Origin and Evolution of Laminin Gene Family Diversity

Bryony Fahey and Bernard M. Degnan*

1 School of Biological Sciences, The University of Queensland, Brisbane, Queensland, Australia
*Corresponding author: E-mail: b.degnan@uq.edu.au.
Associate editor: Claudia Kappen

Abstract

Laminins are a family of multidomain glycoproteins that are important contributors to the structure of metazoan extracellular matrices. To investigate the origin and evolution of the laminin family, we characterized the full complement of laminin-related genes in the genome of the sponge, Amphimedon queenslandica. As a representative of the Demospongiae, a group consistently placed within the earliest diverging branch of animals by molecular phylogenies, Amphimedon is uniquely placed to provide insight into early steps in the evolution of metazoan gene families. Five Amphimedon laminin-related genes possess the conserved molecular features, and most of the domains found in bilaterian laminins, but all display domain architectures distinct from those of the canonical laminin chain types known from model bilaterians. This finding prompted us to perform a comparative genomic analysis of laminins and related genes from a choanoflagellate and diverse metazoans and to conduct phylogenetic analyses using the conserved Laminin N-terminal domain in order to explore the relationships between genes with distinct architectures. Laminin-like genes appear to have originated in the holozoan lineage (choanoflagellates + metazoans + several other unicellular opisthokont taxa), with several laminin domains originating later and appearing only in metazoan (animal) or eumetazoan (placozoans + cnidarians + bilaterians) laminins. Typical bilaterian α, β, and γ laminin chain forms arose in the eumetazoan stem and another chain type that is conserved in Amphimedon, the cnidian, Nematostella vectensis, and the echinoderm, Strongylocentrotus purpuratus, appears to have been lost independently from the placozoan, Trichoplax adhaerens, and from multiple bilaterians. Phylogenetic analysis did not clearly reconstruct relationships between the distinct laminin chain types (with the exception of the α chains) but did reveal how several members of the netrin family were generated independently from within the laminin family by duplication and domain shuffling and by domain loss. Together, our results suggest that gene duplication and loss and domain shuffling and loss all played a role in the evolution of the laminin family and contributed to the generation of lineage-specific diversity in the laminin gene complements of extant metazoans.

Key words: extracellular matrix, Metazoa, netrin, perlecan, Porifera, sponge, usherin.

Introduction

Most animals are covered by a selectively permeable epithelial cell layer. Internal cavities and organs are also often lined by an epithelial sheet. In bilaterians, epithelial cells 1) exhibit an apical–basal polarity, 2) form cohesive belt-form junctions with adjacent epithelial cells, and 3) are connected to an extracellular matrix (ECM) at their basal surface, although there are also cases of apical attachment (Tyler 2003). In general, the genes responsible for determining these epithelial features appear to be conserved between insects, nematodes, and vertebrates (Hutter et al. 2000; Hynes and Zhao 2000; Knust and Bossinger 2002; Hortsch and Margolis 2003; Nance 2005; Humbert et al. 2006). A majority of these genes are present in the genomes of representative species of the earlier branching cnidian and placozoan lineages (Putnam et al. 2007; Srivastava et al. 2008; Chapman et al. 2010; Fahey and Degnan 2010). Recently, it has been shown that most of the genes encoding proteins that establish apical–basal polarity and adherens junctions also exist in the genome of the demosponge, Amphimedon queenslandica, a representative of an even earlier diverging phyletic branch of the Metazoa (Fahey and Degnan 2010; Srivastava et al. 2010).

In contrast, basal ECM (basal lamina) protein orthologues (i.e., collagens IV and XV/XVIII, laminins α, β, and γ, nidogen, and perlecan) are not present in Amphimedon, although laminin-related proteins do exist. In vertebrates and other bilaterians, a laminin network forms part of the basal lamina (Henrikson et al. 1997), providing support and morphogenetic cues to epithelial cells via interactions with cell surface receptors, such as integrin and dystroglycan (Li et al. 2003). In model bilaterians (fly, worm, and mammals), laminins are expressed early in embryogenesis concomitant with the first appearance of basal lamina structures and are thought to play a major role in their initial assembly (Li et al. 2003; Miner 2008). Despite lacking a basal lamina, Amphimedon has a collagen-based ECM (Leys and Degnan 2002; Exposito et al. 2008) and possesses integrin and dystroglycan cell surface receptors (Srivastava et al. 2010), suggesting that laminin-related proteins could play a role in cell–ECM adhesion in this sponge.

In bilaterians, laminins are large multidomain glycoproteins that assemble into heterotrimers consisting of an α, β, and γ chain (fig. 1). Laminin trimers polymerize into a network that provides structural integrity to the basal lamina. Although distinct in structure, α, β and γ laminin subunits share sequence characteristics indicative of a common
ancestry. All chains (with the exception of some truncated vertebrate-specific forms—see below) consist of a Laminin N-terminal domain (LamNT), a series of repeated Laminin-type Epidermal Growth Factor domains (LamEGF), and a region at or near the C-terminus capable of forming a coiled coil. Each of the four conserved bilaterian chain types, \( \alpha \frac{1}{2}, \alpha \frac{3}{5}, \beta, \) and \( \gamma \), differs in the type and number of globular domains found interspersed with the LamEGF repeats or at the C-terminus of the protein (fig. 1A). Laminin G domains (LamG) are found only at the C-terminus of the \( \alpha \)-type chains and appear to be the main sites for interaction between the heterotrimer and the cell surface receptors, integrin and dystroglycan (Tzu and Marinkovich 2008). Laminin \( \alpha \frac{3}{5} \) (LamG) and Laminin IVB (LamIVB; nomenclature according to PROSITE PDOC51115) domains are specific to the LamEGF repeat region of \( \alpha \frac{3}{5} \) and \( \beta \) chains, respectively. Laminin IVA domains (LamIVA; see PROSITE PDOC51115; also often referred to as LamB) are found interspersed with the LamEGF repeats in all chains except those of the \( \beta \) type. Laminin trimerization is mediated by the coiled-coil regions of the constituent chains, which interact to form a triple helix (Beck et al. 1993) (fig. 1B). Conserved cysteines form interchain disulfide bonds at the N- and C-termini of the coiled coil, which help to stabilize this interaction.

Polymerization of laminin heterotrimers into a supportive ECM network appears to involve interactions between the LamNT domains of the three chains and the LamNT domains from surrounding laminin molecules, although little is known about the exact molecular mechanisms for this interaction (Cheng et al. 1997). Further stability is imparted to the basal lamina by interactions between the laminin network and a network of Type IV collagen mediated by the cross-linking molecules, perlecans and nidogens (Timpl and Brown 1996).

As mentioned above, four laminin chain types are conserved amongst model bilaterians, with invertebrates (Drosophila melanogaster and Caenorhabditis elegans) possessing...
one gene encoding for each of the α 1/2, α 3/5, β, and γ subunits (Hutter et al. 2000; Hynes and Zhao 2000). Mammals contain an expanded repertoire of laminin genes (two α 1/2, two α 3/5, one α 4, three or four β, and three γ) including three unique chains (α 4, β 3, and γ 2), which are either N-terminally truncated or missing domains in the LamEGF region (Tzu and Marinkovich 2008). Hence, while worm and fly possess two distinct forms of the laminin heterotrimer, mammals possess as many as 16, some of which contain truncations in one or all of the N-terminal regions (or short arms).

Little is known about the evolution of the laminin gene family. The ECM is a metazoan synapomorphy and, as expected, laminin genes and their characteristic domains are not found in fully sequenced genomes from fungi, plants, and non-opisthokont unicellular eukaryotes (King et al. 2008; Fahey and Degnan 2010). As with several other important ECM domains, the laminin domains, LamNT and LamG, are present in the genome of the choanoflagellate, *Monosiga brevicollis*, a representative of the unicellular lineage most closely related to the Metazoa (King et al. 2008). With the exception of a partial α 3/5-like gene and a full-length β gene isolated from the cnidarian, hydra (*Hydra vulgaris* or *Hydra magnipapillata*) (Sarras et al. 1994; Zhang et al. 2002), members of the laminin gene family have not been characterized in early branching metazoan lineages.

Here, we explore the molecular characteristics of the complete set of laminin-related genes encoded by the genome of the demosponge, *A. queenslandica*. As a member of phylum Porifera, consistently placed at the base of the metazoan phylogenetic tree by molecular analyses (Philippe et al. 2009; Sperling et al. 2009; Pick et al. 2010; Srivastava et al. 2010), *Amphimedon* can provide insight into the early evolution of metazoan-specific gene families. We find that all *Amphimedon* laminins possess unique domain architectures with respect to known bilaterian chain types but still appear to be capable of assembly into higher order structures via formation of coiled coils and interchain disulfide bonds. In order to gain further insight into the evolution of the metazoan laminin family, we characterized the complete complements of laminin-related genes in the genomes of the choanoflagellate, *M. brevicollis*, the placozoan, *Trichoplax adhaerens*, and the cnidarian, *Nematostella vectensis*, and compared these with known laminin chain types from several bilaterians, namely *D. melanogaster*, *Strongylocentrotus purpuratus*, and *Homo sapiens*. We also explored the phylogenetic distribution and origin of conserved metazoan gene families that share unique domains with laminins, including the usherin, perlecan, and netrin families. Finally, domain-specific phylogenetic analyses and alignments allowed us to elucidate relationships between the different metazoan laminin chain types and the nonlaminin gene families that share the relevant domains.

**Materials and Methods**

**Genome Searches**

Genomic trace and expressed sequence tag (EST) data for *A. queenslandica* was generated as part of a collaborative genome project with the Joint Genome Institute (JGI) and is publically available on NCBI (http://www.ncbi.nlm.nih.gov/) and on the metazome (http://spongezome.metazome.net). Assembled genomic contigs, automated gene predictions, bulk annotations of the automated gene predictions, and an annotated genome browser were also kindly provided by the JGI. BLAST searches in *M. brevicollis*, *N. vectensis*, and *T. adhaerens* were carried out using the relevant BLAST tools available at the JGI (http://genome.jgi-psf.org/) and/or the NCBI BLAST website.

The full-length *Drosophila* laminin β sequence (laminin B1, isoform A; NP_476618) was used as a query to search the predicted proteomes of *Amphimedon*, *Monosiga*, *Trichoplax*, and *Nematostella* by BLASTp. tBLASTn searches were also performed against genomic contigs using partial (LamNT domain only) or full-length laminin β (as well as full-length human laminin α 3 isoform 1, NP_937762, and laminin γ 1 isoform 1, NP_002284, for *Monosiga* and *Amphimedon*). As a complementary approach, genomic domain searches were used to detect predicted proteins containing relevant laminin domains (e.g., LamNT, LamEGF, LamIVB, and LamG). For *Monosiga*, *Trichoplax*, and *Nematostella*, these searches were conducted using the advanced search feature at the JGI genome homepages (to detect InterPro domains) and the architecture analysis function on the SMART annotation of *M. brevicollis* website (http://smart.embl.de/Monosiga/index.html) (King et al. 2008; Hunter et al. 2009; Letunic et al. 2009). For *Amphimedon*, genomic domain searches relied on a PFAM annotation of the entire predicted protein set (Finn et al. 2008).

For each identified gene, we chose amongst available gene prediction models by picking the model that best incorporated browser EST or BLASTx alignments and/or possessed the greatest number of relevant domains. Presence of an N-terminal signal peptide for protein secretion (SignalP), as revealed by SMART, was also performed against genomic contigs using partial (LamNT domain only) or full-length laminin β (as well as full-length human laminin α 3 isoform 1, NP_937762, and laminin γ 1 isoform 1, NP_002284, for *Monosiga* and *Amphimedon*) via the SMART annotation of *M. brevicollis* website (http://smart.embl.de/Monosiga/index.html) (King et al. 2008; Hunter et al. 2009; Letunic et al. 2009). For *Amphimedon*, direct translations to edit available models.

For *S. purpuratus*, we relied on the genome analysis already performed by Whittaker et al. (2006). However, sea urchin genome resources available via the NCBI, the sea urchin Genboree site (http://www.genboree.org/java-bin/PurpleUrchin/index.jsp?isPublic=Yes) and SpBase (http://www.spbase.org/SpBase/) were used to improve and extend predictions and check for underlying gaps in genomic contigs. Sequences for *D. melanogaster* and *H. sapiens* laminins and other proteins (netrin, usherin, and perlecan) were taken from the NCBI RefSeq database (Pruitt et al. 2007).

**Bioinformatic Sequence Characterization**

Sequences were annotated for conserved domains by scanning against InterPro (http://www.ebi.ac.uk/Tools/InterProScan/) (Quevillon et al. 2005), SMART (Letunic et al.
genes possess the two sequence characteristics diagnostic for the laminin family, an N-terminal region consisting of a series of LamEGF repeats and a C-terminal region with high probability for coiled-coil formation. Additionally, four of the five genes possess an N-terminal LamNT domain, the only conserved domain besides LamEGF that is common to all chain types. However, none of the Amphimedon laminins displays a domain architecture typical for one of the four conserved bilaterian chain types, α 1/2, α 3/5, β, or γ.

AmqLamα3/5-like resembles laminin α genes in its possession of five LamG domains C-terminal to the coiled coil, a conserved feature of known laminin α chains. Unlike typical full-length α 1/2 and α 3/5 chains, the protein does not possess a LamIVA domain. N-terminal to the LamEGF domains is a stretch of sequence with similarity to the Lam α3/5 domain. However, the Amphimedon protein aligns only to the N-terminal 150 amino acids of the full 500 amino acid domain, suggesting that this may be a degenerate remnant of an ancestral domain that has been maintained in other lineages (supplementary fig. S1, Supplementary Material online). AmqLamα3/5-like also lacks a conserved LamNT domain. A low-scoring hit to the SMART LamNT domain profile is found at the N-terminus of the protein, but the sequence conservation is so poor that it is difficult to distinguish from a false positive (supplementary fig. S1, Supplementary Material online).

AmqLamβ/γ-like1 and AmqLamβ/γ-like2 resemble bilaterian laminin β and γ chains in their domain structure. However, whereas the latter possess either a single LamIVA domain (γ) or a single LamIVB domain (β) interspersed within the LamEGF repeats, the two Amphimedon proteins display a combination of both domains. This combination appears to be unique, with LamIIVA and LamIVB domains not being found together in any of the characterized laminin chains from Drosophila, C. elegans, or mammals (Hutter et al. 2000; Hynes and Zhao 2000; Tzu and Marinovich 2008).

AmqLamγ-like resembles laminin γ in its architecture, and we have previously reported this gene as a putative Amphimedon orthologue of laminin γ (Srivastava et al. 2010). However, close inspection of the coiled-coil region revealed the presence of a short cysteine- and glycine-rich region similar to the Laminin β knob domain found only in bilaterian laminin β proteins (supplementary fig. S1, Supplementary Material online). The sequence of this region is not as well conserved as that of bilaterian and Nematostella laminin β proteins but displays a level of conservation similar to that of the Trichoplax laminin β orthologue. Additionally, phylogenetic analyses (see below) do not provide support for an orthologous relationship between AmqLamγ-like and eumetazoan (placozoans + cnidophores + cnidarians + bilaterians; see also fig. 8) laminin γ proteins. Hence, it is likely that AmqLamγ-like represents a unique chain type distinct from any known bilaterian laminins.

The shortest of the Amphimedon laminins, AmqLam-like, lacks any globular domains at the C-terminus or interspersed with the few LamEGF repeats. A similar protein, laminin β 3, is known from mammals (Tzu and Marinovich 2008). However, sequence affinity and the presence of a Laminin

Results

Amphimedon Encodes a Unique Set of Laminin-Related Proteins

The Amphimedon genome encodes five laminin-related genes with four distinct domain architectures (fig. 2). All
The β knob domain suggests that this protein is likely to represent a laminin β chain that has lost its LamIVB domain. Although distinct from known bilaterian laminin subunit types, it appears that the Amphimedon laminins are capable of forming similar heterotrimer structures. As discussed previously, laminin assembly in bilaterians is driven by the formation of a triple helical coiled coil, with further stability imparted through interchain disulfide bonds. The putative coiled-coil forming regions of the five predicted Amphimedon laminins are all of approximately the same length (582–622 amino acids). They also all contain conserved cysteine residues at the N- and C-termini of the coiled-coil regions in a comparable position to those that appear to form interchain disulfide bonds in the laminins of vertebrates (Beck et al. 1993).

The specificity of coiled-coil interactions is difficult to predict computationally (Mason and Arndt 2004), hence, we are unable to infer confidently from sequence alone which of the Amphimedon laminin chains might be capable of interacting to form heterooligomers. However, while available programs are unable to predict interchain interactions, some are able to classify individual coiled-coil sequences in terms of whether they are more likely to assemble as part of a dimer or a trimer (assuming that they do not assemble into even higher order oligomers). We used PrOCoil (Mahrenholz et al. 2011) to predict oligomerization states for the most confidently assigned coiled-coil segments found in the coiled-coil regions for each trimer.

**Fig. 2.** Amphimedon laminins. (A) Domain diagrams for laminin proteins encoded by the Amphimedon genome. Diagrams are drawn to scale and reflect the locations of features on the primary sequence. Poorly conserved, dubious, or partial domains are denoted by gray outlines, and alignments are provided for these elsewhere (supplementary fig. S1, Supplementary Material online). AmqLamα3/5-like is a single complete gene but has been depicted here as two fragments to conserve space. The locations of conserved cysteines (C) that may form interchain disulfide bonds are indicated with arrows. SignalP, N-terminal signal peptide for protein secretion. Refer to figure 1 legend for domain name abbreviations. (B) Structure of a hypothetical Amphimedon laminin heterotrimer. A hypothetical laminin trimer composed of the three Amphimedon laminin chains that would impart the most similarity to a typical mammalian laminin heterotrimer (i.e., one with full-length short arms) is shown. For illustrative purposes, human Laminin-10 (α5β1γ1) (representative of a typical bilaterian α 3/5, β, and γ heterotrimer) is shown for comparison. Chain features are drawn to scale on the primary sequence, and hence, the diagrams do not represent the dimensions of correctly folded mature trimers. Putative biological functions mediated by the indicated domains are labeled on the Amphimedon trimer. Putative interchain disulfide bonds are depicted by horizontal bars at the N- and C-termini of the coiled-coil regions for each trimer.
all four coiled-coil segments were predicted most likely to form dimers. However, here as in all cases, only the more confidently predicted coiled-coil segments were classified, meaning that contributions to the oligomerization tendency from remaining parts of the coiled-coil forming sequence were not counted. In summary, the results of the PrOCoil analysis do not provide strong evidence in favor of a hypothesis that the *Amphimedon* laminins can assemble as heterotrimers but do not exclude the possibility either.

Although experimental investigation of protein–protein interactions would be required to prove that laminin trimers form in *Amphimedon*, knowledge of the domain composition of *Amphimedon* laminins allows us to make some predictions about the functionality of putative heterotrimers. A hypothetical heterotrimer formed from the *Amphimedon* laminin chains most similar to α (α 3/5-like), β (β-like), and γ (γ-like) would represent the closest approximation to a typical bilaterian heterotrimer (fig. 2B). Were it to exist, this molecule would possess all the domains found in bilaterian laminin chains with the exception of a full Lam α3/5 domain and a well-conserved Laminin β knob. The functional implications of these absences are unclear, however, the lack of a conserved LamNT domain at the N-terminus of the α 3/5-like subunit would be likely to affect the molecule’s capacity for polymerization (see Discussion).

**Phylogenetic Distribution of Laminin Chain Types Reveals Lineage-Specific Variety in Metazoan Laminin Gene Complements**

The differences in domain architecture between *Amphimedon* and bilaterian laminins suggest that the evolution of the metazoan laminin family may have involved extensive domain shuffling and rearrangement. To explore this further, we characterized the complete complements of laminin-related genes in the genomes of the choanoflagellate, *Monosiga brevicollis*, the placozoan, *T. adhaerens*, and the cnidarian, *N. vectensis* (fig. 3). We restricted our analyses to organisms with fully sequenced genomes because of the difficulty of accurately reconstructing domain architectures for large genes from partial EST data.

The *Monosiga* genome encodes a single laminin-related protein, MbLam-like, characterized by a LamNT domain, LamEGF repeats, a putative coiled-coil region, and a long...
region C-terminal to the coiled coil, which does not encode any laminin domains. With the exception of the C-terminal sequence, this gene resembles Amqlam-like in its architecture (fig. 3).

Trichoplax possesses three laminin-related genes, all of which resemble bilaterian forms. Architectures for Talamβ and Talamγ are typical, whereas Talamα1/2-like displays three instead of the five LamG domains that are normally found at the C-terminus of laminin α 1/2 proteins. These data suggest that Trichoplax encodes proteins capable of assembling into a single α, β, and γ heterotrimer, which would possess most of the molecular characteristics of a bilaterian laminin heterotrimer.

Nematostella also possesses a single gene encoding for each of the α, β, and γ chain forms. However, the Nematostella α chain is of the α 3/5 type and an α 1/2 gene is lacking. In addition to these three genes, Nematostella possesses three laminins with the same architecture as the Amphimedon laminin βγ-like proteins. Interestingly, a gene encoding this same domain architecture is also present in the Strongyllocentrotus genome (called Sp-LAMB2 and designated as laminin β-like in Whittaker et al. 2006). The absence of this chain type in Trichoplax, C. elegans, Drosophila, and mammalian genomes suggests that it has been lost independently in multiple lineages (Hutter et al. 2000; Hynes and Zhao 2000; Tzu and Marinkovich 2008). Although not well represented amongst characterized laminins or in the NCBI sequence database, preliminary analyses indicate that similar genes are encoded in other bilaterian genomes, for example, Capitella sp. and Branchiostoma floridæ (data not shown).

As in other bilaterians, Strongyllocentrotus possesses genes for the laminin α 1/2, α 3/5, β, and γ chain types described previously (see Introduction); however, both laminin α genes are C-terminally truncated by breaks in the genome assembly. In addition, a unique laminin α-related gene, Splamz-like (Sp-LAM2_caeel in Whittaker et al. 2006), with only a single LamIVA domain interspersed amongst the LamEGF repeats, is present in the sea urchin genome.

In summary, this analysis reveals that laminin-related genes originated in the holozoan lineage and expanded in metazoans, with metazoan-specific (LamIVA and LamIVB) and ancestral (LamG) domains combining to give rise to multiple chain types. The typical chain types, α 1/2, α 3/5, β, and γ, originated in the eumetazoan and eumetazoan sensu stricto (s.s.; Cnidaria + Bilateria; see also fig. 8) lineages, while the conserved β/γ-like form appears to have been lost independently in multiple groups.

Phylogenetic Distribution of Genes Sharing Laminin Domains Reveals Metazoan and Eumetazoan Gene Families That May Have Originated from Within the Laminin Family

Several important metazoan cell adhesion and ECM proteins share multiple domains with laminins. The netrin family of ECM proteins and the transmembrane protein, usherin, both possess an N-terminal LamNT domain followed by a series of LamEGF repeats. However, the C-terminal regions of members of these protein families contain domains not found in laminins and lack the coiled-coil forming sequence. Perlecan proteins are proteoglycans that contribute to the structure of the basal lamina. They contain a stretch of LamEGF repeats with multiple LamIVA domains distributed amongst them and a variety of nonlaminin domains up and downstream. Our searches for laminins in the genomes of a choanoflagellate and multiple metazoans also allowed us to identify genes that share characteristic laminin domain combinations. In particular, we focused on genes that possess a combination of LamEGF domains with either LamNT or LamIVA domains (LamIVB domains were never found in nonlaminin proteins). The phylogenetic distribution of these genes is shown in figure 4.

None of the gene families mentioned above is represented in the Monosiga genome. Monosiga does possess two unique putative transmembrane proteins, both with an N-terminal LamNT domain and a stretch of LamEGF repeats. Similar genes were not found in metazoan genomes. Usherin orthologues, previously thought to be deuterostome specific (Whittaker et al. 2006), were found in Amphimedon and Nematostella genomes, indicating that their absences in Trichoplax and Drosophila are likely to be a result of lineage-specific losses. In the Nematostella genome, the usherin locus has undergone tandem duplication to give rise to a truncated usherin-related protein with a transmembrane domain directly C-terminal to the LamEGF repeats. Consistent with their important role in the basal lamina, perlecan orthologues appear to be a eumetazoan innovation and have been retained in all of the surveyed eumetazoan genomes. Netrins were found in all eumetazoan s.s. genomes. Genes containing LamEGF repeats and a Netrin C-terminal domain were found in Trichoplax (Srivastava et al. 2010; data not shown), suggesting that the netrin-specific domain architecture may have originated in eumetazoans with loss of the LamNT domain occurring in Trichoplax. Netrin G proteins, which lack a Netrin C-terminal domain, were found only in vertebrates. The Nematostella genome encodes a unique protein with a LamNT domain, LamEGF repeats, a single LamIVA domain, and a transmembrane domain (data not shown; supplementary file 1, Supplementary Material online). The gene, Nematostella Laminin fragment, appears to be a truncated laminin that has lost the coiled-coil region and gained a transmembrane domain instead.

Phylogenetic Analysis of Laminin Domains Provides Insight into the Evolution of Metazoan Laminins and Related Cell Adhesion Proteins

We performed a phylogenetic analysis of the laminin family based on alignment of the shared LamNT domain, with the goal of understanding how laminin genes with different domain architectures are related to each other. The LamNT domain was chosen because of its wide occurrence, whereas the LamEGF repeat and coiled-coil regions were deemed unsuitable due to the repetitive nature and lower complexity of the sequence. Nonlaminin proteins
that possess a LamNT domain (i.e., those discussed in the preceding section) were also included in the analyses. MbLamNTa, MbLamNTb, and usherin proteins were included in the initial alignment but removed later because they were too divergent. Usherin proteins all share a large insertion in the middle of the LamNT domain.

The LamNT phylogeny recovered the expected relationships for all eumetazoan laminin α, β, and γ proteins (fig. 5). MbLam-like appeared to group with the eumetazoan laminin α family, despite lacking any clear molecular synapomorphies with laminin α proteins. By contrast, the analysis did not provide insight into the affinities of the Amphimedon laminins. Two of the laminin β/γ-like genes from Nematostella grouped with Strongylocentrotus laminin β/γ-like and with human netrin G1 and G2 in a well-supported clade. The close relationship between SpLam β/γ-like and the two human netrins suggests that the latter were derived from an ancestral laminin β/γ-like gene with loss of the C-terminal portion of the sequence.

Interestingly, the LamNT phylogeny indicates that all eumetazoan netrins were derived from within the laminin family by domain rearrangement events. Here, as elsewhere, we use the term “laminin family” to refer to the
wider family of proteins displaying typical laminin sequence characteristics, regardless of specific chain type or orthology. The majority of netrins, including those from Nematostella, Drosophila, and Strongylocentrotus, appear most closely related to the laminin $\gamma$ branch, suggesting that they were derived by duplication of a laminin $\gamma$ gene that antedated the divergence of cnidarian and bilaterian lineages. In contrast, the human netrin 4 protein appears more closely related to laminin $\beta$ proteins, indicating a probable independent origin for this gene, despite the shared domain architecture. Although surprising, these results are consistent with a previous analysis of vertebrate netrins and laminins conducted by Koch et al. (2000). Thus, the observation that the three types of netrins display affinity to distinct branches of the laminin family and each arose subsequent to its initial origin suggests that they were derived independently from within the family by domain rearrangement events.

Also included in the LamNT phylogeny was NvLam_frag, a Nematostella protein possessing a LamNT domain, LamEGF repeats, and a single LamIVA domain but with no coiled-coil region. As can be seen, the protein has no clear affinity to any other sequence or group. In addition, two fragmented LamNT domain–containing genes from sea urchin were included. One of these genes appears closely related to sea urchin laminin $\gamma$, whereas the other has no supported relationship with any other laminins.

As mentioned above, our phylogenetic analysis of the LamNT domain failed to provide insight into the affinities of the Amphimedon laminins and did not reveal relationships between any of the distinct laminin chain types (with the exception of the two $\alpha$ chain types). Hence, we performed preliminary phylogenetic analyses using the LamIVA and LamIVB domains, in order to determine whether these could provide additional information. Lam $\alpha$3/5 and LamG domains were not considered suitable for this purpose because of their more restricted occurrence. Initial NJ analyses based on alignment of the LamIVA and LamIVB domain produced trees with only poorly supported nodes and provided few insights further to those obtained from the LamNT phylogeny (supplementary figs. S2 and S3, Supplementary Material online). In light of this, we did not consider further phylogenetic analyses using these domains likely to yield much additional clarity.

However, preliminary phylogenetic analyses of the LamIVA and LamIVB domains, in combination with visual inspection of the alignments, did yield some results of interest (figs. 6 and 7). The LamIVA phylogeny was able to shed light on the affinity of the unique $\alpha$-like laminin from Strongylocentrotus, which could not be included in the LamNT phylogeny because it contains a gap in the LamNT domain. The gene appears to be related to NvLam_frag, a protein with a similar architecture but with no coiled-coil region or LamG repeats. Both proteins fell within the $\alpha$ clade, possessing diagnostic amino acids for this clade at positions 4, 128, and 132 of the LamIVA domain. Within the $\alpha$ clade, there was no obvious sequence differentiation between the LamIVA domains found in laminin $\alpha$ 3/5 or $\alpha$ 1/2 (N- or C-terminal) proteins. Perlecan LamIVA domains were also most closely related to those from laminin $\alpha$ proteins, with the same diagnostic amino acids found at positions 4, 128, and 132 (data not shown). As in the LamNT phylogeny, eumetazoan laminin $\gamma$ proteins grouped together with support and possessed distinctive amino acids at numerous positions in the alignment. Laminin $\beta$/$\gamma$-like proteins did not form a clade but did share amino acids with $\gamma$ sequences to the exclusion of the $\alpha$ clade at some positions (87, 132, 166, and 184).

In the LamIVB domain phylogeny, a split between laminin $\beta$ and laminin $\beta$/$\gamma$-like clades received moderate statistical support. Visual inspection of the alignment adds further support for this split, with diagnostic amino acids for each group found at positions 68, 88, 172, and 225.

**Discussion**

**Amphimedon Laminins**

The Amphimedon genome encodes a set of five laminin-related proteins, including some unique forms. Despite their differences to known bilaterian $\alpha$, $\beta$, and $\gamma$ laminin types, conserved properties of the coiled-coil regions suggest that they are capable of interacting in a similar way to form heterodimers or trimers in the ECM. We infer the molecular capacities of a putative laminin heterotrimer in Amphimedon using known bilaterian trimers as a guide. A heterotrimer formed from Amphimedon $\alpha$ 3/5-like, $\gamma$-like, and $\beta$/$\gamma$-like would represent the molecule with most similarity to a typical mammalian laminin trimer and with the greatest diversity of laminin domains. This structure would be able to interact with the cell membrane receptors for laminin, integrin and dystroglycan (both present in Amphimedon) via the LamG domains of the $\alpha$ 3/5-like subunit (Miner and Yurchenco 2004; Srivastava et al. 2010). However, any laminin trimer incorporating the $\alpha$ 3/5-like chain would lack a LamNT domain at the N-terminus of one of the three short arms, which may affect its capacity for polymerization. Although little is known about the molecular mechanism for laminin network formation via LamNT domain interactions, in vitro experiments suggest that laminin trimers lacking three complete N-terminal arms may not undergo self-polymerization (Cheng et al. 1997). In summary, although Amphimedon laminin chains may be able to form heterotrimers, differences in domain architecture are likely to render them incapable of combining all the key functions of basal lamina forming laminins.

Amphimedon lacks genes coding for the other major molecular constituents of the basal lamina, namely Type IV collagen, perlecan, and nidogen (Fahey and Degnan 2010; Srivastava et al. 2010). This is consistent with the fact that basal lamina–like structures have not been observed in demosponges via electron microscopy (Leys et al. 2009), with the exception of homoscleromorph sponges, which are now thought to represent a lineage distinct from other demosponges (Philippe et al. 2009; Sperling et al. 2009). The combination of molecular properties seen in each of the
The full alignment, spanning the majority of the domain, is provided in supplementary figure S4 (Supplementary Material online). The numbering on this figure corresponds to the numbering of the full alignment, with the section shown here spanning positions 75–169 of the 176 amino acid PROSITE sequence logo for the domain (PS51115). The Amphimedon protein, AmqLamC-like, is not included because it lacked many conserved amino acids in the central region of the domain, making it difficult to align unambiguously. Horizontal lines separate the sequences into distinct groups reflecting their expected and/or apparent phylogenetic relationships. For laminin \( \alpha_1/2 \) proteins, the letters D1 and D2 designate the N- and C-terminal LamIVA domains, respectively. Gray shading is used to mark similar amino acids found across the majority of sequences or across the majority of sequences within a particular group. In sites indicated with an asterisk, the shaded amino acids are not just similar but identical. Phylogenetically uninformative insertions have been removed and are flagged by black arrowheads at the top of the alignment; the deleted insertions are counted in the amino acid numbering.
Alignment of metazoan laminin-related proteins in the region representing the Laminin IVB (LamIVB) domain. The section shown includes positions 1–214 of the 214 amino acid PROSITE sequence logo for the domain (PS51116) and incorporates an additional one amino acid at the N-terminus and five amino acids at the C-terminus. A horizontal line separates the sequences into the laminin b and laminin e domains.

**Fig. 7.** Alignment of metazoan laminin-related proteins in the region representing the Laminin IVB (LamIVB) domain. The section shown includes positions 1–214 of the 214 amino acid PROSITE sequence logo for the domain (PS51116) and incorporates an additional one amino acid at the N-terminus and five amino acids at the C-terminus. A horizontal line separates the sequences into the laminin β and laminin ε domains.
Amphimedon laminins (N-terminal signal peptide, lack of transmembrane domain, and conserved features of the coiled-coil region) strongly suggests that they function as oligomers in the ECM. However, whether these oligomers polymerize and interact with cellular receptors to provide a stabilizing and scaffolding function for sponge tissues remains an open question.

Origin of Laminins and Related Metazoan Gene Families

Our analyses of the laminin family uncovered considerable variety in the gene complements and domain architectures of laminin-related genes found in Monosiga, Amphimedon, Trichoplax, Nematostella, and bilaterian genomes. The commonalities in domains and structure among all laminin genes (LamNT domain, multiple LamEGF repeats, and coiled-coil region) strongly suggest that they all derived from a single precursor by duplication with the addition and shuffling of other key domains. Additionally, three important metazoan cell adhesion and ECM gene families, netrin, perlecan, and usherin, share multiple common domains with laminins and may have evolved by domain shuffling from laminin-related precursors.

A putative scenario for the evolution of the metazoan laminin family based on the results of our phylogenetic analyses is presented in figure 8. Several laminin domain types (LamNT, LamEGF, and LamG) and the molecular architecture typical to metazoan laminins appear to predate the divergence between choanoflagellates and metazoans, with a single laminin-like gene being found in the Monosiga genome. However, LamIVA and LamIVB domains and laminin \( \beta/\gamma \)-like and usherin genes appear to have evolved in the lineage leading to crown metazoans. Additionally, a laminin \( \alpha \)-related gene with C-terminal LamG repeats must have originated in the metazoan lineage, although its exact architecture at the N-terminus cannot be inferred.

\( \beta/\gamma \)-like groups based on domain architecture. Gray shading is used to mark similar amino acids found across the majority of sequences or across the majority of sequences within a particular group. In sites indicated with an asterisk, the shaded amino acids are not just similar but identical. Phylogenetically uninformative insertions have been removed and are flagged by black arrowheads at the top of the alignment; the deleted insertions are counted in the amino acid numbering.
Finally, the well-documented bilaterian laminin chain types, α 1/2, α 3/5, β, and γ as well as perlecan and netrin (not including netrin 4, G1, or G2) genes, are likely to all represent eumetazoan or eumetazoan s.s. innovations.

**Evolution of the Laminin Family Involved Extensive Domain Rearrangement as Well as Whole Gene Loss**

Phylogenetic analysis performed on common domains allowed us to explore the relationships between the identified laminins and laminin relatives and generate hypotheses for how duplication and domain shuffling gave rise to the observed gene complements and domain architectures. However, partly because of the lack of support for relationships between *Amphimedon* laminins and other metazoan laminins, our phylogeny provides little insight into the exact mechanisms by which the various laminin forms were generated in metazoan and eumetazoan last common ancestors. It appears likely that the diversity of metazoan architectures was generated from an ancestral form that contained the three domains common to the majority of laminin forms, LamNT, LamEGF, and LamIVA. At some stage, an ancestral α-related gene must have acquired LamG domains at the C-terminus prior to generating both the laminin α 1/2 and α 3/5 families. Because typical β and γ forms are absent from *Amphimedon*, we speculate that these may have been derived from an original β/γ-like gene through domain loss in the eumetazoan last common ancestor, although it is also possible that laminin γ was generated through domain loss from an α-related precursor.

Our reconstructed history of the laminin family reveals several cases of gene loss, with the loss of at least one laminin type each from the lineages leading to *Trichoplax, Nematostella,* and *Drosophila* (fig. 8). One apparent case of gene loss, the absence of a β/γ-like gene in vertebrates, appears to be an instance of domain/sequence loss, as suggested by the close affinity between the LamNT domain of SpLamβ/γ-like and human netrin G1 and G2.

Our phylogeny of the LamNT domain also reveals that netrin genes with identical domain architectures (HsNetrin4 and the other eumetazoan netrins) have independent origins. A β-related netrin gene is also present in *Ciona intestinalis*, indicating that this domain-shuffling event predated the vertebrate last common ancestor (Sasakura et al. 2003). Finally, our phylogeny of the LamIVA domain shows a relationship between the multiple LamIVA domains of perlecan and those found in laminin α proteins indicative of a common ancestry.

**Conclusion**

In bilaterian model organisms, heterotrimeric composed of four conserved types of α, β, and γ laminin chains form a basic structural unit for the basal lamina, the supportive ECM network that stabilizes epithelia, nerves, and muscles. Our analyses of the laminin gene complements of several early branching metazoans and a choanoflagellate indicate that typical bilaterian α, β, and γ laminin forms arose in the eumetazoan stem, subsequent to the initial origin and expansion of the laminin gene family in ancestral holozoan and metazoan lineages. Laminin genes in the demosponge, *A. queenslandica*, contain typical laminin domains but in combinations and architectures that are not found in characterized α, β, and γ forms, suggesting that the early evolution of the laminin gene family involved duplication and extensive domain rearrangement. Duplication and domain shuffling within the laminin family may also have contributed to the origin of three other important cell adhesion and ECM families: usherin, perlecan, and netrin, all of which appeared in ancestral metazoan or eumetazoan lineages. An analysis of the distribution and phylogenetic relationships of laminins and their relatives revealed considerable variation in the gene complements of different metazoans, with duplication, gene loss, domain loss, and domain shuffling all playing a role in the generation of lineage-specific diversity.

**Supplementary Material**

Supplementary figures S1–S4, table S1, and file 1 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

**Acknowledgments**

This research was supported by grants from the Australian Research Council to B.M.D. We gratefully acknowledge the contribution of The United States Department of Energy Joint Genome Institute in the production of *A. queenslandica* genomic and EST sequences used in this study through the Community Sequencing Program.

**References**


