EVOLUTION & DEVELOPMENT 12:6, 601–617 (2010)
DOI: 10.1111/j.1525-142X.2010.00445.x

Origin of animal epithelia: insights from the sponge genome

Bryony Fahey and Bernard M. Degnan*

School of Biological Sciences, The University of Queensland, Brisbane, Qld 4072, Australia

*Author for correspondence (email: b.degnan@uq.edu.au)

SUMMARY Epithelial tissues are a key metazoan cell type, providing a basic structural unit for the construction of diverse animal body plans. Historically, an epithelial grade of organization was considered to be restricted to the Eumetazoa, with the majority of cell layers described for Porifera lacking any of the conserved ultrastructural characteristics of epithelia. Now with the use of genomic information from the demosponge, *Amphimedon queenslandica*, we identify orthologs of bilaterian genes that determine epithelial cell polarity or encode components of specialized epithelial junctions and extracellular matrix structures. *Amphimedon* possesses orthologs of most bilaterian epithelial polarity and adherens junction genes but few or no tight junction, septate junction, or basal lamina genes. To place this information in an evolutionary context, we extended these analyses to the completed genomes of various fungi, the choanoflagellate, *Monosiga brevicollis*, the placozoan, *Trichoplax adhaerens*, and the cnidarian, *Nematostella vectensis*. The results indicate that the majority of "epithelial" genes originated in metazoan or eumetazoan lineages, with only two genes, Par-1 and Discs large, antedating the choanoflagellate-metazoan split. We further explored the mechanism of evolution for each of these genes by tracking the origin of constituent domains and domain combinations. In general, domain configurations found in contemporary bilaterians are inferred to have evolved early in metazoan evolution and are identical or similar to those present in representatives of modern cnidarians, placozoans, and demosponges.

INTRODUCTION

Epithelial tissues are fundamental units of composition for most eumetazoan body plans. Epithelial sheets provide barriers for the construction of body walls, tissues, and organs, and allow for compartmentalization within the body and between the body and the external environment (Tyler 2003). The epithelial tissue type is not specific to any of the three eumetazoan germ layers, being found in tissues of ectodermal, mesodermal, and endodermal origin (Ruppert et al. 2004).

Three criteria distinguish the "true" epithelial tissue phenotype (Tyler 2003; Ruppert et al. 2004). First, component cells should display an aligned polarity with clearly distinguishable apical and basal surfaces. This characteristic is most visible in columnar cells, in which organelles are differentially distributed along the apical–basal axis. Epithelial cells may also exhibit planar polarity but this is not always as clearly manifested at the morphological level (Zallen 2007). Second, cells should be connected by belt-form junctions that form a continuous structure around the circumference of the cell and include tight (well characterized in vertebrates with possible related structures in ascidians), septate (found in most nonvertebrates but not vertebrates), and zonula adherens (found in both vertebrates and nonvertebrates) junctions (Green and Bergquist 1982; Tyler 2003). Last, cells should be associated with extracellular matrix only at their basal and apical surfaces (i.e., with a basal lamina and sometimes also with an apical cuticle).

Historically, these three morphological features of animal epithelia were best observed by electron microscopy. More recently it has become possible to gain insight into the development and maintenance of epithelial characteristics through analysis of the conserved molecular components, that contribute to epithelial structure and function in bilaterian model organisms. In general, the genes responsible for establishing cellular polarity, forming intercellular junctions, and constructing and regulating adhesion to the basal lamina appear to be conserved between insects, nematodes, and vertebrates (Hutter et al. 2000; Hynes and Zhao 2000; Tepass et al. 2000; Knust and Bossinger 2002; Roh et al. 2002; Deng et al. 2003; Hortsch and Margolis 2003; Li et al. 2003; Cox and Hardin 2004; Segbert et al. 2004; Nance 2005; Oda et al. 2005; Humbert et al. 2006; Johnson et al. 2006).

In differentiating epithelial cells of *Drosophila* and vertebrates the specification of distinct apical and basolateral membrane domains and the positioning of zonula adherens junctions at the boundary between these domains is accomplished by the positive and negative interactions between three cortically localized multiprotein complexes, the Crumbs–Stardust (vertebrate membrane protein, palmitoylated 5 or MPP5)-Patj complex, the Bazooka (vertebrate Par-3)-Par-6-atypical protein kinase C (aPKC) complex and the

Scribbled-Discs large (Dlg)–Lethal (2) giant larvae (Lgl) complex (Nelson 2003; Shin et al. 2006; Martin-Belmonte and Mostov 2008). These proteins maintain restricted patterns of subcellular localization postdifferentiation and can be useful as visual markers for discrete epithelial membrane domains. In both vertebrate and nonvertebrate epithelia the main adhesive junction, the zonula adherens, is positioned as a subapical belt around the circumference of the cell and comprises a physical complex between cell adhesion proteins of the classical cadherin subfamily (cadherins containing the catenin-binding cytoplasmic domain), the cytoskeletal linkers, β- and α-catenin, and the regulatory protein, p120/δ-catenin (Tyler 2003; Cox and Hardin 2004; Oda et al. 2005). By contrast, occluding junctions of vertebrate and nonvertebrate epithelia differ in both molecular composition and localization. Apically localized tight junctions composed of the transmembrane proteins, claudin and occludin, appear to be a vertebrate innovation (the molecular composition of ascidian tight junctions has not been characterized but occludin has not been found in the *Ciona intestinalis* genome) (Sasakura et al. 2003; Shin et al. 2006). The septate junction, the sealing junction of most nonvertebrate epithelia, is distinct from the functionally analogous tight junction, however, its molecular composition in *Drosophila* indicates that it may be homologous with the septate-like paranodal junctions that form between axons and glia in the vertebrate peripheral nervous system (Bhat 2003; Hortsch and Margolis 2003). The main conserved adhesive proteins forming these junctions are Neurexin IV, Contactin, and Neuroglian (and their vertebrate orthologs). Finally, the major components of the basal lamina are conserved between diverse bilaterians and include Type IV and Type XV/XVIII collagen, laminin, nidogen, and perlecan (Timpl and Brown 1996; Hutter et al. 2000; Hynes and Zhao 2000; Whittaker et al. 2006).

The majority of the proteins mentioned above appear to be metazoan-specific (i.e., not found in fungal or plant genomes) and it can be hypothesized that their origin was closely associated with the emergence of the epithelial cell phenotype early in animal evolution. The possession of tissues with a true epithelial grade of organization is generally considered to be a synapomorphy for the Eumetazoa *sensu stricto* (Cnidaria+Ctenophora+Bilateria) (Nielsen 2001; Tyler 2003; Ruppert et al. 2004; Magie and Martindale 2008). The epidermis and gastrodermis of cnidarians fulfil all the criteria of true epithelia. These tissues are composed of columnar cells that are linked by belt-form junctions and rest on a basal lamina enriched with laminin and Type IV collagen (Thomas and Edwards 1991; Sarras et al. 1994; Fowler et al. 2000; Zhang et al. 2002; Magie and Martindale 2008; Shimizu et al. 2008). In placozoans, an early-branching metazoan phylum with an uncertain phylogenetic position (Srivastava et al. 2008, 2010; Philippe et al. 2009; Schierwater et al. 2009; Sperling et al. 2009), upper and lower surface epithelial-like layers consist of polarized cells joined together by zonula adherens but with no underlying basal lamina (Grell and Ruthmann 1991). Structures with similarity to septate junctions have been observed basal to the zonula adherens in some cells of the ventral epithelium but it is unclear whether these should be interpreted as true septate junctions (Ruthmann et al. 1986; Nielsen 2001).

In Porifera, the phylum that most consistently falls outside the Eumetazoa as an early-branching metazoan lineage (or paraphyletic assemblage of lineages) (Philippe et al. 2009; Sperling et al. 2009; Srivastava et al. 2010), most tissues appear to lack key features of epithelia. However, some cell layers, such as the choanoderm (the choanocyte layer), can be considered epithelial-like, displaying an obvious aligned apical–basal cell polarity (Simpson 1984). The pinacoderm or outer layer of some demosponges appears to be integrated by adhesive spot-form junctions but belt-form junctions have not been observed and it is unclear whether sponges possess mechanisms for creating a barrier between the external and internal environment (Pavans de Ceccatty 1986; Leys 2007). Except in the case of the homoscleromorph sponges, which maintain an enrichment of Type IV collagen beneath the pinacoderm, an observable basal lamina appears to be absent (Boute et al. 1996). The ciliated outer layer of some demosponge larvae resembles the ciliated epithelium of cnidarian planula larvae and zonula adherens-like junctions have been observed at the apices between constituent cells (Boury-Esnault et al. 2003; Ereskovsky and Tokina 2004; Usher and Ereskovsky 2005; de Caralt et al. 2007). Again, homoscleromorph larvae appear to possess a thin layer of underlying extracellular matrix, suggesting that this larval epithelium may represent a true epithelium homologous to that of eumetazoans (Boury-Esnault et al. 2003; de Caralt et al. 2007; Maldonado and Riesgo 2008).

We were intrigued by the possibility that the epithelial-like outer layer of demosponge larvae might represent a tissue homologous with eumetazoan epithelia. As a first step toward investigating this problem, we have conducted a survey of the genome of the demosponge, *Amphimedon queenslandica*, for orthologs of bilaterian genes that act as determinants of epithelial structure and function. In addition to identifying orthologs of target genes through sequence similarity searches, we used bulk Pfam domain annotations for the *Amphimedon* predicted proteome to catalogue complete (or near-complete) sets of related genes including those containing PDZ (domain present in *P*SD-95, *D*lg, and *Z*O-1/2), Laminin G (LamG), and Cadherin domains.

To place our results from *Amphimedon* in an evolutionary context, we extended our analyses to include the completed genomes of fungi (multiple genomes—see "Materials and methods"), the choanoflagellate, *Monosiga brevicollis*, the cnidarian, *Nematostella vectensis*, and the placozoan, *Trichoplax adhaerens*. For this purpose we adopt the phylogenetic

hypothesis for metazoan relationships presented in recent phylogenomics analyses (Philippe et al. 2009; Srivastava et al. 2010), although we recognize that relationships among early-branching metazoan phyla are still uncertain (Dunn et al. 2008; Schierwater et al. 2009; Sperling et al. 2009). Unfortunately, instances of gene loss in *Amphimedon* and other early-branching metazoans, along with incomplete genome assembly, will confound our efforts to reconstruct the evolution of these genes and we stress that these analyses remain as hypotheses, which are likely to be reworked as more genomic information becomes available. In particular, a comparative approach using data from other demosponges and other key poriferan clades (i.e., hexactinellid, calcareous, and homoscleromorph sponges) will be necessary to determine whether instances of gene absence in *Amphimedon* might result from lineage-specific loss. Although an argument can be made for the necessity of further extending our analysis to include additional EST data sets from organisms occupying key evolutionary positions among metazoans and their close relatives, the difficulty of making orthology assignments for genes on the basis of partial sequence and domain information precluded their use in this study. The results of our comparative genome analyses also allowed us to generate hypotheses regarding the relative points of origin for domains and architectures relevant to each gene from our target set.

MATERIALS AND METHODS

Amphimedon genome searches

Genomic trace and 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/). 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.

Representative sequences for each bilaterian orthology group were chosen from the NCBI sequence database and used to search *Amphimedon* EST clusters and genomic contigs with tBLASTn. As a complementary approach, bulk analyses of the Pfam, PANTHER, and KOG annotations for the entire predicted protein set were used to detect *Amphimedon* proteins with relevant combinations of domains (Pfam) or matches to gene family profiles (KOG and PANTHER) (Tatusov et al. 2003; Thomas et al. 2003; Mi et al. 2007; Finn et al. 2008). In some cases, an additional search involved the use of keywords to locate regions of the genome that aligned to human proteins from the family of interest. These alignments were generated by BLASTx against the human proteome as part of the genome annotation process (Srivastava et al. 2010).

For each promising hit, we chose among available gene prediction models by picking the model which best incorporated EST or BLASTx alignments and/or produced the greatest number of relevant domain hits when scanned against the InterPro domain database (http://www.ebi.ac.uk/Tools/InterProScan/) (Quevillon et al. 2005). Where necessary, GenomeScan (http://genes.mit.edu/genomescan.html) (Yeh et al. 2001), GeneMark.hmm ES-3.0 (http://exon.gatech.edu/eukhmm.cgi) (Lomsadze et al. 2005), or direct translations were used to predict genes or to edit available models. Candidates were assessed for likelihood of orthology to the family of interest by scoring for domain composition and best BLASTp matches to sequences in the NCBI RefSeq database (http://blast.ncbi.nlm.nih.gov/Blast.cgi) (Altschul et al. 1997; Pruitt et al. 2007). In some cases, an in-house trace assembly pipeline was used to derive missing sequence information for local gaps or breaks in the genomic contigs (Larroux et al. 2007). See supporting information File S1 for the *A. queenslandica* sequences used in this study.

Comparative genomic analyses

We searched fungal, *M. brevicollis*, *T. adhaerens* and *N. vectensis* genomes for orthologs of the genes analyzed in *Amphimedon* using BLASTp (against predicted protein sets), tBLASTn (against genomic scaffolds/contigs) and combinatorial domain searches (against predicted protein sets) (Putnam et al. 2007; King et al. 2008; Srivastava et al. 2008). Fungal BLASTp and tBLASTn searches were conducted using the NCBI fungi genomes BLAST tool (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi?organism=fungi) and domain searches were performed with SMART in genomic mode (http://smart.embl-heidelberg.de/smart/set_mode.cgi?GENOMIC=1) (Letunic et al. 2009). These databases allowed for searches against genomes from multiple fungal species (BLASTp-50+species, tBLASTn-90+species, SMART-14 species). BLAST searches in *Monosiga*, *Nematostella*, and *Trichoplax* were carried out using the relevant BLAST tools available at the JGI (http://genome.jgi-psf.org/) and/or the NCBI BLAST website. Domain searches for these genomes were performed using the advanced search feature on the JGI genome homepages and the architecture analysis function on the SMART annotation of *M. brevicollis* website (http://smart.embl.de/Monosigia/index.html) (King et al. 2008). In some cases, we consulted the *Hydra magnipapillata* genome to confirm the results of surveys in *Nematostella* (Chapman et al. 2010). BLASTp searches were conducted using the NCBI BLAST server and tBLASTn searches using the Hydra BLAST tool on Metazome (http://hydrazome.metazome.net/search.php?show=blast&db=hydra_072606).

Promising hits were scored as for *Amphimedon*. In cases where gene models were obviously truncated or lacking important domains we used GenomeScan (Yeh et al. 2001), GeneMark.hmm ES-3.0 (Lomsadze et al. 2005) or direct translations to edit available models. See supporting information Files S2–S4 for the *M. brevicollis*, *T. adhaerens*, and *N. vectensis* sequences used in this study.

Phylogenetics

Sequences for phylogenetic analysis were obtained from the sources listed above, from the NCBI RefSeq database or from the Broad Institute (http://www.broadinstitute.org/). Alignments were constructed with ClustalX v2.0.10 and manually edited with MacClade 4.07 (Maddison and Maddison 2005; Larkin et al. 2007). Sequences with gaps caused by suspected modeling errors were replaced by alternative models, edited or discarded from the alignment.

Distance neighbor-joining (NJ) analyses with 1000 bootstraps were performed using PHYLIP v3.68 (Felsenstein 2005). For the Stardust/membrane palmitoylated protein (MPP)/Dlg phylogeny, a maximum likelihood (ML) analysis with 100 bootstraps was conducted using the PhyML v3.0 online web server (http://www.atgc-montpellier.fr/phyml/) (Guindon and Gascuel 2003), with the LG model recommended by ProtTest v2.2 (Abascal et al. 2005). For the Stardust/MPP/Dlg phylogeny, Bayesian analyses were conducted using MrBayes v3.1.2 (Ronquist and Huelsenbeck 2003) with the Jones–Taylor–Thornton model, a proportion of invariable sites and a γ-shaped distribution of rates. Two runs were performed for 5 million generations with default sampling parameters, and compared. A burn-in of 25% was discarded.

RESULTS

Amphimedon orthologs of bilaterian "epithelial" genes fit into two main functional categories

Figure 1 summarizes the results of a survey for apical–basal polarity, junction, and basal lamina genes in the *Amphimedon*

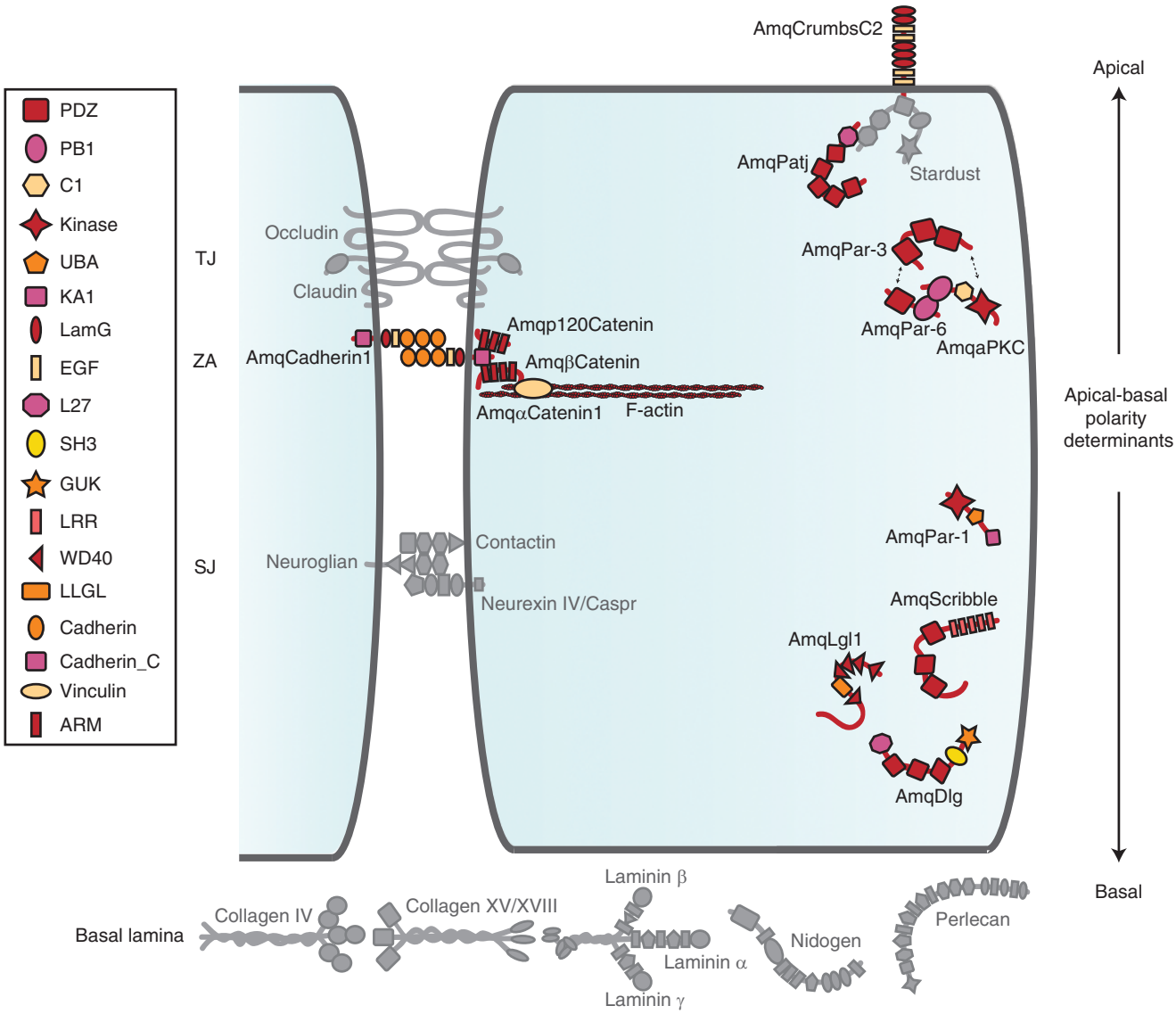


Fig. 1. Cell polarity, junction, and basal lamina genes in *Amphimedon*. Proteins depicted in color are encoded by predicted genes from the *Amphimedon* genome, while those shown in gray represent bilaterian proteins for which no clear *Amphimedon* orthologs were found. The domain diagrams accurately reflect domain composition and order but do not reflect domain numbers. For accurate domain diagrams refer to Fig. 2 and supporting information Figs. S1–S8. Putative protein–protein interactions predicted from studies in bilaterians are depicted where the relevant domain or binding motif is conserved in *Amphimedon*. For Par-3, there is no consensus in the literature as to how the protein interacts with Par-6 and aPKC (dashed arrows), with direct binding to both proteins having been reported (Izumi et al. 1998; Lin et al. 2000; Wodarz et al. 2000; Suzuki et al. 2001). TJ, tight junction; ZA, zonula adherens; SJ, septate junction.

genome. For each discrete functional category (polarity, tight junction, adherens junction, septate junction, basal lamina), *Amphimedon* displays a complete or near-complete presence or absence of the targeted gene complement. In some cases we have detected related *Amphimedon* genes that may be capable of substituting for the functions of genes depicted as absent in Fig. 1.

Cell polarity genes

Amphimedon possesses a single clear ortholog for each protein that functions as part of the Par cell polarity determining pathway in bilaterians (Fig. 1). Proteins encoded by *AmqPar-6* and *AmqaPKC* share unique domain architectures with their bilaterian counterparts (supporting information Figs. S1 and S2, Table S1). AmqPar-3 is distinguishable from numerous other multi-PDZ proteins in the *Amphimedon* genome because of its possession of an N-terminal domain conserved among metazoan Par-3/Bazooka proteins, which has been shown to be essential for proper localization and function of the protein in *Drosophila* and in Madin-Darby canine kidney cells (Benton and St Johnston 2003; Mizuno et al. 2003; Feng et al. 2007) (supporting information Fig. S3, Table S1). Finally, AmqPar-1 forms part of a well-supported clade comprising animal and choanoflagellate Par-1 orthologs in a phylogenetic analysis based on the kinase domain of Par-1 and related CAMKL (nomenclature according to KinBase—http://kinase.com/kinbase/) family kinases (Fig. 2).

In addition to encoding for these four core members of the Par complex, *Amphimedon* possesses orthologs of other genes that act as regulators and effectors of Par signaling. These include Cdc42, which binds to Par-6 and regulates aPKC activity (Yamanaka et al. 2001), and 14-3-3, which interacts with Par-1 (Göransson et al. 2006; Srivastava et al. 2010).

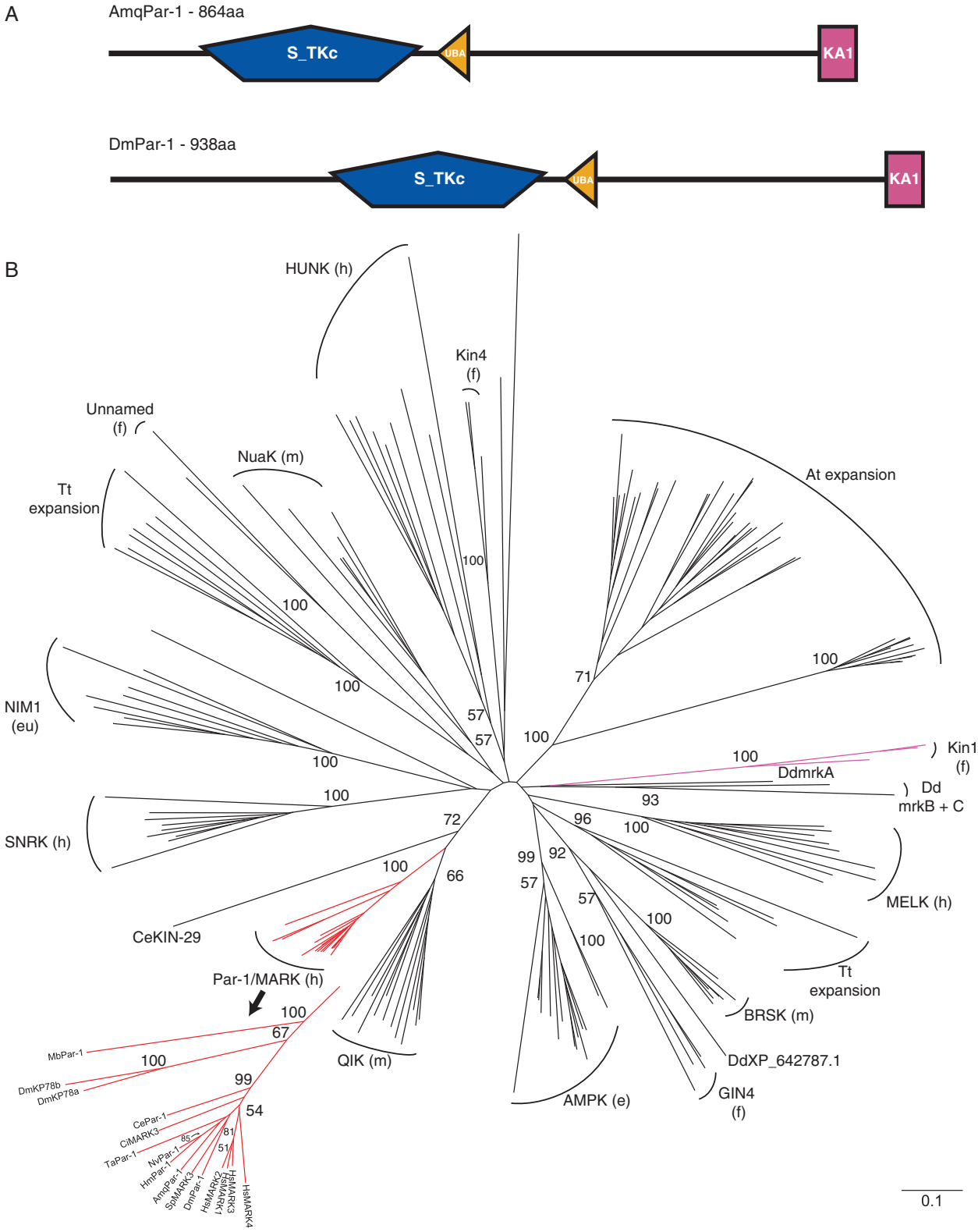
The *Amphimedon* genome contains a gene with similarity to the bilaterian Crumbs family of transmembrane cell polarity determining proteins. The predicted extracellular region of this protein contains a large array of LamG and epidermal growth factor (EGF) repeats, while the short cytoplasmic domain possesses most of the conserved residues that define a functional motif in *Drosophila* Crumbs (supporting information Fig. S4, Table S2) (Klebes and Knust 2000). Interestingly, we identified several loci with nucleotide sequence highly similar to the extreme 3′ end of the full length Crumbs coding sequence, suggesting that this region has undergone duplications in the *Amphimedon* genome. An assembly of relevant genome traces with stringent conditions indicates that there may be as many as nine genes encoding Crumbs C-termini (supporting information Fig. S4). Three of these genes are clustered within 30 kilobase pairs and lack the majority of the Crumbs extracellular region coding sequence, suggesting that they may represent either pseudogenes or genes which give rise to truncated forms of the Crumbs protein. None of the N-terminally truncated Crumbs genes possess obvious signal peptides indicating that they are unlikely to be functional at the cell surface (despite encoding the transmembrane region) (supporting information Fig. S4).

With regards to the Crumbs interacting proteins, Stardust/MPP5 and Patj (orthologous to vertebrate Mpdz and Inadl), we were able to identify a clear ortholog for Patj. AmqPatj is the only multi-PDZ protein in the *Amphimedon* genome that possesses a similar number of PDZ domains (12) to vertebrate and nematode Patj proteins (10 or 13) and the N-terminal L27 domain present in bilaterian Patj proteins (supporting information Fig. S5, Table S1). The *Amphimedon* genome encodes three membrane associated guanylate kinase (MAGUK) proteins with similarity to bilaterian MPP-related proteins (MPP1–MPP7), the wider family to which Stardust/MPP5 orthologs belong (Fig. 3, supporting information Table S1; te Velthuis et al. 2007). In a phylogenetic analysis using the Dlg MAGUK class as an outgroup, only one of these proteins, AmqMPP2/6, is placed into a well-supported clade with other metazoan MPP-related proteins (Fig. 3). Three different phylogenetic methods (distance NJ, ML, and Bayesian) failed to resolve the position of the other two genes. One of these genes, AmqMPP5/7, appeared to be most closely related to the Stardust/MPP5 and Skiff/MPP7 subfamilies and did fall at the base of a eumetazoan Stardust/MPP5 clade in some analyses. Therefore, although it cannot be stated that *Amphimedon* possesses a direct ortholog of Stardust/MPP5, it is possible that AmqMPP5/7 may be similar enough in sequence to substitute for some of the known functions of the Stardust/MPP5 protein.

An *Amphimedon* ortholog of the basolateral polarity determinant, Dlg, has been reported elsewhere (Sakarya et al. 2007) and no paralogous genes were identified in this study. *Amphimedon* possesses a single gene with the domain architecture characteristic of bilaterian Scribble/d proteins (N-terminal Leucine rich repeats followed by multiple PDZ domains), although the predicted protein possesses three instead of four PDZ domains (supporting information Fig. S6, Table S1). Two genes with similarity to bilaterian Lgl proteins were identified in *Amphimedon* (supporting information Fig. S7). Phylogenetic analysis indicates that these genes may be lineage-specific duplicates, with both genes falling at the base of a supported metazoan Lgl clade.

Cell junction genes

The fact that the *Amphimedon* genome contains orthologs of the major components of bilaterian adherens junctions is known from previous studies (Sakarya et al. 2007; Abedin and King 2008; Adamska et al. 2010). Here, we have catalogued the complete complement of *Amphimedon* cadherins and searched for the presence of the cytoplasmic catenin-binding domain, which confers the ability to form linkages between sites of cell contact and the actin-based cytoskeleton. We confirm that of the 17 cadherin genes encoded by the


A
AmqPar-1 - 864aa
S_TKc
UBA
KA1
DmPar-1 - 938aa
S_TKc
UBA
KA1
B
HUNK (h)
Kin4 (f)
Unnamed (f)
NuaK (m)
Tt expansion
At expansion
NIM1 (eu)
SNRK (h)
CeKIN-29
Par-1/MARK (h)
QIK (m)
AMPK (e)
Kin1 (f)
DdmrkA
Dd mrkB + C
MELK (h)
Tt expansion
BRSK (m)
DdXP_642787.1
GIN4 (f)
MbPar-1
DmKP78b
DmKP78a
CePar-1
CiMARK3
TaPar-1
NvPar-1
HmPar-1
AmqPar-1
SpMARK3
DmPar-1
HsMARK2
HsMARK1
HsMARK3
HsMARK4
0.1

Amphimedon genome, only the previously reported sequence (here referred to as AmqCadherin1) contains the catenin-binding domain (Table 1; supporting information Tables S2 and S3). We detected a single additional α-catenin-related gene in the *Amphimedon* genome not reported in previous studies. AmqαCatenin1 is the protein most closely related to bilaterian α-catenin, with phylogenetic analyses placing it at the base of a well-supported clade containing eumetazoan α-catenin-like proteins (Fig. 4, supporting information Fig. S8). AmqαCatenin1 contains a stretch of sequence in the middle of the Vinculin domain that is absent from bilaterian α-catulin proteins, and for this reason we have designated it as α-catenin, despite the fact that it may be directly orthologous to both families. AmqαCatenin2 groups with related proteins from *Trichoplax* and *Nematostella* in a clade, which lacks bilaterian representatives. This clade appears to be more closely related to the Vinculin family—of which there is an *Amphimedon* representative—than to α-catenin and may represent a distinct family that has been lost in all or multiple bilaterian lineages.

As expected, we did not find any *Amphimedon* orthologs of proteins that contribute to vertebrate tight junctions. However, profile-based searches did allow for the identification of a member of the wider PMP-22/EMP/MP-20/Claudin superfamily to which vertebrate Claudins and nonvertebrate Claudin-related proteins belong. AmqClaudinSF contains the four transmembrane regions and conserved motif that characterize the superfamily and is of a similar length to bilaterian PMP-22/EMP/MP-20/Claudin proteins (supporting information Fig. S9). Outside of the conserved motif, members of this family display little sequence similarity to one another, making it difficult to make conclusions regarding orthology. Some *Drosophila* Claudin-related proteins have been claimed to represent direct orthologs of vertebrate Claudins (Behr et al. 2003; Wu et al. 2004) and at least one of these sequences does possess a PDZ-binding motif at the C-terminus that is conserved in the Claudin family. AmqClaudinSF lacks a C-terminal PDZ-binding motif and does not display greater similarity to Claudins than to other PMP-22/EMP/MP-20/Claudin proteins.

No clear orthologs of the major adhesive proteins of nonvertebrate bilaterian septate junctions were located in the *Amphimedon* genome. No known *Amphimedon* predicted proteins possess the domain architecture characteristic of the Neurexin IV/CASPR family. Similarly, none of the many predicted immunoglobulin (IG) and fibronectin type III (FN3) domain containing cell adhesion molecules (CAMS) in the *Amphimedon* genome were considered similar enough in terms of sequence similarity and domain architecture to qualify as clear orthologs of bilaterian Contactin or Neuroglian CAMS.

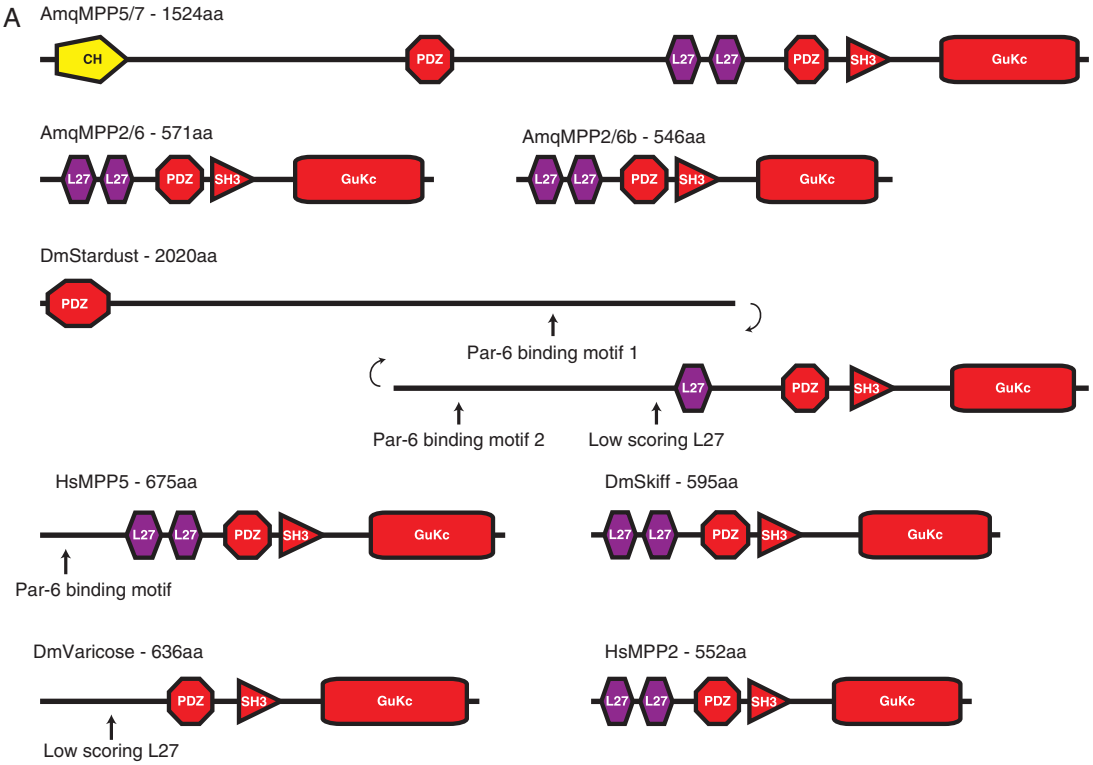
Basal lamina genes

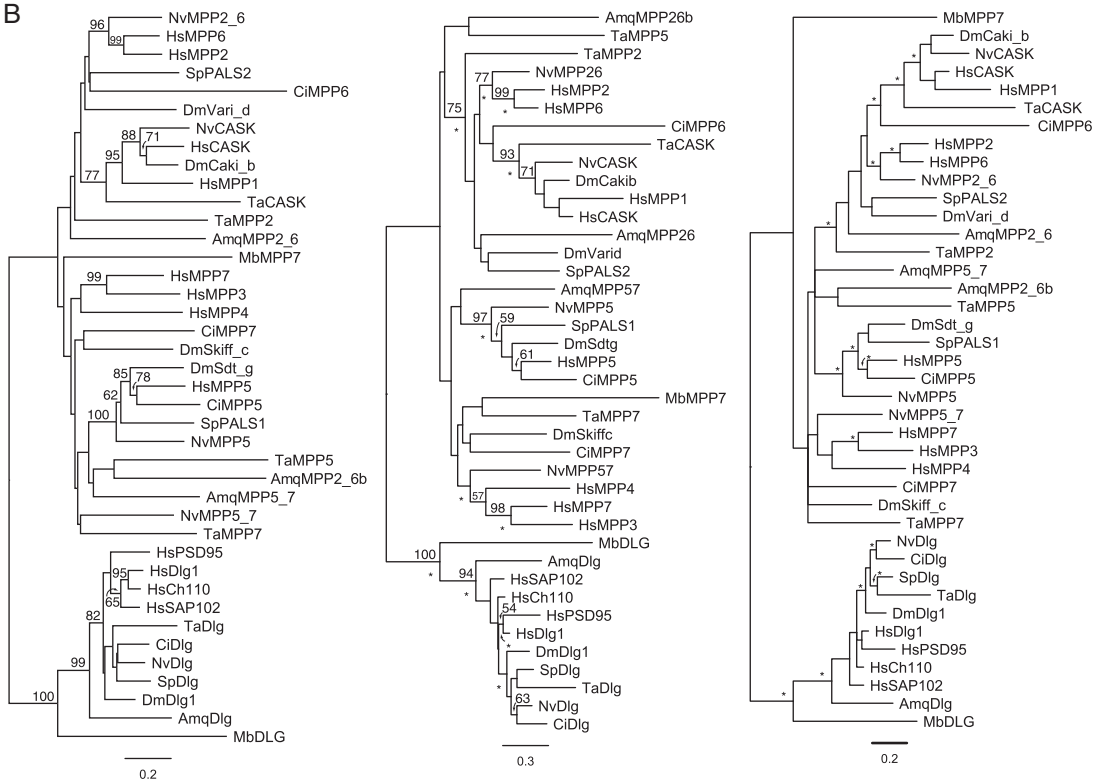
Finally, of the four surveyed basal lamina components (collagens IV and XV/XVIII, laminins α, β, and γ, nidogen, and perlecan) we were only able to detect genes encoding for laminin-related proteins in the *Amphimedon* genome. These encoded predicted secreted proteins with various combinations of laminin domains (Laminin N-terminal, Laminin IVA, Laminin IVB, Laminin α 3/5, Laminin-type EGF, and LamG) and a C-terminal region with similarity to the coiled-coil trimerization region of bilaterian laminins (supporting information Fig. S10, Table S2). Interestingly, none of these proteins possesses a domain architecture that is directly comparable to one of the four well-characterized bilaterian laminin isoforms (called isoforms but encoded by different genes), α 1/2, α 3/5, β, and γ (Srivastava et al. 2010). The architecture of AmqLamγ-like is similar to that of bilaterian laminin γ, however, it possesses a short cysteine and glycine rich stretch in the middle of the coiled-coil region, which strongly resembles the Laminin β-knob domain specific to bilaterian laminin β proteins (Beck et al. 1993; Tzu and Marinkovich 2008). In addition, phylogenetic analyses provide no support for an orthologous relationship between AmqLamγ-like and eumetazoan laminin γ proteins (data not shown).

Comparative genomic analyses reveal that genes contributing to epithelial structure in bilaterians were predominantly metazoan or eumetazoan innovations

In order to track the evolutionary history of the genes shown in Fig. 1 we searched for orthologs in the genomes of

Fig. 2. Structure and molecular phylogeny of Par-1. (A) SMART domain diagrams for AmqPar-1 and *Drosophila* Par-1. S_TKc, Serine/Threonine protein kinases, catalytic domain; UBA, Ubiquitin associated domain; KA1, Kinase associated domain 1. (B) Phylogenetic analysis of the CAMKL kinase family based on alignment of the kinase domain. At key nodes are shown the percentage of bootstrap support by distance neighbor-joining (NJ) with 1000 replicates. Bootstrap values >50% are shown. Text labels indicate subfamilies (nomenclature according to KinBase), species-specific expansions and single sequence names, with the letters in brackets indicating the phylogenetic distribution of sequences within a labeled clade: e, eukaryotic; f, fungal; h, holozoan; m, metazoan; eu, eumetazoan. Relationships within the Par-1/MARK clade are depicted in greater detail in the bottom left of the figure, with all bootstrap values >50% shown. Amq, *Amphimedon queenslandica*; At, *Arabidopsis thaliana*; Ce, *Caenorhabditis elegans*; Ci, *Ciona intestinalis*; Dd, *Dictyostelium discoideum*; Dm, *Drosophila melanogaster*; Hm, *Hydra magnipapillata*; Hs, *Homo sapiens*; Mb, *Monosiga brevicollis*; Nv, *Nematostella vectensis*; Sp, *Strongylocentrotus purpuratus*; Ta, *Trichoplax adhaerens*; Tt, *Tetrahymena thermophila*. Other species included in the analysis but not labeled on the tree include *Neurospora crassa*, *Saccharomyces cerevisiae*, and *Schizosaccharomyces pombe*.


A
AmqMPP5/7 - 1524aa
CH
PDZ
L27
L27
PDZ
SH3
GuKc
AmqMPP2/6 - 571aa
AmqMPP2/6b - 546aa
DmStardust - 2020aa
Par-6 binding motif 1
Par-6 binding motif 2
Low scoring L27
HsMPP5 - 675aa
Par-6 binding motif
DmSkiff - 595aa
DmVaricose - 636aa
Low scoring L27
HsMPP2 - 552aa


B
NvMPP2_6
HsMPP6
HsMPP2
SpPALS2
CiMPP6
DmVari_d
NvCASK
HsCASK
DmCaki_b
HsMPP1
TaCASK
TaMPP2
AmqMPP2_6
MbMPP7
HsMPP7
HsMPP3
HsMPP4
CiMPP7
DmSkiff_c
DmSdt_g
HsMPP5
CiMPP5
SpPALS1
NvMPP5
TaMPP5
AmqMPP2_6b
AmqMPP5_7
NvMPP5_7
TaMPP7
HsPSD95
HsDlg1
HsCh110
HsSAP102
TaDlg
CiDlg
NvDlg
SpDlg
DmDlg1
AmqDlg
MbDLG
0.2
AmqMPP26b
TaMPP5
TaMPP2
NvMPP26
HsMPP2
HsMPP6
CiMPP6
TaCASK
NvCASK
DmCakib
HsMPP1
HsCASK
AmqMPP26
DmVarid
SpPALS2
AmqMPP57
NvMPP5
SpPALS1
DmSdtg
HsMPP5
CiMPP5
MbMPP7
TaMPP7
DmSkiffc
CiMPP7
NvMPP57
HsMPP4
HsMPP7
HsMPP3
MbDLG
AmqDlg
HsSAP102
HsCh110
HsPSD95
HsDlg1
DmDlg1
SpDlg
TaDlg
NvDlg
CiDlg
0.3
MbMPP7
DmCaki_b
NvCASK
HsCASK
HsMPP1
TaCASK
CiMPP6
HsMPP2
HsMPP6
NvMPP2_6
SpPALS2
DmVari_d
AmqMPP2_6
TaMPP2
AmqMPP5_7
AmqMPP2_6b
TaMPP5
DmSdt_g
SpPALS1
HsMPP5
CiMPP5
NvMPP5
NvMPP5_7
HsMPP7
HsMPP3
HsMPP4
CiMPP7
DmSkiff_c
TaMPP7
NvDlg
CiDlg
SpDlg
TaDlg
DmDlg1
HsDlg1
HsPSD95
HsCh110
HsSAP102
AmqDlg
MbDLG
0.2

multiple fungi, a choanoflagellate, a placozoan, and a cnidarian, and compared our results with published data for representative bilaterians (Fig. 5). Our analyses yield a minimal estimate of the ancestral gene complement at each of the nodes depicted in Fig. 5. It is important to note that for genomes other than *Amphimedon* we analyzed only the best matches for unambiguously identified orthologs, and that therefore accompanying sequences (supporting information Files S2–S4) do not represent complete catalogues of the relevant genes from each genome.

We did not find any direct orthologs of the genes from our list in fungal genomes. Fungi have been reported to possess homologs of Par-1 (Drewes et al. 1997; Drewes and Nurse 2003; Elbert et al. 2005), however, our phylogenetic analysis of the kinase domain does not recover a sister group relationship between these proteins and holozoan Par-1 representatives (Fig. 2B). Fungal kin1 and kin2 proteins do share a C-terminal regulatory domain with animal Par-1 proteins (the KA1 domain) but this domain is also found in related kinases that fall into distinct subfamilies of the CAMKL family (e.g., *Arabidopsis* AKIN10 and human Maternal embryonic leucine zipper kinase) (Fig. 2). In our analysis, fungal kin1 and kin2 proteins do not group within any clear subfamily and it remains possible that their actual phylogenetic relationships have been obscured by considerable levels of sequence divergence within the kinase domain. Fungi also possess a Lgl/Tomosyn-related gene (two genes in *Saccharomyces cerevisiae*). In a phylogenetic analysis of this family, the fungal proteins and all other nonmetazoan Lgl/Tomosyn proteins are excluded from two well-supported clades representing the metazoan Lgl and Tomosyn families (supporting information Fig. S7). These data suggest that the metazoan Lgl and Tomosyn families arose through duplication from a single precursor before metazoan cladogenesis, as has been suggested previously (Kloepper et al. 2008).

The *M. brevicollis* genome contains only two genes that are direct orthologs of genes from our target list, *MbPar*-1 and *MbDlg*. Both show strong similarity to representatives from the corresponding animal gene family and were confidently grouped with members of that family in phylogenetic analyses (Figs. 2B and 3B). The *Monosiga* genome also encodes several genes related to our genes of interest, namely *MbMPP7*, *MbLgl*/*Tomosyn*, *MbVinculin*, and *MbLaminin-like*. Phylogenetic analyses failed to resolve the position of MbMPP7, with some methods placing it at the base of all metazoan MPP-related proteins and others placing it within the Skiff/MPP7 subfamily or basal to a clade containing Skiff/MPP7 and Stardust/MPP5 sequences (Fig. 3B). As with other nonmetazoan Lgl/Tomosyn-related proteins, MbLgl/Tomosyn was excluded from both the metazoan Lgl and Tomosyn clades and did not possess greater affinity to either (supporting information Fig. S7). In our phylogenetic analysis of the Vinculin and α-catenin family, MbVinculin grouped with metazoan Vinculin proteins with good support (Fig. 4). The only other nonmetazoan Vinculin/α-catenin-related protein identified in Genbank, *Dictyostelium discoideum* vinculin A, did not cluster with the Vinculin family and therefore it cannot be concluded that metazoan Vinculin and α-catenin proteins arose from a Vinculin-like precursor. Finally, MbLaminin-like resembles metazoan laminin proteins in its possession of an N-terminal LamNT domain, a series of LamEGF repeats and a sequence with likelihood for coiled-coil formation (mediates laminin subunit trimerization) (Srivastava et al. 2010). However, as with the laminin-related proteins identified in *Amphimedon*, MbLaminin-like is not comparable to any of the bilaterian laminins in terms of domain architecture.

A comparison of the gene set from *Amphimedon* with that from fungi and *Monosiga* reveals genes or families that are likely to have arisen along the metazoan stem. Unambiguous metazoan innovations include Par-3, Par-6, aPKC, Crumbs, Patj, Scribble/d, classic/adherens junction type cadherin, and p120/δ-catenin. In addition, Lgl and α-catenin appear to be metazoan-specific families that arose through the duplication and divergence of premetazoan precursors.

Finally, our results suggest that Stardust/MPP5, Neurexin IV/CASPR, Neuroglian, collagen XV/XVIII, laminin α, laminin β, laminin γ, nidogen, and perlecan are all eumetazoan

Fig. 3. *Amphimedon* membrane palmitoylated protein (MPP)-related proteins. (A) SMART domain diagrams for *Amphimedon* MPP-related proteins and relevant MPP-related proteins from *Drosophila* and humans. *AmqMPP2/6b* is so named because it is located adjacent to the MPP2/6-related gene, *AmqMPP2/6*, on a contig. It is unclear whether the Calponin homology (CH) and PDZ domains at the N-terminus of AmqMPP5/7 belong to the same gene or to an adjacent gene. One isoform of *Drosophila* Stardust (Isoform G) also encodes an N-terminal PDZ (pictured). Pictured *Drosophila* proteins represent single orthologs for each of the MPP5 (Stardust), MPP7 (Skiff), and MPP2/6 (Varicose) subfamilies. In cases where the domain architecture of the *Drosophila* protein is not representative of the subfamily, one of the human paralogs is also displayed. L27, domain in receptor targeting proteins Lin-2 and Lin-7; SH3, Src homology 3 domains; GuKc, guanylate kinase domain. (B) Phylogenetic analyses of the MPP-related MAGUK class based on alignment of the SH3 and guanylate kinase (GUK or GuKc) domains. Trees generated by distance neighbor-joining (NJ) (left), maximum likelihood (ML) (middle), and Bayesian (right) analyses are shown. Trees were rooted using the Discs large (Dlg) class as the outgroup. For each node, values above the branch represent the percentage of bootstrap support obtained by distance NJ with 1000 replicates or by ML with 100 replicates. For the ML tree, asterisks represent aLRT values (SH-like method) >0.9. For the Bayesian tree, asterisks represent posterior probabilities >0.95. For *Monosiga*, *Trichoplax*, and *Nematostella* MPP-related proteins, sequence names were based on initial observations of BLAST similarity and may not accurately reflect the observed placement in the phylogenetic analyses. Refer to Fig. 2 legend for species name abbreviations.

Table 1. *Amphimedon* Cadherin domain containing genes

Gene model[1]	Domains[2,3]	Top BLAST hit
g13768.t1 (edited) (Aqu1.212079)	SignalP, 14 × Cadherin, 13 × EGF, 2 × LamG, TM, Cadherin_C	Vertebrate FAT tumor suppressor 1
snap.38132 (Aqu1.216652)	SignalP, very many Cadherin, 2 × EGF, 2 × LamG, TM	Vertebrate FAT tumor suppressor homolog 1
ren.15543 (Aqu1.217020)	Hedgehog signal, VWA, 12 × Cadherin, IGc2, IG, 2 × EGF, TM	*Nematostella* FAT Tumor suppressor 4 like
snap.54997 (Aqu1.224307)	5–6 × Cadherin, 3 × EGF, 2 × LamG, LamEGF, GPCR family 2 extracellular domain (HRM), GPS, 7 × TM	Insect Protocadherin-like wing polarity protein stan precursor (Protein starry night) (Protein flamingo)
Aqu1.227599	7 × Cadherin, TM	Vertebrate cadherin, EGF LAG seven-pass G-type receptor 2
aq_ka13411x00790 (Aqu1.220982)	SignalP, 10 × Cadherin, TM	*Nematostella* FAT tumor suppressor homolog 4 like
aq_ka13273x00330 (Aqu1.216632)	5 × Cadherin, 2 × LamG, EGF, TM	Vertebrate protocadherin 2 alpha b 5
aq_ka13433x00220 (Aqu1.221884)	SignalP, very many Cadherin, TM	Vertebrate FAT tumor suppressor homolog 4
snap.61946 (Aqu1.227600)	SignalP, very many Cadherin, TM	Vertebrate FAT tumor suppressor homolog 4
ren.15682 (Aqu1.225106)	SignalP, very many Cadherin, TM	Vertebrate FAT tumor suppressor homolog 4
aq_ka13478x00870 (Aqu1.224598)	Very many Cadherin, TM	Vertebrate cadherin-like 23
aq_ka13434x00540 (Aqu1.221984)	SignalP, very many Cadherin, TIG, FN3, TM, ZF-MYND	Vertebrate Fat tumor suppressor homolog 4
aq_ka13448x00660	SignalP, very many Cadherin, 2 × EGF	*Nematostella* Fat4 like
aq_ka13514x00590 (Aqu1.228397)	Very many Cadherin, TM	Vertebrate Fat tumor suppressor homolog 4
aq_ka13490x00420 (Aqu1.225649)	SignalP, very many Cadherin, 2 × LamG	Vertebrate Fat tumor suppressor homolog 4
aq_ka13423x00300 (Aqu1.221460)	EGF, very many Cadherin, TM	*Nematostella* Fat4 like
snap_lcmask.38703 (Aqu1.222292)	SignalP, 2 × Cadherin, IG, FN3, TM	Vertebrate protocadherin 19

[1]Gene model identifiers are listed for the chosen gene model and for the corresponding Aqu1 model in brackets.
[2]Domains and features are listed in order of appearance in the protein from N-terminus to C-terminus. The lists do not accurately reflect the domain architecture of each protein.
[3]SignalP, N-terminal signal peptide for secretion.
TM, transmembrane domain.

innovations. Although not found in *Amphimedon*, a collagen IV gene has been found in the homoscleromorph sponge, *Pseudocorticium jarrei*, suggesting that the origin of this type of collagen predates the Eumetazoa (Boute et al. 1996). Proposed sponge paraphyletic phylogenies have the homoscleromorph lineage distinct from other demosponges and more closely related to eumetazoans (Sperling et al. 2009), raising the possibility that the collagen IV gene is a eumetazoan+homoscleromorph (i.e., epitheliozoan) synapomorphy. Although we were able to locate putative orthologs for the majority of polarity, adherens junction, septate junction, and basal lamina genes in *Nematostella*, we were unable to detect an unambiguous *Nematostella* Contactin gene. This was also the case with *Hydra* and *Trichoplax* genomes. All three genomes encode a large number of IG and FN3 domain containing CAMs but none were considered similar enough to Contactin to pass as a clearly identified ortholog for that gene. Considering that cnidarians are thought to possess septate junctions (Magie and Martindale 2008), it is possible that some other CAM is able to substitute for the function of bilaterian Contactin in these organisms.

Epithelial genes were predominantly assembled from preexisting domains with the occasional incorporation of novel domains and motifs

We further investigated the evolution of our genes of interest by classifying each gene on the basis of mode of origin (supporting information Fig. S11). We used the classification system defined in Putnam et al. (2007) for the description of novel eumetazoan genes. Briefly, Type I novelties represent entirely novel sequences, Type II novelties represent the combining of ancestral and new domains, and Type III novelties represent new domain architectures assembled from ancestral domains. We expanded these categories to include Type 0 and Type IV. Type 0 novelties represent genes and families that are not distinguished by unique domain architectures and that are likely to have arisen through duplication and divergence of similar ancestral genes. Type IV novelties are a special case in which the representatives from early-branching metazoan lineages are recognizable as orthologs of characterized bilaterian genes but contain additional or missing domains.

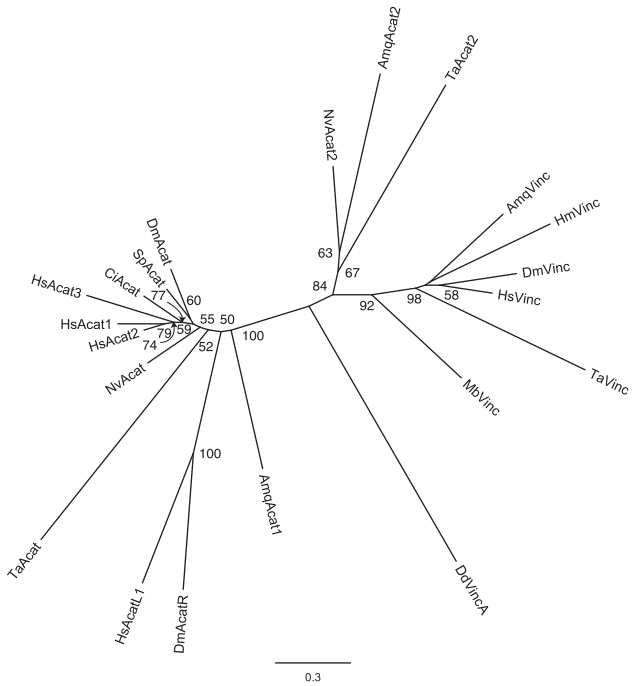


Fig. 4. Phylogenetic analysis of α-catenin- and Vinculin-related proteins based on alignment of conserved regions within the Vinculin domain. Values at each node represent the percentage of bootstrap support obtained by distance neighbor-joining (NJ) with 1000 replicates. Refer to Fig. 2 legend for species name abbreviations.

The majority of genes analyzed here (all except Neurexin IV/CASPR) were classified as Type 0, Type II, or Type III novelties. Evolution of these gene families involved the incorporation of preexisting ancestral domains with novelty arising from new domains and motifs, new domain combinations or sequence differentiation.

Only Type II novelties involve the addition of novel domains. Type II gene families include those such as Par-3 and Crumbs for which novel family-specific domains or motifs differentiate them from genes with similar domains. Other Type II gene families originated when evolutionarily novel domains were assembled with preexisting domains into new combinations. For example, while *Amphimedon* possesses all of the constituent domains for perlecan except for the eumetazoan-specific SEA domain (domain found in *s*ea urchin sperm protein, *e*nterokinase, *a*grin), these domains are not found together in a perlecan-like gene.

Type 0 novelties include genes, which evolved through duplication and divergence, like Stardust/MPP5 and Lgl, for which domain architectures are indistinguishable from those of paralogous families. They also include genes like p120/δ-catenin and Neuroglian that probably evolved through duplication and sequence divergence from an ancestral gene with a similar architecture.

Genes such as Par-6 and aPKC, in which ancestral domains are combined together for the first time, represent typical Type III novelties. Also included as a Type III novelty is the Dlg family. Although Dlg shares a set of domains with MPP-related proteins, it possesses a unique number and arrangement of these domains.

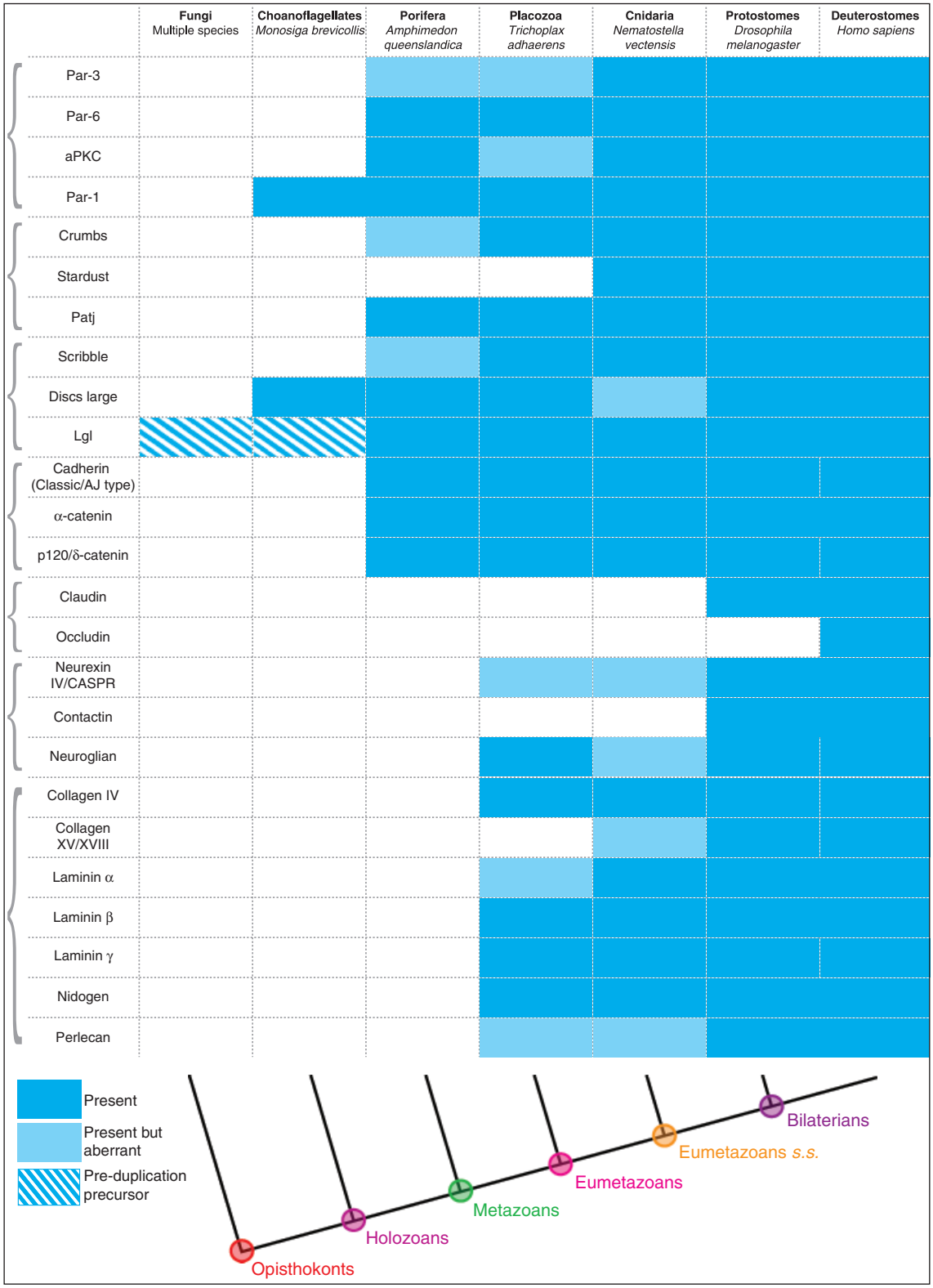
We identified only one case of a Type IV novelty. Neurexin IV/CASPR orthologs from both *Nematostella* and *Trichoplax* genomes lack an N-terminal Discoidin (FA58C) domain but resemble bilaterian Neurexin IV orthologs in all other respects. The Discoidin domain is an extracellular carbohydrate-binding domain and therefore its absence is likely to affect the adhesive properties of the protein.

DISCUSSION

Demosponges encode molecular components necessary for epithelial cell adhesion and polarity

In comparison with a typical bilaterian epithelium, *Amphimedon* possesses a near-complete complement of the known proteins that act to establish cellular apical–basal polarity and form adherens junctions. Despite sponge and eumetazoan lineages having diverged over 600 million years ago, many of the genes have maintained conserved sequence elements that have been shown to be necessary for protein function, localization, and protein–protein interactions in model bilaterians. This suggests that these orthologs, pathways, and complexes have the potential to operate in a conserved fashion in *Amphimedon*. However, it would be premature to conclude on the basis of sequence data alone that the *Amphimedon* genes identified here must be involved in establishing cell polarity and forming junctional structures resembling the zonula adherens of bilaterian epithelial cells. As with many metazoan developmentally regulated genes, members of the Crumbs, Par, and Scribble/d complexes, have multiple context-dependent roles in the animals in which they have been studied. The majority of these roles are indeed polarity-related (e.g., the Par complex in asymmetric division of the oocyte in *C. elegans*, the Par complex and Lgl in asymmetric division of neuroblast precursors in *Drosophila*, the Crumbs complex in polarity of photoreceptor cells in *Drosophila* and vertebrates), however, without data from nonbilaterian outgroups, it is impossible to know which were the ancestral functions and which are derived (Omori and Malicki 2006; Suzuki and Ohno 2006). Similarly, the proteins that form the zonula adherens of bilaterian epithelial cells, can give rise to spot-form adherens junctions in nonepithelial cells like neurones (synaptic junctions) and fibroblasts (Yagi and Takeichi 2000; El Sayegh et al. 2007).

We failed to identify a clear ortholog of the Crumbs interacting protein, Stardust/MPP5, in *Amphimedon* by phylogenetic analysis. The *Amphimedon* protein that appears most


Fungi
Multiple species
Choanoflagellates
Monosiga brevicollis
Porifera
Amphimedon queenslandica
Placozoa
Trichoplax adhaerens
Cnidaria
Nematostella vectensis
Protostomes
Drosophila melanogaster
Deuterostomes
Homo sapiens
Par-3
Par-6
aPKC
Par-1
Crumbs
Stardust
Patj
Scribble
Discs large
Lgl
Cadherin (Classic/AJ type)
α-catenin
p120/δ-catenin
Claudin
Occludin
Neurexin IV/CASPR
Contactin
Neuroglian
Collagen IV
Collagen XV/XVIII
Laminin α
Laminin β
Laminin γ
Nidogen
Perlecan
Present
Present but aberrant
Pre-duplication precursor
Opisthokonts
Holozoans
Metazoans
Eumetazoans
Eumetazoans s.s.
Bilaterians

closely related to the Stardust/MPP5 family, AmqMPP5/7, possesses the domains which mediate binding to Patj, Lin-7, and Crumbs in bilaterian Stardust/MPP5 proteins (Fig. 3; Bit-Avragim et al. 2008). However, these do not appear obviously more similar in sequence to Stardust/MPP5 proteins than to the same domains found in Skiff/MPP7 proteins. The protein also lacks the N-terminal Par-6 binding motif that is conserved among bilaterian Stardust/MPP5 proteins (Fig. 3; Penkert et al. 2004). In summary, although it is possible that one of the *Amphimedon* MPP5/7-related proteins may substitute for the functions of Stardust/MPP5 in apical–basal polarity determination, there is no clear evidence based on sequence data to suggest this.

In contrast to *Trichoplax* and *Nematostella*, we were unable to find convincing *Amphimedon* orthologs for the majority of septate junction and basal lamina genes. These results are consistent with the fact that these structures have not been widely observed by electron microscopy in demosponges. Reports of septate-like junctions in demosponges involve various cell types (Leys et al. 2009), and are often unconvincing. Clear septate junctions have been observed in calcareous sponges, where they appear to be involved in sealing between cells involved in extracellular spicule calcite deposition (Ledger 1975). As discussed previously, convincing evidence for the presence of a basal lamina is also restricted to certain taxa within the Porifera, in this case the homoscleromorphs, which are traditionally assigned as a group within the Demospongiae. The evolutionary significance of the absence of septate junctions, basal lamina, and their associated molecular components in *Amphimedon* hinges on the phylogenetic position of Calcarea and Homoscleromorpha with respect to other demosponges. If the Porifera is monophyletic, as most recently proposed by Philippe et al. (2009), then the absence of these traits may be derived in *Amphimedon* and most other demosponge clades. If these sponges are paraphyletic (e.g., Sperling et al. 2009), then these may represent characters that evolved in the clade containing the Calcarea, Homoscleromorpha, and Eumetazoa after it diverged from the lineage comprising the majority of demosponges. Whatever the evolutionary implications of these results, we can present a conservative estimate of the gene complement of the last common ancestor (LCA) to all extant animals and a working estimate of the total genes present in *Amphimedon*.

Despite lacking some key basal lamina components, *Amphimedon* does possess a set of unique laminin-related genes. Although none of these are quite reconcilable with the laminin α, β, and γ forms of bilaterians, it appears that they may be able to form similar heterotrimer structures with one another. Conserved cysteine residues at the N- and C-termini of the coiled-coil regions are in comparable positions to those that appear to form interchain disulfide bonds in vertebrate laminins (Beck et al. 1993). An *Amphimedon* laminin heterotrimer could interact with cell surface receptors like integrin and dystroglycan, which are both present in *Amphimedon*, through the LamG domains at the C-terminus of the AmqLamα3/5-like chain (Miner and Yurchenco 2004; Srivastava et al. 2010). However, it is possible that the lack of a well-conserved LamNT domain in the α 3/5-like chain may negatively affect the ability of the heterotrimer to self-polymerize into a stable supportive network (Cheng et al. 1997).

In summary, we find that the demosponge, *Amphimedon*, contains the necessary molecular components to form polarized layers of cells with stabilizing adherens junctions, but does not contain the machinery needed to make structures homologous with the occluding junctions or basal lamina of bilaterian epithelia. These data suggest that epithelial-like layers of *Amphimedon* and other demosponges may share some but not all of the conserved characteristics of eumetazoan epithelia. It remains to be explored whether sponge tissues make use of a unique set of proteins and subcellular structures to achieve functions such as sealing and integration.

Origins of epithelial genes and domains

Genome surveys conducted on representatives from fungi, choanoflagellates, placozoans, and cnidarians, allowed us to generate hypotheses on the points of origin for key epithelial genes. We stress that these hypotheses remain tentative because instances of lineage-specific gene loss, as well as the possibility of gaps in genome sequencing projects, may confound attempts to accurately reconstruct the history of genes and their domains.

Fig. 5. Evolutionary origin of bilaterian epithelial genes. The distribution of orthologs of bilaterian epithelial genes is shown superimposed on a species tree for the organisms in which genomes were surveyed. Brackets on the left represent functional interrelationships between the proteins as indicated by studies in bilaterian model organisms. Aberrant orthologs are defined as those with obvious minor differences in domain composition or domain numbers with respect to bilaterian representatives or those with gaps in underlying genome contigs that compromise the prediction. In summary: AmqPar-3, gaps in prediction in PDZ encoding region; TaPar-3, missing one of three PDZ domains; TaaPKC, gap in prediction in presumed protein kinase C conserved region 1 (C1) encoding sequence; AmqCrumbsC2, extra LamG domain at N-terminus; AmqScribble, missing one of four PDZ domains; NvDlg, extra SH3 domain; TaNeurexinIV and NvNeurexinIV, missing N-terminal Discoidin domain; NvNeuroglian, missing one of six IG domains; NvCollagenXV/XVIII, missing a LamG/Thrombospondin N-terminal-like (TSPN) domain; TaLamα1/2-like, missing two of five C-terminal LamG domains; TaPerlecan, missing one of three C-terminal LamG domains; NvPerlecan, in two fragments that can't be assembled (both located at the end of their respective contigs). Eumetazoans *s.s.*, eumetazoans *sensu strictu* (Bilateria+Cnidaria).

In general the genes surveyed here were found to be metazoan-specific, with only Par-1 and Dlg, observed to predate the divergence of choanoflagellate and metazoan lineages. Almost all genes contained domains that antedate the Metazoa, with novelty being generated through sequence divergence, domain shuffling, and addition of novel metazoan- or eumetazoan-specific domains (supporting information Fig. S11). Premetazoan domains included protein–protein interaction and signaling domains such as PDZ and Serine/Threonine Kinase as well as cell adhesion and extracellular matrix-related domains such as Cadherin, LamG, EGF, and IG.

The fact that most of the polarity genes evolved after the divergence of metazoan and choanoflagellate lineages suggests that animals use a largely novel set of proteins for generating and regulating cell polarity. However, these proteins undoubtedly made use of preexisting mechanisms for generating polarity. For example, Par-6 binds to and requires Cdc42, an opisthokont-specific protein that is involved in generation of cell polarity in yeast (Etienne-Manneville 2004).

Our results indicate that the majority of genes included in our analysis were already in place before the divergence of the early eumetazoan lineages, Placozoa and Cnidaria, from the stem leading to the crown bilaterians, indicating that animals within these groups contain most of the necessary components needed to make true epithelial tissues. This is consistent with observations of ultrastructure and immunochemical reactivity in cnidarians (Thomas and Edwards 1991; Sarras et al. 1994; Shimizu et al. 2008). The fact that the *Trichoplax* genome has maintained a copy of all the main basal lamina genes, suggests that these are used at some stage of the life cycle, despite the fact that no basal lamina or extracellular matrix of any kind has been observed in adult forms (Grell and Ruthmann 1991). Although we did not find all components of septate junctions in the genome of *Nematostella* and *Trichoplax*, it is possible that a related molecule may substitute for the function of the missing genes in the septate junctions of these organisms. The absence of a Contactin gene in these genomes and of the N-terminal Discoidin domain in the *Nematostella* and *Trichoplax* Neurexin IV/CASPR orthologs, suggests that the adhesive and barrier properties of cnidarian and placozoan septate junctions may be quite different to those of bilaterians.

CONCLUSIONS

As an outgroup to the Eumetazoa, sponges are uniquely placed to provide insight into the evolution of epithelial tissues and other shared derived eumetazoan traits. As with all other metazoans, morphogenesis, and morphostasis of the sponge body plan appears to be reliant on entrained cell–cell and cell–matrix interactions largely mediated by proteins on the cell surface. Sponges clearly possess cohesive and integrated cell layers in both larval and adult forms. With the *Amphimedon* genome encoding components to (i) confer apical–basal cellular polarity, (ii) build adherens junctions, (iii) synthesise a collagen-based extracellular matrix (Exposito et al. 2008) consisting of proteoglycans (Fernandez-Busquets and Burger 2003; Srivastava et al. 2010) and a suite of laminin-like factors, and (iv) anchor to the extracellular matrix (e.g., integrins), we infer that the metazoan LCA had the capacity to form cell layers that functioned in a manner similar to extant epithelia. Nonetheless, it remains unclear if all aspects of sponge epithelial-like layers can be considered homologous to eumetazoan epithelia or whether they represent a parallel approach to the problem of integration and compartmentalization.

Acknowledgments

This work was supported by Australian Research Council grants 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

Abascal, F., Zardoya, R., and Posada, D. 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21: 2104–2105.

Abedin, M., and King, N. 2008. The premetazoan ancestry of cadherins. *Science* 319: 946–948.

Adamska, M., et al. 2010. Structure and expression of conserved Wnt pathway components in the demosponge *Amphimedon queenslandica*. *Evol. Dev.* 12: 494–518.

Altschul, S. F., et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389–3402.

Beck, K., Dixon, T. W., Engel, J., and Parry, D. A. 1993. Ionic interactions in the coiled-coil domain of laminin determine the specificity of chain assembly. *J. Mol. Biol.* 231: 311–323.

Behr, M., Riedel, D., and Schu, R. 2003. The claudin-like megatrachea is essential in septate junctions for the epithelial barrier function in *Drosophila*. *Dev. Cell* 5: 611–620.

Benton, R., and St Johnston, D. 2003. A conserved oligomerization domain in *Drosophila* Bazooka/PAR-3 is important for apical localization and epithelial polarity. *Curr. Biol.* 13: 1330–1334.

Bhat, M. A. 2003. Molecular organization of axo-glial junctions. *Curr. Opin. Neurobiol.* 13: 552–559.

Bit-Avragim, N., Hellwig, N., Rudolph, F., Munson, C., Stainier, D. Y., and Abdelilah-Seyfried, S. 2008. Divergent polarization mechanisms during vertebrate epithelial development mediated by the Crumbs complex protein Nagie oko. *J. Cell Sci.* 121: 2503–2510.

Boury-Esnault, N., Ereskovsky, A. V., Bezac, C., and Tokina, D. 2003. Larval development in the Homoscleromorpha (Porifera, Demospongiae). *Invertebr. Biol.* 122: 187–202.

Boute, N., et al. 1996. Type IV collagen in sponges, the missing link in basement membrane ubiquity. *Biol. Cell.* 88: 37–44.

Chapman, J. A., et al. 2010. The dynamic genome of *Hydra*. *Nature* 464: 592–596.

Cheng, Y. S., Champliaud, M. F., Burgeson, R. E., Marinkovich, M. P., and Yurchenco, P. D. 1997. Self-assembly of laminin isoforms. *J. Biol. Chem.* 272: 31525–31532.

Cox, E. A., and Hardin, J. 2004. Sticky worms: adhesion complexes in *C. elegans*. *J. Cell Sci.* 117: 1885–1897.

de Caralt, S., Uriz, M. J., Ereskovsky, A. V., and Wijffels, R. H. 2007. Embryo development of *Corticium candelabrum* (Demospongiae: Homosclerophorida). *Invertebr. Biol.* 126: 211–219.

Deng, W. M., et al. 2003. Dystroglycan is required for polarizing the epithelial cells and the oocyte in *Drosophila*. *Development* 130: 173–184.

Drewes, G., Ebneth, A., Preuss, U., Mandelkow, E. M., and Mandelkow, E. 1997. MARK, a novel family of protein kinases that phosphorylate microtubule-associated proteins and trigger microtubule disruption. *Cell* 89: 297–308.

Drewes, G., and Nurse, P. 2003. The protein kinase kin1, the fission yeast orthologue of mammalian MARK/PAR-1, localises to new cell ends after mitosis and is important for bipolar growth. *FEBS Lett*. 554: 45–49.

Dunn, C. W., et al. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* 452: 745–749.

Elbert, M., Rossi, G., and Brennwald, P. 2005. The yeast par-1 homologs kin1 and kin2 show genetic and physical interactions with components of the exocytic machinery. *Mol. Biol. Cell* 16: 532–549.

El Sayegh, T. Y., Kapus, A., and McCulloch, C. A. 2007. Beyond the epithelium: cadherin function in fibrous connective tissues. *FEBS Lett*. 581: 167–174.

Ereskovsky, A. V., and Tokina, D. B. 2004. Morphology and fine structure of the swimming larvae of *Ircinia oros* (Porifera, Demospongiae, Dictyoceratida). *Invertebr. Reprod. Dev*. 45: 137–150.

Etienne-Manneville, S. 2004. Cdc42-the centre of polarity. *J. Cell Sci*. 117: 1291–1300.

Exposito, J. Y., Larroux, C., Cluzel, C., Valcourt, U., Lethias, C., and Degnan, B. M. 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: 28226–28235.

Felsenstein, J. 2005. *PHYLIP (Phylogeny Inference Package) Version 3.6*. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.

Feng, W., Wu, H., Chan, L. N., and Zhang, M. J. 2007. The Par-3 NTD adopts a PB1-like structure required for Par-3 oligomerization and membrane localization. *Embo. J*. 26: 2786–2796.

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

Finn, R. D., et al. 2008. The Pfam protein families database. *Nucleic Acids Res*. 36: D281–D288.

Fowler, S. J., Jose, S., Zhang, X., Deutzmann, R., Sarras, M. P. J., and Boot-Handford, R. P. 2000. Characterization of hydra type IV collagen. *J. Biol. Chem*. 275: 39589–39599.

Göransson, O., Deak, M., Wullschleger, S., Morrice, N. A., Prescott, A. R., and Alessi, D. R. 2006. Regulation of the polarity kinases PAR-1/MARK by 14-3-3 interaction and phosphorylation. *J. Cell Sci*. 119: 4059–4070.

Green, C. R., and Bergquist, P. R. 1982. Phylogenetic relationships within the Invertebrata in relation to the structure of septate junctions and the development of occluding junctional types. *J. Cell Sci*. 53: 279–305.

Grell, K. G., and Ruthmann, A. 1991. Placozoa. In F. W. Harrison and J. A. Westfall (eds.). *Microscopic Anatomy of Invertebrates*. Wiley-Liss, New York, pp. 13–27.

Guindon, S., and Gascuel, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol*. 52: 696–704.

Hortsch, M., and Margolis, B. 2003. Septate and paranodal junctions: kissing cousins. *Trends Cell Biol*. 13: 557–561.

Humbert, P. O., Dow, L. E., and Russell, S. M. 2006. The Scribble and Par complexes in polarity and migration: friends or foes? *Trends Cell Biol*. 16: 622–630.

Hutter, H., et al. 2000. Conservation and novelty in the evolution of cell adhesion and extracellular matrix genes. *Science* 287: 989–994.

Hynes, R. O., and Zhao, Q. 2000. The evolution of cell adhesion. *J. Cell Biol*. 150: F89–F95.

Izumi, Y., et al. 1998. An atypical PKC directly associates and colocalizes at the epithelial tight junction with ASIP, a mammalian homologue of *Caenorhabditis elegans* polarity protein PAR-3. *J. Cell Biol*. 143: 95–106.

Johnson, R. P., Kang, S. H., and Kramer, J. M. 2006. *C. elegans* dystroglycan DGN-1 functions in epithelia and neurons, but not muscle, and independently of dystrophin. *Development* 133: 1911–1921.

King, N., et al. 2008. The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* 451: 783–788.

Klebes, A., and Knust, E. 2000. A conserved motif in Crumbs is required for E-cadherin localisation and zonula adherens formation in *Drosophila*. *Curr. Biol*. 10: 76–85.

Kloepper, T. H., Kienle, C. N., and Fasshauer, D. 2008. SNAREing the basis of multicellularity: consequences of protein family expansion during evolution. *Mol. Biol. Evol*. 25: 2055–2068.

Knust, E., and Bossinger, O. 2002. Composition and formation of intercellular junctions in epithelial cells. *Science* 298: 1955–1959.

Larkin, M. A., et al. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947–2948.

Larroux, C., Fahey, B., Degnan, S. M., Adamski, M., Rokhsar, D. S., and Degnan, B. M. 2007. The NK homeobox gene cluster predates the origin of Hox genes. *Curr. Biol*. 17: 706–710.

Ledger, P. W. 1975. Septate junctions in the calcareous sponge *Sycon ciliatum*. *Tissue Cell* 7: 13–18.

Letunic, I., Doerks, T., and Bork, P. 2009. SMART 6: recent updates and new developments. *Nucleic Acids Res*. 37: D229–D232.

Leys, S. P. 2007. Sponge coordination, tissues, and the evolution of gastrulation. In M. R. Custodio, G. Lobo-Hajdu, E. Hajdu, and G. Muricy (eds.). *Porifera Research: Biodiversity, Innovation and Sustainability*. *[Serie Livros 28]*. Museu Nacional, Rio de Janeiro, pp. 53–59.

Leys, S. P., Nichols, S. A., and Adams, E. D. M. 2009. Epithelia and integration in sponges. *Integr. Comp. Biol*. 49: 167–177.

Li, S. H., Edgar, D., Fassler, R., Wadsworth, W., and Yurchenco, P. D. 2003. The role of laminin in embryonic cell polarization and tissue organization. *Dev. Cell* 4: 613–624.

Lin, D., Edwards, A. S., Fawcett, J. P., Mbamalu, G., Scott, J. D., and Pawson, T. 2000. A mammalian PAR-3-PAR-6 complex implicated in Cdc42/Rac1 and aPKC signalling and cell polarity. *Nat. Cell Biol*. 2: 540–547.

Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y. O., and Borodovsky, M. 2005. Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Res*. 33: 6494–6506.

Maddison, D. R., and Maddison, W. P. 2005. *MacClade 4: Analysis of Phylogeny and Character Evolution, Version 4.07*. Sinauer Associates, Sunderland.

Magie, C. R., and Martindale, M. Q. 2008. Cell–cell adhesion in the Cnidaria: nsights into the evolution of tissue morphogenesis. *Biol. Bull*. 214: 218–232.

Maldonado, M., and Riesgo, A. 2008. Reproductive output in a Mediterranean population of the homosclerophorid *Corticium candelabrum* (Porifera, Demospongiae), with notes on the ultrastructure and behavior of the larva. *Mar. Ecol*. 29: 298–316.

Martin-Belmonte, F., and Mostov, K. 2008. Regulation of cell polarity during epithelial morphogenesis. *Curr. Opin. Cell Biol*. 20: 227–234.

Mi, H. Y., Guo, N., Kejariwal, A., and Thomas, P. D. 2007. PANTHER version 6: protein sequence and function evolution data with expanded representation of biological pathways. *Nucleic Acids Res*. 35: D247–D252.

Miner, J. H., and Yurchenco, P. D. 2004. Laminin functions in tissue morphogenesis. *Annu. Rev. Cell Dev. Biol*. 20: 255–284.

Mizuno, K., et al. 2003. Self-association of PAR-3-mediated by the conserved N-terminal domain contributes to the development of epithelial tight junctions. *J. Biol. Chem*. 278: 31240–31250.

Nance, J. 2005. PAR proteins and the establishment of cell polarity during *C. elegans* development. *Bioessays* 27: 126–135.

Nelson, W. J. 2003. Adaptation of core mechanisms to generate cell polarity. *Nature* 422: 766–774.

Nielsen, C. 2001. *Animal Evolution: Interrelationships of the Living Phyla*. 2nd Ed. Oxford University Press, New York.

Oda, H., Tagawa, K., and Akiyama-Oda, Y. 2005. Diversification of epithelial adherens junctions with independent reductive changes in cadherin form: identification of potential molecular synapomorphies among bilaterians. *Evol. Dev*. 7: 376–389.

Omori, Y., and Malicki, J. 2006. *oko meduzy* and related *crumbs* genes are determinants of apical cell features in the vertebrate embryo. *Curr. Biol*. 16: 945–957.

Pavans de Ceccatty, M. 1986. Cytoskeletal organization and tissue patterns of epithelia in the sponge *Ephydatia mulleri*. *J. Morphol.* 189: 45–65.
Penkert, R. R., DiVittorio, H. M., and Prehoda, K. E. 2004. Internal recognition through PDZ domain plasticity in the Par-6-Pals1 complex. *Nat. Struct. Mol. Biol.* 11: 1122–1127.
Philippe, H., et al. 2009. Phylogenomics revives traditional views on deep animal relationships. *Curr. Biol.* 19: 706–712.
Pruitt, K. D., Tatusova, T., and Maglott, D. R. 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* 35: D61–D65.
Putnam, N. H., et al. 2007. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* 317: 86–94.
Quevillon, E., et al. 2005. InterProScan: protein domains identifier. *Nucleic Acids Res.* 33: W116–W120.
Roh, M. H., et al. 2002. The Maguk protein, Pals1, functions as an adapter, linking mammalian homologues of Crumbs and Discs Lost. *J. Cell Biol.* 157: 161–172.
Ronquist, F., and Huelsenbeck, J. P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.
Ruppert, E. E., Fox, R. S., and Barnes, R. D. 2004. *Invertebrate Zoology: A Functional Evolutionary Approach*. 7th Ed. Thomson-Brooks/Cole, Belmont.
Ruthmann, A., Behrendt, G., and Wahl, R. 1986. The ventral epithelium of *Trichoplax adhaerens* (Placozoa): cytoskeletal structures, cell contacts and endocytosis. *Zoomorphology* 106: 115–122.
Sakarya, O., et al. 2007. A post-synaptic scaffold at the origin of the animal kingdom. *PLoS One* 2: e506.
Sarras, M. P. J., et al. 1994. Cloning and biological function of laminin in *Hydra vulgaris*. *Dev. Biol.* 164: 312–324.
Sasakura, Y., et al. 2003. A genomewide survey of developmentally relevant genes in *Ciona intestinalis* - X. Genes for cell junctions and extracellular matrix. *Dev. Genes Evol.* 213: 303–313.
Schierwater, B., et al. 2009. Concatenated analysis sheds light on early metazoan evolution and fuels a modern "urmetazoon" hypothesis. *PLoS Biol* 7: e1000020.
Segbert, C., Johnson, K., Theres, C., van Furden, D., and Bossinger, O. 2004. Molecular and functional analysis of apical junction formation in the gut epithelium of *Caenorhabditis elegans*. *Dev. Biol.* 266: 17–26.
Shimizu, H., et al. 2008. The extracellular matrix of hydra is a porous sheet and contains type IV collagen. *Zoology* 111: 410–418.
Shin, K., Fogg, V. C., and Margolis, B. 2006. Tight junctions and cell polarity. *Annu. Rev. Cell Dev. Biol.* 22: 207–235.
Simpson, T. L. 1984. *The Cell Biology of Sponges*. Springer-Verlag, New York.
Sperling, E. A., Peterson, K. J., and Pisani, D. 2009. Phylogenetic-signal dissection of nuclear housekeeping genes supports the paraphyly of sponges and the monophyly of Eumetazoa. *Mol. Biol. Evol.* 26: 2261–2274.
Srivastava, M., et al. 2008. The *Trichoplax* genome and the nature of placozoans. *Nature* 454: 955–960.
Srivastava, M., et al. 2010. The *Amphimedon queenslandica* genome and the evolution of animal complexity. *Nature* 466: 720–727.
Suzuki, A., and Ohno, S. 2006. The PAR-aPKC system: lessons in polarity. *J. Cell Sci.* 119: 979–987.
Suzuki, A., et al. 2001. Atypical protein kinase C is involved in the evolutionarily conserved PAR protein complex and plays a critical role in establishing epithelia-specific junctional structures. *J. Cell Biol.* 152: 1183–1196.
Tatusov, R. L., et al. 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4: 41.
Tepass, U., Truong, K., Godt, D., Ikura, M., and Peifer, M. 2000. Cadherins in embryonic and neural morphogenesis. *Nat. Rev. Mol. Cell Biol.* 1: 91–100.
te Velthuis, A. J., Admiraal, J. F., and Bagowski, C. P. 2007. Molecular evolution of the MAGUK family in metazoan genomes. *BMC Evol. Biol.* 7: 129.
Thomas, M. B., and Edwards, N. C. 1991. Cnidaria: hydrozoa. In F. W. Harrison and J. A. Westfall (eds.). *Microscopic Anatomy of Invertebrates*. Wiley-Liss, New York, pp. 91–183.
Thomas, P. D., et al. 2003. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* 13: 2129–2141.
Timpl, R., and Brown, J. C. 1996. Supramolecular assembly of basement membranes. *Bioessays* 18: 123–132.
Tyler, S. 2003. Epithelium - the primary building block for metazoan complexity. *Integr. Comp. Biol.* 43: 55–63.
Tzu, J., and Marinkovich, M. P. 2008. Bridging structure with function: structural, regulatory, and developmental role of laminins. *Int. J. Biochem. Cell Biol.* 40: 199–214.
Usher, K., and Ereskovsky, A. V. 2005. Larval development, ultrastructure and metamorphosis in *Chondrilla australiensis* Carter, 1873 (Demospongiae, Chondrosida, Chondrillidae). *Invertebr. Reprod. Dev.* 47: 51–62.
Whittaker, C. A., Bergeron, K.-F., Whittle, J., Brandhorst, B. P., Burke, R. D., and Hynes, R. O. 2006. The echinoderm adhesome. *Dev. Biol.* 300: 252–266.
Wodarz, A., Ramrath, A., Grimm, A., and Knust, E. 2000. *Drosophila* atypical protein kinase C associates with Bazooka and controls polarity of epithelia and neuroblasts. *J. Cell Biol.* 150: 1361–1374.
Wu, V. M., Schulte, J., Hirschi, A., Tepass, U., and Beitel, G. J. 2004. Sinuous is a *Drosophila* claudin required for septate junction organization and epithelial tube size control. *J. Cell Biol.* 164: 313–323.
Yagi, T., and Takeichi, M. 2000. Cadherin superfamily genes: functions, genomic organization, and neurologic diversity. *Genes Dev.* 14: 1169–1180.
Yamanaka, T., et al. 2001. PAR-6 regulates aPKC activity in a novel way and mediates cell–cell contact-induced formation of the epithelial junctional complex. *Genes Cells* 6: 721–731.
Yeh, R. F., Lim, L. P., and Burge, C. B. 2001. Computational inference of homologous gene structures in the human genome. *Genome Res.* 11: 803–816.
Zallen, J. A. 2007. Planar polarity and tissue morphogenesis. *Cell* 129: 1051–1063.
Zhang, X., et al. 2002. Structure and function of an early divergent form of laminin in hydra: a structurally conserved ECM component that is essential for epithelial morphogenesis. *Dev. Genes Evol.* 212: 159–172.

SUPPLEMENTARY MATERIAL

Additional Supporting Information may be found in the online version of this article:

Fig. S1. AmqPar-6.
Fig. S2. AmqaPKC.
Fig. S3. AmqPar-3.
Fig. S4. *Amphimedon* Crumbs orthologues.
Fig. S5. AmqPatj.
Fig. S6. AmqScribble.
Fig. S7. *Amphimedon* Lgl orthologues.
Fig. S8. *Amphimedon* α-catenin-related proteins.
Fig. S9. AmqClaudinSF.
Fig. S10. *Amphimedon* laminin-related proteins.
Fig. S11. Evolutionary origin of domains, domain architectures and epithelial genes.
Table S1. *Amphimedon* PDZ domain containing genes.
Table S2. *Amphimedon* Laminin G domain containing genes.
Table S3. *Amphimedon* Cadherin domain containing genes (full table).
File S1. *Amphimedon queenslandica* protein sequences FASTA.
File S2. *Monosiga brevicollis* protein sequences FASTA.

File S3. *Trichoplax adhaerens* protein sequences FASTA.
File S4. *Nematostella vectensis* protein sequences FASTA.

References

Joberty, G., Petersen, C., Gao, L., and Macara, I. G. 2000. The cell-polarity protein Par6 links Par3 and atypical protein kinase C to Cdc42. *Nat. Cell Biol.* 2: 531–539.

Lupas, A., Vandyke, M., and Stock, J. 1991. Predicting coiled coils from protein sequences. *Science* 252: 1162–1164.

Nagai-Tamai, Y., Mizuno, K., Hirose, T., Suzuki, A., and Ohno, S. 2002. Regulated protein-protein interaction between aPKC and PAR-3 plays an essential role in the polarization of epithelial cells. *Genes Cells* 7: 1161–1171.