



Draper, G. W., Shoemark, D. K., & Adams, J. C. (2019). Modelling the early evolution of extracellular matrix from modern Ctenophores and Sponges. *Essays in Biochemistry*, 63(3), 389-405.
<https://doi.org/10.1042/EBC20180048>

Peer reviewed version

Link to published version (if available):
[10.1042/EBC20180048](https://doi.org/10.1042/EBC20180048)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Portland Press at <https://portlandpress.com/essaysbiochem/article-lookup/doi/10.1042/EBC20180048>. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

Modelling the Early Evolution of Extracellular Matrix from Modern Ctenophores and Sponges

Graham Draper, Deborah K. Shoemark and Josephine C. Adams

School of Biochemistry, University of Bristol, Bristol BS8 1TD, U.K.

*Corresponding author. Email: jo.adams@bristol.ac.uk

Keywords: Ctenophora, Integrins, Metazoa, Multicellularity, Porifera, Protein Domains

Running Title: Early evolution of Metazoan ECM

Abbreviations: ADAMTS, A disintegrin and metalloproteinase with thrombospondin motifs; BLAST, basic local alignment search tool; BMP1; Bone morphogenetic protein 1; ECM, extracellular matrix; HIF, Hypoxia inducible factor; MMP, Matrix metalloproteinase; SPARC, Small protein acidic and rich in cysteine; TGM, Transglutaminase; TSA, Transcriptome shotgun assembly.

ABSTRACT

Animals (metazoans) include some of the most complex living organisms on Earth, with regard to their multicellularity, numbers of differentiated cell types and lifecycles. The metazoan extracellular matrix (ECM) is well-known to have major roles in the development of tissues during embryogenesis and in maintaining homeostasis throughout life, yet insight into the ECM proteins which may have contributed to the transition from unicellular eukaryotes to multicellular animals remains sparse. Recent phylogenetic studies place either ctenophores or poriferans as the closest modern relatives of the earliest-emerging metazoans. Here, we review the literature and representative genomic and transcriptomic databases for evidence of ECM and ECM-affiliated components known to be conserved in bilaterians, that are also present in ctenophores and/or poriferans. Whereas an extensive set of related proteins are identifiable in poriferans, there is a strikingly lack of conservation in ctenophores. From this perspective, much remains to be learnt about the composition of ctenophore mesoglea. The principal ECM-related proteins conserved between ctenophores, poriferans and bilaterians include collagen IV, laminin-like proteins, thrombospondin superfamily members, integrins, membrane-associated proteoglycans and tissue transglutaminase. These are candidates for a putative ancestral ECM that may have contributed to the emergence of the metazoans.

INTRODUCTION

Some of the most fascinating and mysterious steps in the evolution of life on earth involve the debut of multicellular organisms from single-celled ancestors. Modern multicellular lifeforms are present in both the bacterial and eukaryotic domains of life and there is evidence that multicellularity has emerged independently multiple times [1-5]. For eukaryotes, particularly the Metazoa (animals), the transition from unicellularity to multicellularity is complex to consider because ancestral forms are either not represented in the fossil record, or are extremely rare and difficult to identify. Genome and transcriptome sequencing projects offer new routes to consider these evolutionary transitions more systematically. For example, an analysis of

genomic data from many modern prokaryotes to identify commonalities in protein-coding sequences allowed inference of the possible repertoire of proteins in a last common ancestral cell [6].

In considering central attributes of multicellular organisms, the evolution of stable mechanisms for organised cell-to-cell attachments is a key requirement of multicellularity. The most complex modern multicellular organisms, including between 3 to about 122 different cell types, are found amongst the Metazoa [5]. The single-celled eukaryotes most closely related to animals are the choanoflagellates [7] and filastereans such as *Capsaspora owczarzaki*, [8,9]. From comparative studies of the transcriptomes and predicted proteomes of these protists, it has been deduced that the origin of metazoans probably involved the functional adaptation of pre-existing gene products, for example new or adapted roles of integrin and cadherin receptors, both of which have been identified in certain unicellular eukaryotes [10, 11], as well as genetic rearrangement events that led to the origin of new types of gene products with novel functional capacities for cell interactions and inter-cellular communications [12].

A central mediator of metazoan multicellularity is the extracellular matrix (ECM), a structured extracellular network of collagens, glycoproteins, proteoglycans and associated carbohydrates such as glycosaminoglycans. The secreted proteins that build the ECM appear to fall within the novel category of metazoan gene products, because many ECM proteins are highly-conserved throughout animals and yet are not represented in choanoflagellates or filastereans [13, 14]. Williams et al. [13] established that three species of these protists express distinct sets of predicted secreted proteins (identified by presence of a N-terminal secretory signal peptide and no transmembrane domain), none of which have a domain composition equivalent to a metazoan ECM protein, although individual domains common in ECM proteins are present. Specific examples are the separate domains of fibrillar collagens [15] or thrombospondins [16]. These data suggest that gene rearrangement and domain shuffling had an important role in the emergence of the large, multidomain, secreted proteins that characterise the metazoan

ECM. In modern metazoans, the secretion and extracellular assembly of structural proteins of the ECM depends on many ECM-affiliated proteins, both intracellular and extracellular: for example, to effect post-translational modifications, proteolytic processing, or interactions with non-structural and matricellular proteins within the ECM [17, 18]. Thus, consideration of the phylogeny of these affiliated proteins is also relevant to constructing models for the evolution of metazoan ECM.

The distinct domain architectures of secreted proteins of choanoflagellates and metazoans suggest that additional insights into metazoan ECM evolution could be ascertained from careful comparative analysis of ECM and affiliated proteins encoded in modern species from the earliest-diverging metazoan phyla. By analogy with the analysis of a prokaryotic ancestral cell [6], proteins in common between extant species in the earliest-diverging phyla would be candidates for membership of “ancestral ECM”. Of the early-diverging metazoan phyla (Ctenophora (comb jellies), Porifera (sponges), Placozoa and Cnidaria), cnidarians are, to date, by far the most-studied with regard to their ECM and cell-adhesion mechanisms. This relates to the lengthy history of Hydra as an experimental model [18], the phylogenetic position of cnidarians as the sister group to bilaterians, and the presence of a morphologically well-defined ECM, the mesoglea, that, in Hydra, can be isolated away from the cell layers of the body wall as an acellular structure and is thus suitable for biochemical study [19, 20].

Knowledge of the ECM of the other early-diverging phyla is much more sparse. Placozoa comprise an enigmatic phylum that to date includes only a few species and will not be considered further here [21]. A mesoglea between the epithelial cell layers is apparent in ctenophores, but information on its molecular composition is very limited (discussed further below). Sponges have bio-mineralised extracellular structures (spicules) embedded in a fibrillar meshwork, but most classes lack ECM as recognised morphologically in bilaterians. A limitation for the study of ctenophores and sponges until recent years has been a lack of laboratory model species or cell culture [22, 23].

The limited knowledge is significant because both Porifera and Ctenophora are considered to be of earlier evolutionary origin than Cnidaria, as evidenced by the fossil record and molecular phylogeny reconstructions [24, 25]. However, there remains considerable controversy over whether the sponges or the ctenophores are of earliest evolutionary origin, i.e., which phylum represents the sister group to all other animals. Historically, sponges were placed at the base of the animal evolutionary tree due to their simple morphological organisation, limited number of cell types, and the absence of recognisable nerve or muscle structures: both of the latter are present in ctenophores and cnidarians [26] (Fig. 1A). With the expansion of molecular phylogenetics, several studies have surprisingly placed ctenophores as the sister group to all other animals [e.g., 27] (Fig. 1B), whereas others continue to support the traditional “poriferans-sister” model [e.g., 28, 29]. Genome sequencing of two species of ctenophores indicated major differences in the categories of encoded proteins in comparison to all other metazoans, with many proteins of bilaterians noted to be absent [30, 31, 32]. Thus the “ctenophore first” hypothesis remains under active discussion and investigation [33, 34].

The advent of genomic and transcriptomic sequencing projects for an increasing number of poriferan and ctenophore species has revolutionised the possibility to gain insight into ECM content in sponges and ctenophores, through analysis of the predicted proteomes of individual species. This article will review the published literature and discuss our findings from a recent detailed survey of public genomic and transcriptomic databases for ECM proteins in species representing three classes of sponges and three species of ctenophores.

Known Components of ECM in Ctenophores and Poriferans

Ctenophores. The unique anatomy and ultrastructure of ctenophores has been studied by light and electron microscopy, with major interest in developmental processes, nerve and muscle tissues, and the specialised rows of locomotary ciliated combs (Fig. 1C, 1D) [35, 36]. Prey capture is carried out by specific colloblast cells on extensible tentacles (most species)

or by direct engulfment (*Beroe* species, that lack tentacles) [27, 37]. The mesoglea is typically described as transparent and jelly-like. From transmission electron microscopy studies of the tentacles of *Euplokamis*, the mesoglea was observed to contain networks of striated fibrils, interpreted as collagen fibrils, as well as muscle fibres, mesenchymal cells and a network of nerve cells. Curious box-like, acellular, extracellular structures were also observed [38]. Later immunofluorescent staining studies of *Pleurobrachia* species or *Beroe abyssicola* also identified many cell types within the bodywall mesoglea, including networks of nerve cells, muscle and other cell types [39-41]. Transmission electron microscopy has also provided views of a basement membrane-like layer that underlies ectodermal cells in *Pleurobrachia bachei* and *B. ovata*, but is not visible in *Mnemiopsis leidyi* [14]. There is very little direct knowledge of the composition of ctenophore mesoglea, but a phylogenetic study of the basement membrane proteoglycan, perlecan, concluded that perlecan is absent from *M. leidyi* [42]. Fidler et al. [14] detected collagen IV by immunohistochemistry as diffuse arrays in proximity to ectodermal cells in *M. leidyi*, and with more appearance of linear elements in *Beroe* and *Pleurobrachia*. Genomic and transcriptomic analyses identified many collagen IV paralogues, whereas small collagenous proteins (designated spongins from their initial identification in Porifera) were absent and a unique type of secreted protein, containing only a non-collagenous (NC) domain, was identified and designated NC1 protein [14].

Porifera. Adult sponges are sessile, vase-shaped animals with pores that filter water into the body cavity for food uptake by specialised choanocytes (Fig. 1E) [43, 44]. The body wall consists of an epithelial bilayer supported by mineralised spicules and a meshwork of extracellular fibrils termed the mesohyl (Fig. 1F). Unlike ctenophores, overt cell-cell junctions are apparent between epithelial cells [45]. There are four extant classes of sponges (Fig. 1G) and different classes have different processes of biomineralisation. In calcareous sponges, calcium carbonate-based spicules are assembled extracellularly through carbonic anhydrase activity and possibly in association with acidic extracellular proteins [46-48]. In siliceous sponges, silicon dioxide spicules are templated through intracellular and extracellular processes

involving (in many species) the polymerising enzyme, silicatein, and frequently with templating onto collagen fibres [49-51]. Silicateins are related to the cathepsin family of intracellular processing and degrading enzymes and are thought to have arisen in the sponge lineage through ancestral gene duplication and point mutation of cathepsin L [52]. Chitin has also been identified as a spicule-associated, possible template in demosponges and a glass sponge [53, 54].

By electron microscopy, class Homoscleromorpha is distinguished by the presence of a basement membrane structure [55, 56]. Indeed, a collagen IV cloned from a homoscleromorph sponge was shown to have a basement membrane-like localisation [55]. In addition, sponges (along with various other invertebrates) encode short-chain spongins that contribute to 10nm microfibrils within the mesohyl. Spongins comprise around 79-100 Gly-Xaa-Yaa triplets and 3 noncollagenous regions, with the C-terminal non-collagenous regions having homology and a proposed shared evolutionary origin with the NC1 domain of collagen IV [57, 58]. Collagen fibril structures have been identified by ultrastructural criteria in several sponge species [59-61] and molecular cloning led the recognition of a diversity of molecular forms of collagens of the fibrillar or interrupted-triple-helix types in addition to spongins ([62-64] and reviewed by [65]).

Species-specific, carbohydrate-based cell-to-cell adhesion mechanisms have also been characterized functionally and biochemically in sponges. Sulphated polysaccharides have major roles in species-specific cell-to-cell adhesion, in conjunction with proteins termed aggregation factor complex, spongican or glyconectins [66-71]. These proteins can adopt unusual, ring-like conformations and appear to be specific to the demosponge lineage [72, 73]. Other mechanisms may involve C-type lectins and a calcium-dependent lectin, clathrilectin [74, 75], as well as self-association of carbohydrates [76]. Chemically, the sulphated polysaccharides appear very varied, with varying amounts of sulphated galactose, fucose, arabinose or hexuronic acid identified across species [77]. Examination of the structures of the acid-labile carbohydrates of glyconectins from several species identified these to

include heterogeneous sulphated oligosaccharides with variable amounts of fucose, arabinose, or py(4,6)Galactose residues, and thus distinct from the repeated disaccharide units of bilaterian glycosaminoglycans [78].

With regards to other mechanisms of cell-ECM associations, integrin subunits have been cloned from several sponge species [79, 80, 81]. Integrin(s) have been implicated functionally in autograft fusion in *Geodia cydonium* [82] and the response of *Microciona prolifera* cells to depletion of extracellular sulphate [83]. The identification of vinculin in *Oscarella pearsei* and the localisation of this protein to cell-cell and cell-ECM contact sites further supports that poriferan integrins are likely to function in adhesion and cell signaling, as in bilaterians [84]. Dystroglycan-like proteins with possible laminin-binding capacity have been recognised in several sponges in addition to the dystroglycans of cnidarians and bilaterians [85].

Insights from Genomics and Transcriptomics: A Structured Survey of ECM Proteins Encoded in Poriferans and Ctenophores

The sequencing of the genome of the demosponge *Amphimedon queenslandica* expanded the view on candidate ECM proteins of sponges. Analysis of the predicted proteins indicated that, even in the absence of overt basement membrane-like structures, laminin-like proteins are encoded [86, 87]. Since 2010, transcriptomes for sponges of other classes and genomes and transcriptomes for several species of ctenophores have been published [31, 32, 88, 89].

To obtain a wider view of ECM and ECM-associated proteins in ctenophores and poriferans, we surveyed sponge and ctenophore genome- and/or transcriptome-predicted proteins for selected ECM and ECM-affiliated proteins. The ECM proteins chosen for study are highly conserved in invertebrate and vertebrate bilaterians [90, 91] and have known functional roles in the fibrils and meshworks of the ECM. A range of collagens were included as search tools to assist identification of possible disparate forms. Major glycoprotein and proteoglycan receptors that tether ECM proteins at cell surfaces and extracellular proteases important for ECM dynamics in

bilaterians were also included (Fig. 2A), along with intracellular proteins that are important for procollagen assembly, processing and collagen fibril formation (Fig. 2B) [92], or for the post-translational assembly of the core tetrasaccharide linker for glycosaminoglycan substitution on proteoglycan core proteins [93] (Fig. 2C). Spongin and silicatein are not present in vertebrates but were included because of their known importance in poriferan ECM. The suite of 41 proteins analysed is listed in Supplementary Table 1. The sponges species included represent the classes Demospongiae (the most abundant in terms of species), Homoscleromorpha and Calcarea and the three species of ctenophores have sequenced genomes and/or a large transcriptome dataset.

From initial BLASTP or TBLASTN searches, protein sequences of highest homology (determined by e -value $<1e-30$ and extensive % coverage of the query sequence) were compiled and validated by best-reciprocal BLAST search. Sequences passing this step were subjected to additional quality control analysis for secretory signal peptides, transmembrane domains, overall domain architecture and, where relevant, enzyme active site. The results obtained are summarised in Tables 1 and 2, respectively. The underlying data are presented in Supplementary Tables 2-8.

Collectively, the data demonstrate dramatic differences in the profile of conserved proteins in ctenophores versus poriferans (Fig. 3, Table 1). Many more proteins in common with bilaterian ECM are encoded in poriferans than in ctenophores. Nevertheless, the ctenophore list does include a repertoire for a basic cell-ECM adhesion system: cell-surface receptors, ECM proteins, a cross-linking enzyme and a potential ECM-proteolytic enzyme (Fig. 3). The proteins identified in ctenophores were for the most part present in all three species, with the exception of syndecan, identified only in *P. bachei*, and potential matrix metalloproteases, identified in *M. leidyi* and *H. californensis*. In agreement with Fidler et al. [14], many collagen IV-like paralogues were identified. Post-translational modifications of proline and lysine residues contribute to the stability of collagen triple helices in vertebrates, however only prolyl-4-hydroxylase and not pro-collagen lysine dioxygenase (lysine

hydroxylase) was identified in these ctenophores (Table 1, Fig. 3). The encoding of multiple integrin alpha and beta subunits (Table 1) indicates potential for diverse specificities of integrin-mediated cell adhesion, perhaps in line with the relatively large number of cell types now documented in ctenophores [94]. Silicatein-like proteins were identified in all three species; however, given the recognised general divergence of protein sequences in ctenophores [31, 32], in-depth studies will be needed to determine the relationship of these to the cathepsin family (Table 1).

We examined the laminin-like proteins in more detail in view of the early evolution of collagen IV [14] and the central role of laminin in basement membrane assembly in bilaterians [95]. With the caveat that some of the identified sequences are incomplete, the laminin subunits identified present a complex picture. Although all are large proteins with many of the characteristic domains of laminins, many variations in domain organisation are apparent, including atypical domains such as thrombospondin or fibronectin III domains. Overall, the laminin proteins of sponges are more similar to those of bilaterians, yet distinctions between beta and gamma subunits are blurred at both the sequence and domain levels. Notably, the laminin N-domain is lacking from (apparently full-length) ctenophore proteins and two alpha-like subunits but no beta- or gamma-like subunits were identified in *P. bachei* (Fig. 4). The numbers of alpha-like and beta/gamma-like subunits varied between species and the alpha-like subunits included at most three laminin-G domains. Biochemical experiments will be needed to determine if these proteins are capable of forming stable heterotrimers and undergoing extracellular polymerisation or integrin-binding. For the ctenophore proteins, it is of interest whether heteromers including two different alpha subunits can be assembled.

We also examined the tissue transglutaminase 2- (TGM2) like proteins. In vertebrates, TGM2 has major roles in cross-linking fibrillin or fibronectin molecules into extracellular fibrils via redox-dependent processes [96, 97]. The TGM-like proteins identified in the ctenophores and *A. queenslandica* sponge each include all the major domains of TGM2 but have only 30%-35% sequence identity to human TGM2. However, at the active site of TGM2, the

identity is 65%-80% and the cysteine residue is completely conserved (Fig 5A). Molecular models of the active site region for four of the sequences, constructed against secondary structure alignments of four TGM2 structures from Protein database (4KTY_B, 1G0D_A, 1LM9_A and 2Q3Z_A), are presented overlaid with the structure of human TGM2 (4PYG) that was not used for modelling (Fig. 5B, 5C). The models demonstrate that residues within 60nm of the active site cysteine align very well with the known structure (Fig. 5B), as do the highly-conserved residues at the active site (Fig. 5C). We predict that the ctenophore and sponge proteins should be active transglutaminases.

ECM-related proteins *not* found in the ctenophores raise other intriguing questions about the biochemical nature of ctenophore mesoglea. It was previously noted that core proteins of secreted proteoglycans of bilaterians are not conserved in early-diverging metazoans [19, 90]. However, carbohydrate has been reported as <1% of dry weight of ctenophores [98].

With the exception of glucuronyl-transferase, homologues of the bilaterian enzymes for addition of the core O-linked saccharides (Fig. 2C), are not identifiable (Table 1, Fig 3). Whether this pathway mechanism evolved later or has been lost through lineage-specific gene losses in ctenophores is unclear. With regard to ECM structure, no fibrillar-like collagens were identified and the collagen cross-linking enzyme lysyl oxidase was also absent, raising questions over the nature of observed striated fibrils in ctenophore mesoglea [38]. However, bone morphogenetic protein 1 (BMP1), which cleaves the C-propeptide of fibrillar procollagen [99] (Fig. 2B), was present. In bilaterians, BMP1 has many other substrates including a laminin gamma chain [99] and it may be expected that the ctenophore protein can target other substrates. In agreement with [14], spongins were not identified.

In contrast, poriferans were confirmed to encode a wider repertoire of ECM proteins including fibrillar-like collagens of various domain architectures and fibrillin, as well as SPARC and one or more thrombospondin superfamily members (see [16] for details of the thrombospondin superfamily). Collagen IV proteins were apparent in the calcareous sponges as well as (as expected)

in the homoscleromorph sponge. The encoding of lysyl oxidase is in agreement with the detection of striated collagen fibrils in sponges [59-61]. Notwithstanding the unusual carbohydrate structures reported in sponges (see section above), the suite of carbohydrate-addition enzymes encoded indicate potential for addition of the O-linked core tetrasaccharide of glycosaminoglycans (Table 2, Fig. 3). Other aspects of ECM in common with bilaterian ECM include the encoding of A Disintegrin and Metalloproteinase with Thrombospondin motifs (ADAMTS) proteases and a wider repertoire of cell-ECM attachment receptor types, including membrane-bound proteoglycan core proteins, dystroglycan and integrin subunits. Spongins were identified in *A. queenslandica* and *O. carmela*, but not in the calcereous sponges. Silecateins were identified as expected in the demosponge *A. queenslandica* and also in the other species examined (Table 3).

Perspective

Current laboratory experiments and analyses of genome-predicted proteins indicate that ctenophore ECM has a very different protein composition to other metazoans. This cannot be interpreted as a result of the early phylogenetic emergence of this phylum because poriferans, traditionally considered the earliest-diverging metazoans, are found to have an array of ECM and ECM-affiliated proteins that is clearly closer to the conserved repertoire of cnidarians and bilaterians. The difference in ECM might be considered an indication that ctenophores evolved prior to poriferans, in which case the limited set of proteins conserved between ctenophores, poriferans and bilaterians can be taken to represent a prototypic “toolkit” for a minimal metazoan ECM [100]. The combination of collagen IV, laminin-like proteins and thrombospondin superfamily members is of great interest, as the concept of coordinated function of these three proteins within ECM has received little consideration.

However, other factors also need to be taken into consideration. There are estimated to be around 5,000 species of extant poriferans, yet only about 150 known species of ctenophores. This may reflect that the deep-sea lifestyle of many ctenophores makes it difficult to identify the true number of species, or

alternatively could indicate very different evolutionary histories of poriferans and ctenophores. Sponges and ctenophores evolved when oxygen levels on Earth were far lower than at present [101]. Indeed, members of both phyla lack hypoxia-inducible factor α (HIF) indicating that oxygen availability does not drive gene expression through the HIF pathway as in cnidarians and bilaterians [102]. In a “poriferan-first” evolutionary scenario, the limited repertoire of known ECM proteins in ctenophores would represent secondary gene losses, leading to a proposal of a relatively complex ECM in the metazoan ancestor. Many ctenophores live in the deep sea and the environment of low oxygen and sparse food sources [101, 103], and high hydrostatic pressure may have driven selection for a unique form of ECM. Clearly, the anatomy of ctenophores does include a mesoglea and, to date, to our knowledge, an unbiased study of ctenophore mesoglea by proteomic methods has not been carried out. Only through this type of approach will a clear view of ctenophore mesoglea composition be gained. A limitation of focusing on the ECM proteins conserved with bilaterians is that possible ctenophore- or poriferan-specific ECM proteins remain undisclosed. As discussed by others, it is very likely that a considerable “hidden biology” of ctenophores and poriferans remains to be discovered [104]. Nevertheless, the positive identification of certain ECM proteins conserved between these early-emerging phyla and other metazoans increases the precision of models for an ancestral metazoan ECM.

SUMMARY POINTS

1. Ctenophores and poriferans have distinct sets of ECM-related proteins in comparison to the most highly-conserved ECM and ECM-affiliated proteins of bilaterians.
2. In particular, ctenophores lack many of the structural ECM proteins and enzymes for addition of the core O-linked tetrasaccharide that is characteristic of bilaterian glycosaminoglycan substitutions.
3. Proteomic studies are needed to gain more comprehensive views of the composition of ECM in ctenophores and poriferans.
4. The very limited set of conserved ECM and ECM-affiliated proteins in ctenophores may represent a minimal toolkit that reflects the earliest emerging form of metazoan ECM. Given the current controversies over the relative phylogenetic placements of ctenophores and poriferans it cannot be excluded that this apparent simplicity may result from

secondary gene losses that are specific to the evolution of the ctenophore lineage.

FIGURES AND TABLES

Figure 1. Ctenophores and Poriferans. A, B, Schematic evolutionary trees of the poriferan-first (A), or ctenophore-first (B), models of metazoan evolution. Choanoflagellates are included as the closest outgroup. C, Morphology of a ctenophore, *Pleurobrachia pileus*. Internal organs are visible through the transparent mesoglea (asterisk). Key: Mouth (mo); aboral neuro-sensory complex (nsc); comb rows (cr); tentacles (tcl); lateral branches of tentacles (tentillae), (tt); tentacle root (tr), and pharynx (ph). Scale bar: 1cm. (reproduced from [105] under CC-BY). D, Schematic diagram of a ctenophore. Mesoglea is present within the tentacles and body wall (reproduced from [106] under CC-BY). E, Morphology of a demosponge, *Echinoclathria gibbosa* (reproduced from [107] under CC-BY). F, Diagram of the body wall and cell types of a sponge, (reproduced from [108]; permission in progress). G, Tree diagram of the four classes of extant sponges.

Figure 2. The cell-ECM adhesion proteins included in our genomic and transcriptomic survey. A, Schematic diagram of molecular processes involved in cell-ECM adhesion; B, The major enzymes involved in procollagen processing and collagen fibril assembly; C, The enzymes involved in O-linkage of the core tetrasaccharide for glycosaminoglycan addition.

Figure 3. ECM and ECM-affiliated proteins of poriferans and ctenophores that are conserved with bilaterians, as identified by our survey. The Venn diagram includes all categories of proteins that were identified in at least one of the species studied. See Tables 1 and 2 and supplementary Tables for details of the protein identifications.

Figure 4. Domain architectures of the laminin-like proteins identified in the species studied. Domains were identified in InterProScan 5. Representative examples of human laminin α , β , and γ chains are shown at the top. The number of amino acids is given in small font at the C-terminus of each protein. Dashed lines indicate incomplete sequence. Accession numbers are given on the righthand side, see Tables 1 and 2 for details. The IV-B domain (in coral)

is specific to the β subunit. Key: SP, signal peptide; N, laminin N-terminal domain; EGF, epidermal growth factor-like domain, with numbers referring to the number of repeated domains; IV, laminin domain IV (in blue); G, laminin G domain; FNIII, fibronectin type III domain; GFR, growth factor-cysteine-rich domain; TSR, thrombospondin type 1 domain. The G domain is structurally related to the Concanavalin A domain (Con A), which was identified in some sequences. Not to scale.

Figure 5. Molecular modelling of the active sites of the predicted tissue transglutaminases identified in ctenophores and a sponge. A, CLUSTAL sequence alignment of the active site of human TGM2 (Hs) with the homologous regions of the identified ctenophore (Ml, *M. leidyi*; Pb, *P. bachei*, and poriferan (Aq, *A. queenslandica*) TGMs. See Tables 1 and 2 for full accession numbers. B, Structure of the active site of 4PYG.pdb (human TGM2) [96]. The helix is highlighted in yellow as orientation for the overlay models in Fig. 5C. C, Models were prepared by HHPRED and MODELLER [110] and are shown as Aq (pink), ML03126a (green), ML25826a (salmon) and Pb3462531 (silver) overlaid with the crystal structure of human TGM2 from 4PYG.pdb, (black, with the catalytic cysteine (C277) labelled). The overlays show high conservation of the sidechains of residues within a 6 Angstrom radius around the catalytic cysteine.

TABLES

Table 1. The ECM-related proteins conserved with bilaterians identified from the ctenophore species studied.

Table 2. The ECM-related proteins conserved with bilaterians identified from the sponge species studied.

SUPPLEMENTARY DATA

Supplementary Table 1. The ECM and ECM-affiliated proteins of bilaterians included in our survey.

Supplementary Tables 2-8. Summary of BLAST top hits and validations for the 7 species surveyed.

REFERENCES

1. Arias Del Angel, J. A, Escalante, A., Martínez-Castilla, L., Benítez, M. (2017) An Evo-Devo perspective on multicellular development of Myxobacteria. *J Exp Zool Part B Mol Dev Evol*, **328**:165-178
2. A. Knoll (2011). The multiple origins of complex multicellularity. *Annu Rev Earth Planet Sci*, **39**: 217-239
3. Brunet T, King N. (2017) The Origin of Animal Multicellularity and Cell Differentiation. *Dev Cell*. **43**(2):124-140
4. Sebé-Pedrós A, Degnan BM, Ruiz-Trillo I. (2017) The origin of Metazoa: a unicellular perspective. *Nat Rev Genet*. **18**(8):498-512
5. Rokas A. (2008) The origins of multicellularity and the early history of the genetic toolkit for animal development. *Annu Rev Genet*. **42**:235-51
6. Weiss MC, Sousa FL, Mrnjavac N, Neukirchen S, Roettger M, Nelson-Sathi S, Martin WF. (2016) The physiology and habitat of the last universal common ancestor. *Nat Microbiol*. **1**(9):16116
7. Carr M, Leadbeater BS, Hassan R, Nelson M, Baldauf SL. (2008) Molecular phylogeny of choanoflagellates, the sister group to Metazoa. *Proc Natl Acad Sci U S A*. **105**(43):16641-6
8. Suga H, Chen Z, de Mendoza A, Sebé-Pedrós A, Brown MW, Kramer E, Carr M, Kerner P, Vervoort M, Sánchez-Pons N, Torruella G, Derelle R, Manning G, Lang BF, Russ C, Haas BJ, Roger AJ, Nusbaum C, Ruiz-Trillo I.(2013) The *Capsaspora* genome reveals a complex unicellular prehistory of animals. *Nat Commun*. **4**:2325
9. Paps J, Medina-Chacón LA, Marshall W, Suga H, Ruiz-Trillo I. (2013) Molecular phylogeny of unikonts: new insights into the position of apusomonads and ancyromonads and the internal relationships of opisthokonts. *Protist* **164**(1):2-12.
10. Sebé-Pedrós A, Roger AJ, Lang FB, King N, Ruiz-Trillo I. (2010). Ancient origin of the integrin-mediated adhesion and signaling machinery. *Proc Natl Acad Sci U S A*. **107**(22):10142-7
11. Nichols SA, Roberts BW, Richter DJ, Fairclough SR, King N. (2012) Origin of metazoan cadherin diversity and the antiquity of the classical cadherin/ β -catenin complex. *Proc Natl Acad Sci U S A*. **109**(32):13046-51
12. Paps J, Holland PWH. (2018) Reconstruction of the ancestral metazoan genome reveals an increase in genomic novelty. *Nat Commun*. **9**(1):1730
13. Williams, F., Tew, H.A., Paul, C.E. and Adams J.C. (2014) The predicted secretomes of *Monosiga brevicollis* and *Capsaspora owczarzaki*, close unicellular relatives of metazoans, reveal new insights into the evolution of the metazoan extracellular matrix. *Matrix Biology* **37**: 60-68.
14. Fidler AL, Darris CE, Chetyrkin SV, Pedchenko VK, Boudko SP, Brown KL, Gray Jerome W, Hudson JK, Rokas A, Hudson BG. (2017) Collagen IV and basement membrane at the evolutionary dawn of metazoan tissues. *Elife* **6**. pii: e24176. doi: 10.7554/eLife.24176

15. Exposito JY, Larroux C, Cluzel C, Valcourt U, Lethias C, Degnan BM. (2008) 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. *J Biol Chem* **283**(42):28226-35.
16. Shoemark, D.K, Ziegler, B., Watanabe, H., Strompen, J., Tucker, R.P., Özbek, S., Adams, J.C. (2019). Emergence of a Thrombospondin Superfamily at the Origin of Metazoans. *Molecular Biology and Evolution* **36**: 1220-1238
17. Marini JC, Forlino A, Bächinger HP, Bishop NJ, Byers PH, Paepe A, Fassier F, Fratzi-Zelman N, Kozloff KM, Krakow D, Montpetit K, Semler O. (2017). Osteogenesis imperfecta. *Nat Rev Dis Primers* **3**:17052. doi: 10.1038/nrdp.2017.52
18. Rousselle P, Beck K. (2013) Laminin 332 processing impacts cellular behavior. *Cell Adh Migr*. **7**(1):122-34
18. Galliot B. (2012) Hydra, a fruitful model system for 270 years. *Int J Dev Biol*. **56**(6-8):411-23
19. Tucker RP, Adams JC. (2014) Adhesion networks of cnidarians: a postgenomic view. *Int Rev Cell Mol Biol*. **308**:323-77
20. Lommel, M+, Strompen, J.+ , Hellewell, A.L.+ , Balasubramanian G.P., Christofidou E.D., Thomson, A.R., Boyle, A.L., Woolfson, D.N., Puglisi, K, Hartl, M., Holstein T.W., Adams, J.C. and Ozbek, S. (2018). Hydra Mesoglea Proteome Identifies Thrombospondin as a Conserved Component Active in Head Organizer Restriction. *Scientific Reports* **8**:11753. +joint first authors.
21. Eitel M, Osigus HJ, DeSalle R, Schierwater B. (2013) Global diversity of the Placozoa. *PLoS One* **8**(4):e57131
22. Vandepas LE, Warren KJ, Amemiya CT, Browne WE. (2017) Establishing and maintaining primary cell cultures derived from the ctenophore *Mnemiopsis leidyi*. *J Exp Biol*. **220**:1197-1201
23. de Caralt S1, Uriz MJ, Wijffels RH. (2007) Cell culture from sponges: pluripotency and immortality. *Trends Biotechnol*. **25**(10):467-71.
24. Cunningham JA, Liu AG, Bengtson S, Donoghue PC. (2017) The origin of animals: Can molecular clocks and the fossil record be reconciled? *Bioessays* **39**(1):1-12
25. Sperling EA, Stockey RG. (2018) The Temporal and Environmental Context of Early Animal Evolution: Considering All the Ingredients of an "Explosion". *Integr Comp Biol*. **58**(4):605-622
26. Paps J. (2018) What Makes an Animal? The Molecular Quest for the Origin of the Animal Kingdom. *Integr Comp Biol*. **58**(4):654-665
27. Whelan NV, Kocot KM, Moroz TP, Mukherjee K, Williams P, Paulay G, Moroz LL, Halanych KM. (2017) Ctenophore relationships and their placement as the sister group to all other animals. *Nat Ecol Evol*. **1**(11):1737-1746.

28. Feuda R, Dohrmann M, Pett W, Philippe H, Rota-Stabelli O, Lartillot N, Wörheide G, Pisani D.(2017). Improved Modeling of Compositional Heterogeneity Supports Sponges as Sister to All Other Animals. *Curr Biol.* **27**(24):3864-3870.e4
29. Simion P, Philippe H, Baurain D, Jager M, Richter DJ, Di Franco A, Roure B, Satoh N, Quéinnec É, Ereskovsky A, Lapébie P, Corre E, Delsuc F, King N, Wörheide G, Manuel M. (2017) A Large and Consistent Phylogenomic Dataset Supports Sponges as the Sister Group to All Other Animals. *Curr Biol.* **27**(7):958-967
30. Schenkelaars Q, Fierro-Constain L, Renard E, Borchiellini C.(2018) Retracing the path of planar cell polarity. *BMC Evol Biol.* **16**:69
31. Ryan, J.F., Pang, K., Schnitzler, C.E., Nguyen, A.D., Moreland, R.T., Simmons, D.K., Koch, B.J., Francis, W.R., Havlak, P., NISC Comparative Sequencing Program, Smith, S.A., Putnam, N.H., Haddock, S.H.D., Dunn, C.W., Wolfsberg, T.G., Mullikin, J.C., Martindale, M.Q., Baxevanis, A.D. (2013). The genome of the ctenophore *Mnemiopsis leidyi* and its implications for cell type evolution. *Science* **342**(6164):1242592
32. Moroz LL, Kocot KM, Citarella MR, Dosung S, Norekian TP, Povolotskaya IS, Grigorenko AP, Dailey C, Berezikov E, Buckley KM, Ptitsyn A, Reshetov D, Mukherjee K, Moroz TP, Bobkova Y, Yu F, Kapitonov VV, Jurka J, Bobkov YV, Swore JJ, Girardo DO, Fodor A, Gusev F, Sanford R, Bruders R, Kittler E, Mills CE, Rast JP, Derelle R, Solovyev VV, Kondrashov FA, Swalla BJ, Sweedler JV, Rogaev EI, Halanych KM, Kohn AB. (2014) The ctenophore genome and the evolutionary origins of neural systems. *Nature* **510**(7503):109-14
33. King N, Rokas A. (2017) Embracing Uncertainty in Reconstructing Early Animal Evolution. *Curr Biol.* **27**(19):R1081-R1088
34. Belahbib H, Renard E, Santini S, Jourda C, Claverie JM, Borchiellini C, Le Bivic A. (2018) New genomic data and analyses challenge the traditional vision of animal epithelium evolution. *BMC Genomics* **19**(1):393
35. Pang, K. and Martindale, M.Q. (2008) Comb jellies (Ctenophora): a model for Basal metazoan evolution and development. *CSH Protoc.* pdb.emo106 <http://dx.doi.org/10.1101/pdb.emo106>
36. Jager M, Manuel M. (2016) Ctenophores: an evolutionary-developmental perspective. *Curr Opin Genet Dev.* **39**:85-92
37. Babonis LS, DeBiasse MB, Francis WR, Christianson LM, Moss AG, Haddock SHD, Martindale MQ, Ryan JF. (2018) Integrating Embryonic Development and Evolutionary History to Characterize Tentacle-Specific Cell Types in a Ctenophore. *Mol Biol Evol.* **35**(12):2940-2956
38. G.O. Mackie 1, C.E. Mills 2, and C.L. Singla (1988). Structure and function of the prehensile tentilla of *Euplokamis* (Ctenophora, Cydippida). *Zoomorphology* **107**:319-337
39. Jager M, Chiori R, Alie´ A, Dayraud C, Que´innec E, Manuel M. (2011). New insights on ctenophore neural anatomy: immunofluorescence study in *Pleurobrachia pileus* (Muller, 1776). *J. Exp. Zool. (Mol. Dev. Evol.)* **316**:171–187

40. Norekian TP, Moroz LL. (2018) Neuromuscular organization of the Ctenophore *Pleurobrachia bachei*. *J Comp Neurol*. doi: 10.1002/cne.24546
41. Norekian TP, Moroz LL. (2019) Neural system and receptor diversity in the ctenophore *Beroë abyssicola*. *J Comp Neurol*. doi: 10.1002/cne.24633
42. Warren CR, Kassir E, Spurlin J, Martinez J, Putnam NH, Farach-Carson MC.(2015) Evolution of the perlecan/HSPG2 gene and its activation in regenerating *Nematostella vectensis*. *PLoS One*. **10**(4):e0124578. doi: 10.1371/journal.pone.0124578
43. Godefroy N, Le Goff E, Martinand-Mari C, Belkhir K, Vacelet J, Baghdiguian S. (2019) Sponge digestive system diversity and evolution: filter feeding to carnivory. *Cell Tissue Res*. 2019 doi: 10.1007/s00441-019-03032-8
44. Leys SP, Hill A. (2012) The physiology and molecular biology of sponge tissues. *Adv Mar Biol*. **62**:1-56
45. Adams EDM, Goss GG, Leys SP (2010) Freshwater Sponges Have Functional, Sealing Epithelia with High Transepithelial Resistance and Negative Transepithelial Potential. *PLoS ONE* **5**(11): e15040
46. Ledger PW, Jones WC. (1977) Spicule formation in the calcareous sponge *Sycon ciliatum*. *Cell Tissue Res*. **181**(4):553-67
47. Vaceletia SP, Germer J, Mann K, Wörheide G, Jackson DJ. (2015) The Skeleton Forming Proteome of an Early Branching Metazoan: A Molecular Survey of the Biomineralization Components Employed by the Coralline Sponge *PLoS One*. **10**(11):e0140100
48. Voigt O, Adamska M, Adamski M, Kittelmann A, Wencker L, Wörheide G. (2017) Spicule formation in calcareous sponges: Coordinated expression of biomineralization genes and spicule-type specific genes. *Sci Rep*. **7**:45658
49. Müller WE, Boreiko A, Wang X, Belikov SI, Wiens M, Grebenjuk VA, Schlossmacher U, Schröder HC. (2007) Silicateins, the major biosilica forming enzymes present in demosponges: protein analysis and phylogenetic relationship. *Gene* **395**(1-2):62-71
50. Annenkov VV, Danilovtseva EN. (2016) Spiculogenesis in the siliceous sponge *Lubomirskia baicalensis* studied with fluorescent staining. *J Struct Biol*. **194**(1):29-37
51. Aguilar-Camacho JM, Doonan L, McCormack GP. (2019) Evolution of the main skeleton-forming genes in sponges (phylum Porifera) with special focus on the marine Haplosclerida (class Demospongiae). *Mol Phylogenet Evol*. **131**:245-253
52. Riesgo A, Maldonado M, López-Legentil S, Giribet G. (2015) A Proposal for the Evolution of Cathepsin and Silicatein in Sponges. *J Mol Evol*. **80**(5-6):278-91
53. Ehrlich H, Maldonado M, Spindler KD, Eckert C, Hanke T, Born R, et al. (2007) First Evidence of Chitin as a Component of the Skeletal Fibers of Marine Sponges. Part I. Verongidae (Demospongia: Porifera). *J Exp Zool B Mol Dev Evol*. **308**(4):347–356
54. Ehrlich H, Krautter M, Hanke T, Simon P, Knieb C, Heinemann S, Worch H. (2007) First evidence of the presence of chitin in skeletons of marine sponges. Part II. Glass sponges (Hexactinellida: Porifera). *J Exp Zool B Mol Dev Evol*. **15**:308(4):473-83

55. Boute N, Exposito JY, Boury-Esnault N, Vacelet J, Noro N, Miyazaki K, Yoshizato K, Garrone R. (1996) Type IV collagen in sponges, the missing link in basement membrane ubiquity. *Biol Cell* **88**(1-2):37-44
56. Ereskovsky AV, Borchiellini C, Gazave E, Ivanisevic J, Lapébie P, Perez T, Renard E, Vacelet J. (2009) The Homoscleromorph sponge *Oscarella lobularis*, a promising sponge model in evolutionary and developmental biology: model sponge *Oscarella lobularis*. *Bioessays* **31**(1):89-97
57. Exposito JY, Le Guellec D, Lu Q, Garrone R. (1991) Short chain collagens in sponges are encoded by a family of closely related genes. *J Biol Chem* **266**:21923–21928
58. Aouacheria A, Geourjon C, Aghajari N, Navratil V, Deléage G, Lethias C, Exposito JY. (2006) Insights into early extracellular matrix evolution: spongin short chain collagen-related proteins are homologous to basement membrane type IV collagens and form a novel family widely distributed in invertebrates. *Mol Biol Evol.* **23**(12):2288-302
59. Ledger PW. (1974) Types of collagen fibres in the calcareous sponges *Sycon* and *Leucandra*. *Tissue Cell* **6**(3):385-9
60. Heinemann S, Ehrlich H, Douglas T, Heinemann C, Worch H, Schatton W, Hanke T. (2007). Ultrastructural studies on the collagen of the marine sponge *Chondrosia reniformis* Nardo. *Biomacromolecules* **8**(11):3452-7
61. Pallela R, Bojja S, Janapala VR. (2011) Biochemical and biophysical characterization of collagens of marine sponge, *Ircinia fusca* (Porifera: Demospongiae: Irciniidae). *Int J Biol Macromol.* **49**(1):85-92.
62. Exposito JY, Garrone R. (1990) Characterization of a fibrillar collagen gene in sponges reveals the early evolutionary appearance of two collagen gene families. *Proc Natl Acad Sci USA* **87**:6669–6673
63. Exposito JY, Larroux C, Cluzel C, Valcourt U, Lethias C, Degnan BM. (2008) 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. *J Biol Chem.* **283**(42):28226-35
64. Pozzolini M, Bruzzone F, Berilli V, Mussino F, Cerrano C, Benatti U, Giovine M. (2012) Molecular characterization of a nonfibrillar collagen from the marine sponge *Chondrosia reniformis* Nardo 1847 and positive effects of soluble silicates on its expression. *Mar Biotechnol (NY)* **14**(3):281-93
65. Ehrlich H, Wysokowski M, Żółtowska-Aksamitowska S, Petrenko I, Jesionowski T. (2018) Collagens of Poriferan Origin. *Mar Drugs* **16**(3). pii: E79
66. Muller WEG, Zahn RK. (1973) Purification and characterization of a species-specific aggregation factor in sponges. *Exp Cell Res.* **80**:95–104
67. Misevic GN, Burger MM. (1993) Carbohydrate-carbohydrate interactions of a novel acidic glycan can mediate sponge cell adhesion. *J Biol Chem.* **268**(7):4922-9
68. Jarchow, J., Fritz, J., Anselmetti, D., Calabro, A., Hascall, V. C., Gerosa, D., Burger, M. M., and Fernandez-Busquets, X. (2000) Supramolecular structure

of a new family of circular proteoglycans mediating cell adhesion in sponges. *J. Struct. Biol.* **132**:95–105

69. Misevic GN, Guerardel Y, Sumanovski LT, Slomianny MC, Demarty M, Ripoll C, Karamanos Y, Maes E, Popescu O, Strecker G. (2004) Molecular recognition between glyconectins as an adhesion self-assembly pathway to multicellularity. *J Biol Chem.* **279**(15):15579-90

70. Vilanova E, Coutinho C, Maia G, Mourão PA. (2010) Sulfated polysaccharides from marine sponges: conspicuous distribution among different cell types and involvement on formation of in vitro cell aggregates. *Cell Tissue Res.* **340**(3):523-31

71. Vilanova E, Santos GR, Aquino RS, Valle-Delgado JJ, Anselmetti D, Fernández-Busquets X, Mourão PA. (2016) Carbohydrate-Carbohydrate Interactions Mediated by Sulfate Esters and Calcium Provide the Cell Adhesion Required for the Emergence of Early Metazoans. *J Biol Chem.* **291**(18):9425-37

72. Fernandez-Busquets X, Burger MM. (2003) Circular proteoglycans from sponges: first members of the spongican family. *Cell Mol Life Sci.* **60**:88–112

73. Grice LF, Gauthier MEA, Roper KE, Fernández-Busquets X, Degnan SM, Degnan BM. (2017) Origin and Evolution of the Sponge Aggregation Factor Gene Family. *Mol Biol Evol.* **34**(5):1083-1099

74. Gundacker, D., Leys, S.P., Schröder, H.C., Müller, I.M., Müller, W.E. (2001) Isolation and cloning of a C-type lectin from the hexactinellid sponge *Aphrocallistes vastus*: a putative aggregation factor. *Glycobiology* **11**:21–29

75. Gardères J, Domart-Coulon I, Marie A, Hamer B, Batel R, Müller WE, Bourguet-Kondracki ML. (2016) Purification and partial characterization of a lectin protein complex, the clathrilectin, from the calcareous sponge *Clathrina clathrus*. *Comp Biochem Physiol B Biochem Mol Biol.* **200**:17-27

76. Haseley SR, Vermeer HJ, Kamerling JP, Vliegenthart JF. (2001) Carbohydrate self-recognition mediates marine sponge cellular adhesion. *Proc Natl Acad Sci U S A.* **98**(16):9419-24

77. Zierer MS, Mourão PA. (2000) A wide diversity of sulfated polysaccharides are synthesized by different species of marine sponges. *Carbohydr Res.* **328**(2):209-16

78. Guerardel Y, Czeszak X, Sumanovski LT, Karamanos Y, Popescu O, Strecker G, Misevic GN. (2004) Molecular fingerprinting of carbohydrate structure phenotypes of three porifera proteoglycan-like glyconectins. *J Biol Chem.* **279**(15):15591-603

79. Brower DL, Brower SM, Hayward DC, Ball EE. (1997) Molecular evolution of integrins: genes encoding integrin beta subunits from a coral and a sponge. *Proc Natl Acad Sci U S A.* **94**(17):9182-7

80. Müller WE. (1997) Origin of metazoan adhesion molecules and adhesion receptors as deduced from cDNA analyses in the marine sponge *Geodia cydonium*: review. *Cell Tissue Res.* **289**(3):383-95

81. Wimmer W, Perovic S, Kruse M, Schröder HC, Krasko A, Batel R, Müller WE. (1999a) Origin of the integrin-mediated signal transduction. Functional studies with cell cultures from the sponge *Suberites domuncula*. *Eur J Biochem.* **260**(1):156-65
82. Wimmer W, Blumbach B, Diehl-Seifert B, Koziol C, Batel R, Steffen R, Müller IM, Müller WE. (1999b) Increased expression of integrin and receptor tyrosine kinase genes during autograft fusion in the sponge *Geodia cydonium*. *Cell Adhes Commun.* **7**(2):111-24
83. Kuhns WJ, Rusciano D, Kaltenbach J, Ho M, Burger M, Fernandez-Busquets X. (2001) Up-regulation of integrins alpha(3) beta(1) in sulfate-starved marine sponge cells: functional correlates. *Biol Bull.* **201**(2):238-9
84. Miller PW, Pokutta S, Mitchell JM, Chodaparambil JV, Clarke DN, Nelson WJ, Weis WI, Nichols SA. (2018) Analysis of a vinculin homolog in a sponge (phylum Porifera) reveals that vertebrate-like cell adhesions emerged early in animal evolution. *J Biol Chem.* **293**(30):11674-11686
85. Adams, J.C. and Brancaccio, A. (2015) The evolution of the dystroglycan complex, a major mediator of muscle integrity. *Biology Open* **4**:1163-1179
86. Srivastava M, Simakov O, Chapman J, Fahey B, Gauthier ME, Mitros T, Richards GS, Conaco C, Dacre M, Hellsten U, Larroux C, Putnam NH, Stanke M, Adamska M, Darling A, Degnan SM, Oakley TH, Plachetzki DC, Zhai Y, Adamski M, Calcino A, Cummins SF, Goodstein DM, Harris C, Jackson DJ, Leys SP, Shu S, Woodcroft BJ, Vervoort M, Kosik KS, Manning G, Degnan BM, Rokhsar DS. (2010) The *Amphimedon queenslandica* genome and the evolution of animal complexity. *Nature* **466**(7307):720-6.
87. Fernandez-Valverde SL, Calcino AD, Degnan BM. (2015) Deep developmental transcriptome sequencing uncovers numerous new genes and enhances gene annotation in the sponge *Amphimedon queenslandica*. *BMC Genomics* **16**:387
88. Riesgo A, Farrar N, Windsor PJ, Giribet G, Leys SP. (2014) The analysis of eight transcriptomes from all poriferan classes reveals surprising genetic complexity in sponges. *Mol Biol Evol.* **31**(5):1102-20
89. Hemmrich, G., Bosch, T.C.G. (2010). Compagen, a comparative genomics platform for early branching metazoan animals reveals early origins of genes regulating stem cell differentiation *BioEssays* **20**(10):1010-1018
90. Ozbek, S., Balasubramanian, P., Chiquet-Ehrismann, R., Tucker, R.P. and Adams, J.C. (2010). The evolution of extracellular matrix. *Mol. Biol. Cell* **21**: 4300-4306
91. Hynes RO. (2012) The evolution of metazoan extracellular matrix. *J Cell Biol.* **196**(6):671-679
92. Bella J, Hulmes DJ. (2017) Fibrillar Collagens. *Subcell Biochem.* **82**:457-490
93. Sasarman F, Maftei C, Campeau PM, Brunel-Guitton C, Mitchell GA, Allard P. (2016) Biosynthesis of glycosaminoglycans: associated disorders and biochemical tests. *J Inherit Metab Dis.* **39**(2):173-88

94. Seb -Pedr s A, Chomsky E, Pang K, Lara-Astiaso D, Gaiti F, Mukamel Z, Amit I, Hejnl A, Degnan BM, Tanay A. (2018) Early metazoan cell type diversity and the evolution of multicellular gene regulation. *Nat Ecol Evol.* **2**(7):1176-1181
95. Pozzi A, Yurchenco PD, Iozzo RV. (2017) The nature and biology of basement membranes. *Matrix Biol.* **57-58**:1-11
96. Kanchan K1, Fuxreiter M, F s s L. (2015) Physiological, pathological, and structural implications of non-enzymatic protein-protein interactions of the multifunctional human transglutaminase 2. *Cell Mol Life Sci.* **72**(16):3009-35.
97. Pinkas DM, Strop P, Brunger AT, Khosla C (2007) Transglutaminase 2 undergoes a large conformational change upon activation *PLoS Biol.* **5** p.e327
98. Percy, J.A., Fife, F.J. (1981) The biochemical composition and energy content of arctic marine macrozooplankton. *Arctic* **34**: 307-313
99. Moali C, Font B, Ruggiero F, Eichenberger D, Rousselle P, Fran ois V, Oldberg A, Bruckner-Tuderman L, Hulmes DJ. (2005) Substrate-specific modulation of a multisubstrate proteinase. C-terminal processing of fibrillar procollagens is the only BMP-1-dependent activity to be enhanced by PCPE-1. *J Biol Chem.* **280**(25):24188-94
100. Revilla-I-Domingo R, Simakov O. (2018) The Diversification of Early Emerging Metazoans: A Window into the Evolution of Animal Multicellularity. *Bioessays* **40**(5):e1800029.
101. Leys SP, Kahn AS. (2018) Oxygen and the Energetic Requirements of the First Multicellular Animals. *Integr Comp Biol.* **58**(4):666-676
102. Mills DB, Francis WR, Vargas S, Larsen M, Elemans CP, Canfield DE, W rheide G. (2018) The last common ancestor of animals lacked the HIF pathway and respired in low-oxygen environments. *Elife* **7**. pii: e31176
103. Winnikoff JR, Francis WR, Thuesen EV, Haddock SHD. (2019) Combining transcriptomes for secrets of deep-sea survival: Environmental diversity drives patterns of protein evolution. *Integr Comp Biol.* pii: icz063
104. Dunn CW, Leys SP, Haddock SH. (2015) The hidden biology of sponges and ctenophores. *Trends Ecol Evol.* **30**(5):282-91
105. Jager M, Dayraud C, Mialot A, Qu innec E, le Guyader H, Manuel M. (2013) Evidence for involvement of Wnt signalling in body polarities, cell proliferation, and the neuro-sensory system in an adult ctenophore. *PLoS One* **8**(12):e84363
106. Ehrlich H, Shaala LA, Youssef DTA,  oltowska-Aksamitowska S, Tsurkan M, Galli R, Meissner H, Wysokowski M, Petrenko I, Tabachnick KR, Ivanenko VN, Bechmann N, Joseph Y, Jesionowski T. (2018) Discovery of chitin in skeletons of non-verongiid Red Sea demosponges. *PLoS One* **13**(5):e0195803
107. Dayraud C, Ali  A, Jager M, Chang P, Le Guyader H, Manuel M, Qu innec E. (2012) Independent specialisation of myosin II paralogues in muscle vs. non-muscle functions during early animal evolution: a ctenophore perspective. *BMC Evol Biol.* **12**:107

108. Barnes, R.D. (1974) *Invertebrate Zoology*. ISBN 0-7216-1562-7. Publishers W. B. Saunders, PA, USA
109. Francis, W.R., Shaner, N.C., Christianson, L.M., Powers, M.L, Haddock,S.H. (2015) Occurrence of isopenicillin-N-Synthase homologs in bioluminescent ctenophores and implications for coelenterazine biosynthesis. *PLoS ONE* **10** (6):e0128742
110. Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V. (2018) A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J Mol Biol.* **430**:2237-2243

ECM/ECM-affiliated Protein	Accession numbers of homologues identified in ctenophore species		
	<i>M. leidyi</i>	<i>P. bachei</i>	<i>H. californensis</i>
ADAMTS	-	-	-
Agrin	-	-	-
beta-1,4-galactosyltransferase 1	-	-	-
beta-1,4-galactosyltransferase 2	-	-	-
Bone morphogenetic protein 1	ML31401a	PB3463711	GGLO01058316.1
Bifunctional heparan sulphate N-deacetylase/N-sulphotransferase	-	-	-
Collagen I	-	-	-
Collagen IV	ML034334a ML034337a ML18175a ML17501a ML034336a ML18198a ML16441a ML25069a ML18197a ML18176a ML17504a ML17502a ML034335a	PB3460513 PB3460485 PB3479585 PB3460474 PB3479630	GGLO01050633.1 GGLO01050605.1 GGLO01033148.1 GGLO01027773.1 GGLO01030859.1 GGLO01027525.1 GGLO01030860.1 GGLO01014639.1 GGLO01023628.1 GGLO01020933.1 GGLO01020934.1
Collagen VI	-	-	-
Collagen X	-	-	-
Collagen XIV	-	-	-
Collagen XVIII	-	-	-
Collagen XXIII	-	-	-
Collagen XXV	-	-	-
Dystroglycan	-	-	-
Fibrillin	-	-	-
Fibulin	-	-	-
Glucuronyl-transferase	ML271519a ML073039a	PB3467704	GGLO01045111.1
Glypican	ML14244a	PB3464484	GGLO01048445.1
Heparan sulphate 2-O-sulphotransferase	-	-	-
Hsp47 (SERPINH1)	-	-	-
Integrin alpha	ML09514a ML463513a ML30811a ML27362a	PB3464239	GGLO01029506.1 GGLO01029505.1 GGLO01050007.1 GGLO01050010.1
Integrin beta	ML068314a ML073028a ML04042a ML05098a	PB3461633 PB3460305 PB3465728 PB3460908	GGLO01026249.1 GGLO01070288.1
Laminin alpha	ML097515a	-	GGLO01023715.1
Laminin beta	-	-	GGLO01023769.1
Laminin gamma	ML05951a ML035910a	PB3460962 PB3479626	GGLO01031685.1
Lysyl hydroxylase	-	-	-
Lysyl Oxidase	-	-	-
Matrix metalloproteinase	ML030215a	-	GGLO01044249.1
Nidogen	-	-	-
Perlecan	-	-	-
Procollagen C-proteinase enhancer	-	-	-
Prolyl 4-hydroxylase	GFAT01094412.1, ML257632	PB3469803 (p)	GGLO01022174
Silicatein	ML017310a ML005020a ML10109a ML007332a	PB3466562 PB3465824	GGLO01072762.1 GGLO01008148.1 GGLO01041937.1 GGLO01007610.1+
SPARC	-	-	-
Spongin	-	-	-
Syndecan	-	PB12832424	-
Thrombospondin superfamily	ML34222a	PB11599022 (p)	GGLO01050054(p)

Tissue Transglutaminase 2	ML03126a	PB 3462531	GGLO01036061
Xylosyltransferase	-	-	-

Table 1. Homologues of ECM and affiliated proteins of bilaterians identified in three species of ctenophores. Accession numbers are from *M. leidy* genome browser [31], *P. bachei* transcriptome at Neurobase [32], and GenBank TSA for the *H. californensis* transcriptome [109]. Underlying BLAST results are reported in Supplementary Tables 2-4. Key. p = partial sequence. + BLASTP for silicatein in *H. californensis* yielded >100 hits with many redundant sequences. Four top hits are included here.

ECM /ECM-affiliated protein	Accession numbers of top homologues identified in poriferan species			
	<i>O. carmella</i>	<i>A. queenslandica</i>	<i>L. complicata</i>	<i>S. ciliatum</i>
ADAMTS	g302.t1, g3943.t1, g5893.t1	XP_011406035.1	lcpid17910, lcgid3864 lcpid26189, lcgid1420 lcpid16435, lcgid3864 lcpid14744, lcgid48263 lcpid8832, lcgid1429	scpid8465, scgid0298 scpid14367, scgid31153 scpid11730, scgid3308 scpid7699, scgid3308 scpid11197, scgid3308
Aggrin	-	-	-	-
beta-1,4-galactosyltransferase 1	g6154.t1 g11091.t1	XP_003383478.3	lcpid72445, lcgid57516 lcpid108583, lcgid4190	scpid62588, scgid16118 scpid83531, scgid23687
beta-1,4-galactosyltransferase 2	g11091	XP_003383478.3	lcpid72445, lcgid57516	scpid83531
Bone morphogenetic protein 1	g3280.t1	XP_003387878.1	lcpid53711	scpid73058
Bifunctional heparan sulphate N-deacetylase/N-sulphotransferase	g9342.t1	XP_019854532.1	lcpid25736, lcgid12663 lcpid35336, lcgid53985 lcpid65133, lcgid5933 lcpid44097, lcgid5933	scpid22021, scgid1378 scpid11705, scgid14663 scpid24542, scgid2276 scpid19876, scgid2276 scpid27630, scgid5678
Fibrillar Collagen-like *	g11085.t1 g426.t1 g10272.t1	XP_003388783.2 XP_011405131.1 XP_011405304.1 XP_019864450.1	lcpid17276, lcgid40511, lcpid13783, lcgid6124	scpid13476, scgid0358 scpid12346, scgid28066, scpid10799, scgid31760, scpid12114, scgid35048
Collagen IV	g1079.t1 g10977.t1	-	lcpid52947, lcgid75348 lcpid4786, lcgid64311 lcpid13694, lcgid38464 lcpid16924, lcgid10403	scpid10989, scgid8635 scpid11320, scgid9595 scpid8958, scgid15448
Dystroglycan	g5213.t1	-	-	scpid43419, scgid4112 scpid78051, scgid4112
Fibrillin	g2667.t1 g3633.t1	XP_019850772.1 XP_019859670	lcpid4195, lcgid79342	scpid6157, scgid20153
Fibulin	-	-	-	-
Glypican	g1352.t1	XP_003386976.1 XP_019852741.1	lcpid67857, lcgid45685 lcpid32714, lcgid59020	-
Glucuronyl-transferase	g2277.t1	XP_019851580.1 XP_019851583.1	lcpid40065, lcgid0547	scpid33216, scgid11109
Heparan sulphate 2-O-sulphotransferase	g2593.t1	-	lcpid67263, lcgid13423	scpid75685, scgid31021 scpid39942, scgid12865
Integrin alpha	g1989.t1 g9424.t1 g551.t1	XP_019848674.1 XP_019864024.1 XP_019864028.1 XP_019855406.1 XP_019855408.1 XP_019856433.1 XP_019850770.1	lcpid14662, lcgid5141 lcpid30951, lcgid20056 lcpid24895, lcgid25747 lcpid28736, scgid0236 lcpid24944, lcgid0796 lcpid38700, lcgid0796 lcpid25998, lcgid57014 lcpid29144, lcgid59731 lcpid53986, lcgid5141	scpid27434, scgid16145 scpid34588, scgid4727 scpid23074, scgid2395 scpid12646, scgid0799 scpid27708, scgid2395 scpid23633, scgid1153 scpid26968, scgid0799 scpid13516, scgid3686 scpid25262, scgid3686 scpid47429, scgid3686 scpid17601, scgid24595 scpid19223, scgid26959
Integrin beta	g9310.t1 g9748.t1	XP_011409775.1 XP_003388422.1 XP_019851540.1 XP_019864033.1 XP_011403560.2 XP_011404206.1	lcpid40636, lcgid16859 lcpid44623, lcgid59216 lcpid32340, lcgid73219 lcpid170227, lcgid60939	scpid4557, scgid16671 scpid20168, scgid21580 scpid4556, scgid3103 scpid38528, scgid18306 scpid22880, scgid21705 scpid49462, scgid3434 scpid22880, scgid21705 scpid41506, scgid3434 scpid47976, scgid3144 scpid29534, scgid3083 scpid19946, scgid33446 scpid27672, scgid3144
Laminin alpha	g7319.t1	XP_019852585.1	lcpid6587, lcgid43813	scpid4187, scgid33329
Laminin beta	-	XP_019854276.1 XP_019849782.1	lcpid31379, lcgid51207	scpid4527, scgid23872
Laminin gamma	g7968.t1 g4177.t1	XP_019855647.1	lcpid13654, lcgid40757	scpid4187, scgid33329
Lysyl hydroxylase	g1142.t1	XP_019854114.1	lcpid37848, lcgid5516	scpid32678, scgid32033

Lysyl Oxidase	g7137.t1	XP_003387025.1	lcpid22308, lcgid72170 lcpid23110, lcgid74733	scpid30388, scgid0815 scpid36618, cgid24646 scpid31988, scgid0815
Matrix metalloproteinase	-	-	-	-
Nidogen	-	-	-	-
Perlecan	-	-	-	-
Procollagen C-proteinase enhancer	-	-	-	-
Prolyl 4-hydroxylase	g1383.t1	XP_003383442.1	lcpid38875, lcgid64402	scpid58888 scgid3871 scpid40070, scgid33742
Silicatein	g4492.t1 g6175.t1 g6176.t1	XP_003383103.1	lcpid104541, lcgid12146 lcpid113056, lcgid19166	scpid95074, scgid11171 scpid63063, scgid35043
SPARC	g4839.t1	-	lcpid126889, lcgid36337, lcpid74779, lcgid64875	-
Spongin	g1095.t1 g988.t1 g2075.t1 g3758.t1	XP_011405650.1	-	-
Syndecan	-	-	lcpid156268, lcgid70156	-
Thrombospondin Superfamily	m162353a/g.1623 53	XP_011406237.1	lcpid9057a, lcpid8282, lcpid36553	scpid30246, scpid2291a, scpid12552
Tissue Transglutaminase 2	-	XP_019853849.1	-	-
Xylosyltransferase	g25200.t1	XP_011404143.1	lcpid41679, lcgid69457	scpid26149, scgid20327

Table 2. Homologues of ECM and affiliated proteins of bilaterians identified in four species of poriferans. Accession numbers are from genomes and transcriptome-predicted proteins of *O. carmela*, *S. ciliatum* and *L. complicata* [88] at Compagen [89], and GenBank entries of the *A. queenslandica* genome and transcriptome projects [86, 87]. Apart from collagen IV-like collagens and spongins, the various collagens identified have been grouped as fibrillar-like collagens. Underlying BLAST results and validations are reported in Supplementary Tables 5-8.

Figure 1

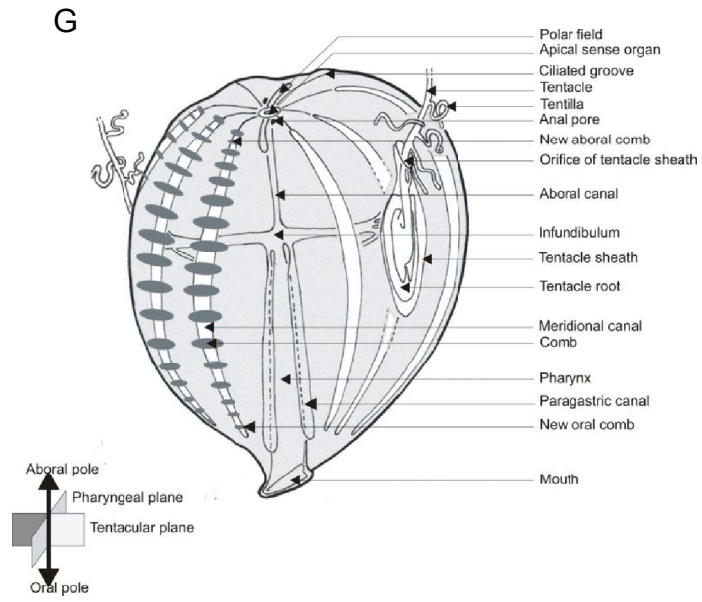
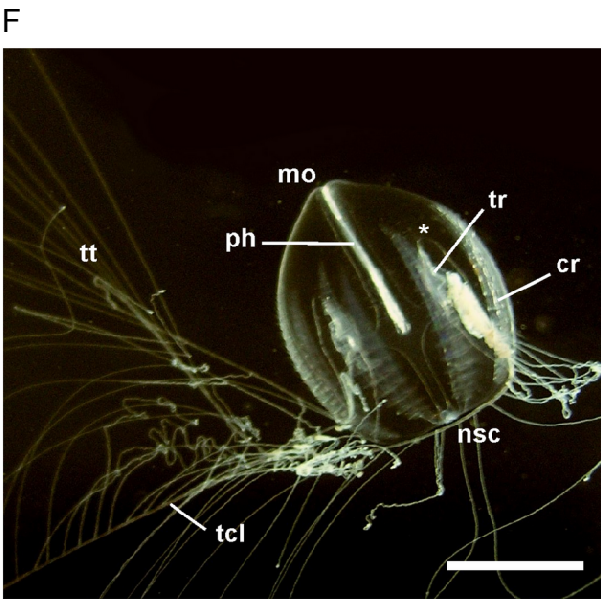
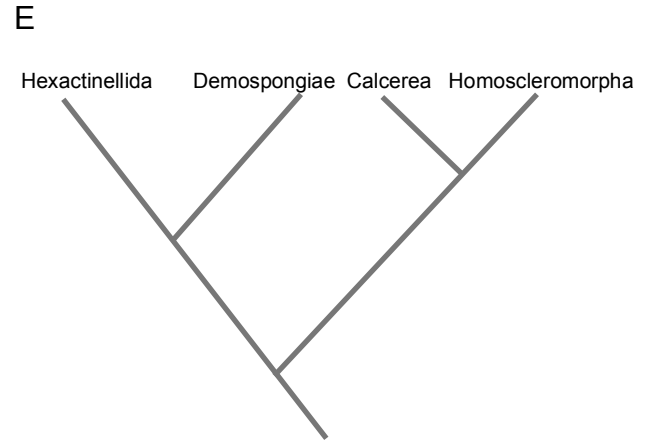
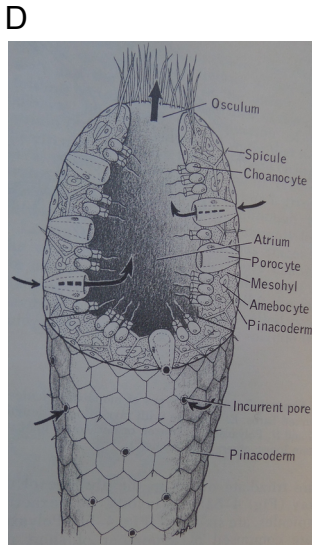
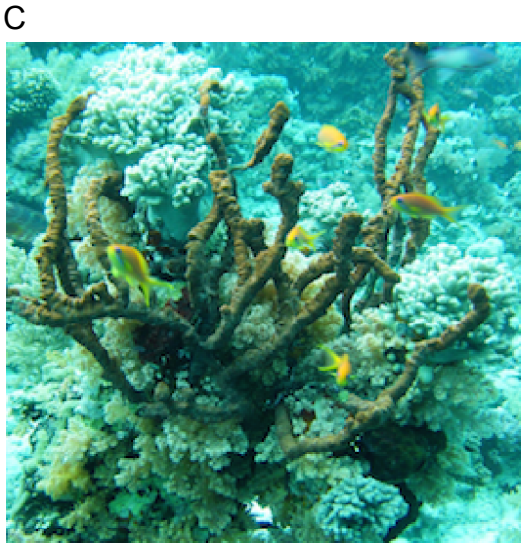
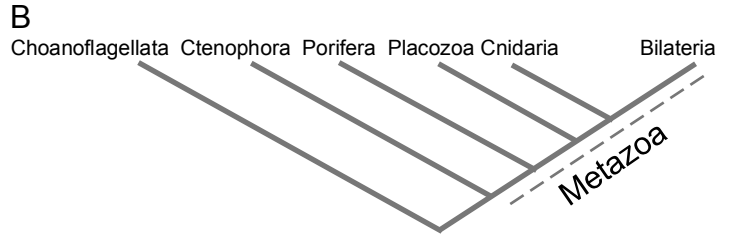
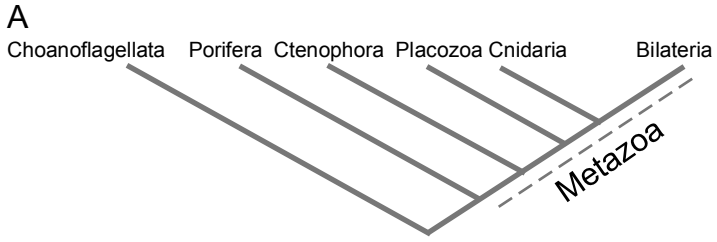
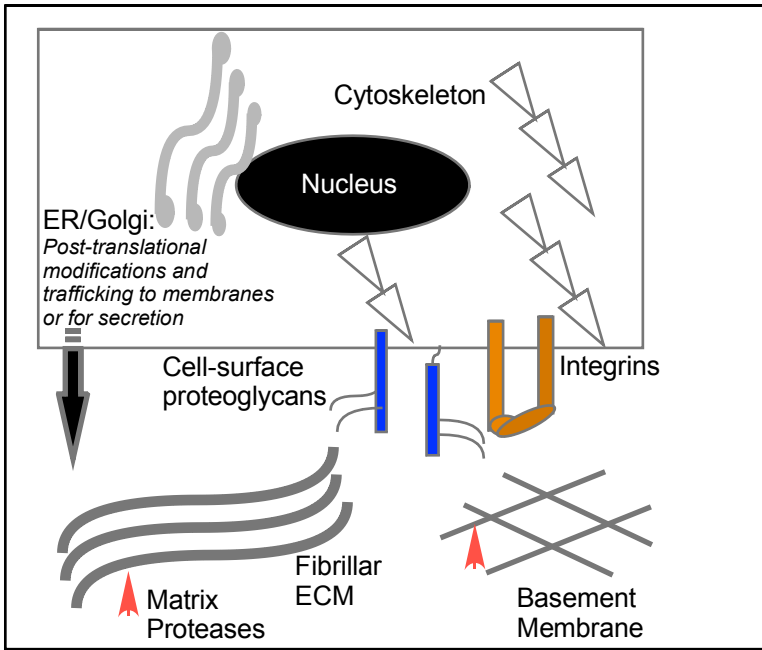
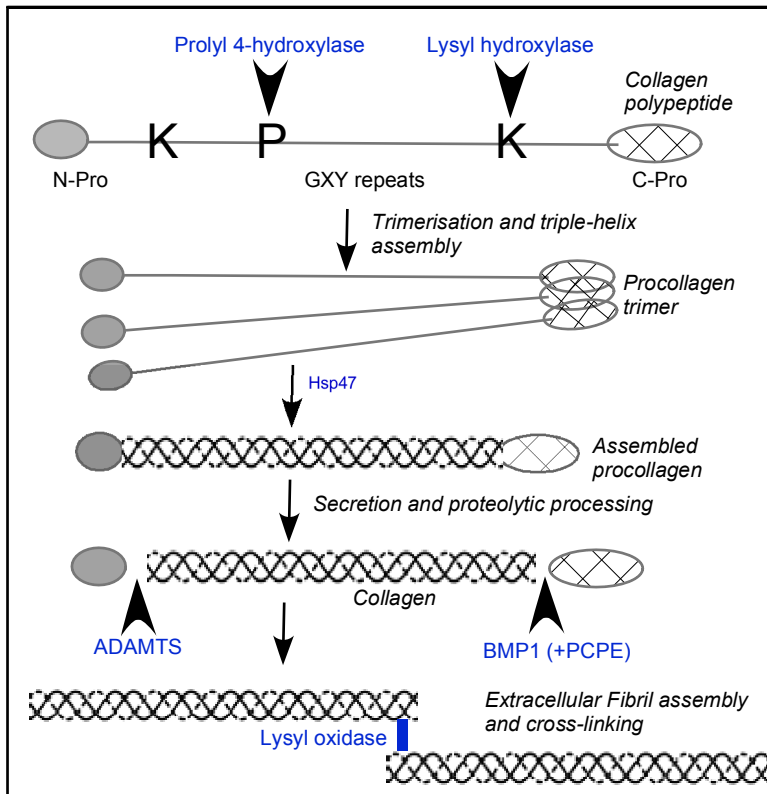


Figure 2

A



B



C

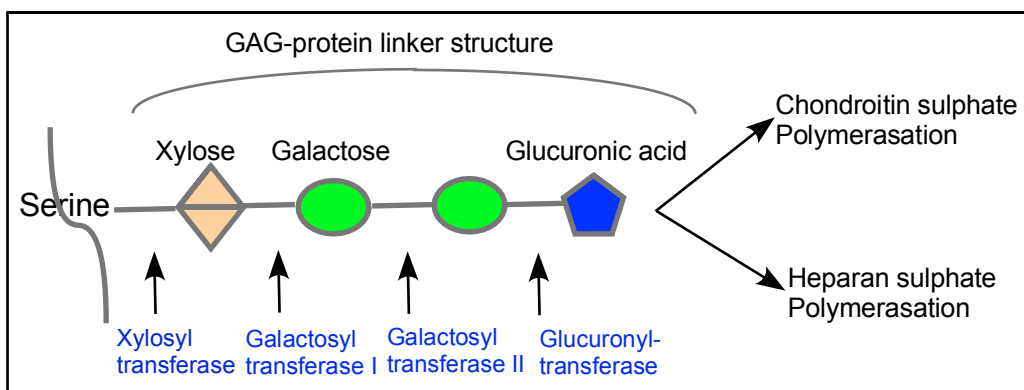


Figure 3

Porifera

Ctenophora

MMP

BMP1
Collagen IV
Glypican
Glucuronyl-transferase
Integrin α , β
Laminin chains
Prolyl-4-hydroxylase
Silicatein
Syndecan
Thrombospondin superfamily
Tissue transglutaminase

ADAMTS
 β 1,4 galactosyltransferase
Collagens (Fibrillar)
Collagen XXIII-like
Dystroglycan
Fibrillin
HS-deacetylase
HS 2-O-sulphotransferase
Lysyl hydroxylase
Lysyl oxidase
SPARC
Spongin
Xylosyl-transferase

Figure 4

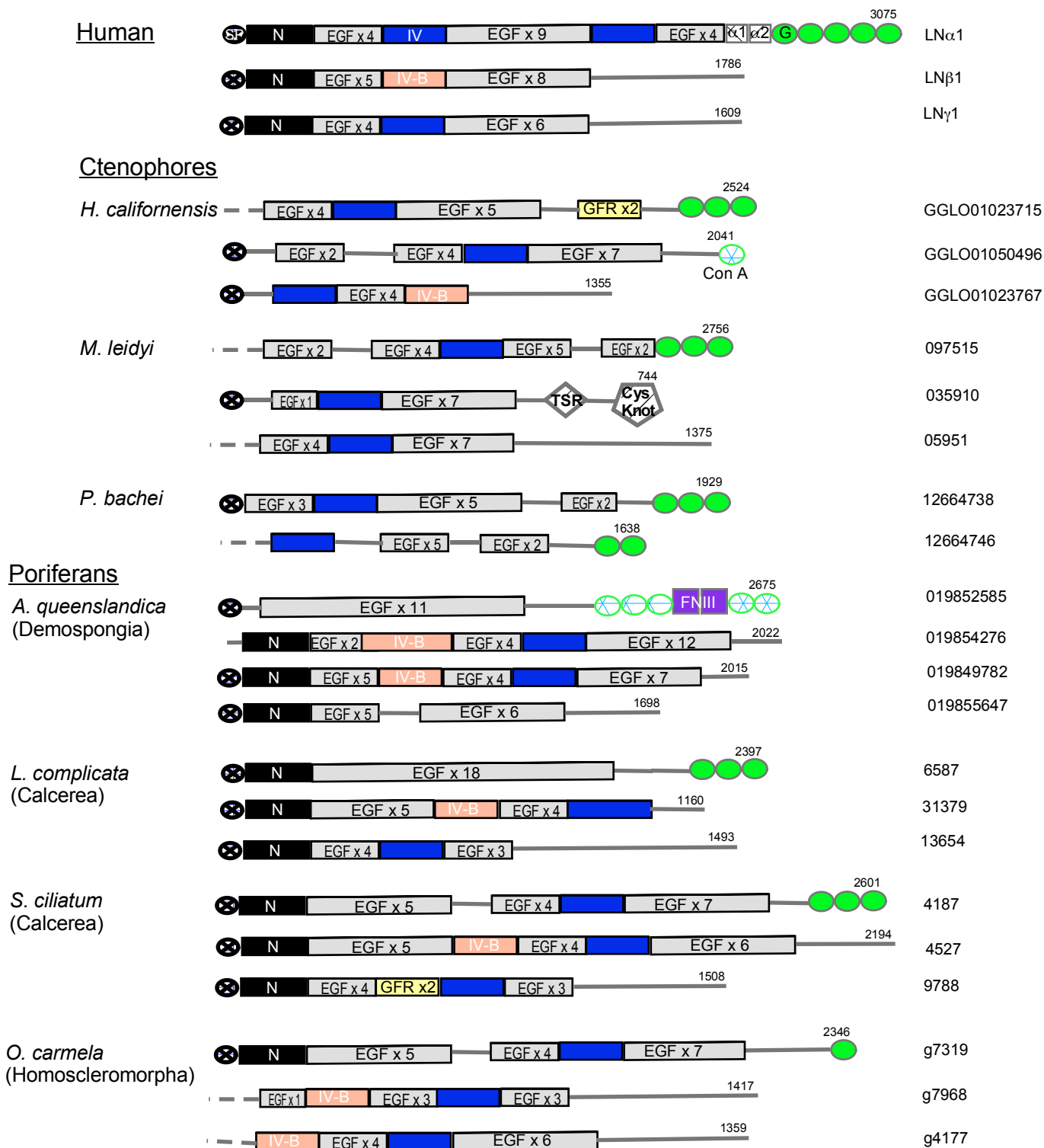


Figure 5

A

```
Aq      AQCWVFGGTLTTVLRTIGIP
4PYG_active_site. GQCWVFAAVACTVLRCLGIP
ML03126a GQCWVFAGVLTSLYRTLIGIP
Pb      GQCWVFSGVLCSLYRTLIGIP
ML25826a GQCWVFSGVLTLLRCLGIP
        .*****...  :: * :***
```

B



C

