

RESEARCH ARTICLE

Pre-metazoan origins and evolution of the cadherin adhesome

Paul S. Murray^{1,2} and Ronen Zaidel-Bar^{3,4,*}

ABSTRACT

Vertebrate adherens junctions mediate cell–cell adhesion via a “classical” cadherin–catenin “core” complex, which is associated with and regulated by a functional network of proteins, collectively named the cadherin adhesome (“cadhesome”). The most basal metazoans have been shown to conserve the cadherin–catenin “core”, but little is known about the evolution of the cadhesome. Using a bioinformatics approach based on both sequence and structural analysis, we have traced the evolution of this larger network in 26 organisms, from the uni-cellular ancestors of metazoans, through basal metazoans, to vertebrates. Surprisingly, we show that approximately 70% of the cadhesome, including proteins with similarity to the catenins, predate metazoans. We found that the transition to multicellularity was accompanied by the appearance of a small number of adaptor proteins, and we show how these proteins may have helped to integrate pre-metazoan sub-networks via PDZ domain–peptide interactions. Finally, we found the increase in network complexity in higher metazoans to have been driven primarily by expansion of paralogs. In summary, our analysis helps to explain how the complex protein network associated with cadherin at adherens junctions first came together in the first metazoan and how it evolved into the even more complex mammalian cadhesome.

KEY WORDS: cadherin, adherens junction, multicellularity, evolution, protein interaction network

INTRODUCTION

Adherens junctions (AJs) are multi-protein structures that form a bridge between the actin cytoskeletons of adjacent cells, enabling their integration into higher order organizations such as epithelial sheets. In this capacity, they play a critical role in coordinating cellular mechanics and tissue dynamics during development and homeostasis, as well as in diseases such as cancer (reviewed by Harris and Tepass, 2010). Adhesion between cells by AJs in vertebrates is primarily mediated by transmembrane receptors of the type I “classical” cadherin family. The structure and function of vertebrate “classical” cadherin extracellular regions is now well understood (reviewed by Brasch et al., 2012). Briefly, the

cadherin ectodomain is comprised of five consecutively-linked extracellular-cadherin (EC) domains. Binding between apposed cells is primarily, but not exclusively, homophilic and is dependent on Ca^{2+} . Trans- and cis-interactions of the ectodomains facilitate the oligomerization of cadherins into higher order structures, which are further stabilized by interactions of the cytoplasmic domain, and its binding partners, with F-actin (Harrison et al., 2011; Hong et al., 2013).

The cytoplasmic domain, or tail, of “classical” cadherins is approximately 150 amino acids long and natively unstructured (Huber et al., 2001). Within the tail, tandem unstructured domains, the juxtamembrane domain (JMD) and catenin-binding domain (CBD), bind P120 and β -catenin, respectively (Thoreson et al., 2000; Stappert and Kemler, 1994). Both P120 and β -catenin belong to the armadillo (ARM) repeat family of proteins, which is characterized by forty-amino acid repeats of three α -helices (Huber and Weis, 2001; Ishiyama et al., 2010). An electro-positive groove formed by tandem ARM repeats serves as a binding site for the JMD and CBD, which can be postrationally modified at multiple sites; depending on the site, the interaction between cadherin and catenins is either strengthened or weakened (Dupre-Crochet et al., 2007). α -catenin interacts with β -catenin (Pokutta and Weis, 2000), and may link the complex to the actin cytoskeleton (Desai et al., 2013; Drees et al., 2005). The binding of cadherins to catenins appears to be a feature that is common to all metazoans, from vertebrates to sponge, but absent from pre-metazoans (Nichols et al., 2012; Miller et al., 2013). Herein, we refer to any cadherin that contains both EC domains and a CBD as “classical”-like.

Like metazoans, the choanoflagellates, which are the closest extant unicellular relatives of metazoans, contain many cadherins (Nichols et al., 2012; King et al., 2008; Abedin and King, 2008; reviewed by Richter and King, 2013). However, the two choanoflagellates with sequenced genomes, *Salpingoeca rosetta* and *Monosiga brevicollis*, do not contain a “classical”-like cadherin in which a clearly defined JMD or CBD can be recognized in the cytoplasmic tail. Choanoflagellate cadherins are not known to participate in cell–cell adhesion and little is known about their function in these organisms (Dayel et al., 2011; Fairclough et al., 2013), though two cadherins from *M. brevicollis* have been localized to the feeding collar (Abedin and King, 2008). The slime mold, *Dictyostelium discoideum*, has proteins related to β - and α -catenin (Aardvark and *Ddx*-catenin, respectively), which form a complex and are important in establishing epithelial-like cell polarity, along with *Dictyostelium* IQGAP (*rgaA*) and myosin II (Dickinson et al., 2012b; Dickinson et al., 2011; Grimson et al., 2000). Though Aardvark and *Ddx*-catenin are not known to participate in cell–cell adhesion, they share many functional similarities with β -catenin and α -catenin, respectively, and are therefore considered close relatives of these catenins (reviewed by Miller et al., 2013). Aardvark-like proteins have also been reported in the choanoflagellate *S. rosetta* and in the filasterea *Capsaspora owczarzaki* (Nichols et al., 2012; Suga

¹Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032, USA. ²Center of Computational Biology and Bioinformatics, Department of Systems Biology, Columbia University, Irving Cancer Research Center, New York, NY 10032, USA. ³Mechanobiology Institute Singapore, National University of Singapore, Singapore 117411. ⁴Department of Biomedical Engineering, National University of Singapore, Singapore 117575.

*Corresponding author (biezbr@nus.edu.sg)

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

et al., 2013). Thus, at least some unicellular holozoa contain both cadherins and proteins related to the catenins (Aardvark), but there is currently no evidence for a physical interaction between the two protein families in these unicellular organisms.

P120, β - and α -catenin serve as hubs for a diverse network of associated proteins (reviewed by Braga, 2002; Hartsock and Nelson, 2008; Stepniak et al., 2009; Cavallaro and Dejana, 2011; Bertocchi et al., 2012; Ratheesh and Yap, 2012). In addition to providing physical connections with the cytoskeleton, the network of proteins associated with AJs regulates the location and strength of adhesion complexes in response to internal and external signals and forces (reviewed by Lecuit, 2005; Baum and Georgiou, 2011; Niessen et al., 2011). We refer henceforth to this larger functional network, including the cadherin–catenin “core”, as the “cadherin adhesome” or “cadhesome”. Based on defined criteria and a systematic analysis of the literature, we previously compiled a list of all known human cadhesome components and their interactions (Zaidel-Bar, 2013). The literature-based cadhesome has 173 proteins, which can interact with each other in up to ~400 different combinations, and is populated by a variety of structural, catalytic, and regulatory proteins. Our goal here has been to understand which cadhesome components were incorporated from pre-metazoan systems, which were innovated at the onset of multicellularity, and which types of components expanded in higher metazoans. Starting with the cadherin–catenin “core” of AJs, we then expand our analysis to the larger cadhesome network, tracing its evolution from pre-metazoans to vertebrates, both in terms of its components and functional interactions.

MATERIALS AND METHODS

Reciprocal best-hits analysis

To identify analogous proteins with similar protein domain architecture to the 173 human cadhesome components, we performed a reciprocal best-hits analysis of twenty-five unikont (the unity of holozoa, fungi, and amoebozoa/apusozoa) genomes: *Homo Sapiens* (Hs); *Mus musculus* (Mm); *Gallus gallus* (Gg); *Xenopus tropicalis* (Xt); *Danio rerio* (Dr); *Ciona intestinalis* (Ci); *Branchiostoma floridae* (Bf); *Strongylocentrotus purpuratus* (Sp); *Drosophila melanogaster* (Dm); *Caenorhabditis elegans* (Ce); *Hydra vulgaris* (Hv); *Nematostella vectensis* (Nv); *Trichoplax adhaerens* (Ta); *Amphimedon queenslandica* (Aq); *Oscarella carmela* (Oc); *Monosiga brevicollis* (Mb); *Salpingoeca rosetta* (Sr); *Sphaeroforma arctica* (Sa); *Capsaspora owczarzaki* (Co); *Allomyces macrogynus* (Am); *Mortierella verticillata* (Mv); *Spizellomyces punctatus* (Spu); *Saccharomyces cerevisiae* (Sc); *Schizosaccharomyces pombe* (Spo); *Thecamonas trahens* (Tt); and *Dictyostelium discoideum* (Dd). These species were selected for analysis because they span the ancient to recent ancestry of metazoans, including critical innovations such as cell–ECM and cell–cell adhesion, and their complete or draft genomes are available (Ruiz-Trillo et al., 2007). The phylogenetic relationship between the twenty-five species and *Homo sapiens* is depicted in Fig. 1.

The reciprocal best-hits analysis was carried out as follows: 1) each of the 173 human cadhesome components was used as a protein BLAST (Altschul et al., 1997) query in a search of the sequence databases of twenty-five organisms (Ruiz-Trillo et al., 2007; Apweiler et al., 2004; Hemmrich and Bosch, 2008). 2) The sequence with the lowest E-value was retrieved for each query, from every organism, if the E-value was below 10 (bit score of ~30). 3) Each qualifying hit was submitted to both SMART (Letunic et al., 2012; Schultz et al., 1998) and CDD (Marchler-Bauer et al., 2013; Marchler-Bauer et al., 2011) in order to determine its protein domain content. The two databases are largely redundant, but in a number of cases proved to be complementary. 4) Each identified sequence was used as a BLAST query in a reciprocal search (re-BLAST) of the human sequence database. A protein was considered as analogous

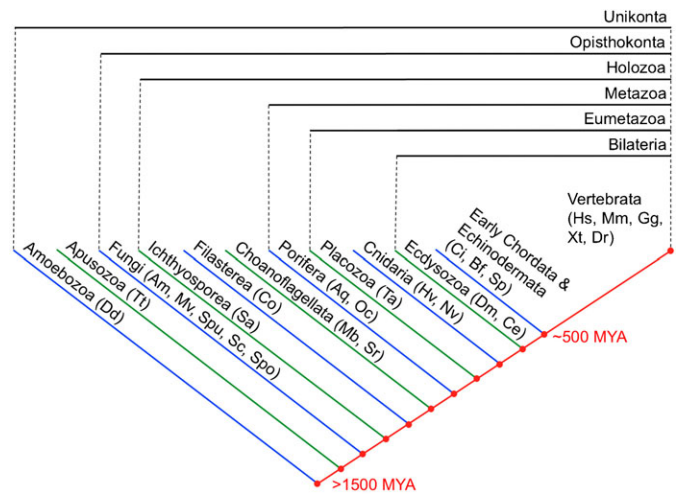


Fig. 1. Phylogeny of metazoans and their uni-cellular ancestors. The phylogenetic relationship between the twenty-six metazoan and pre-metazoan species we analyzed is depicted. Each species is positioned according to its phylum or kingdom; species are also grouped into broader taxonomic categories, delimited by black bars. “MYA” stands for million years ago.

if it retrieved the original human component, or a highly related component, in the re-BLAST step, and was comprised of the same major domains – hence the terminology “reciprocal best-hit”.

We also applied the reciprocal best-hits approach to the sponge and *T. adhaerens* analogous proteins we identified when we searched with the human cadhesome, using them to BLAST search all pre-metazoans, then re-BLAST search the pre-metazoan sequences to humans. Similarly, we used all unicellular holozoa analogous proteins identified in BLAST searches with human, sponge, or *T. adhaerens* analogous proteins to BLAST search all pre-metazoans, then re-BLAST against humans. These added levels of searching identified many analogous proteins that were not identified with the human analogous proteins alone, e.g. *Dictyostelium* sibC, which we identified in a search with *O. carmela* integrin, and has been shown experimentally to function in a manner similar to vertebrate integrin (Cornillon et al., 2006). In general, analogous proteins had a high bit score/low E-value and high percent query coverage in the BLAST round, but this was not always the case (e.g. *DdAardvark*, which has been experimentally defined as related to β -catenin (Dickinson et al., 2011)). Since AJs might differ significantly across species, our primary objective was to determine whether or not the fundamental building blocks of the human junction exist. A full list of analogous proteins in each of the organisms can be found in supplementary material Table S4.

“Classical” and “classical”-like cadherins differ significantly in their extracellular regions from organism to organism (Oda and Takeichi, 2011; Hulpiau and van Roy, 2009; Hulpiau and van Roy, 2011). However, all share a conserved cytoplasmic tail that binds catenins. For this reason, we queried the organism sequence databases with both A) full-length type I and II human cadherins, and B) their cytoplasmic tails alone. This search method resulted in good coverage of both “classical” and “classical”-like cadherins in all organisms analyzed. We then also searched all pre-metazoans with the full-length “classical”-like cadherins from all the non-vertebrate metazoans we analyzed, as well as their cytoplasmic tails alone.

Searching choanoflagellata, filasterea and ichthyosporea with the P120 and β -catenin families identified sequences that were not readily identified by the re-BLAST step as from a known human family. Many of the same sequences were retrieved when we BLAST searched choanoflagellata, filasterea and ichthyosporea with *Dictyostelium* Aardvark. In order to better characterize these sequences, we used each in a re-BLAST search of slime mold, yeast, and the green algae, *Volvox*

carteri. We also used each to search for similar sequences in the other unicellular holozoa. We iteratively performed this BLAST, re-BLAST process until each sequence could be confidently defined by its sequence similarity and domains as related to the catenins/Aardvark, members of a known family (e.g. importin- α), or unrelated to the catenins/Aardvark. This all-against-all BLAST search of eight species helped to characterize many of the choanoflagellata, filasterea and ichthyosporea proteins as Aardvark-like; it also identified in them many sequences in addition to the original hits based on our BLAST search with the human proteins (supplementary material Table S5).

Sequence analysis and structure prediction

SMART and TMHMM (Krogh et al., 2001; Sonnhammer et al., 1998) were used to predict transmembrane helices, other domains in the cytoplasmic tail, and membrane topology of the metazoan “classical”-like cadherins, the 23 *M. brevicollis* and 29 *S. rosetta* cadherins, and the two *C. owczarzaki* and *T. trahens* cadherins (supplementary material Table S4). Cadherin PDZ-binding motifs were identified via visual inspection. The T-coffee suite of methods (Di Tommaso et al., 2011; Notredame et al., 2000) was used to produce multiple sequence alignments.

Using BLAST, SMART, and CDD, it was not always clear how many ARM repeats were formed by the choanoflagellata, filasterea and ichthyosporea Aardvark-like sequences. Therefore, we used a number of additional sequence- and structure-based methods to predict ARM repeats, including Superfamily (Gough et al., 2001), ARD (Palidwor et al., 2009), HHrepID (Biegert and Söding, 2008), and secondary structure prediction (McGuffin et al., 2000; Pollastri et al., 2002; Cole et al., 2008). We also used the servers Pudge (Norel et al., 2010), PHYRE2 (Kelley and Sternberg, 2009), and I-tasser (Zhang, 2008; Roy et al., 2010) to predict the tertiary structure of the sequences. Only in the union of these methods were we able to confidently predict ARM repeats (supplementary material Table S6). For all modeled sequences, there was good agreement between Pudge, PHYRE2, and I-tasser, in terms of template selection, and query to template alignment. We scored all the models constructed with a number of model evaluation tools: Prosa (Sippl, 1993), DFIRE (Zhou and Zhou, 2002), and Verify3D (Bowie et al., 1991; Lüthy et al., 1992) (supplementary material Table S7).

Domain architecture-based clustering

The 173 human cadhesome components were first clustered into families, by eye, based on matching SMART- and CDD-predicted domains. For example, we clustered the membrane-associated guanylate kinase (MAGUK) proteins, ZO-1 (TJP1) and ZO-2 (TJP2) in the same group because they contain the same number and order of domains while other MAGUK proteins, e.g. MAG11 were assigned to separate families. This clustering yielded a “simplified” cadhesome, consisting of 112 protein families. A full list of the protein clusters in the “simplified” cadhesome can be found in supplementary material Table S4. Using the “simplified” cadhesome, we determined when, in the evolutionary timeline suggested by the species analyzed here, families of proteins appeared. To analyze paralog expansion, we clustered components the same as the “simplified” cadhesome, except for ACTN1/4/SPTBN1 and the MAGUK proteins, which were more broadly clustered based on similar, but not exact type, number, and order of domains.

RESULTS

Conserved motifs in the cytoplasmic tail of metazoans cadherins

Since a large part of the cadhesome involves binding or regulating the cytoplasmic tail of cadherin we started our evolutionary investigation by examining the degree of conservation of various motifs along the tail of cadherin. To this end we performed BLAST searches with human “classical” cadherins against the genomes of twenty-five species ranging across unikonts (Fig. 1). Our searches retrieved “classical” or “classical”-like cadherins in all metazoans analyzed. Fig. 2 shows the alignment of the cytoplasmic tails of some of these cadherins, from human E-cad (*HsCDH1*) at the top, down to the most basal metazoans, the sponges *Oscarella carmela* and *Amphimedon queenslandica*, in order of increasing evolutionary age. The full-length JMD and CBD span most of the length of the “classical” cadherin tail, but we show only the most crucial catenin-binding regions, as indicated in the figure. Depicted as

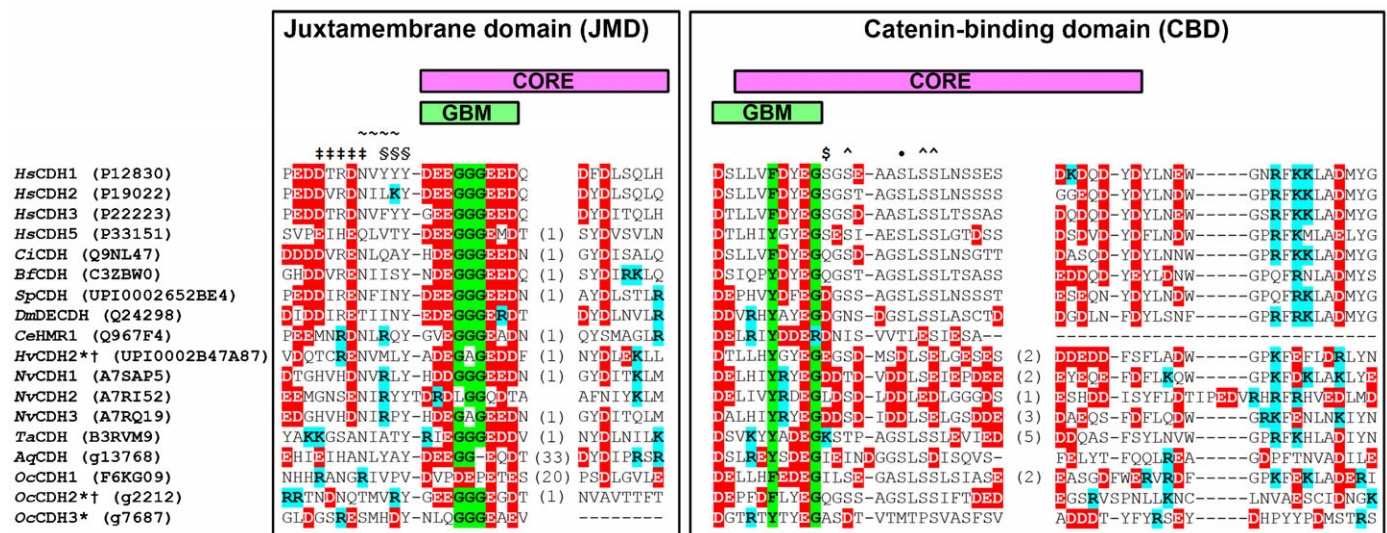


Fig. 2. Sequence alignment of cadherin cytoplasmic domains from metazoa. The cytoplasmic tails of the metazoan “classical”-like cadherins are aligned with human E-cad (*HsCDH1*), and divided up into JMD and CBD. Magenta bars indicate core regions known via experiment to be required for cadherin–catenin binding. Green bars indicate groove-binding motifs (GBMs), which are known to bind in an extended peptide conformation to the ARM repeats of catenins, based on structures of E-cad-P120 and E-cad- β -catenin. If inserts exist relative to E-cad, the number of intervening amino acids is indicated. Cadherins denoted by “*” are named out of convenience, as no general naming standard exists (UniProt or Compagen identifiers are also included). “†” indicates that the protein has no EC domains. Residues are colored according to their properties, with acidic residues (E/D) in red and basic residues (K/R) in cyan. Amino acids in the JMD and CBD are annotated based on known functions in vertebrates: “†” indicates a CASP3 cleavage site; “~” a NUMB PTB-binding site; “\$” a Hakai binding site; “.” a CSNK1 phosphorylation site; “\$” a CSNK2 site, and “^” a GSK3 β site.

magenta bars are two regions, the JMD core and CBD core, which are known via mutagenesis to be absolutely required for binding P120 and β -catenin, respectively (Thoreson et al., 2000; Stappert and Kemler, 1994). Crystal structures of E-cad JMD-P120 (Ishiyama et al., 2010) and E-cad CBD- β -catenin (Huber and Weis, 2001) demonstrate that, within both core regions, a motif of approximately ten residues (green bars) binds in the electro-positive groove formed by the ARM repeats of its cognate catenin. Mutagenesis of residues in P120 and β -catenin that contact these motifs abrogates cadherin binding, which suggests that they are crucial to this interaction (Ishiyama et al., 2010; Graham et al., 2000). We refer to these two motifs, xx[ED]GGGExx (where “x” is any amino acid) in the JMD, and Dxxxx ϕ xxEG (where “ ϕ ” is aromatic) in the CBD, as groove binding motifs (GBMs).

Three consecutive glycines in the JMD GBM anchor the region in a small hydrophobic pocket in the ARM repeats of P120, while acidic residues on both sides contact basic residues in P120 (Ishiyama et al., 2010). In the GBM of the CBD, a phenylalanine anchors the region in a small pocket of β -catenin, while acidic residues on both ends contact basic residues in β -catenin (Huber and Weis, 2001). Thus, both GBMs use a combination of acidic residues and glycine/hydrophobic residues to bind catenins. Fig. 2 demonstrates that the acidic catenin-binding residues of both the JMD GBM and CBD GBM are well conserved in metazoa, as are the intervening glycine/hydrophobic residues. As noted previously, the known “classical”-like cadherin in *O. carmela* (*OcCDH1*) has a functioning CBD, but no apparent JMD (Nichols et al., 2012). However, we found at least one sequence in *O. carmela* (*OcCDH2**) that contains both a JMD GBM and CBD GBM. Though *OcCDH2** lacks EC domains, it does contain EGF and LamG domains, which suggests it is related to early metazoan cadherins. Thus, both GBMs appear together in at least one sequence from each sponge, which suggests that both innovations were present in tandem in the last common ancestor of the sponges and other metazoans.

Most of the residues in the E-cad JMD and CBD do not form secondary structure (Ishiyama et al., 2010; Huber and Weis, 2001). The “classical”-like cadherins we analyzed also have a low propensity to form secondary structure, and are devoid of SMART-predicted globular domains (supplementary material Table S4). Such long, unstructured regions leave the E-cad JMD and CBD available for posttranslational modification. For example, binding of β -catenin is inhibited by CSNK1 phosphorylation of human E-cad Ser844 (designated by “•” in the CBD of Fig. 2A) (Dupre-Crochet et al., 2007), which we found to be well conserved in metazoans. CSNK2 phosphorylates Ser838 (“\$”) (Serres et al., 2000), which enhances binding to β -catenin, but this serine is not conserved outside of vertebrates. We found at least one GSK3 β phosphorylation site (“^”) (Huber and Weis, 2001) to be conserved in all metazoans we analyzed. The presence of known sequence elements in the cadherins from a wide variety of metazoans suggests that a subset of phospho-regulatory switches were present in the AJs of the most ancestral metazoans. This contrasts with other regulatory sites in the JMD, which we found to be later innovations: a CASP3 apoptotic cleavage site ([DSTE]xxD[GSAN], “‡”) (Steinhusen et al., 2001); a NUMB PTB domain-binding site (NVYY, “~”) (Wang et al., 2009); and a YYY motif (“\$”), which binds the ubiquitin ligase Hakai (CBL1) (Mukherjee et al., 2012; Fujita et al., 2002).

Previous studies based on searches for EC domains have identified numerous cadherins in choanoflagellata and filasterea:

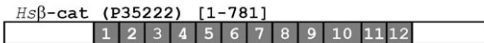
23 in *M. brevicollis*, 29 in *S. rosetta*, and one in *C. owczarzaki* (Abedin and King, 2008; Nichols et al., 2012). However, using SMART, we found no “classical” cadherin tail in any of these cadherins (supplementary material Table S4). Furthermore, we found no JMD core, CBD core, JMD GBM (xx[ED]GGGExx), or CBD GBM (Dxxxx ϕ xxEG) when we visually inspected the known unicellular holozoa cadherins, as well as a second cadherin in *C. owczarzaki* (E9CEE8) and two cadherins in *Thecamonas trahens* (AMSG_00600 and AMSG_08522) that we identified here for the first time. Interestingly, we found PDZ-binding motifs in some of the choanoflagellate cadherins, similar to metazoan cadherins (supplementary material Table S4) (Demontis et al., 2006).

β -catenin/Aardvark and the *Dictyostelium* polarity network in unicellular holozoa

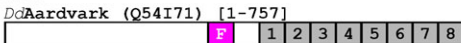
In order to determine if pre-metazoans have ARM repeat proteins related to β -catenin/Aardvark, we performed BLAST searches using human β -catenin and plakoglobin (JUP) and *Dictyostelium* Aardvark as queries. Our searches retrieved a variety of ARM proteins with significant sequence similarity (high BLAST E-value) to *DdAardvark* in *C. owczarzaki* (*CoE9CA59*) and *S. rosetta*, consistent with previous work (Nichols et al., 2012), and also in *Sphaeroforma arctica* (*SaSARC_01382*, *SaSARC_01383* and *SaSARC_02298*) (Fig. 3). In total we found at least six sequences similar to Aardvark in *S. rosetta* (*SrF2UHX6*, *SrF2UHT4*, *SrF2UNI1*, *SrF2UQB1*, *SrF2UQ79* and *SrF2UQC3*), only one of which was previously identified. We found no proteins analogous to Aardvark in the other choanoflagellate we analyzed, *M. brevicollis*. Consistent with *DdAardvark*, all but one of these sequences (*SaSARC_01383*) returns ARM6 in the re-BLAST search of humans, and Aardvark in a re-BLAST search of *Dictyostelium*, the latter with a higher E-value (supplementary material Table S5). To further characterize these sequences, we employed a number of additional sequence- and structured-based methods (see Materials and Methods). Though the choanoflagellata, filasterea and ichthyosporia sequences lack an F-box domain, and some have added Toll/interleukin-1 receptor (TIR) domains, our analysis suggests their ARM repeats have significant sequence similarity to *DdAardvark*. Hence, we refer to these ten unicellular holozoa sequences as Aardvark-like.

Together with *Dd α -catenin*, IQGAP1 (*rgaA*), and myosin II, *Dictyostelium* Aardvark helps to establish cell polarity in this organism (Dickinson et al., 2012b; Dickinson et al., 2011). In addition to Aardvark-like sequences, each of the unicellular holozoa we analyzed also possessed proteins analogous to α -catenin or the closely-related vinculin, IQGAP, and myosin II. Thus, *S. rosetta*, *C. owczarzaki*, and *S. arctica*, all seem to have conserved the components of the *Dictyostelium* polarity network. Furthermore, using *S. rosetta* transcriptomics data (Fairclough et al., 2013) we found that one of the Aardvark-like proteins we identified (*F2UQ79*), α -catenin/vinculin (*F2UM91*), IQGAP (*F2TWA6*), and myosin II (*F2U9L1*) are all upregulated in thecate cells (supplementary material Table S1), which suggests that these four proteins function together in this organism, perhaps to establish polarity in a manner akin to *Dictyostelium*. As in *D. discoideum*, the binding partner of the *S. rosetta* Aardvark ARM repeats is not known, but it is interesting to note that six cadherins were also found to be upregulated in thecate cells (Fairclough et al., 2013).

Compared to *rgaA*, the proteins analogous to IQGAP from *S. rosetta*, *C. owczarzaki*, and *S. arctica* have gained the WW and

β -cat/plakoglobin

Aardvark



Aardvark-like

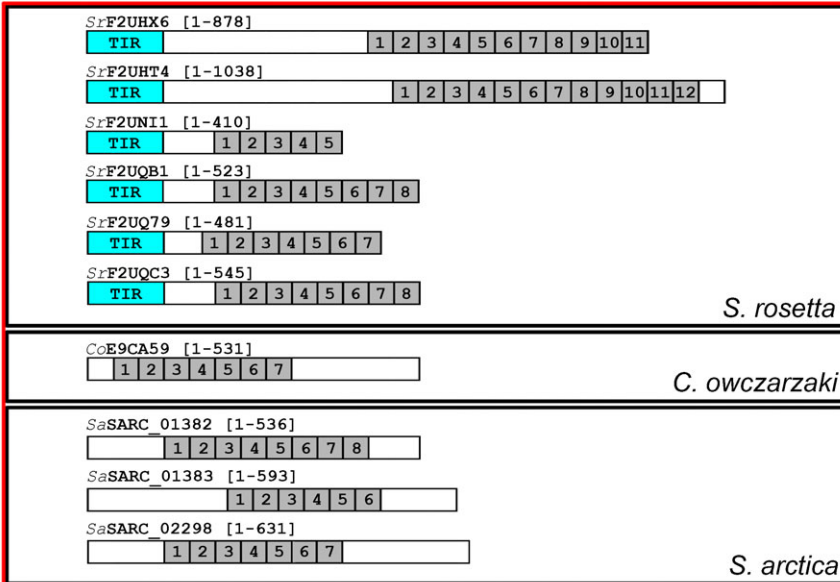


Fig. 3. Aardvark-like analogous proteins in choanoflagellata, filasterea, and ichthyosporea.

Unicellular holozoa sequences identified as related to Aardvark/ β -catenin are depicted in relation to vertebrate β -catenin and Aardvark. ARM repeats are numbered and represented as boxes in the context of the full sequence, with structurally resolved repeats colored dark grey and predicted repeats colored light grey (see Materials and Methods). Boxes are also used to represent other predicted domains, which are labeled and colored: “F” is F-box and “TIR” is toll interleukin-1 receptor. Species of origin, identifier (Uniprot/Broad Institute gene), and length of each sequence are also depicted.

IQ domains (supplementary material Table S4), which suggests that more of its ten known vertebrate interactions may have been in place prior to multicellularity. In this way, incorporation by metazoans of a single adaptor, e.g. Aardvark/ β -catenin, and its pre-metazoan interaction network may have added the functionality of multiple proteins to the newly-formed AJ. Proteins analogs of the *Dictyostelium* polarity network are also present in all the metazoans we analyzed (except *Drosophila* that is missing an IQGAP1; other insects, such as honey bees, have IQGAP1 analogs). In the context of vertebrate AJs, IQGAP binds β -catenin and myosin (Weissbach et al., 1998; Fukata et al., 1999) and can be either a positive or negative regulator of E-cadherin-mediated adhesion (reviewed by Noritake et al., 2005).

Incorporating a pre-metazoan cadhesome “tool kit”

Hundreds of proteins interact with the cadherin–catenin “core” of human AJs (Zaidel-Bar, 2013). In order to trace the evolution of this protein network (the cadhesome), we first clustered its components based on similar domain arrangement (e.g. proteins with three LIM domains, such as LPP and TRIP6), which yielded 112 protein families ranging in size from 1 to 18 members (supplementary material Table S4). We plotted each protein family of this “simplified” cadhesome as a function of its implied evolutionary age, i.e. in what phylum a representative analogous protein first appeared (Fig. 4). For example, the origin of “classical”-like cadherin is placed after the transition to multicellularity, while β -catenin/Aardvark is placed before the transition because *DdAardvark* was shown to function in a similar manner to β -catenin (Dickinson et al., 2011; Dickinson et al., 2012b). What is strikingly apparent from Fig. 4 is that the majority of cadhesome protein families predate the transition to multicellularity. Summing up the groups to the right of the transition, we found that ~71% (80/112) of the “simplified” cadhesome predates metazoans.

The pre-metazoan proteins incorporated into the metazoan cadhesome belong to a variety of categories, including kinases, phosphatases, small GTPases and their regulators, actin regulators, and adaptor proteins (Fig. 4). There are 45 families with origins in *D. discoideum*/*T. trahens* or earlier organisms. Though the majority of these ancient proteins are regulatory, there are a few interesting exceptions, e.g. the transmembrane protein VEZT, which is characterized by the Vezatin domain. In vertebrates, VEZT interacts with MYO7A and possibly α -catenin, thereby recruiting MYO7A to AJs where it functions to strengthen cell–cell adhesions (Küssel-Andermann et al., 2000). We found proteins analogous to VEZT, MYO7A, and α -catenin/vinculin in both *Dictyostelium* and *T. trahens*. Though *DdVEZT* lacks a SMART-defined Vezatin domain, we found it to have two adjacent transmembrane domains followed by a region with high sequence similarity to the other VEZT analogs (supplementary material Fig. S1), which suggests this protein may function in the same network as far back as amoebzoa/apusozoa. VEZT exists in only two other pre-metazoan organisms – *S. rosetta*, and *Spizellomyces punctatus* – indicating a loss in most non-metazoans analyzed.

Noteworthy, the majority of the 27 protein families first appearing in ichthyosporea, filasterea, and choanoflagellata are PDZ-containing adaptor and regulatory proteins (denoted by “+” in Fig. 4), including TIAM1, SIPA1L3, RAPGEF2, Syntenin-1 (SDCBP), LIN7A, MAGI1, Discs large (DLG1), and possibly Discs large 5 (DLG5). According to published transcriptomic data we analyzed, three of these – SDCBP (F2UJY6), DLG1 (F2U0X6), and DLG5 (F2UG17) – are upregulated in *S. rosetta* thecate cells, along with another 23 other PDZ-containing proteins and the *Dictyostelium* polarity network (Fairclough et al., 2013) (supplementary material Tables S1, S2). Of particular interest (see below) is the PDZ-containing protein

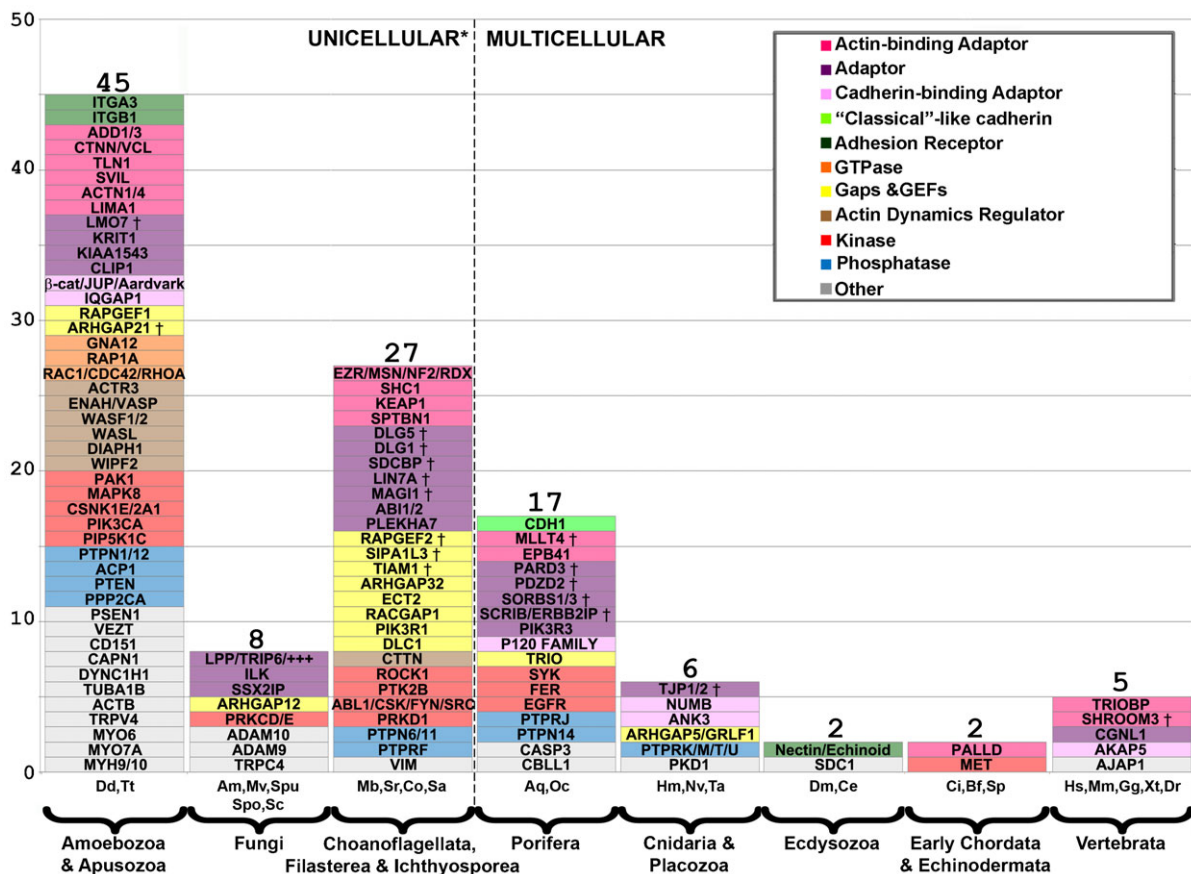


Fig. 4. Innovation of cadhesome gene families throughout evolution. The 173 cadhesome components were clustered into a "simplified" cadhesome of 112 families. Novel proteins/protein families (*y*-axis) are plotted according to when – in which phylum/kingdom – a representative of that family first appears, based on the organisms we analyzed (*x*-axis). Some protein families are represented by one member of the family (LPP/++++), with each "+" signifying the existence of an additional family member. Phyla/kingdoms are ordered from left to right by increasing evolutionary age, with the number of novel proteins appearing in each noted above the bars. The dashed line represents the transition from uni- to multicellularity. Gene families are colored according to the legend. PDZ proteins are denoted with a "†".

MAGI. In addition to the MAGI orthologs previously identified in *M. brevicollis* and *C. owczarzaki* (de Mendoza et al., 2010; Ruiz-Trillo et al., 2008), we also found a MAGI analogous protein in *S. rosetta* (F2U2A7). Like *Drosophila* and *Caenorhabditis elegans* MAGI, SMART predicts two WW domains for *Sr*MAGI, which seem to overlap with a predicted GuKc domain (supplementary material Fig. S2). This is in contrast to vertebrate MAGI orthologs in which there is a clear separation between the WW and GuKc domains.

PDZ-peptide interactions play an important role at the unicellular to multicellular transition

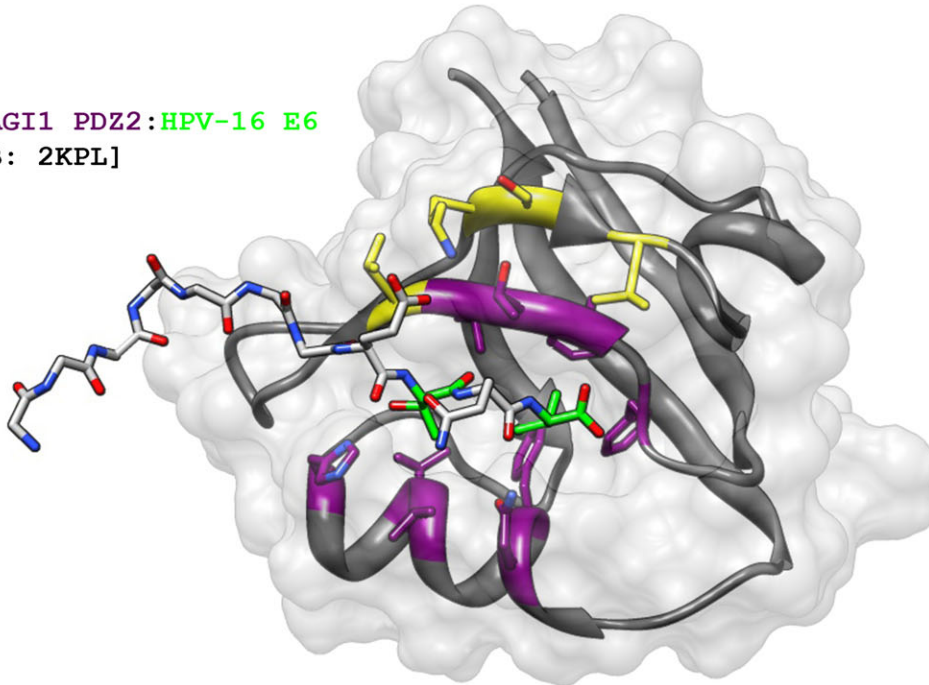
The most basal organism with morphological AJs and a clearly defined interaction between a "classical"-like cadherin and β-catenin is the sponge, *O. carmela*. Fig. 4 demonstrates that the difference in cadhesome content between sponges and pre-metazoans is only 17 protein families, including "classical"-like cadherins with clear catenin-binding tails. Most of the novel metazoan proteins are adaptors, including afadin (MLLT4), ponsin/vinexin (SORBS1/3), PDZD2, Protein 4.1 (EPB41), Par3 (PARD3), and Scribble (SCRIB), many of which bind catenins via their PDZ domains. Importantly, the PDZ-binding motifs of β-catenin and P120 family members are well conserved from sponge to humans, whereas they are absent from all the Aardvark-like

proteins we identified in unicellular holozoa, except for *Sa*SARC_01382 (supplementary material Table S3). Except for the cadherins from *Hydra Vulgaris* and *Trichoplax adhaerens*, PDZ-binding motifs are absent from metazoa "classical"-like cadherins outside of bilateria. Despite this, between "classical"-like cadherins and catenins, at least two components of the "core" have PDZ-binding motifs in every metazoan we analyzed. Thus, it appears that novel PDZ-peptide interactions were important innovations at the dawn of metazoa.

The innovation of PDZ-binding motifs in the catenins not only served to link them to novel metazoa PDZ proteins, but also to pre-metazoa PDZ proteins. As an example, vertebrate MAGI1 employs its six PDZ domains to form complexes with E-cadherin, β-catenin, afadin, and PTEN during the formation of cell-cell junctions in epithelial cells (Vogelmann et al., 2005; Kotelevets et al., 2005; Dobrosotskaya and James, 2000). Fig. 5A shows the structure of human MAGI1 PDZ2, bound to a viral peptide, with the residues that define peptide-binding specificity mapped onto the structure (in yellow). As depicted in Fig. 5B, PDZ6, which binds β-catenin (Dobrosotskaya and James, 2000), is well conserved in all metazoa analyzed, including the residues that define specificity (Tonikian et al., 2008), suggesting that the MAGI1-β-catenin link is conserved as far back as sponges. MAGI PDZ3, which recruits PTEN is also well conserved, as is

A

HsMAGI1 PDZ2:HPV-16 E6
[PDB: 2KPL]



HsMAGI1-2 (Q96QZ7): GKF IHTKLRKSSRGEFGFTTVVGGDEPDEF LQKSLVLDGPAALDGMETGDVIVSVNDTCVLGHTHAQVVKIEFQSIPIGASVDLELCRGY

B

<i>HsMAGI1-6</i> (Q96QZ7)	:	QDFYTVLELRG-AKGF ⁰ GF ⁻² SRGGREY-NMDLYVLR ⁰ LAEDGPAERCGKMRIGDEILEINGETT ⁰ KNM ⁰ SRATEL ⁰ IKNGGRRVRLFLKR
<i>DzMAGI1-6</i> (Q5WPB2)	:	AEFYVDLERD-NKGF ⁰ GF ⁻² SRGGREY-NMDLYVLR ⁰ LAEDGAAVRNGKMRVGEILEINGESTKGMK ⁰ HARATEL ⁰ IKNGGRRVHLVLR
<i>CiMAGI1-6</i> (UPI000180D281)	:	SECFYVELPRS-NKGF ⁰ FG ⁰ IRGGKEY-NLPLYVLR ⁰ VAKDGPAAESGQIRVGDILMEINGKSTIDIPESQAINL ⁰ IKQGGSVVKKLLKH
<i>SpMAGI1-6</i> (UPI0000E45C8E)	:	GDYNNVDLKKG-TTGF ⁰ GF ⁻² SRGGREYDNTPLFVLR ⁰ MDGGPAAQSI ⁰ LMRVGDELIEINSQSTEGML ⁰ HSDA ⁰ IAIRNGGDTITLVLR
<i>CeMAGI1-6</i> (C1P634)	:	GNLIDVILERG-TKGF ⁰ GF ⁻² SRGGQEF ⁰ SGMPLFVLR ⁰ IAADGPAKADGRLQVGDQITTINGQSTKGMSEDDAIR ⁰ IKQHT-MVNLTVLR
<i>BmMAGI1-6</i> (UPI0001924712)	:	TETFSVVLKKG-ETGF ⁰ GF ⁻² SRKGGEG--VPI ⁰ SILRLAENGAAWN ⁰ DGRLKVGDEILEINNADSEVMS ⁰ HSQAVNA ⁰ IRNSGKSCCELLIR
<i>NvMAGI1-6</i> (A7S157)	:	TQVLEIELIRK-NQGF ⁰ GF ⁻² SRGGQEM-DMP ⁰ IYVLR ⁰ KIAEGGVADLDGRIKVGDEVLEINGRSTQHML ⁰ HTDA ⁰ ISMIR-GGSKVRLVLR
<i>AqMAGI1-6</i> (UPI00021A41C9)	:	EEIHHIEFQRD-ENG ⁰ GF ⁻² SRGAEY-NSHL ⁰ FVLR ⁰ IAENGAADK ⁰ SKSLKVGDEILEINGNSTEGML ⁰ HEADAT ⁰ TIK ⁰ HGGK ⁰ VS ⁰ LIR
<i>OcMAGI1-6</i> (g10427)	:	GEVVDVLR ⁰ RHQKQ ⁰ GF ⁻² SRGGREYQNTPL ⁰ FILK ⁰ IAEDGVAARDG ⁰ LQVGEILEINGAGT ⁰ MDMR ⁰ TEA ⁰ SL ⁰ KRGGD ⁰ TVRLK ⁰ VYR

C

<i>HsMAGI1-3</i> (Q96QZ7)	:	ELITVHIVK ⁰ GPMCF ⁰ GF ⁻² TIADSPGGGQ ⁰ RVR ⁰ QIVDS ⁰ PRCRLKEGDLIVE ⁰ VNKNVQAL ⁰ TENQ ⁰ VVD ⁰ DMVECPKGSEVTLVQR
<i>DzMAGI1-3</i> (Q5WPB2)	:	ELITVHMEKGD ⁰ KCF ⁰ GF ⁻² TIADSPGGGQ ⁰ RVR ⁰ QIVDY ⁰ PRCRLREGDI ⁰ IMEV ⁰ NKRVNQL ⁰ TEHQ ⁰ VVD ⁰ DL ⁰ SKCPKGSEVTLVQR
<i>CiMAGI1-3</i> (UPI000180D281)	:	LVMQV ⁰ IHKKEQCF ⁰ GF ⁻² TIADSPQ--GQ ⁰ RVR ⁰ QVLEAQR ⁰ CHELM ⁰ EGDILLE ⁰ INGTTVQ ⁰ NMB ⁰ PNV ⁰ HH ⁰ KDCPSNAAA ⁰ IFV ⁰ IQR
<i>BfMAGI1-3</i> (C3Z2K2)	:	EYVTVHIVK ⁰ GDRCF ⁰ GF ⁻² TIAD ⁰ SAY--GQ ⁰ RVR ⁰ QILDE ⁰ PRCK ⁰ NLCVGDV ⁰ LVE ⁰ INGT ⁰ DVR ⁰ QCT ⁰ NEI ⁰ VG ⁰ IK ⁰ ECTK ⁰ GETT ⁰ IIVQR
<i>SpMAGI1-3</i> (UPI0000E45C8E)	:	EKHDI ⁰ VEVKS ⁰ NQCF ⁰ GF ⁻² TVADSPH--GQ ⁰ KVR ⁰ QILAP ⁰ QRCK ⁰ T ⁰ LR ⁰ EGDLLE ⁰ INRES ⁰ ITDL ⁰ PNQ ⁰ VV ⁰ QR ⁰ KDCPCQ ⁰ QDT ⁰ IISVQR
<i>DmMAGI1-3</i> (Q9W2L2)	:	EIYTF ⁰ SIMK ⁰ SGMCF ⁰ GF ⁻² TIAD ⁰ SAC--GQ ⁰ IVK ⁰ ILDRN ⁰ CCT ⁰ LM ⁰ EGDV ⁰ LLE ⁰ INGLN ⁰ VR ⁰ NK ⁰ BE ⁰ FY ⁰ VV ⁰ EL ⁰ K ⁰ CSQ ⁰ TP ⁰ TA ⁰ VKIQR
<i>CeMAGI1-3</i> (C1P634)	:	DLHEID ⁰ IFK ⁰ GESECF ⁰ GF ⁻² TIAD ⁰ NLN--GQ ⁰ IKK ⁰ ILF ⁰ PSOCP ⁰ NLM ⁰ EGDT ⁰ IVE ⁰ LDGR ⁰ NVR ⁰ PI ⁰ BT ⁰ Q ⁰ V ⁰ DM ⁰ RER ⁰ PI ⁰ GYR ⁰ GK ⁰ LV ⁰ KR
<i>BmMAGI1-3</i> (UPI0001924712)	:	QYIVV ⁰ PVVR ⁰ GNCF ⁰ GF ⁻² TIAD ⁰ SPY--GQ ⁰ RVR ⁰ LD ⁰ V ⁰ QRR ⁰ CQ ⁰ LQ ⁰ Q ⁰ GDV ⁰ IIQ ⁰ INNS ⁰ VV ⁰ HN ⁰ LD ⁰ SE ⁰ VV ⁰ EL ⁰ K ⁰ SCPR ⁰ G ⁰ IPT ⁰ N ⁰ FQ ⁰ QR
<i>NvMAGI1-3</i> (A7S157)	:	EMHNVR ⁰ IVK ⁰ GPMCF ⁰ GF ⁻² TIAD ⁰ SPY--GQ ⁰ K ⁰ IK ⁰ ILDA ⁰ QR ⁰ CEN ⁰ LM ⁰ EGDL ⁰ LM ⁰ ELNG ⁰ YN ⁰ LR ⁰ NS ⁰ SET ⁰ ET ⁰ VD ⁰ M ⁰ K ⁰ NC ⁰ PKG ⁰ REAV ⁰ IL ⁰ QR
<i>TaMAGI1-3</i> (B3RUE9)	:	EII ⁰ ITS ⁰ NI ⁰ AK ⁰ GT ⁰ MC ⁰ GF ⁻² TIAD ⁰ STF--GQ ⁰ IR ⁰ S ⁰ IM ⁰ D ⁰ NR ⁰ CR ⁰ QL ⁰ REGDL ⁰ LLD ⁰ IDG ⁰ Q ⁰ N ⁰ V ⁰ KN ⁰ F ⁰ NE ⁰ F ⁰ Q ⁰ LV ⁰ N ⁰ ML ⁰ K ⁰ SY ⁰ TI ⁰ G ⁰ YV ⁰ NIT ⁰ VQR
<i>AqMAGI1-3</i> (UPI00021A41C9)	:	EKVVV ⁰ GIVK ⁰ GPL ⁰ CF ⁰ GF ⁻² SL ⁰ ET ⁰ PQ--GP ⁰ IVK ⁰ Q ⁰ IM ⁰ DI ⁰ PRCA ⁰ QL ⁰ REGDL ⁰ IT ⁰ ELNG ⁰ Q ⁰ N ⁰ V ⁰ LS ⁰ LS ⁰ TS ⁰ DL ⁰ IT ⁰ LR ⁰ CK ⁰ PK ⁰ G ⁰ NTAN ⁰ FL ⁰ VSR
<i>OcMAGI1-3</i> (g10427)	:	EIIDAS ⁰ VVK ⁰ GPY ⁰ CF ⁰ GF ⁻² TIAD ⁰ TQR--GQ ⁰ K ⁰ VR ⁰ Q ⁰ IL ⁰ D ⁰ HER ⁰ CE ⁰ LQ ⁰ EK ⁰ DL ⁰ LLE ⁰ IDG ⁰ EL ⁰ VR ⁰ NY ⁰ DE ⁰ ET ⁰ V ⁰ TV ⁰ IK ⁰ G ⁰ QR ⁰ GN ⁰ K ⁰ VR ⁰ FL ⁰ VR

D

<i>SrMAGI1-1</i> (F2U2A7)	:	HVHLTHEVKRPVEGG ⁰ CF ⁰ FTVAGMYPEKM--LI ⁰ H ⁰ TV ⁰ FK ⁰ TSP--AYG ⁰ KLQ ⁰ PGD ⁰ HLLAV ⁰ NGV ⁰ DT ⁰ KL ⁰ THEML ⁰ DL ⁰ SRV ⁰ PR ⁰ GLV ⁰ LF ⁰ SI
<i>SrMAGI1-2</i> (F2U2A7)	:	PARLKLTIQRPPSGTYGFA ⁰ ITYDEATDC-YRV ⁰ SH ⁰ LK ⁰ SGSV ⁰ VDNEGLV ⁰ RIGDK ⁰ IVK ⁰ NGR ⁰ DVSN ⁰ CD ⁰ HE ⁰ VEL ⁰ RQ ⁰ FPD ⁰ HV ⁰ LV ⁰ LR
<i>SrMAGI1-3</i> (F2U2A7)	:	ASRVEIVTFQRLSGSE ⁰ CF ⁰ TVVGGY ⁰ GSSV ⁰ PPM ⁰ GK ⁰ VL ⁰ PN ⁰ SAA-LHAD ⁰ LRV ⁰ G ⁰ D ⁰ L ⁰ LN ⁰ ED ⁰ V ⁰ G ⁰ SL ⁰ SH ⁰ Q ⁰ EV ⁰ LD ⁰ R ⁰ IKAS ⁰ PER ⁰ IV ⁰ LT ⁰ ISR
<i>CoMAGI1-1</i> (E9C6R5)	:	DVGR ⁰ TVK ⁰ LNRR ⁰ EGDV ⁰ CF ⁰ FG ⁰ FG ⁰ GGSE ⁰ HD ⁰ LT ⁰ LSI ⁰ SK ⁰ LV ⁰ PGAVAD ⁰ ASQ ⁰ VLIG ⁰ DEV ⁰ QSINN ⁰ Q ⁰ IVV ⁰ GATH ⁰ DEA ⁰ SL ⁰ IK ⁰ QS ⁰ ST ⁰ TL ⁰ LS ⁰ LR ⁰ Y

Fig. 5. Conservation of selected PDZ domains from MAGI. (A) A peptide derived from the C-terminus of the human papilloma virus E6 protein is shown in complex with human MAGI1 PDZ2 (PDB: 2KPL). E6 residues “0” and “-2” (green) bind in a cleft formed by the PDZ fold; residues well conserved in all PDZ domains (purple), and residues found to be important for defining peptide-binding specificity (yellow) are shown. Also shown are the sequence and corresponding secondary structure of the PDZ domain. (B,C) Sequence alignments for selected PDZ domains from Metazoan MAGI are depicted, with the same colors as in panel A. Analogous proteins are represented by their Uniprot/Compagen identifiers. (D) The PDZ sequences of the pre-metazoan MAGI analogs are depicted, with the same colors as in panels A–C.

the PDZ-binding motif of PTEN from *O. carmela* (Fig. 5C). Thus, via its multiple PDZ domains, even the most basal metazoan MAGI analogous protein seems capable of integrating multiple proteins, e.g. β -catenin and PTEN, at AJs. This is in stark contrast to the pre-metazoan MAGI proteins – from *C. owczarzewski* and *S. rosetta* – which possess one and three PDZ domains, respectively, none of which have similarity to HsMAGI PDZ3 or PDZ6 (Fig. 5D). Thus, the innovation of metazoan catenins with PDZ binding motifs is coupled to the appearance in the most basal metazoa of PDZ domains, both in the form of novel proteins, and in new or altered PDZ domains within existing proteins.

Expansion of the cadhesome in metazoa

We found that ~87% (97/112) of the “simplified” cadhesome existed in the most basal metazoa. Fig. 4 shows that only 15 novel proteins or protein families appeared since sponges diverged from the other metazoa, including Tara (TRIOBP), SHROOM3, palladin (PALLD), and paracingulin (CGNL1), all of which are either actin- or myosin-binding proteins. This suggests an increased variety of cytoskeletal attachment. Though only a few novel protein families appeared in higher metazoa, the total number of known cadhesome components nearly doubled, from ~100 in basal metazoa to ~170 in vertebrates

(supplementary material Fig. S3). To analyze this expansion in more detail, we focused our attention on cadherins, cadherin-binding adaptors, actin-binding adaptors, and other adaptors. Fig. 6 shows the total number of members per group in each phylum, from amoebozoa/apusozoa to vertebrates. Significant vertebrate expansion is exhibited by all three adaptor groups, including proteins characterized by three LIM domains (LPP, TRIP6, FBLIM1, JUB, and ZYX), which have been implicated in mechano-sensing (Schiller et al., 2011). The most striking example of expansion is experienced by the cadherins. Outside of vertebrates, all metazoa examined have at most three “classical”-like cadherins; within vertebrates there are between 12 and 18 “classical” cadherins of type I and II.

DISCUSSION

It is now well established that cadherins predate metazoa, existing in choanoflagellata and filasterea (Abedin and King, 2008; Nichols et al., 2006), early unikonts such as the apusozoa, *T. trahens* (identified herein), and earlier eukaryotes such as the plant pathogen, *Pythium ultimum* (Nichols et al., 2012), but only metazoa have cadherins with a clearly conserved JMD and CBD (Fig. 2) (Miller et al., 2013). Our search of pre-metazoa for “classical” cadherins seems to confirm this: we found no clear JMD or CBD outside of metazoa, and no JMD GBM or CBD

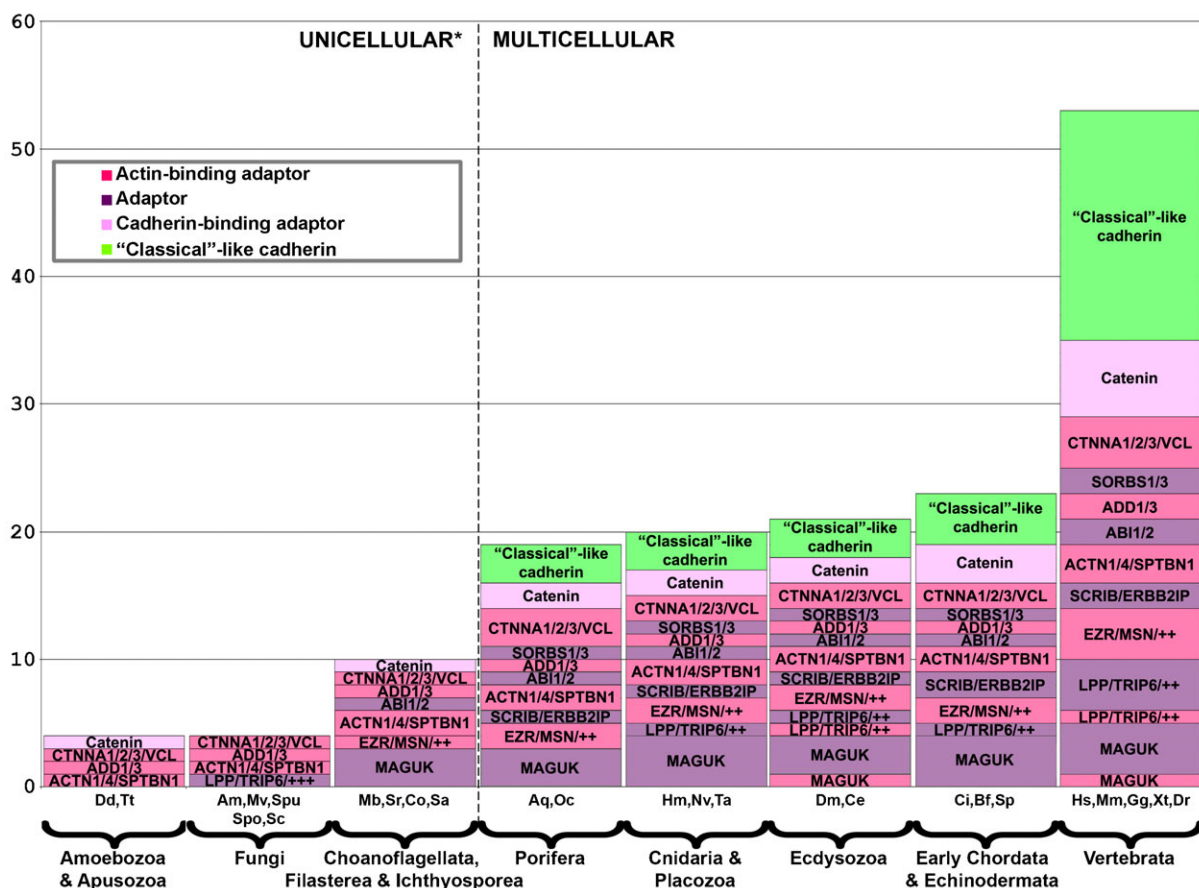


Fig. 6. Expansion of cadhesome gene families throughout evolution. “Classical” cadherin-like proteins, cadherin-binding adaptors, actin-binding adaptors, and adaptors are portrayed. These categories were divided up into the same groups as the “simplified” cadhesome, except for ACTN1/4/SPTBN1 and the MAGUK proteins, which were more broadly clustered (see Materials and Methods). The total number of members in each group (y-axis) was plotted over evolutionary time (x-axis), as represented by the evolutionary ages of the phyla/kingdoms, which are ordered from left to right by increasing age. The dashed line represents the transition from uni- to multicellularity. Gene families are colored according to the legend.

GBM. It has recently been suggested that cadherin-based adhesion may derive from a cadherin-based bacterial feeding system, incorporated from a unicellular ancestor, and perhaps still present in modern choanoflagellates (Abedin and King, 2008). In this system, choanoflagellate cadherins may bind to extra-cellular receptors on bacteria, which triggers phagocytosis of the prey-bound cadherin, similar in principle to how *Listeria* triggers endocytosis via E-cadherin in humans (Mengaud et al., 1996). Consistent with this, the choanoflagellate cadherins, *MbCDH1* and 2, have been localized to actin-filled microvilli in the feeding collar of *M. brevicollis* (Abedin and King, 2008). Furthermore, transcriptomic analysis has demonstrated that a subset of cadherins is upregulated in *S. rosetta* thecate cells, perhaps linking them to feeding or substrate adhesion (Fairclough et al., 2013). A different set of cadherins is upregulated in colonies, in which cells have been observed to be linked by intercellular bridges and a shared extracellular matrix (Dayel et al., 2011).

Our search of unicellular holozoa with β -catenin and Aardvark identified at least six Aardvark-like sequences in *S. rosetta*, three in *S. arctica*, and one in *C. owczarzaki* (Fig. 3). When we analyzed recent *S. rosetta* transcriptomics data (Fairclough et al., 2013) we found that one of these Aardvark-like sequences is upregulated in thecate cells, along with protein analogs of α -catenin/vinculin, IQGAP, and myosin II, which together function to establish epithelial-like polarity in *Dictyostelium* (Dickinson et al., 2012a), as well as four membrane-spanning cadherins. As noted above, none of these cadherins have a clear catenin-binding JMD or CBD, but, in theory, this places the “core” elements of AJs together in the same cell type of the same organism. Furthermore, the organization of the *Dictyostelium* polarity proteins (Fig. 7A) (Dickinson et al., 2012b) fits well with a model layout of an *S. rosetta* thecate cell, with Aardvark on the collar end, possibly co-localized with cadherins (Abedin and King, 2008), and myosin II on the opposite end, from which cellulose is secreted to form the theca (Fig. 7B). Though this suggests *S. rosetta* thecate cells are more relevant to metazoa

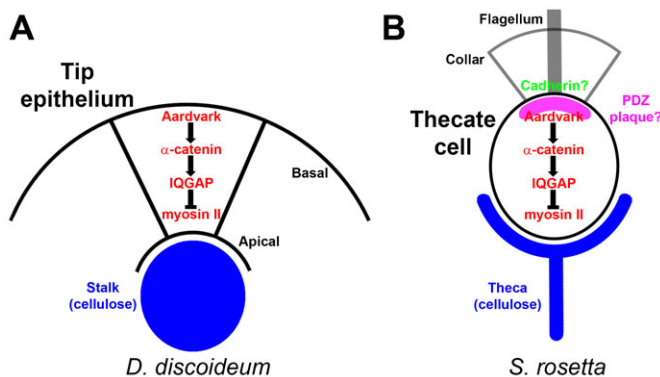


Fig. 7. Model of *Dictyostelium* polarity network in *S. rosetta* thecate cells. (A) The proposed organization of the *Dictyostelium* polarity network is depicted for cells of the *Dictyostelium* tip epithelium (Dickinson et al., 2012b). Aardvark is located at the basal end of the cell, with myosin II at the apical end, where cellulose is secreted to form the stalk (in blue). (B) Our hypothetical model of polarity network organization is depicted for *S. rosetta* thecate cells. Analogous to *Dictyostelium* tip cells, the theca (blue), which is comprised of cellulose, is secreted from the “apical” end, where myosin II is located. Aardvark is located at the “basal” end by the feeding collar, where choanoflagellate cadherins are known to localize. Some of these cadherins possess PDZ-binding motifs, so we also placed the theoretical thecate PDZ plaque (magenta) on the basal end with cadherins.

multicellularity than the colonial state, we also note co-upregulation of cadherins and Aardvark-like sequences in colonies, perhaps suggesting a greater diversity in these organisms of the link between cadherins and proteins related to the catenins. It is interesting to note that the Aardvark-like proteins we identified in *S. rosetta* contain Toll/interleukin-1 receptor (TIR) domains, which have a conserved role in environmental sensing and innate immunity (O’Neill, 2000), raising the intriguing possibility that early cell–cell adhesion structures recruited signaling modules from the innate immunity pathway. Of the other unicellular holozoa we analyzed, only *C. owczarzaki* has all the basic elements of the cadherin–catenin “core”, though it has only two cadherins.

Many studies have demonstrated that AJs are more than simply the cadherin–catenin “core”. Searching the literature, we previously found over 170 proteins reported to be associated with the “core”. We call this network of proteins and their interactions the cadhesome (Zaidel-Bar, 2013). To determine when in evolution cadhesome proteins first appeared we first clustered similar proteins into families, yielding the “simplified” cadhesome, and then determined in what phyla a representative of each family first appeared (Fig. 4). This is depicted in graphical form in Fig. 8, where evolutionary age is color-coded, function is shape-coded, and position is a function of whether the human protein interacts directly or indirectly with the “core”. As evident in Fig. 8, protein families with pre-metazoa origins (colored red, orange, and yellow) make up ~70% of this “simplified” cadhesome, and many of these interact with the “core” (innermost shell). That so many ancient proteins interact with the “core” directly (“shell 1”), coupled with the pre-metazoa origins of the catenins, seems to suggest that much of the AJ regulatory network may have been in place prior to metazoa. This helps to explain how the transition from uni- to multicellular life forms could take place, but it also elicits important questions: for example, what was the function of cadhesome proteins before AJs existed? How were so many pre-existing proteins and networks incorporated and integrated by so few novel proteins?

Forty proteins or protein families of the “simplified” cadhesome (black box in Fig. 8) overlap with the similarly compiled integrin adhesome (Zaidel-Bar and Geiger, 2010; Zaidel-Bar et al., 2007; Zaidel-Bar, 2009). According to our analysis, 35 of these 40 families predate metazoa. As focal adhesions predate AJs (Seb e-Pedr os et al., 2010), inclusion of these components in cell–matrix adhesion may explain the pre-metazoa role of some cell–cell adhesion components. Other cadhesome components may have performed similar roles in pre-metazoa, albeit under very different circumstances. Vertebrate VEZT serves to link the cadherin–catenin “core” to MYO7A, thereby strengthening adhesions (K ussel-Andermann et al., 2000), but it is also crucial for *Listeria* entry into epithelial cells (Sousa et al., 2004). We also found VEZT and MYO7A in *S. rosetta* and *Dictyostelium* where they may serve to internalize bacteria for feeding or other non-infection purposes. Thus, by recruiting VEZT to AJs, the first metazoa may have also recruited MYO7A. Indeed, it may have been possible to incorporate small multi-protein networks into the cadhesome, by way of only one of the proteins in the complexes, e.g. Aardvark or VEZT.

We find that over half of the novel proteins in unicellular holozoa and basal metazoa were modular adaptor proteins, specifically PDZ adaptors (Figs 4 and 5). This innovation occurred concurrently with the development in metazoa of catenins with PDZ-binding motifs (supplementary material

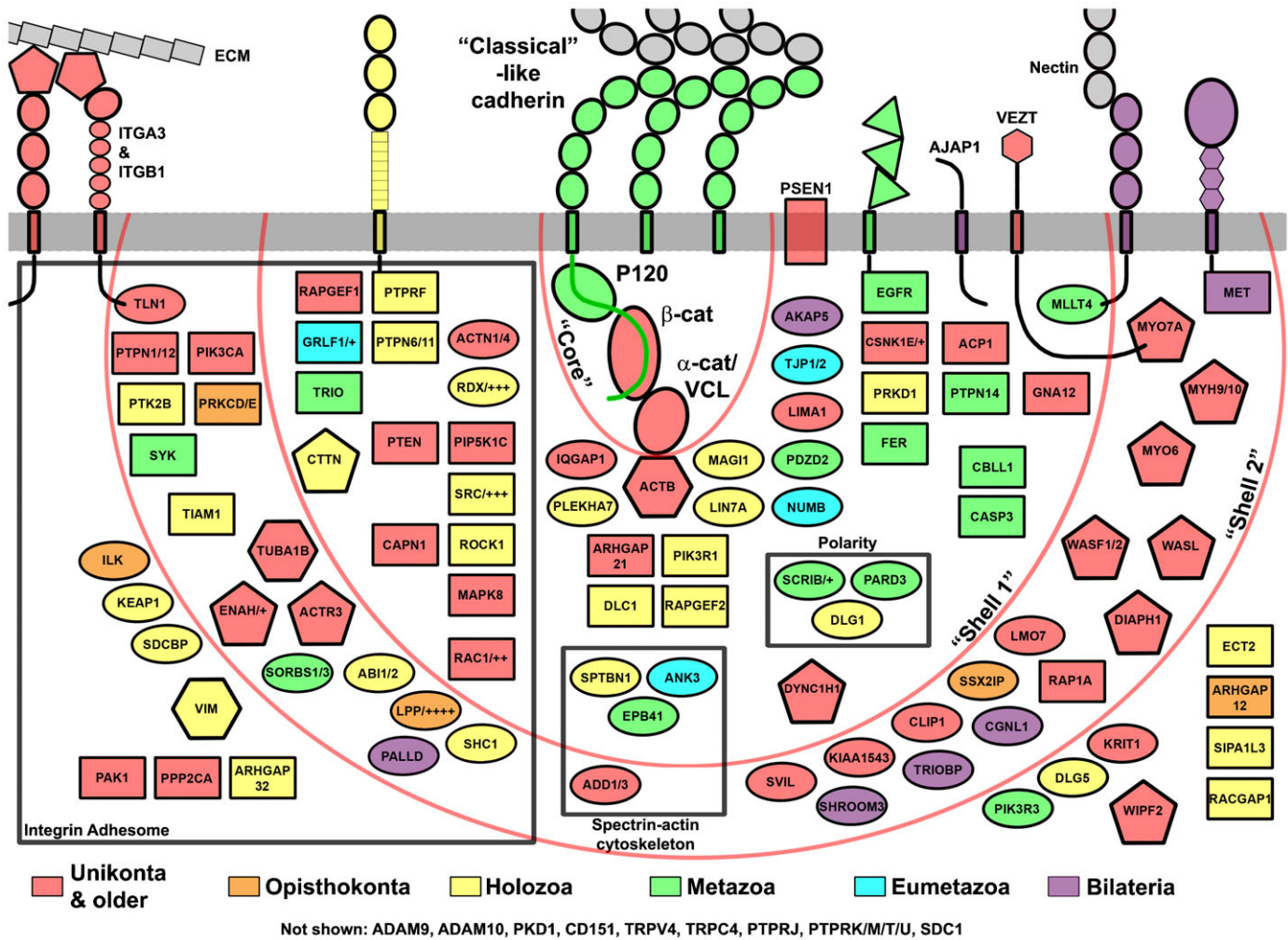


Fig. 8. Evolutionary age of cadhesome components. Proteins and protein families of the “simplified” cadhesome are colored according to when in evolution they appeared, and organized based on how they interact with the “classical” cadherin–catenin “core” (central semi-circle). Protein age is distinguished by color: unikont origin or older (red), opisthokont (orange), holozoa (yellow), metazoan (green), eumetazoa (cyan), and bilateria/vertebrata (violet). Some protein families are represented by one member of the family (LPP/++++), with each “+” signifying an additional family member. Components that interact directly with the cytoplasmic tail of cadherin or one of the catenins in vertebrates are located within “Shell 1”; components that interact indirectly with cadherin–catenin – i.e. via components in “Shell 1” – are located within “Shell 2”. All remaining components are located outside the two shells. Component types are distinguished by shape: adaptors (oval), regulatory proteins (rectangle), actin dynamics regulators (pentagon), cytoskeleton (hexagon), with receptors spanning the plasma membrane (and extracellular domains drawn to schematically show their domain structure). Cadhesome receptors are colored based on when in our analysis they appear as full membrane-spanning proteins with both cytoplasmic and extracellular domains. Boxed are components shared with other systems, including the spectrin–actin cytoskeleton, metazoan polarity complexes, and the integrin adhesome.

Table S1). In theory, these two innovations may have served to link the cadherin–catenin “core” to many other proteins in the cadhesome, thereby integrating multiple proteins at the origin of metazoa. The principle of domain modularity is well appreciated in the context of evolution: domain shuffling, the addition of pre-existing domains to proteins, and the innovation of novel domains, all open up new sets of interactions and functions (Pawson and Nash, 2003; Jin et al., 2009). Indeed, many of these processes seem to have occurred in choanoflagellates (King et al., 2008; Lim and Pawson, 2010; Fairclough et al., 2013). Judging by the overall rise in PDZs around the transition to multicellularity, and their use as protein–protein interaction modules in adaptors from large complexes, e.g. the post-synaptic density, PDZ domains seem well suited to this purpose (Fahey and Degnan, 2010; Sakarya et al., 2010; Sakarya et al., 2007). Furthermore, few changes are required to transform the unstructured C-terminus of a given

protein into a PDZ-binding motif (Kim et al., 2012), as seems likely to have taken place with many cadherins (Demontis et al., 2006), including some in choanoflagellates (supplementary material Table S4). Interestingly, we found that many PDZ domains, including three from the cadhesome, are upregulated in *S. rosetta* thecate cells together with specific cadherins and the *Dictyostelium* polarity network. This may hint at the existence of a pre-metazoa PDZ–protein plaque, similar to that of AJs or the post-synaptic density, which may interact with choanoflagellate cadherins, some of which have PDZ-binding motifs (Fig. 7B). What is clear from our analysis is that multiple-PDZ adaptor proteins were a critical development at the dawn of multicellularity, as they allowed for the incorporation and integration of pre-metazoa proteins in the first metazoa.

The primary mode of cadhesome expansion in the 600 million years since the emergence of metazoa was gene duplication

(Fig. 6). This expansion likely reflects both a growing need for complexity, within and between cells, and the increased specialization of cells and tissue; furthermore, it coincides with a well-known whole-genome duplication event in early vertebrates (Lundin, 1999). In many cases, the paralogs conserve the primary functionality of the original component, while gaining or losing secondary functions. For example, all members of the vertebrate P120 family bind to the JMD of “classical” cadherins. However, unlike the most ancestral of the P120 family (Zhao et al., 2011; Carnahan et al., 2010), P120 has no PDZ binding motif at its C-terminus, which suggests it has been decoupled from PDZ adaptor proteins and their functionalities. The functional diversity of catenins is thus enabled by modularity: the N- and C-termini of the P120 family are free to mutate and sample new functions, while the cadherin-binding ARM repeats remain relatively unchanged.

The evolution of the cadhesome network appears to be the product of a variety of genetic processes: incorporation, integration, innovation, and expansion. In elucidating these processes, we have identified a number of interactions conserved in sequence and possibly structure across hundreds of millions of years, and constructed a plausible evolutionary history of cadherin-based adhesion systems. This reconstruction makes specific testable predictions by evolutionary inference. For example, we would expect the MAGI and β -catenin analogs from *O. carmela* to bind each other. The next step will be to experimentally test these hypotheses, as has been done for related questions in *D. discoideum* and the sponge, *O. carmela* (Dickinson et al., 2011; Nichols et al., 2012). Much of the integrin adhesome (81) has been recently validated by proteomic approaches (reviewed by Geiger and Zaidel-Bar, 2012). On top of validating a systems-level approach for studying adhesion junctions, hundreds of additional components were identified, suggesting our current view of junctions is still overly simplistic. As more experimental data will expand our view of the cadherin adhesome, an evolutionary perspective, such as is presented here, coupled with advances in methods such as transcriptomics and proteomics, will undoubtedly help make sense of the complex biology presented to us.

Acknowledgements

The authors thank Barry Honig and Larry Shapiro (Columbia University) for many helpful discussions and their continuous support.

Competing interests

The authors have no competing interests to declare.

Author contributions

P.S.M. generated all the data, performed analysis and modeling and wrote the paper; R.Z.-B. conceived the project, analyzed the data and wrote the paper.

Funding

This work was supported in part by the National Research Foundation Singapore under its NRF fellowship (NRF-RF2009-RF001-074) awarded to R.Z.-B.

References

- Abedin, M. and King, N. (2008). The premetazoan ancestry of cadherins. *Science* **319**, 946–948.
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.
- Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M. et al. (2004). UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res.* **32**, D115–D119.
- Baum, B. and Georgiou, M. (2011). Dynamics of adherens junctions in epithelial establishment, maintenance, and remodeling. *J. Cell Biol.* **192**, 907–917.
- Bertocchi, C., Vaman Rao, M. and Zaidel-Bar, R. (2012). Regulation of adherens junction dynamics by phosphorylation switches. *J. Signal Transduct.* **2012**, 125295.
- Biegert, A. and Söding, J. (2008). De novo identification of highly diverged protein repeats by probabilistic consistency. *Bioinformatics* **24**, 807–814.
- Bowie, J. U., Lüthy, R. and Eisenberg, D. (1991). A method to identify protein sequences that fold into a known three-dimensional structure. *Science* **253**, 164–170.
- Braga, V. M. (2002). Cell–cell adhesion and signalling. *Curr. Opin. Cell Biol.* **14**, 546–556.
- Brasch, J., Harrison, O. J., Honig, B. and Shapiro, L. (2012). Thinking outside the cell: how cadherins drive adhesion. *Trends Cell Biol.* **22**, 299–310.
- Carnahan, R. H., Rokas, A., Gaucher, E. A. and Reynolds, A. B. (2010). The molecular evolution of the p120-catenin subfamily and its functional associations. *PLoS ONE* **5**, e15747.
- Cavallaro, U. and Dejana, E. (2011). Adhesion molecule signalling: not always a sticky business. *Nat. Rev. Mol. Cell Biol.* **12**, 189–197.
- Cole, C., Barber, J. D. and Barton, G. J. (2008). The Jpred 3 secondary structure prediction server. *Nucleic Acids Res.* **36**, W197–W201.
- Cornillon, S., Gebbie, L., Benghezal, M., Nair, P., Keller, S., Wehrle-Haller, B., Charette, S. J., Brückert, F., Letourneur, F. and Cosson, P. (2006). An adhesion molecule in free-living Dictyostelium amoebae with integrin beta features. *EMBO Rep.* **7**, 617–621.
- Dayel, M. J., Alegado, R. A., Fairclough, S. R., Levin, T. C., Nichols, S. A., McDonald, K. and King, N. (2011). Cell differentiation and morphogenesis in the colony-forming choanoflagellate *Salpingoeca rosetta*. *Dev. Biol.* **357**, 73–82.
- de Mendoza, A., Suga, H. and Ruiz-Trillo, I. (2010). Evolution of the MAGUK protein gene family in premetazoan lineages. *BMC Evol. Biol.* **10**, 93.
- Demontis, F., Habermann, B. and Dahmann, C. (2006). PDZ-domain-binding sites are common among cadherins. *Dev. Genes Evol.* **216**, 737–741.
- Desai, R., Sarpal, R., Ishiyama, N., Pellikka, M., Ikura, M. and Tepass, U. (2013). Monomeric α -catenin links cadherin to the actin cytoskeleton. *Nat. Cell Biol.* **15**, 261–273.
- Di Tommaso, P., Moretti, S., Xenarios, I., Orobítz, M., Montanyola, A., Chang, J. M., Taly, J. F. and Notredame, C. (2011). T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res.* **39**, W13–W17.
- Dickinson, D. J., Nelson, W. J. and Weis, W. I. (2011). A polarized epithelium organized by beta- and alpha-catenin predates cadherin and metazoan origins. *Science* **331**, 1336–1339.
- Dickinson, D. J., Nelson, W. J. and Weis, W. I. (2012a). An epithelial tissue in Dictyostelium challenges the traditional origin of metazoan multicellularity. *BioEssays* **34**, 833–840.
- Dickinson, D. J., Robinson, D. N., Nelson, W. J. and Weis, W. I. (2012b). α -catenin and IQGAP regulate myosin localization to control epithelial tube morphogenesis in Dictyostelium. *Dev. Cell* **23**, 533–546.
- Dobrosotskaya, I. Y. and James, G. L. (2000). MAGI-1 interacts with beta-catenin and is associated with cell–cell adhesion structures. *Biochem. Biophys. Res. Commun.* **270**, 903–909.
- Drees, F., Pokutta, S., Yamada, S., Nelson, W. J. and Weis, W. I. (2005). Alpha-catenin is a molecular switch that binds E-cadherin-beta-catenin and regulates actin-filament assembly. *Cell* **123**, 903–915.
- Dupre-Crochet, S., Figueroa, A., Hogan, C., Ferber, E. C., Bialucha, C. U., Adams, J., Richardson, E. C. and Fujita, Y. (2007). Casein kinase 1 is a novel negative regulator of E-cadherin-based cell–cell contacts. *Mol. Cell Biol.* **27**, 3804–3816.
- Fahey, B. and Degnan, B. M. (2010). Origin of animal epithelia: insights from the sponge genome. *Evol. Dev.* **12**, 601–617.
- Fairclough, S. R., Chen, Z., Kramer, E., Zeng, Q., Young, S., Robertson, H. M., Begovic, E., Richter, D. J., Russ, C., Westbrook, M. J. et al. (2013). Premetazoan genome evolution and the regulation of cell differentiation in the choanoflagellate *Salpingoeca rosetta*. *Genome Biol.* **14**, R15.
- Fujita, Y., Krause, G., Scheffner, M., Zechner, D., Leddy, H. E., Behrens, J., Sommer, T. and Birchmeier, W. (2002). Hakai, a c-Cbl-like protein, ubiquitinates and induces endocytosis of the E-cadherin complex. *Nat. Cell Biol.* **4**, 222–231.
- Fukata, M., Kuroda, S., Nakagawa, M., Kawajiri, A., Itoh, N., Shoji, I., Matsuura, Y., Yonehara, S., Fujisawa, H., Kikuchi, A. et al. (1999). Cdc42 and Rac1 regulate the interaction of IQGAP1 with beta-catenin. *J. Biol. Chem.* **274**, 26044–26050.
- Geiger, T. and Zaidel-Bar, R. (2012). Opening the floodgates: proteomics and the integrin adhesome. *Curr. Opin. Cell Biol.* **24**, 562–568.
- Gough, J., Karplus, K., Hughey, R. and Chothia, C. (2001). Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *J. Mol. Biol.* **313**, 903–919.
- Graham, T. A., Weaver, C., Mao, F., Kimelman, D. and Xu, W. (2000). Crystal structure of a beta-catenin/Tcf complex. *Cell* **103**, 885–896.
- Grimson, M. J., Coates, J. C., Reynolds, J. P., Shipman, M., Blanton, R. L. and Harwood, A. J. (2000). Adherens junctions and beta-catenin-mediated cell signalling in a non-metazoan organism. *Nature* **408**, 727–731.
- Harris, T. J. and Tepass, U. (2010). Adherens junctions: from molecules to morphogenesis. *Nat. Rev. Mol. Cell Biol.* **11**, 502–514.
- Harrison, O. J., Jin, X., Hong, S., Bahna, F., Ahlsen, G., Brasch, J., Wu, Y., Vendome, J., Felsővalyi, K., Hampton, C. M. et al. (2011). The extracellular architecture of adherens junctions revealed by crystal structures of type I cadherins. *Structure* **19**, 244–256.

- Hartscock, A. and Nelson, W. J. (2008). Adherens and tight junctions: structure, function and connections to the actin cytoskeleton. *Biochim. Biophys. Acta* **1778**, 660–669.
- Hemrich, G. and Bosch, T. C. (2008). Compagen, a comparative genomics platform for early branching metazoan animals, reveals early origins of genes regulating stem-cell differentiation. *BioEssays* **30**, 1010–1018.
- Hong, S., Troyanovsky, R. B. and Troyanovsky, S. M. (2013). Binding to F-actin guides cadherin cluster assembly, stability, and movement. *J. Cell Biol.* **201**, 131–143.
- Huber, A. H. and Weis, W. I. (2001). The structure of the beta-catenin/E-cadherin complex and the molecular basis of diverse ligand recognition by beta-catenin. *Cell* **105**, 391–402.
- Huber, A. H., Stewart, D. B., Laurents, D. V., Nelson, W. J. and Weis, W. I. (2001). The cadherin cytoplasmic domain is unstructured in the absence of beta-catenin. A possible mechanism for regulating cadherin turnover. *J. Biol. Chem.* **276**, 12301–12309.
- Hulpiau, P. and van Roy, F. (2009). Molecular evolution of the cadherin superfamily. *Int. J. Biochem. Cell Biol.* **41**, 349–369.
- Hulpiau, P. and van Roy, F. (2011). New insights into the evolution of metazoan cadherins. *Mol. Biol. Evol.* **28**, 647–657.
- Ishiyama, N., Lee, S. H., Liu, S., Li, G. Y., Smith, M. J., Reichardt, L. F. and Ikura, M. (2010). Dynamic and static interactions between p120 catenin and E-cadherin regulate the stability of cell–cell adhesion. *Cell* **141**, 117–128.
- Jin, J., Xie, X., Chen, C., Park, J. G., Stark, C., James, D. A., Olhovskiy, M., Linding, R., Mao, Y. and Pawson, T. (2009). Eukaryotic protein domains as functional units of cellular evolution. *Sci. Signal.* **2**, ra76.
- Kelley, L. A. and Sternberg, M. J. (2009). Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc.* **4**, 363–371.
- Kim, J., Kim, I., Yang, J. S., Shin, Y. E., Hwang, J., Park, S., Choi, Y. S. and Kim, S. (2012). Rewiring of PDZ domain–ligand interaction network contributed to eukaryotic evolution. *PLoS Genet.* **8**, e1002510.
- King, N., Westbrook, M. J., Young, S. L., Kuo, A., Abedin, M., Chapman, J., Fairclough, S., Hellsten, U., Isogai, Y., Letunic, I. et al. (2008). The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* **451**, 783–788.
- Kotelevets, L., van Hengel, J., Bruyneel, E., Mareel, M., van Roy, F. and Chastre, E. (2005). Implication of the MAGI-1b/PTEN signalosome in stabilization of adherens junctions and suppression of invasiveness. *FASEB J.* **19**, 115–117.
- Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580.
- Küssel-Andermann, P., El-Amraoui, A., Safieddine, S., Nouaille, S., Perfettini, I., Lecuit, M., Cossart, P., Wolfmuth, U. and Petit, C. (2000). Vezatin, a novel transmembrane protein, bridges myosin VIIA to the cadherin–catenins complex. *EMBO J.* **19**, 6020–6029.
- Lecuit, T. (2005). Adhesion remodeling underlying tissue morphogenesis. *Trends Cell Biol.* **15**, 34–42.
- Letunic, I., Doerks, T. and Bork, P. (2012). SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res.* **40**, D302–D305.
- Lim, W. A. and Pawson, T. (2010). Phosphotyrosine signaling: evolving a new cellular communication system. *Cell* **142**, 661–667.
- Lundin, L. G. (1999). Gene duplications in early metazoan evolution. *Semin. Cell Dev. Biol.* **10**, 523–530.
- Lüthy, R., Bowie, J. U. and Eisenberg, D. (1992). Assessment of protein models with three-dimensional profiles. *Nature* **356**, 83–85.
- Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., Fong, J. H., Geer, L. Y., Geer, R. C., Gonzales, N. R. et al. (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res.* **39**, D225–D229.
- Marchler-Bauer, A., Zheng, C., Chitsaz, F., Derbyshire, M. K., Geer, L. Y., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Lanczycki, C. J. et al. (2013). CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Res.* **41**, D348–D352.
- McGuffin, L. J., Bryson, K. and Jones, D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404–405.
- Mengaud, J., Ohayon, H., Gounon, P., Mege R-M, and Cossart, P. (1996). E-cadherin is the receptor for internalin, a surface protein required for entry of *L. monocytogenes* into epithelial cells. *Cell* **84**, 923–932.
- Miller, P. W., Clarke, D. N., Weis, W. I., Lowe, C. J. and Nelson, W. J. (2013). The evolutionary origin of epithelial cell–cell adhesion mechanisms. *Curr Top Membr.* **72**, 267–311.
- Mukherjee, M., Chow, S. Y., Yusoff, P., Seetharaman, J., Ng, C., Sinniah, S., Koh, X. W., Asgar, N. F., Li, D., Yim, D. et al. (2012). Structure of a novel phosphotyrosine-binding domain in Hakai that targets E-cadherin. *EMBO J.* **31**, 1308–1319.
- Nichols, S. A., Dirks, W., Pearse, J. S. and King, N. (2006). Early evolution of animal cell signaling and adhesion genes. *Proc. Natl. Acad. Sci. USA* **103**, 12451–12456.
- Nichols, S. A., Roberts, B. W., Richter, D. J., Fairclough, S. R. and King, N. (2012). Origin of metazoan cadherin diversity and the antiquity of the classical cadherin/β-catenin complex. *Proc. Natl. Acad. Sci. USA* **109**, 13046–13051.
- Niessen, C. M., Leckband, D. and Yap, A. S. (2011). Tissue organization by cadherin adhesion molecules: dynamic molecular and cellular mechanisms of morphogenetic regulation. *Physiol. Rev.* **91**, 691–731.
- Norel, R., Petrey, D. and Honig, B. (2010). PUDGE: a flexible, interactive server for protein structure prediction. *Nucleic Acids Res.* **38**, W550–W554.
- Noritake, J., Watanabe, T., Sato, K., Wang, S. and Kaibuchi, K. (2005). IQGAP1: a key regulator of adhesion and migration. *J. Cell Sci.* **118**, 2085–2092.
- Notredame, C., Higgins, D. G. and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217.
- O’Neill, L. A. (2000). The interleukin-1 receptor/Toll-like receptor superfamily: signal transduction during inflammation and host defense. *Sci. STKE* **2000**, re1.
- Oda, H. and Takeichi, M. (2011). Evolution: structural and functional diversity of cadherin at the adherens junction. *J. Cell Biol.* **193**, 1137–1146.
- Palidwor, G. A., Shcherbinin, S., Huska, M. R., Rasko, T., Stelzl, U., Arumughan, A., Foulle, R., Porras, P., Sanchez-Pulido, L., Wanker, E. E. et al. (2009). Detection of alpha-rod protein repeats using a neural network and application to huntingtin. *PLoS Comput. Biol.* **5**, e1000304.
- Pawson, T. and Nash, P. (2003). Assembly of cell regulatory systems through protein interaction domains. *Science* **300**, 445–452.
- Pokutta, S. and Weis, W. I. (2000). Structure of the dimerization and beta-catenin-binding region of alpha-catenin. *Mol. Cell* **5**, 533–543.
- Pollastri, G., Przybylski, D., Rost, B. and Baldi, P. (2002). Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. *Proteins* **47**, 228–235.
- Ratheesh, A. and Yap, A. S. (2012). A bigger picture: classical cadherins and the dynamic actin cytoskeleton. *Nat. Rev. Mol. Cell Biol.* **13**, 673–679.
- Richter, D. J. and King, N. (2013). The genomic and cellular foundations of animal origins. *Annu. Rev. Genet.* **47**, 509–537.
- Roy, A., Kucukural, A. and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5**, 725–738.
- Ruiz-Trillo, I., Burger, G., Holland, P. W., King, N., Lang, B. F., Roger, A. J. and Gray, M. W. (2007). The origins of multicellularity: a multi-taxon genome initiative. *Trends Genet.* **23**, 113–118.
- Ruiz-Trillo, I., Roger, A. J., Burger, G., Gray, M. W. and Lang, B. F. (2008). A phylogenomic investigation into the origin of metazoa. *Mol. Biol. Evol.* **25**, 664–672.
- Sakarya, O., Armstrong, K. A., Adamska, M., Adamski, M., Wang, I. F., Tidor, B., Degnan, B. M., Oakley, T. H. and Kosik, K. S. (2007). A post-synaptic scaffold at the origin of the animal kingdom. *PLoS ONE* **2**, e506.
- Sakarya, O., Conaco, C., Egecioglu, O., Solla, S. A., Oakley, T. H. and Kosik, K. S. (2010). Evolutionary expansion and specialization of the PDZ domains. *Mol. Biol. Evol.* **27**, 1058–1069.
- Schiller, H. B., Friedel, C. C., Boulegue, C. and Fässler, R. (2011). Quantitative proteomics of the integrin adhesome show a myosin II-dependent recruitment of LIM domain proteins. *EMBO Rep.* **12**, 259–266.
- Schultz, J., Milpetz, F., Bork, P. and Ponting, C. P. (1998). SMART, a simple modular architecture research tool: identification of signaling domains. *Proc. Natl. Acad. Sci. USA* **95**, 5857–5864.
- Sebé-Pedrós, A., Roger, A. J., Lang, F. B., King, N. and Ruiz-Trillo, I. (2010). Ancient origin of the integrin-mediated adhesion and signaling machinery. *Proc. Natl. Acad. Sci. USA* **107**, 10142–10147.
- Serres, M., Filhol, L., Lickert, H., Grangeasse, C., Chambaz, E. M., Stappert, J., Vincent, C. and Schmitt, D. (2000). The disruption of adherens junctions is associated with a decrease of E-cadherin phosphorylation by protein kinase CK2. *Exp. Cell Res.* **257**, 255–264.
- Sippl, M. J. (1993). Recognition of errors in three-dimensional structures of proteins. *Proteins* **17**, 355–362.
- Sonnhammer, E. L., von Heijne, G. and Krogh, A. (1998). A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **6**, 175–182.
- Sousa, S., Cabanes, D., El-Amraoui, A., Petit, C., Lecuit, M. and Cossart, P. (2004). Unconventional myosin VIIa and vezatin, two proteins crucial for *Listeria* entry into epithelial cells. *J. Cell Sci.* **117**, 2121–2130.
- Stappert, J. and Kemler, R. (1994). A short core region of E-cadherin is essential for catenin binding and is highly phosphorylated. *Cell Adhes. Commun.* **2**, 319–327.
- Steinhilber, U., Weiske, J., Badock, V., Tauber, R., Bommert, K. and Huber, O. (2001). Cleavage and shedding of E-cadherin after induction of apoptosis. *J. Biol. Chem.* **276**, 4972–4980.
- Stepniak, E., Radice, G. L. and Vasioukhin, V. (2009). Adhesive and signaling functions of cadherins and catenins in vertebrate development. *Cold Spring Harb. Perspect. Biol.* **1**, a002949.
- Suga, H., Chen, Z., de Mendoza, A., Sebé-Pedrós, A., Brown, M. W., Kramer, E., Carr, M., Kerner, P., Vervoort, M., Sánchez-Pons, N. et al. (2013). The *Capsaspora* genome reveals a complex unicellular prehistory of animals. *Nat. Commun.* **4**, 2325.
- Thoreson, M. A., Anastasiadis, P. Z., Daniel, J. M., Ireton, R. C., Wheelock, M. J., Johnson, K. R., Hummingbird, D. K. and Reynolds, A. B. (2000). Selective uncoupling of p120(ctn) from E-cadherin disrupts strong adhesion. *J. Cell Biol.* **148**, 189–202.
- Tonikian, R., Zhang, Y., Szinsky, S. L., Currell, B., Yeh, J. H., Reva, B., Held, H. A., Appleton, B. A., Evangelista, M., Wu, Y. et al. (2008). A specificity map for the PDZ domain family. *PLoS Biol.* **6**, e239.
- Vogelmann, R., Nguyen-Tat, M. D., Giehli, K., Adler, G., Wedlich, D. and Menke, A. (2005). TGFβ-induced downregulation of E-cadherin-based cell–cell adhesion depends on PI3-kinase and PTEN. *J. Cell Sci.* **118**, 4901–4912.

- Wang, Z., Sandiford, S., Wu, C. and Li, S. S.** (2009). Numb regulates cell–cell adhesion and polarity in response to tyrosine kinase signalling. *EMBO J.* **28**, 2360–2373.
- Weissbach, L., Bernards, A. and Herion, D. W.** (1998). Binding of myosin essential light chain to the cytoskeleton-associated protein IQGAP1. *Biochem. Biophys. Res. Commun.* **251**, 269–276.
- Zaidel-Bar, R.** (2009). Evolution of complexity in the integrin adhesome. *J. Cell Biol.* **186**, 317–321.
- Zaidel-Bar, R.** (2013). Cadherin adhesome at a glance. *J. Cell Sci.* **126**, 373–378.
- Zaidel-Bar, R. and Geiger, B.** (2010). The switchable integrin adhesome. *J. Cell Sci.* **123**, 1385–1388.
- Zaidel-Bar, R., Itzkovitz, S., Ma'ayan, A., Iyengar, R. and Geiger, B.** (2007). Functional atlas of the integrin adhesome. *Nat. Cell Biol.* **9**, 858–867.
- Zhang, Y.** (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**, 40.
- Zhao, Z. M., Reynolds, A. B. and Gaucher, E. A.** (2011). The evolutionary history of the catenin gene family during metazoan evolution. *BMC Evol. Biol.* **11**, 198.
- Zhou, H. and Zhou, Y.** (2002). Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. *Protein Sci.* **11**, 2714–2726.