge to human) or at the sequence level (from sea anemone to human).

including B and C clade collagen members, as found in the NJ analysis (low bootstrap support, 43%).

Early Evolution of Metazoan Collagen Chains—Altogether, these data suggest that the divergence between ancestral A and B/C clade fibrillar collagen chains occurred very early during metazoan evolution, before the separation of poriferan and eumetazoan lineages. Based on our analyses, a model for the evolution of fibrillar collagen chains is presented in Fig. 6. The presence of proteins with either triple helical domains or COLF1 module in the choanoflagellate *M. brevicollis* suggests that an ancestral fibrillar collagen harboring both modules arose at the very dawn of metazoans. However, we cannot reject the possibility that this choanoflagellate has secondarily lost genomic sequence encoding a fibrillar collagen chain during its evolution.

In this model, the emergence of A and B/C clades predated the divergence of Parazoa and Eumetazoa, as illustrated by the

diversity of fibrillar collagen chains present in the marine demosponge *A. queenslandica*. Indeed, sponge TSPN-minus α chains are related to the A clade. However, the addition of a VWC module in the N-propeptide of A clade fibrillar collagens seems to have occurred after the cnidarian/bilateria split. Interestingly, the WAP module present in some cnidarian A clade-related chains corresponds to a cysteine-rich domain (8 cysteine residues) that is reminiscent of the VWC module (10 cysteine residues) in its length (~50–60 residues), its location in the N-propeptide, and the presence of two successive cysteine residues near its C terminus. This suggests that a VWC module may have evolved from a WAP module after cnidarians had departed from the main eumetazoan lineage. However, from the StellaBase web server, several predicted sea anemone proteins contain VWC modules. From SMART analysis, all of these sequences are positively recognized as VWC modules (*E* values comprised between 10^{-5} and 10^{-1}), this result being in contradiction to a close relationship between WAP and VWC domains.

As shown in our model, two distinct scenarios can be proposed for the evolution of the B and C clades. In the first scenario, which is supported by our phylogenetic analyses, sponge TSPN-plus collagens are descended from a B/C clade ancestral fibrillar chain. In this case, the divergence of B and C clades occurred between Parazoa-Eumetazoa and Cnidaria-Eumetazoa splits, as suggested by the presence of B and C clade-related fibrillar collagen chains in the sea anemone. In the second scenario, supported by the modular organization of sponge collagen N-propeptides, the emergence of B and C clades from a B/C ancestor occurred earlier, before the emergence of sponges.

DISCUSSION

The sea anemone genome has revealed that in regard to transcription factors and other protein families, the last common ancestor to all modern eumetazoans had a complex genome, which was more similar to vertebrate genomes than those of flies and nematodes (40, 41). In contrast, the genome of the demosponge *A. queenslandica* indicates that gene families may have not yet fully diversified in the first metazoan animals (42–44). Here we use the cnidarian and sponge genomes to explore

the origin and evolution of fibrillar collagen. Based on phylogenetic analyses, the emergence of the three fibrillar collagen clades predated the emergence of the Eumetazoa. Although not strongly supported, the general congruence between this analysis and the modular structure of sponge fibrillar collagens is compelling evidence of the early emergence of the A, B, and C clades, possibly prior to the divergence of poriferan lineage. The major finding that will be discussed below is the great conservation of the B clade during metazoan evolution and its existence in the first multicellular animals.

Early Emergence of the Three Fibrillar Collagen Clades—The use of newly available genomic and EST sequence data from early branching animals gave us insights into the early evolution of the fibrillar collagen family, although some issues remain unresolved. We have shown that demosponges already have fibrillar collagen chains related either to the A clade or the B/C clade. Although the modular organization of the TSPN-Plus sponge fibrillar collagen chains suggests that B and C clade-related chains are present in demosponges, phylogenetic studies do not lend more support to this scenario than to one that sees these two fibrillar collagen clades emerging between Parazoa-Eumetazoa and Cnidaria-Bilateria splits (Fig. 6). The difficulty in supporting either of these scenarios in our phylogenetic analysis has several origins. First, the long branch lengths indicate increased levels of sequence divergence in sponge and cnidarian fibrillar collagen chains, probably reflecting the ancient divergence of these phyla. Several studies suggest that the eumetazoan ancestor lived ~630–830 million years ago (40, 45). However, the cladistic distribution observed in the phylogenetic trees (Fig. 5) does not appear to be due to long branch attraction events because the same clustering was obtained with all methods used, and the sponge chains were retrieved at the base of A and B-C clades. The second point results from the lack of resolution of basal metazoan lineages. The three sponge classes Hexactinellida, Demospongia, and Calcarea are generally considered to be paraphyletic (46). It has also been proposed that Homoscleromorpha, a group of siliceous sponges that possess a basement membrane-like structure, may form a fourth sponge class and should be considered closer to eumetazoans than the demosponges (47). The future availability of genome data from other animal groups at the base of the metazoan tree (all different sponge classes and ctenophores) should improve our model of evolution.

The lack of fibrillar collagen gene but the presence of sequences encoding proteins harboring either collagenous sequence or a COLF1 module in the newly published genome of the unicellular *M. brevicollis* (25), a choanoflagellate representing a sister group to the Metazoa, suggest that fibrillar collagens are metazoan-specific extracellular matrix proteins. Our analyses suggest that these genes evolved at the dawn of the Metazoa to give rise to three clades, which were originally defined in mammals. One of the paradoxes we found in this study is the high number of fibrillar collagen genes in sponges and cnidarians in comparison with protostomes (two genes in honeybee and mosquito) or the invertebrate chordate *Ciona intestinalis* (four genes). As previously indicated for *Hydra* (9), some lineage-specific gene duplications might explain the high number of TSPN-minus fibrillar collagen chains in Porifera and Cni-

daria. The clustering of certain collagen genes in the demosponge and sea anemone genomes supports this scenario. The relationships of *Hydra* and sea anemone fibrillar collagens, despite the presence in some of these α chains of a WAP module, is not obvious. As previously indicated, the depth of the *Hydra-Nematostella* split is comparable with the protostome-deuterostome divergence and emphasizes the distant relationship between anthozoans and hydrozoans (40).

B Clade Collagens and Metazoan Evolution—An important result of this study is the conservation of B clade fibrillar collagens, with the same modular organization and triple helix characteristics from sponge to human. More precisely, in our model and with the available data, cnidarians are the earliest branching metazoan phylum to possess definitive B clade fibrillar collagen ortholog. Moreover, phylogenetic analyses support a relationship between sponge TSPN-plus and B clade fibrillar collagen chains with low statistical support (Fig. 5); although based on modular presence and organization, the sponge possesses B (Amq5 α and Amq6 α) and C (Amq7 α) clade α chains (Fig. 1). Interestingly, Amq5 α seems to be more highly expressed during embryogenesis than in adult, a result reminiscent of vertebrate type V collagen (48). These results are in agreement with previous works suggesting some similarities between invertebrate and types V/XI fibrillar collagen chains (2, 7).

Recent data support a pivotal function of type V collagen molecules ($\alpha 1(V)_2$ and $\alpha 2(V)$) in the nucleation of fibril assembly in mice noncartilaginous tissues (49). In these tissues, mature fibrils are generally heterotypic and are made of types I, III, and V collagens. In *col5a1*^{-/-} mice, in the absence of $\alpha 1(V)$ chain, fibrils are virtually absent in most noncartilaginous tissues, although embryonic fibroblasts synthesize and secrete normal amounts of type I collagen. It has also been reported that type XI is probably buried inside types II/XI heterotypic fibrils in cartilage (50), suggesting a common role for types V and XI collagens (mostly B clade fibrillar collagens) in early fibril initiation. More recently, it has been demonstrated that vertebrate type XXVII C clade fibrillar collagen form 10-nm-thick nonstriated fibrils (11, 12), an ultrastructural network that is entirely distinct from the well known striated fibrils made of A and B clade collagens. In this respect, C clade collagens (not demonstrated for type XXIV collagen) also have the capacity to initiate the formation of fibrils. Moreover, C clade collagens form thin nonstriated fibrils, whereas types V and XI also play an important function in the regulation of fibril diameter. One of the mechanisms regulating the growth of fibrils is the retention of the N-propeptide for vertebrate types V and XI collagens, the sea urchin 5 α chain, and the hydra Hcol1 α chain (51–53). From *in vitro* fibrillogenesis and molecular modeling of triple helical peptides, it has also been argued that the presence of bulky hydrophobic amino acids and glycosylated hydroxylysine residues in the major triple helix might be the main contributor to limiting the lateral accretion of fibrillar collagen molecules, with hydroxylation of lysine residues occurring in the Yaa position of repeating collagen Gly-Xaa-Yaa triplets (54). C clade and to a lesser extent B clade collagen chains have generally reduced levels of alanine and higher levels of lysine and subsequently glycosylated hydroxylysine residues in com-

Early Evolution of Fibrillar Collagen Genes

parison with A clade collagens (9, 12). From these analyses, it has been suggested that the relatively low levels of alanine and high levels of lysine in the major triple helix of a fibrillar collagen may be a predictor of its capacity to form thin fibrils.

In agreement with this hypothesis, the Hydra fibrillar collagen chains present alanine and lysine levels comparable with B and C clade vertebrate collagens and are involved in the formation of thin, 10-nm nonstriated fibrils. In demosponges, striated fibrils have a uniform diameter of 25 nm (2). However, in their major triple helix, the demosponge collagen chains Amq1 α to Amq6 α have an alanine level (6.5–11.05%) comparable with human A clade collagens (I–III, 8.8–11.2%), and a lysine level (3.4–5.75%) comparable either to human A clade collagens (I–III, 3–3.7%) or to types V and XI (3.6–5.4%) collagens. However, in these sponge fibrillar collagens, the lysine residues are more often observed in the Yaa position of collagenous Gly-Xaa-Yaa triplets. Hence, their lysine levels in the Yaa position (2.9–4.8%) are comparable with those observed in Hydra and in human B and C clade fibrillar α chains (2.4–4.7%). The last demosponge fibrillar collagen chain, Amq7 α , presents in its triple helix a low level of alanine (3.4%) and a very high level of lysine residues (9.7%). These observations indicate that demosponges have triple helical characteristics favoring the formation of thin fibrils, as is the case with Hydra fibrillar collagen chains. Altogether, these data suggest that the B/C clade ancestral fibrillar collagen chain might have had the potential to primarily initiate the formation of fibrils and to secondarily regulate the fibril diameter. Moreover, the presence in the first multicellular animals of members of the three fibrillar collagen clades indicates their common functional importance in the formation of collagen-based extracellular matrices.

REFERENCES

1. Myllyharju, J., and Kivirikko, K. I. (2004) *Trends Genet.* **20**, 33–43
2. Exposito, J. Y., Cluzel, C., Garrone, R., and Lethias, C. (2002) *Anat. Rec.* **268**, 302–316
3. Rasmussen, M., Jacobsson, M., and Björck, L. (2003) *J. Biol. Chem.* **278**, 32313–32316
4. Wang, C., and St. Leger, R. J. (2006) *Proc. Natl. Acad. Sci. U. S. A.* **103**, 6647–6652
5. Veit, G., Kobbe, B., Keene, D. R., Paulsson, M., Koch, M., and Wagener, R. (2006) *J. Biol. Chem.* **281**, 3494–3504
6. Söderhäll, C., Marenholz, I., Kerschler, T., Rüschemdorf, F., Esparza-Gordillo, J., Worm, M., Gruber, C., Mayr, G., Albrecht, M., Rohde, K., Schulz, H., Wahn, U., Hubner, N., and Lee, Y. A. (2007) *PLoS Biol.* **5**, e242
7. Exposito, J. Y., and Garrone, R. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87**, 6669–6673
8. Boute, N., Exposito, J. Y., Boury-Esnault, N., Vacelet, J., Noro, N., Miyazaki, K., Yoshizato, K., and Garrone, R. (1996) *Biol. Cell* **88**, 37–44
9. Zhang, X., Boot-Handford, R. P., Huxley-Jones, J., Forse, L. N., Mould, A. P., Robertson, D. L., Li, L., Athiyal, M., and Sarras, M. P., Jr. (2007) *J. Biol. Chem.* **282**, 6792–6802
10. Aouacheria, A., Geourjon, C., Aghajari, N., Navratil, V., Deleage, G., Lethias, C., and Exposito, J. Y. (2006) *Mol. Biol. Evol.* **23**, 2288–2302
11. Hjorten, R., Hansen, U., Underwood, R. A., Telfer, H. E., Fernandes, R. J., Krakow, D., Sebald, E., Wachsmann-Hogiu, S., Bruckner, P., Jacquet, R., Landis, W. J., Byers, P. H., and Pace, J. M. (2007) *Bone* **41**, 535–542
12. Plumb, D. A., Dhir, V., Mironov, A., Ferrara, L., Poulsom, R., Kadler, K. E., Thornton, D. J., Briggs, M. D., and Boot-Handford, R. P. (2007) *J. Biol. Chem.* **282**, 12791–12795
13. Koch, M., Laub, F., Zhou, P., Hahn, R. A., Tanaka, S., Burgeson, R. E., Gerecke, D. R., Ramirez, F., and Gordon, M. K. (2003) *J. Biol. Chem.* **278**, 43236–43244
14. Boot-Handford, R. P., Tuckwell, D. S., Plumb, D. A., Rock, C. F., and Poulsom, R. (2003) *J. Biol. Chem.* **278**, 31067–31077
15. Pace, J. M., Corrado, M., Missero, C., and Byers, P. H. (2003) *Matrix Biol.* **22**, 3–14
16. Sicot, F. X., Exposito, J. Y., Masselot, M., Garrone, R., Deutsch, J., and Gaill, F. (1997) *Eur. J. Biochem.* **246**, 50–58
17. Aouacheria, A., Cluzel, C., Lethias, C., Gouy, M., Garrone, R., and Exposito, J. Y. (2004) *J. Biol. Chem.* **279**, 47711–47719
18. Wada, H., Okuyama, M., Satoh, N., and Zhang, S. (2006) *Evol. Dev.* **8**, 370–377
19. Rychel, A. L., Smith, S. E., Shimamoto, H. T., and Swalla, B. J. (2006) *Mol. Biol. Evol.* **23**, 541–549
20. Nichols, S. A., Dirks, W., Pearce, J. S., and King, N. (2006) *Proc. Natl. Acad. Sci. U. S. A.* **103**, 12451–12456
21. Exposito, J. Y., van der Rest, M., and Garrone, R. (1993) *J. Mol. Evol.* **37**, 254–259
22. Schröder, H. C., Boreiko, A., Korzhev, M., Tahir, M. N., Tremel, W., Eckert, C., Ushijima, H., Müller, I. M., and Müller, W. E. (2006) *J. Biol. Chem.* **281**, 12001–12009
23. Burge, C., and Karlin, S. (1997) *J. Mol. Biol.* **268**, 78–94
24. Schultz, J., Milpetz, F., Bork, P., and Ponting, C. P. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 5857–5864
25. King, N., Westbrook, M. J., Young, S. L., Kuo, A., Abedin, M., Chapman, J., Fairclough, S., Hellsten, U., Isogai, Y., Letunic, I., Marr, M., Pincus, D., Putnam, N., Rokas, A., Wright, K. J., Zuzow, R., Dirks, W., Good, M., Goodstein, D., Lemons, D., Li, W., Lyons, J. B., Morris, A., Nichols, S., Richter, D. J., Salamov, A., JGI Sequencing, Bork, P., Lim, W. A., Manning, G., Miller, W. T., McGinnis, W., Shapiro, H., Tjian, R., Grigoriev, I. V., and Rokhsar, D. (2008) *Nature* **451**, 783–788
26. Leys, S. P., and Degnan, B. M. (2001) *Biol. Bull.* **201**, 323–338
27. Leys, S. P., and Degnan, B. M. (2002) *Invertebr. Biol.* **121**, 171–189
28. Larroux, C., Fahey, B., Liubicich, D., Hinman, V. F., Gauthier, M., Gongora, M., Green, K., Wörheide, G., Leys, S. P., and Degnan, B. M. (2006) *Evol. Dev.* **8**, 150–173
29. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673–4680
30. Galtier, N., Gouy, M., and Gautier, C. (1996) *Comput. Appl. Biosci.* **12**, 543–548
31. Thompson, J. D., Thierry, J. C., and Poch, O. (2003) *Bioinformatics* **19**, 1155–1161
32. Felsenstein, J. (1996) *Methods Enzymol.* **266**, 418–427
33. Guindon, S., Lethiec, F., Duroux, P., and Gascuel, O. (2005) *Nucleic Acids Res.* **33**, W557–W559
34. Ronquist, F., and Huelsenbeck, J. P. (2003) *Bioinformatics* **19**, 1572–1574
35. Huelsenbeck, J. P., and Ronquist, F. (2001) *Bioinformatics* **17**, 754–755
36. Page, R. D. (1996) *Comput. Appl. Biosci.* **12**, 357–358
37. Exposito, J. Y., D'Alessio, M., Solorsh, M., and Ramirez, F. (1992) *J. Biol. Chem.* **267**, 15559–15562
38. Olsen, B. R. (1982) in *New Trends in Basement Membrane Research* (Kühn, K., Schöene, H., and Timpl, R., eds) pp. 225–236, Raven Press, New York
39. Exposito, J. Y., Cluzel, C., Lethias, C., and Garrone, R. (2000) *Matrix Biol.* **19**, 275–279
40. Putnam, N. H., Srivastava, M., Hellsten, U., Dirks, B., Chapman, J., Salamov, A., Terry, A., Shapiro, H., Lindquist, E., Kapitonov, V. V., Jurka, J., Genikhovich, G., Grigoriev, I. V., Lucas, S. M., Steele, R. E., Finnerty, J. R., Technau, U., Martindale, M. Q., and Rokhsar, D. S. (2007) *Science* **317**, 86–94
41. Kortschak, R. D., Samuel, G., Saint, R., and Miller, D. J. (2003) *Curr. Biol.* **13**, 2190–2195
42. Larroux, C., Fahey, B., Degnan, S. M., Adamski, M., Rokhsar, D. S., and Degnan, B. M. (2007) *Curr. Biol.* **17**, 706–710
43. Larroux, C., Luke, G. N., Koopman, P., Rokhsar, D. S., Shimeld, S. M., and Degnan, B. M. (2008) *Mol. Biol. Evol.* **25**, 980–996
44. Simionato, E., Ledent, V., Richards, G., Thomas-Chollier, M., Kerner, P., Coornaert, D., Degnan, B. M., and Vervoort, M. (2007) *BMC Evol. Biol.* **7**, 33

45. Peterson, K. J., and Butterfield, N. J. (2005) *Proc. Natl. Acad. Sci. U. S. A.* **102**, 9547–9552
46. Borchiellini, C., Manuel, M., Alivon, E., Boury-Esnault, N., Vacelet, J., and Le Parco, Y. (2001) *J. Evol. Biol.* **14**, 171–179
47. Borchiellini, C., Chombard, C., Manuel, M., Alivon, E., Vacelet, J., and Boury-Esnault, N. (2004) *Mol. Phylogenet. Evol.* **32**, 823–837
48. Roulet, M., Ruggiero, F., Karsenty, G., and Le Guellec, D. (2007) *Cell Tissue Res.* **327**, 323–332
49. Wenstrup, R. J., Florer, J. B., Brunskill, E. W., Bell, S. M., Chervoneva, I., and Birk, D. E. (2004) *J. Biol. Chem.* **279**, 53331–53337
50. Bos, K. J., Holmes, D. F., Kadler, K. E., McLeod, D., Morris, N. P., and Bishop, P. N. (2001) *J. Mol. Biol.* **306**, 1011–1022
51. Birk, D. E. (2001) *Micron* **32**, 223–237
52. Cluzel, C., Lethias, C., Garrone, R., and Exposito, J. Y. (2004) *J. Biol. Chem.* **279**, 9811–9817
53. Deutzmann, R., Fowler, S., Zhang, X., Boone, K., Dexter, S., Boot-Handford, R. P., Rachel, R., and Sarras, M. P., Jr. (2000) *Development* **127**, 4669–4680
54. Mizuno, K., Adachi, E., Imamura, Y., Katsumata, O., and Hayashi, T. (2001) *Micron* **32**, 317–323