New Insights into the Evolution of Metazoan Cadherins

Paco Hulpiau[1] and Frans van Roy*[,1,2]

[1]Department for Molecular Biomedical Research, Flanders Institute for Biotechnology (VIB), Ghent, Belgium

[2]Department of Biomedical Molecular Biology, Ghent University, Ghent, Belgium

***Corresponding author:** E-mail: f.vanroy@dmbr.vib-ugent.be.

Associate editor: William Jeffery

Abstract

Mining newly sequenced genomes of basal metazoan organisms reveals the evolutionary origin of modern protein families. Specific cell–cell adhesion and intracellular communication are key processes in multicellular animals, and members of the cadherin superfamily are essential players in these processes. Mammalian genomes contain over 100 genes belonging to this superfamily. By a combination of tBLASTn and profile hidden Markov model analyses, we made an exhaustive search for cadherins and compiled the cadherin repertoires in key organisms, including *Branchiostoma floridae* (amphioxus), the sea anemone *Nematostella vectensis*, and the placozoan *Trichoplax adhaerens*. Comparative analyses of multiple protein domains within known and novel cadherins enabled us to reconstruct the complex evolution in metazoa of this large superfamily. Five main cadherin branches are represented in the primitive metazoan Trichoplax: classical (CDH), flamingo (CELSR), dachsous (DCHS), FAT, and FAT-like. Classical cadherins, such as E-cadherin, arose from an Urmetazoan cadherin, which progressively lost N-terminal extracellular cadherin repeats, whereas its cytoplasmic domain, which binds the armadillo proteins p120ctn and β-catenin, remained quite conserved from placozoa to man. The origin of protocadherins predates the Bilateria and is likely rooted in an ancestral FAT cadherin. Several but not all protostomians lost protocadherins. The emergence of chordates coincided with a great expansion of the protocadherin repertoire. The evolution of ancient metazoan cadherins points to their unique and crucial roles in multicellular animal life.

Key words: metazoan evolution, cadherins, protocadherins, key organisms, protein superfamily.

Introduction

Unicellular to multicellular transition occurred more than once during evolution. The conditions and genetic mechanisms were not the same for each kingdom in the tree of life. For instance, in animals, the cell adhesion molecules of the cadherin and integrin gene families played a crucial role in this transition (Rokas 2008). In plants, which lack cadherins or integrins, other molecules, such as pectins, are important for cell adhesion (Iwai et al. 2002).

Cadherins are calcium-dependent transmembrane proteins hallmarked by the extracellular cadherin repeat (EC). The biological relevance of these remarkable proteins is widely appreciated, as evidenced by the numerous studies on the functions of cadherins during early embryonic development and morphogenesis and their roles in many genetic diseases and cancer dissemination (for recent reviews, see Nelson and Fuchs 2010). Classically, cadherins are known as mediators of specific cell–cell adhesion, as demonstrated by the pioneering work of Masatoshi Takeichi and others (Takeichi 1990; van Roy and Berx 2008). However, over the years, it has become clear that cadherin family members are also very active in intracellular signal transduction. For example, they sequester the signaling armadillo proteins β-catenin and p120ctn, and they trigger important signaling cascades upon cell recognition (Nelson and Nusse 2004; van Roy and Berx 2008; McCrea and Gu 2010). Specific cell recognition influences key processes, such as planar cell polarity (Saburi and McNeill 2005) and control of cell proliferation versus cell death (Berx and van Roy 2009). Nevertheless, the unique functions of most cadherin superfamily members have not been elucidated.

Mammalian genomes contain over 100 cadherin and cadherin-related genes (Hulpiau and van Roy 2009). Surprisingly, up to 23 putative cadherin-like genes have been identified in the genome of *Monosiga brevicollis* (Abedin and King 2008). This unicellular nonmetazoan choanoflagellate is the closest known relative of the Metazoa. However, the variety in size and domain composition of the predicted choanoflagellate cadherin-like proteins makes it difficult to identify the evolutionary links with the many cadherins in the metazoan lineages. FAT, FAT-like, and Flamingo/CELSR cadherins seem to be conserved among bilaterians, but classical type-I and type-II cadherins and protocadherins, which represent more than half of the mammalian cadherin repertoires, appear to have arisen only in deuterostomians (Hulpiau and van Roy 2009).

The details of the premetazoan and metazoan evolution of the cadherin superfamily are largely unknown, but the many newly sequenced genomes of key metazoan organisms provide a rich resource for unraveling the evolution of complex protein families. We therefore scrutinized the cadherin repertoires of the recently published genomes of three basal model organisms: the lancelet *Branchiostoma floridae*, the sea anemone *Nematostella vectensis*, and the very primitive placozoan *Trichoplax adhaerens*. These organisms occupy key positions in the evolution of Metazoa.

To determine the origin and evolution of the numerous members of the cadherin superfamily, we made an extensive comparison of the domain organization and amino acid sequences of the predicted cadherin and cadherin-related proteins of ten model organisms from different metazoan lineages as well as selected cadherins of nine other organisms. Our analyses enabled us to construct reliable models for the evolution of the cadherin superfamily members in Metazoa.

Materials and Methods

Identification and Annotation of Cadherins in *B. floridae*

A dual approach was used to identify the cadherin repertoire in the cephalochordate *B. floridae* (amphioxus or lancelet). The assembly release version 2 of its genome (May, 2008), downloaded from DOE Joint Genome Institute (JGI: http://genome.jgi-psf.org/), contains 398 scaffolds (Putnam et al. 2008). First, the six-frame translation of the amphioxus genome was searched using the five existing profile hidden Markov models (HMM) in the Pfam database (Finn et al. 2010) and one newly built profile HMM based on an alignment of the first cadherin repeat (EC1), which we described previously (Hulpiau and van Roy 2009) (supplementary table S1, Supplementary Material online). All 1,050 significant domain hits, sorted by scaffold, are shown in supplementary table S2 (Supplementary Material online) and summarized in supplementary table S3 (Supplementary Material online). Second, 292 cadherin sequences from a wide variety of metazoan taxa were aligned to the amphioxus genome by using tBLASTn (Johnson et al. 2008) to identify potential orthologs (supplementary table S4 and summarized in supplementary table S5, Supplementary Material online). Finally, the results of both analyses were merged to yield the full list of cadherin superfamily members in amphioxus (supplementary fig. S1, supplementary table S6, and supplementary note S1, Supplementary Material online). In some cases, additional gene predictions using GenScan (Burge and Karlin 1998) were performed in and around the genomic region in which the putative genes are located, and these predictions were then compared with the available RefSeq data and JGI gene models. Finally, the domains in all candidate genes were annotated based on CD search (Marchler-Bauer et al. 2009) and Phobius (Kall et al. 2007). The start and end of every cadherin repeat were corrected manually. The start was taken at the beginning of the adhesion arm featuring a conserved Glu residue as position 11; the end was the DxNDxxPxF motif, which is, together with Glu11, important for calcium binding.

Identification and Annotation of Cadherins in *N. vectensis*

The approach described above for identification of cadherins in amphioxus was also used on the genome of the non-bilaterian sea anemone *N. vectensis*. The *N. vectensis* genome assembly 1.0 contains 10,804 genome scaffolds (Putnam et al. 2007). Profile HMM analysis yielded 1,016 hits (supplementary table S7, Supplementary Material online), which were grouped by scaffold and summarized in supplementary table S8 (Supplementary Material online). The results of the tBLASTn analysis of 322 bilaterian cadherin sequences (the 292 sequences listed in supplementary table S4, Supplementary Material online, plus 30 from *B. floridae*) versus the *N. vectensis* genome are listed in supplemenatary table S9 (Supplementary Material online) and summarized in supplementary table S10 (Supplementary Material online). All the results were merged and annotated into the *N. vectensis* cadherin repertoire, as described for amphioxus (supplementary fig. S2, supplementary table S11, and supplementary note S2, Supplementary Material online).

Identification and Annotation of Cadherins in *T. adhaerens*

Again, both the profile HMM and tBLASTn methods were used to identify putative placozoan cadherins encoded by the *T. adhaerens* genome. The draft release v1.0 of *T. adhaerens* Grell-BS-1999 is assembled into 1,415 scaffolds (Srivastava et al. 2008). Only 196 profile HMM hits, shown in supplementary table S12 (Supplementary Material online), were found using the six HMMs listed in supplementary table S1 (Supplementary Material online). This HMM analysis was combined with a tBLASTn analysis of 338 metazoan cadherins versus the *T. adhaerens* genome (see list in supplementary table S13, Supplementary Material online). The metazoan cadherins analyzed were the 322 sequences listed in supplementary table S9 (Supplementary Material online) plus 16 from *N. vectensis*. The results were merged and annotated into the *T. adhaerens* cadherin repertoire as described above (supplementary fig. S3, supplementary table S14, and supplementary note S3, Supplementary Material online).

Comparative and Phylogenetic Study of the Metazoan Cadherin Superfamily Members

For pairwise homology analysis of protein domains, we used "basic local alignment search tool two sequences" (bl2seq) (Johnson et al. 2008). The analysis included sequence comparison of EC blocks, EC-by-EC analyses, and comparison of non-EC domains.

For multisequence homology analysis, sequences were aligned by ClustalX2 (Larkin et al. 2007) using the PAM protein weight matrix, a gap open penalty of 5 and a gap extension penalty of 0.05 in both the pairwise and the multiple parameter settings. A neighbor joining (NJ) tree was constructed with 1,000 bootstrap replicates, and a Bayesian inference (BI) consensus tree was built by using MrBayes 3 (Ronquist and Huelsenbeck 2003) (100,000 generations, sample frequency 100, and burnin 25%). Both types of trees were drawn by using Dendroscope (Huson et al. 2007). The NJ tree is represented as a radial cladogram with bootstrap values and the BI tree as a radial phylogram with Bayesian posterior probabilities.

We used VectorNTI (Lu and Moriyama 2004) to generate Clustal W alignments of the 7EC and the 6EC ectodomains of *Ciona intestinalis* protocadherins and also of the human protocadherins. By using the AlignX module of VectorNTI, we added protein domain annotations below the relevant sequence alignment blocks. To compare individual EC domains from *Ciona* protocadherins with the corresponding genomic sequences, we performed BLAT analyses (Kuhn et al. 2009) in the University of California Santa Cruz genome browser (http://genome.ucsc.edu/cgi-bin/hgBlat?command=start). The full protein sequences of selected members in each family in the phylogenetic tree were aligned similarly by the Clustal W algorithm in VectorNTI.

Results and Discussion

Cadherin Repertoires of Key Metazoan Organisms

We first identified the cadherin repertoires of the recently sequenced genomes of *B. floridae* (amphioxus = Amphi) (Putnam et al. 2008), the sea anemone *N. vectensis* (Nv) (Putnam et al. 2007), and the placozoan *T. adhaerens* (Ta) (Srivastava et al. 2008). We used a combination of tBLASTn and profile HMM analyses to identify the cadherin superfamily members in these three organisms (supplementary tables S2–S14, Supplementary Material online). The tBLASTn method can identify an ortholog and the most homologous hits in a specific organism by using the organism-restricted search option. However, tBLASTn analysis might not detect cadherins with limited similarity to known cadherins, such as lineage-specific cadherins. Therefore, we also used the five cadherin-based HMMs available in the Pfam database (Finn et al. 2010) and a sensitive custom built HMM to detect every cadherin domain in the genome. These domains were then interpreted and manually annotated on the basis of current knowledge of the cadherin superfamily.

After merging the tBLASTn and HMM results, followed by detailed annotation, we identified 54 cadherin or cadherin-related genes: 30 in *B. floridae* (supplementary fig. S1, supplementary table S6, and supplementary note S1, Supplementary Material online), 16 in *N. vectensis* (supplementary fig. S2, supplementary table S11, and supplementary note S2, Supplementary Material online), and 8 in *T. adhaerens* (supplementary fig. S3, supplementary table S14, and supplementary note S3, Supplementary Material online). The cadherin repertoire of *B. floridae* is surprisingly richer than the repertoires of the vase tunicate *C. intestinalis* (table 1; 15 genes) (Noda and Satoh 2008) or the sea urchin *Strongylocentrotus purpuratus* (table 1; 14 genes) (Whittaker et al. 2006). On the other hand, Trichoplax contains only eight cadherin superfamily members, which is remarkably few compared with the 23 putative cadherins in *M. brevicollis* (Abedin and King 2008).

Table 1. The Cadherin Repertoires of Sequenced Model Organisms.

Model Organisms (acronym)	No. of Cadherin Genes	References
Homo sapiens (Hs)	113	(Hulpiau and van Roy 2009)
Mus musculus (Mm)	119	(Hulpiau and van Roy 2009)
Ciona intestinalis (Ci)	15	(Noda and Satoh 2008)
Branchiostoma floridae (Amphi)	30	This paper
Strongylocentrotus purpuratus (Sp)	14	(Whittaker et al. 2006)
Drosophila melanogaster (Dm)	17	(Hill et al. 2001; Fung et al. 2008)
Caenorhabditis elegans (Ce)	12	(Hill et al. 2001; Pettitt 2005)
Nematostella vectensis (Nv)	16	This paper
Trichoplax adhaerens (Ta)	8	This paper
Monosiga brevicollis (Mb)	23	(Abedin and King 2008)

Identification and Comparative Analysis of Bilaterian Classical Cadherins

The strict definition of classical cadherins (CDH) refers to type-I cadherins, of which E-cadherin is the prototype (Tanihara et al. 1994; Nollet et al. 2000). Some authors use a broader definition and include the type-II cadherins and even the so-called nonchordate classic cadherins with a "primitive classic cadherin domain", and so this broader definition also includes type-III and type-IV cadherins (Takeichi 1995; Oda and Tsukita 1999; Tanabe et al. 2004; Hulpiau and van Roy 2009). All these different classic cadherins share a conserved classical cadherin cytoplasmic domain (CCD) that binds catenins of the armadillo family (fig. 1), and many of them are found in intercellular junctions of bilaterian epithelia. In amphioxus, we found AmphiCDH with a CCD sequence and AmphiDCHS with a CCD-like sequence (supplementary fig. S1, Supplementary Material online). Type-I, type-II, and desmosomal cadherins, which together we designated the C1-family (Hulpiau and van Roy 2009), typically have five EC repeats, whereas type-III cadherins have more ECs and also laminin G (LamG) and epidermal growth factor (EGF)–like domains in their ectodomains. AmphiCDH appears to be a classical type-III cadherin (fig. 1 and supplementary fig. S1 and supplementary table S6, Supplementary Material online). AmphiDCHS is a cadherin-related homolog of Dachsous (see below).

To determine the most likely origin of modern mammalian cadherins (C1 family), we initially compared their 5EC ectodomains with a sliding 5EC block in five amphioxus cadherins: AmphiCDH, AmphiCELSR, AmphiFAT, AmphiFAT-like, and AmphiDCHS (supplementary table S15, Supplementary Material online). AmphiCDH has 18 ECs. The 5EC ectodomains of mouse Cdh1, Cdh5, Cdh13, Cdh15, Cdh26, and Dsc1 as well as those of *C. intestinalis* C-1 family members CiCadherin and CiCadherin-II (Noda and Satoh 2008) show the highest homology (lowest e-values), with the penultimate ECs (EC13 to EC17) from AmphiCDH. In addition, we compared the 5EC ectodomains of modern cadherins with a sliding 5EC block in the 13 ECs of the type-III cadherin of chicken (Tanabe

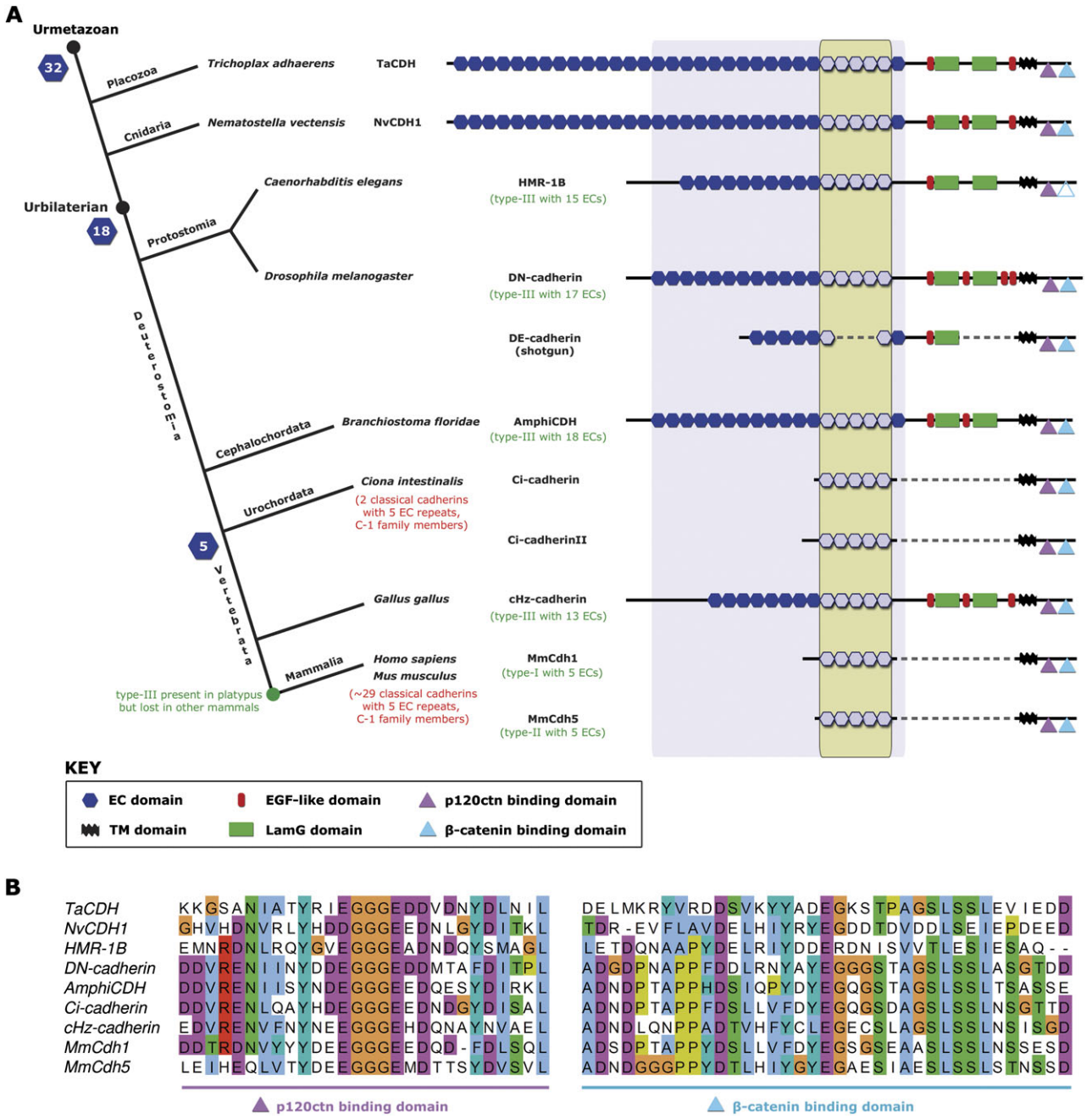

FIG. 1 Metazoan evolution of the classical cadherins from placozoa to man. (*A*) On the left is a cladogram of the organisms for which cadherin superfamily members are analyzed here or were analyzed previously (Hulpiau and van Roy 2009). Urmetazoan and Urbilaterian are hypothetical last common ancestors. Numbers in blue hexagons indicate the typical number of EC repeats in the respective ectodomains. On the right, representative proteins are depicted with their amino termini toward the left. Cadherins were classified into families and types as previously reported (Hulpiau and van Roy 2009). Alignment is at the transmembrane (TM) domains. The first major ectodomain reduction during evolution is represented by the blue rectangle and the second major reduction by the green rectangle. Dotted lines indicate internal deletions. An open triangle indicates imperfect sequence conservation. Key: domain symbols used. (*B*) Amino acid alignment of the conserved motifs in the cytoplasmic cadherin domains that bind p120ctn and β-catenin. The interaction of classical cadherins with the armadillo proteins p120ctn and β-catenin is a universal theme in animal life.

et al. 2004). We also compared them with the type-III cadherin of platypus (*Ornithorhynchus anatinus* OaCDH; sequence in supplementary note S4, Supplementary Material online), which we had predicted from its genomic sequence (Warren et al. 2008). Platypus is exceptional because type-III cadherins have apparently been lost in non-monotreme mammals, that is, marsupials and placental mammals. The five ECs of classical cadherins are most homologous to the last cadherin repeats (EC9 to EC13) of type-III cadherins and less homologous to the penultimate repeats (supplementary table S15, Supplementary Material online). From this analysis, we deduced that the 13 ECs of vertebrate type-III cadherins match the penultimate ECs in AmphiCDH.

All the above analyses are based on the assumption that the cadherin domains were passed to the cadherin descendants as consecutive or mainly consecutive EC blocks. To detect any gains, losses, or retentions in such EC blocks, we also performed pairwise EC-by-EC analyses. Comparative EC-by-EC analysis has been used before to examine the EC relationships between selected arthropod and deuterostomian cadherins (Oda et al. 2005). In an interesting study, these authors were the first to propose the hypothesis of loss of domains from classical cadherins in the bilaterian lineage. We first compared each of the 5 ECs of selected C-1 family members with each of the 18 ECs of AmphiCDH and each of the 13 ECs of the platypus type-III cadherin (supplementary table S16, Supplementary Material online). We also compared the individual ECs of selected type-III cadherins with those of AmphiCDH (supplementary tables S17 and S18, Supplementary Material online). The results of EC-by-EC analyses confirmed those of EC block analyses by showing that the cadherin domains were passed through evolution as blocks of consecutive ECs and that N-terminal EC repeats were lost.

Drosophila melanogaster expresses not only the type-III DN-cadherin, which contains 17 ECs, but also a type-IV DE-cadherin (shotgun) with only 8 ECs (Hulpiau and van Roy 2009). We compared the EC domains of DN-cadherin, DE-cadherin, and AmphiCDH and also their respective LamG and EGF-like domains (supplementary table S19, Supplementary Material online). This analysis revealed that DE-cadherin had lost several N-terminal ECs, three internal ECs, and the second block of LamG/EGF-like domains (summarized in fig. 1). Interestingly, it was recently reported that deletion of the membrane-proximal half of the ectodomain of DE-cadherin does not affect its cell–cell adhesion properties but instead causes a defect in myosin-dependent apical constriction in mutant fly embryos (Haruta et al. 2010).

Domain detection using CD search revealed 18 ECs in AmphiCDH, 17 ECs in DN-cadherin, and 14 ECs in *C. elegans* HMR-1B. On the other hand, the alignment in supplementary fig. S4 (Supplementary Material online) of metazoan classical type-III cadherins showed similarity between EC18 of AmphiCDH and the sequence following the last EC in DN-cadherin and in HMR-1B, which indicates the presence of an EC-like sequence. Likewise, similarity between the first ECs of AmphiCDH and the sequence preceding the first EC of HMR-1B suggests the presence of EC-like sequences. These observations concur with our evolutionary model depicted in figure 1. We hypothesize that the classical cadherin of the urbilaterian ancestor contains 18 EC domains and that during evolution, several N-terminal ECs in HMR-1B and the C-terminal EC in HMR-1B and in DN-cadherin progressively lost their typical cadherin domain characteristics. Consequently, they are not detected by the cadherin domain model used by CD search.

Early Metazoan Evolution of Classical Cadherins

In contrast to the choanoflagellate *M. brevicollis*, basal metazoan phyla, which precede the Bilateria, do have a classical type cadherin. Indeed, we identified classical type-III cadherins in the genomes of *N. vectensis* (NvCDH1, -2, and -3) and *T. adhaerens* (TaCDH) (fig. 1 and supplementary figs. S2 and S3 and supplementary tables S11 and S14, Supplementary Material online). Interestingly, these cadherins possess even more EC repeats (NvCDH1 and TaCDH each has 32 ECs), which indicates that the first loss of ECs preceded the EC losses in the bilaterian lineages. So we compared NvCDH1 and TaCDH with the classical type-III cadherins of vertebrates (supplementary table S20, Supplementary Material online) and nonvertebrates (supplementary table S21, Supplementary Material online). These analyses indicated that the origination of Bilateria was accompanied by an initial loss of N-terminal ECs in very long classical type cadherins (fig. 1). This was confirmed by comparing the individual ECs of selected type-III cadherins with those of TaCDH (supplementary tables S17 and S18, Supplementary Material online). Remarkably, the classical CCD too is conserved from placozoa to man; this CCD contains both the juxtamembrane domain for binding p120ctn and the more carboxy-terminal β-catenin–binding domain (fig. 1). Upon searching the Trichoplax genome, we indeed found homologs for both p120ctn and β-catenin (supplementary note S5, Supplementary Material online). These data allow the definition of classical cadherins to be expanded to include transmembrane proteins with up to 32 consecutive extracellular cadherin domains, with or without LamG and EGF-like domains in the membrane-proximal part of the ectodomain, and with a single transmembrane domain and a typical conserved cytoplasmic domain (CCD) (table 2). It thus seems that a classical type cadherin in complex with two armadillo-type catenins was a key element in the origination of multicellular animals and is a universal feature in the animal kingdom.

Nonclassical Cadherins and Members of the Cadherin-Related Family

The cadherin major branch (for classification, see Hulpiau and van Roy 2009) includes besides "classic" cadherins (CDH) also nonclassical Flamingo/CELSR cadherins with a very peculiar 7-transmembrane (7 TM) domain and additional motifs in their ectodomain. Furthermore, cadherin-related (CDHR) family members include calsyntenins (CLSTN), FAT, FAT-like, Dachsous (DCHS), and the inner ear tip-link cadherin-related family members cadherin-related 23 (CDHR23) and protocadherin-15 (CDHR15). We thus performed phylogenetic analyses to identify within the cadherin superfamily the major families that evolved during evolution from very basal to modern animals. Both *N. vectensis* and *B. floridae* encode a short calsyntenin-like protein but apparently *T. adhaerens* does not (supplementary figs. S1–S3, Supplementary Material online). For the other cadherins, we compared the protein sequences in mouse *Mus musculus* (Mm), amphioxus (Amphi), *D. melanogaster* (Dm), *N. vectensis* (Nv), and *T. adhaerens* (Ta). Multisequence alignments and

Table 2. Major Cadherin Families and Their Specific Domains and Features.

Family Symbols	Family Names	Specific Domains[a] and Features	Alignment[b]
CDH	Classical cadherins	(13–32 ECs for type-III or 5 ECs for type-I and -II) + (2 LamG + several EGF like for type-III) + TM + CCD	Supplementary fig. S4
CELSR	Flamingo cadherins	9 ECs + 2 LamG + several EGF like + HRM + GPS + 7 TMs + CD	Supplementary fig. S6
DCHS	Dachsous cadherins	27 ECs + TM + CCD like	Supplementary fig. S7
FAT	FAT cadherins	34 ECs + 2 LamG + several EGF + TM + CD	Supplementary fig. S8
FAT-like	FAT-like cadherins	34 ECs + LamG + several EGF + TM + CD	Supplementary fig. S8
PCDH	Protocadherins	(7 or 6 ECs) + TM + (CM for nonclustered protocadherins)	Supplementary figs. S10–S12
CDHR	Other cadherin related	Different from those defined above, at least two consecutive ECs and one TM	n.d.

[a] NOTE.—Domain abbreviations: CCD, classical cadherin cytoplasmic domain; CD, cytoplasmic domain; CM, cytoplasmic domain with protocadherin-specific conserved motifs; EGF-like, epidermal growth factor–like domain; GPS, latrophilin/CL-1-like GPCR proteolytic site domain; HRM, hormone receptor domain; LamG, laminin G domain; EC, extracellular cadherin domain; and TM, transmembrane domain. n.d., not done.

[b] Supplementary figures in Supplementary Material online.

phylogenetic tree building were performed as described in Materials and Methods. The resulting NJ tree is shown in supplementary fig. S5 (Supplementary Material online). This tree was validated by building a BI consensus tree (fig. 2). The two trees agree very well with each other and elegantly show the different ancient cadherin families and the members of each family all across the metazoan kingdom. The detailed alignments of entire protein sequences of representative metazoan cadherins from each family (supplementary figs. S4 and S6–S9, Supplementary Material online) revealed clearly distinguishable features typical for each family (table 2).

Our phylogenetic analyses revealed that highly related orthologs of Flamingo/CELSR are encoded by the genomes of *B. floridae* and *N. vectensis*, and even in the basal metazoan *T. adhaerens* (fig. 2 and supplementary figs. S1–S4, Supplementary Material online). This remarkable conservation could indicate that these proteins play essential roles in all Metazoa and that these roles are different from those of classic cadherins. CDH23/CDHR23 and PCDH15/CDHR15 have orthologs in amphioxus (AmphiCdhr23 and AmphiCdhr15, respectively) and thus originated before the appearance of vertebrates. For the CDHR23 branch, we even found an ortholog in the nonbilaterian sea anemone (NvCdhr1) (fig. 2 and supplementary fig. S5, Supplementary Material online) but not in Trichoplax. In contrast, we found FAT, FAT-like, and DCHS orthologs in *B. floridae*, *N. vectensis*, and *T. adhaerens* (fig. 2 and supplementary figs. S1–S4, Supplementary Material online). Both the length and the multidomain composition of the extracellular parts of "ancient" classical cadherins in *N. vectensis* and *T. adhaerens*, which consist of more than 30 EC domains and several LamG and EGF-like domains, closely resemble those of the FAT and FAT-like cadherin superfamily members in these organisms (supplementary figs. S1–S3, Supplementary Material online). The cytoplasmic domains of classic cadherins and DCHS proteins also show similarity. This suggests that these cadherin superfamily members had a common ancestor in an organism even more ancient than placozoa.

Origin and Metazoan Evolution of the Protocadherins

Protocadherins, which constitute the largest family within the cadherin superfamily, contain six or seven extracellular cadherin repeats, a single TM region, and a protocadherin-specific cytoplasmic domain (Morishita and Yagi 2007; Hulpiau and van Roy 2009). They are expressed mainly in the nervous system, and their involvement in synaptic development is indicated by the profound neurological defects in mice lacking the Pcdh-γ cluster (Weiner et al. 2005). However, their functions are not entirely clear (Takeichi 2007). Protocadherins have not been found in worms or flies, and until now, they have been considered a chordate innovation (Pettitt 2005).

We identified two protocadherins in *B. floridae* (AmphiPcdh1 and AmphiPcdh2; fig. 3 and supplementary fig. S1 and supplementary table S6, Supplementary Material online) and, surprisingly, also one in *N. vectensis* (NvPcdh; fig. 3 and supplementary fig. S2 and supplementary table S11, Supplementary Material online). We also searched the genome of a mollusk, the California sea slug or *Aplysia californica*, for the presence of protocadherin genes. Thornton et al. (2003) have shown that *Aplysia* has an estrogen receptor ortholog, in contrast to *Drosophila* and *C. elegans*, which have no orthologs at all for steroid receptors. Likewise, we identified in *Aplysia* a genuine protocadherin (AcPCDH) that had been reported as a cadherin-related molecule (GenBank accession number: AAO84370). Comparative analysis of four basal protocadherins, two from amphioxus (AmphiPCDH1 and AmphiPCDH2), one from sea anemone (NvPCDH), and one from California sea hare (AcPCDH), revealed that they all have seven EC repeats and possess conserved cytoplasmic motifs (CMs; fig. 3). These conserved motifs were first observed in the cytoplasmic domains of long isoforms of vertebrate nonclustered δ-protocadherins (Vanhalst et al. 2005). This suggests that protocadherins predate the bilaterians and were lost in the protostomian lineages leading to arthropods and nematodes but not in all protostomian lineages. In chordates, the protocadherin family greatly expanded, with only one

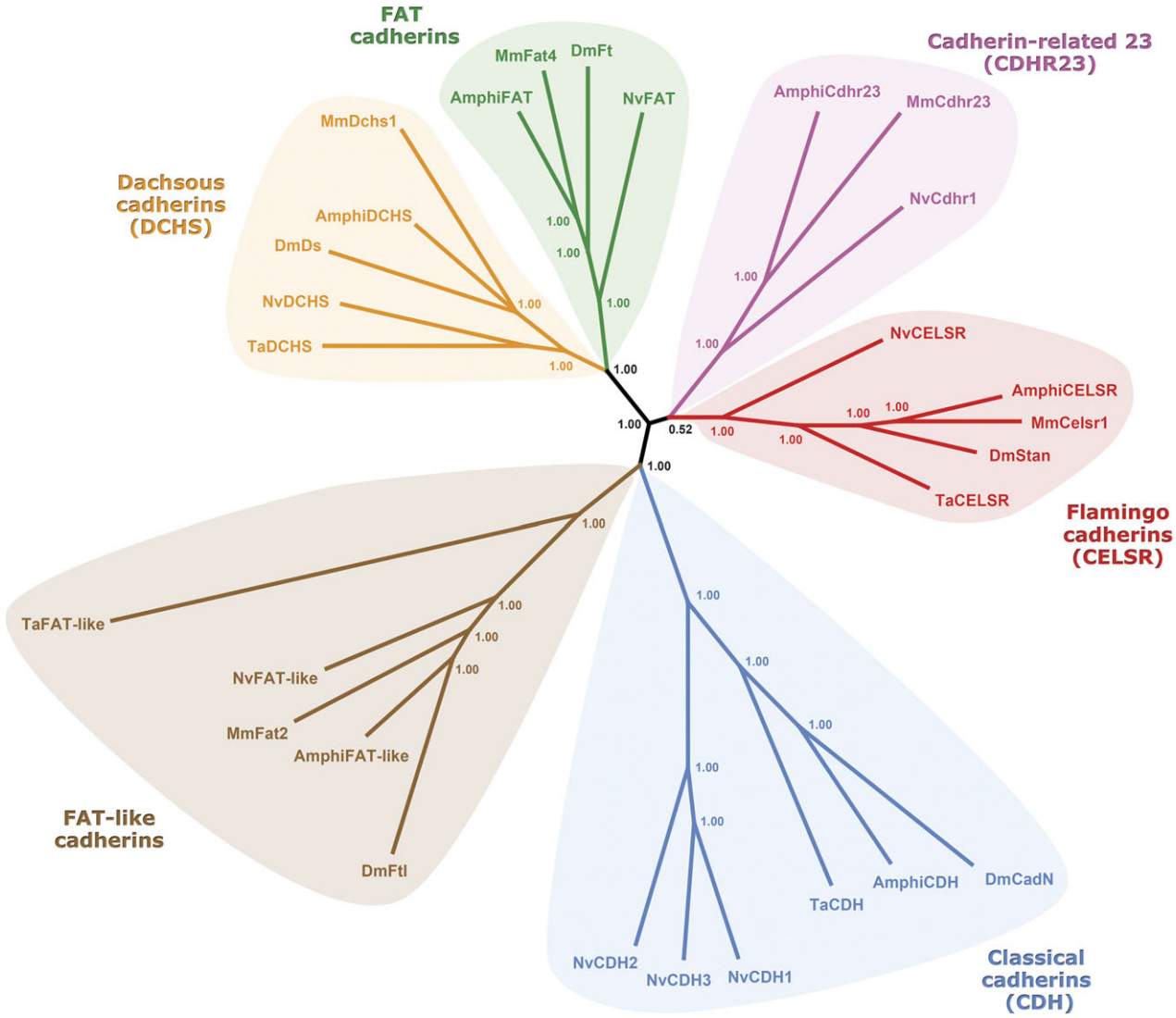

FIG. 2 Metazoan cadherin families: phylogenetic tree of cadherin superfamily members with several highly conserved branches. We used protein sequence blocks comprising all ECs of selected cadherins of mouse *Mus musculus* (Mm), amphioxus (Amphi), fruit fly *D. melanogaster* (Dm), sea anemone *N. vectensis* (Nv), and the placozoan *T. adhaerens* (Ta) (see also table 1 and supplementary figs. S1–S3, Supplementary Material online). The proteins analyzed belong to branches of classical (CDH), flamingo/CELSR, dachsous (DCHS), FAT, and FAT-like cadherins or cadherin-related member 23 (CDH23/CDHR23) (see also table 2). This consensus tree is based on BI analysis and drawn as a radial phylogram. Numbers at branch points indicate Bayesian posterior probabilities. The structure of this tree is supported by a NJ tree (supplementary fig. S5, Supplementary Material online).

protocadherin in sea urchin (Whittaker et al. 2006), two paralogs in amphioxus, six members in *C. intestinalis* (Noda and Satoh 2008), and between 60 and 80 genes in mammals, fishes, and reptiles (Noonan et al. 2004; Yagi 2008; Jiang et al. 2009) (fig. 3). In vertebrates, most protocadherin genes seem to have been generated by repeated lineage-specific gene duplications, gene conversions, and adaptive variation, which led to a huge gene cluster comprising typical repetitions of different as well as shared constant exons (Noonan et al. 2004; Yagi 2008; Jiang et al. 2009). It is conceivable that this protocadherin boom was one of the factors underlying the increase in central nervous complexity in the chordate lineage. The origin of the synapse lies in the last common ancestor of cnidarians and bilaterians, which would have assembled an ursynapse (Ryan and Grant 2009). Thus, the appearance of an ancient protocadherin in the cnidarian *N. vectensis* is concordant with an innovative protocadherin role of mediating synapse development.

To determine the origin of the seven EC repeats present in these protocadherins, we compared a selection of 7EC protocadherin ectodomains from various organisms with a sliding 7EC block in several *N. vectensis* cadherins and additionally with a sliding 7EC block in AmphiFAT (supplementary table S22, Supplementary Material online). The cadherin repeats of protocadherins matched best with the first seven repeats in NvFAT, as indicated by the significantly lower e-values. This match was also suggested by the tBLASTn analysis (supplementary table S9, Supplementary Material online) in which NvFAT, represented by RefSeq XP_001627512, was the best hit for most of the mammalian protocadherins. A comparison of 6EC

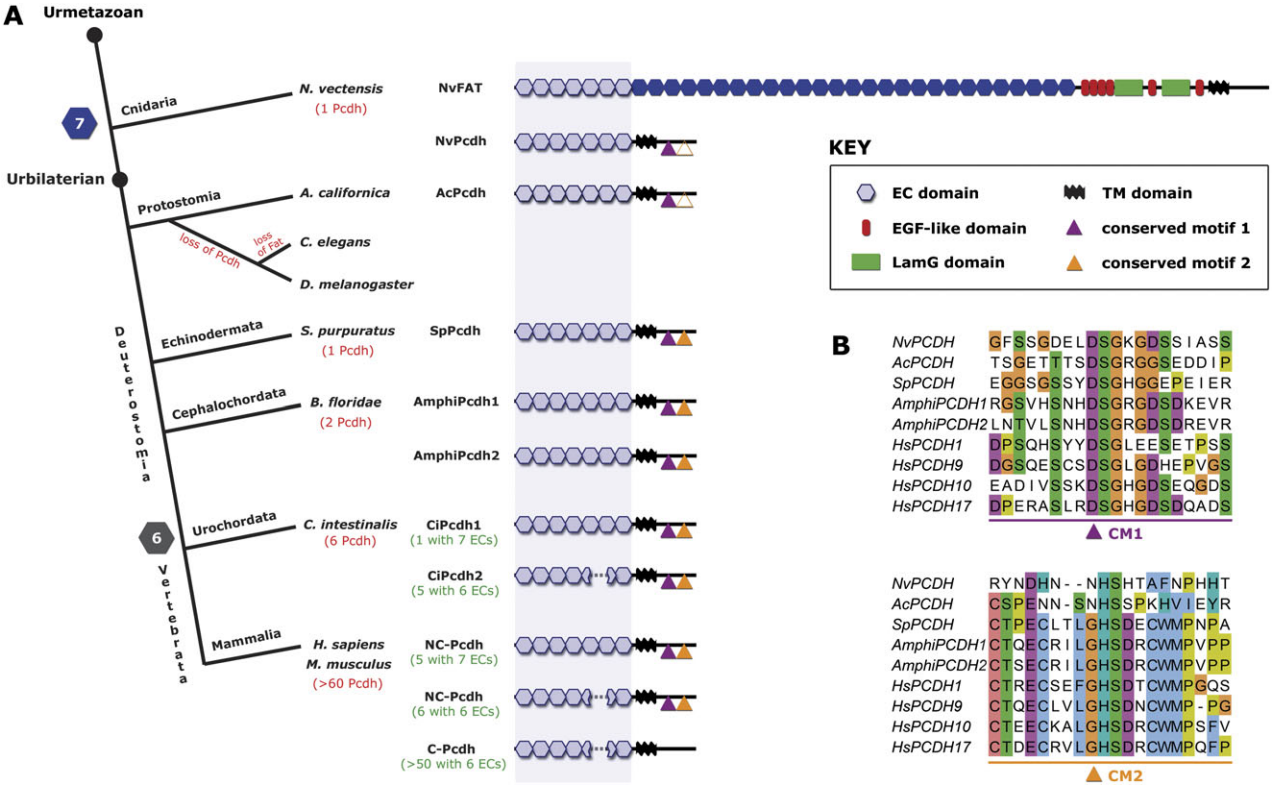

FIG. 3 Metazoan evolution of protocadherins from the sea anemone to man. (*A*) On the left is a cladogram of organisms for which cadherin superfamily members are analyzed here or were analyzed previously (Hulpiau and van Roy 2009). Numbers in blue hexagons indicate the typical number of EC repeats in the respective ectodomains of protocadherins. Representative proteins are depicted on the right. NvFAT represents the proposed FAT-related ancestor. In vertebrates, nonclustered protocadherins (NC-Pcdh or δ-protocadherins) (Hulpiau and van Roy 2009) have seven or six ECs in their ectodomains, whereas all clustered protocadherins (C-Pcdh) have six. Alignment is at the amino termini on the left. The first abrupt ectodomain reduction is represented by the blue rectangle. The reduction from a 7-EC protocadherin to the 6-EC protocadherins occurred by an internal deletion depicted by dotted lines. An open triangle indicates imperfect sequence conservation. Key: domain symbols used. (*B*) Amino acid alignment of the conserved binding motifs 1 (CM1) and 2 (CM2) in the cytoplasmic domains of representative protocadherins.

blocks from a selection of mouse protocadherins (MmPcdha1, MmPcdhb1, MmPcdh8, and MmPcdh10) with a sliding 6EC block in NvFAT and AmphiFAT confirmed these findings (supplementary table S23, Supplementary Material online). Also here, an EC-by-EC analysis was performed to evaluate the possibility of the loss or gain of one or more ECs (supplementary table S24, Supplementary Material online). The hypothesis that the ectodomain of protocadherins evolved from the N-terminal ECs of an ancient FAT protein (fig. 3), rather than from another ancestor, agrees with the comprehensive EC1-based phylogenetic analysis we previously described (Hulpiau and van Roy 2009), which positioned FAT and DCHS closer to the protocadherins than to other cadherins or cadherin-related proteins. When the 7EC domains of several protocadherins were used for multiple sequence alignment with the 34 ECs of sea anemone FAT (NvFAT) and amphioxus FAT (AmphiFAT), they convincingly aligned with the N-terminal EC domains of the FAT cadherins (supplementary fig. S9, Supplementary Material online).

Protocadherins with six instead of seven ECs arose early in the olfactores clade, possibly in the last common ancestor of tunicates and vertebrates (fig. 3). The vase tunicate *C. intestinalis* has six protocadherins (Noda and Satoh 2008), of which five have six EC repeats and one has seven ECs. To investigate the origin of 6EC protocadherins, we compared the ECs of protocadherins with 7ECs and 6ECs in *C. intestinalis*, *Mus musculus*, and *Homo sapiens* (supplementary table S25, Supplementary Material online). The first four ECs of 6EC protocadherins clearly match the first four ECs of 7EC protocadherins. Also, the last EC of 6EC protocadherins matches the last EC of 7EC protocadherins. These data suggest that loss of parts of EC5 and EC6 from a 7EC protocadherin produced a 6EC protocadherin. To confirm this hypothesis, we aligned the protein sequences of a selection of protocadherins of either *C. intestinalis* (supplementary fig. S10, Supplementary Material online) or *H. sapiens* (supplementary fig. S11, Supplementary Material online). In both alignments, the ECs match until the beginning of the fifth EC. The rest of the EC5 of 6EC protocadherins (CiPcdh2 to CiPcdh6 in supplementary fig. S10, Supplementary Material online, and HsPCDH10 and following in supplementary fig. S11, Supplementary Material online) matches part of the sixth EC of 7EC protocadherins (CiPcdh1 in supplementary fig. S10, Supplementary

Material online, and HsPCDH1 and the three following sequences in supplementary fig. S11, Supplementary Material online). There is convincing alignment for the last EC of both protocadherin types. This means that 6EC protocadherins were generated by loss of the C-terminal part of EC5 and the N-terminal part of EC6 from a 7EC protocadherin, as depicted in figure 3. We also investigated the intron–exon structure of all six Ciona protocadherin genes (supplementary fig. S12, Supplementary Material online). The five genes encoding 6EC CiPcdhs clearly have more introns in the genomic region encoding ECs than the gene encoding the 7EC protocadherin CiPcdh1. This could have facilitated the internal loss of one or more exons encoding the C-terminal part of EC5 and the N-terminal part of EC6 in a 7EC protocadherin.

Finally, clustered protocadherins with six ECs probably originated in the early days of vertebrates from a nonclustered 6EC type ancestral protocadherin. This hypothesis has to be confirmed by examining other cadherin repertoires, such as that of the lamprey (*Petromyzon marinus*).

Ancient Long Cadherins

The amphioxus genome encodes the strikingly large number of up to 30 cadherin and cadherin-related genes (supplementary fig. S1, Supplementary Material online). Noteworthy, only about ten of them have homologs in vertebrates, whereas most seem to be lineage specific. They often have long ectodomains and sometimes unusual protein domains but generally no LamG, EGF-like, or CCD domain. One of these peculiar cadherins is AmphiCdhr1 with 40 cadherin repeats. The *N. vectensis* genome also encodes very long cadherins, namely NvCdhr2 and NvCdhr3, which have 43 and 51 ECs, respectively, but lack other typical domains (supplementary fig. S2, Supplementary Material online). We hypothesize that these unusual cadherins are ancient remnants from the Urmetazoan ancestor that had been lost during evolution into cadherins of either Protostomia or vertebrates. This type of cadherin has not been reported for metazoans, whereas the cadherin repertoire of *M. brevicollis* includes several somewhat similar proteins (Abedin and King 2008). Nonetheless, we were unable to detect convincing homology between individual ECs or EC blocks of any of the *N. vectensis* or *T. adhaerens* cadherins or between ECs or EC blocks of several long *M. brevicollis* cadherins (data not shown). Also, we found no evidence for internal tandem duplications to explain the origin of these long ectodomains (data not shown). This implies that long EC stretches in cadherins of basal organisms are really ancient structures that had existed before many lineage splits.

Conclusions

In summary, we propose that the Urmetazoan, the last common ancestor of animals, had expressed at least five members of the cadherin superfamily that are found in nearly all its descendants (fig. 2): a classical cadherin (CDH), a nonclassical Flamingo cadherin (CELSR), and three cadherin-related members, namely FAT, FAT-like, and DCHS. Except for DCHS, each of these ancestral cadherin superfamily members has multiple EGF-like domains and one or more LamG domains inserted between the ECs and the TM region. Their similar domain composition suggests that they were originally paralogs. They differ substantially from the many diverse proteins containing EC repeats encoded by the *M. brevicollis* genome (Abedin and King 2008). Therefore, the choanoflagellate cadherin repertoire might have been generated from a separate evolution from one or more ancient cadherins that had been present in the unicellular Urmetazoan/choanoflagellate concestor more than 600 Ma.

The classical type cadherins have been evolving from the Urmetazoan to man by progressive loss of N-terminal ECs from the ectodomain in combination with loss of membrane-proximal motifs (fig. 1). The domain composition of AmphiCDH corresponds to the archetypal representation of the Urbilaterian ancestor of protostomian and deuterostomian classical cadherins. The five ECs of the C-1 cadherin family members correspond with the membrane-proximal ECs of vertebrate type-III cadherins and with the penultimate EC repeats of AmphiCDH. The 13 ECs of vertebrate type-III cadherins, the 15 ECs of worm HMR-1B, and the 17 ECs of fruit fly DN-cadherin match extensively with the penultimate ECs of AmphiCDH. When the chordate descendants emerged, the type-III cadherin gave rise to "modern" type-I and type-II cadherins having five to seven ECs in their ectodomains. The last five EC repeats and the CCD have been conserved, whereas other ECs and the LamG and EGF domains were lost. *C. intestinalis* has two modern cadherins, and the more recently evolved chordate species have many more, with up to 29 major branch family members in man. Increasing the diversity of these 5EC-type cadherins might have been fundamental in increasing tissue complexity during chordate evolution.

In contrast, the structures of the Flamingo/CELSR, the FAT-like, and the DCHS cadherins have remained unchanged in at least three monophyletic groups: Bilateria, Cnidaria, and Placozoa (fig. 2 and supplementary figs. S1–S3, Supplementary Material online) (Hulpiau and van Roy 2009). The same is largely true for the FAT cadherins, except that the ectodomain is shorter in the Trichoplax FAT protein and that *Caenorhabditis elegans* has no FAT proteins (Pettitt 2005). This remarkable ancient origin of several cadherin types suggests that each of them has fulfilled separate and essential needs throughout the evolution of Metazoa. Flamingo/CELSR, DCHS, and FAT cadherins play established roles in planar cell polarity (Simons and Mlodzik 2008), but such polarity is probably not essential in the two-layered placozoans, which lack any kind of symmetry. However, the recently discovered involvement of DCHS–FAT interactions in the Hippo kinase pathway, which coordinates cell proliferation with cell death in flies and mammals (Badouel et al. 2009), indicates that these highly conserved cadherin superfamily members play important roles in basal metazoan life. In agreement

with this, we obtained evidence that several key molecules in the Hippo signaling pathway have homologs in Trichoplax (data not shown). Along the same line is the essentially unchanged cytoplasmic domain of classical cadherins (CCD), which has two regions for binding armadillo proteins. The multiples roles of the armadillo proteins p120ctn and β-catenin, which act both as stabilizers of cell junctions and as cytoplasmic and nuclear signaling proteins (McCrea and Gu 2010), might be the reason for their remarkable evolutionary conservation.

The shortening of the ectodomain of classical cadherins during evolution has been stepwise: The reduction from 32 ECs to ~18 ECs at the dawn of the Bilateria was followed later by further reduction to 5 ECs at the dawn of vertebrates (fig. 1). In contrast, protocadherins apparently emerged by an abrupt C-terminal deletion in a FAT-like precursor (fig. 3), probably in combination with gene fusion. A protocadherin is not present in the "living fossil" Trichoplax, although it can be identified as early as cnidarians (Nematostella). Several but not all prostostomians have lost protocadherins. Nonclustered protocadherin genes, which are more ancient than the renowned clustered protocadherin genes, have peculiar conserved motifs (CM) in their cytoplasmic domains but only in the long isoforms generated by alternative splicing of 3′ exons (Vanhalst et al. 2005). In transcripts of clustered protocadherin genes, splicing occurs to constant 3′ exons, which are shared by many genes in the cluster and differ from those of the nonclustered ancient protocadherin genes (Yagi 2008). Therefore, the introduction of "new" cytoplasmic domains, which differ among protocadherins and also from that of the FAT-related ancestor, might be due to exon-swapping events. It is intriguing that both modern classical cadherins and modern protocadherins have short ectodomains, with at most five to seven ECs. One may speculate that the role of particular cadherins has changed from intercellular signaling via loose contacts to tight interactions in carefully composed junctional complexes. This might reflect the evolution from the limited requirements of a colonial unicellular eukaryote to the strict morphogenetic programs essential for metazoan organs.

Supplementary Material

Supplementary figures S1–S12, supplementary tables S1–S25, and supplementary notes (sequences) S1–S5 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank A. Bredan for critical reading and editing of the manuscript and D. Adriaens for advice on taxonomy. Supported by the Research Foundation Flanders (FWO) and the Geconcerteerde Onderzoeksacties of Ghent University.

References

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

Badouel C, Garg A, McNeill H. 2009. Herding hippos: regulating growth in flies and man. *Curr Opin Cell Biol*. 21:837–843.

Berx G, van Roy F. 2009. Involvement of members of the cadherin superfamily in cancer. In: Nelson J, Fuchs E, editors. Cold Spring Harbor perspectives in biology. Vol. 1. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press. p. a003129.

Burge CB, Karlin S. 1998. Finding the genes in genomic DNA. *Curr Opin Struct Biol*. 8:346–354.

Finn RD, Mistry J, Tate J, et al. (14 co-authors). 2010. The Pfam protein families database. *Nucleic Acids Res*. 38:D211–D222.

Fung S, Wang F, Chase M, Godt D, Hartenstein V. 2008. Expression profile of the cadherin family in the developing Drosophila brain. *J Comp Neurol*. 506:469–488.

Haruta T, Warrior R, Yonemura S, Oda H. 2010. The proximal half of the *Drosophila* E-cadherin extracellular region is dispensable for many cadherin-dependent events but required for ventral furrow formation. *Genes Cells*. 15:193–208.

Hill E, Broadbent ID, Chothia C, Pettitt J. 2001. Cadherin superfamily proteins in Caenorhabditis elegans and Drosophila melanogaster. *J Mol Biol*. 305:1011–1024.

Hulpiau P, van Roy F. 2009. Molecular evolution of the cadherin superfamily. *Int J Biochem Cell Biol*. 41:349–369.

Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, Rupp R. 2007. Dendroscope: an interactive viewer for large phylogenetic trees. *BMC Bioinformatics*. 8:460.

Iwai H, Masaoka N, Ishii T, Satoh S. 2002. A pectin glucuronyltransferase gene is essential for intercellular attachment in the plant meristem. *Proc Natl Acad Sci U S A*. 99:16319–16324.

Jiang XJ, Li S, Ravi V, Venkatesh B, Yu WP. 2009. Identification and comparative analysis of the protocadherin cluster in a reptile, the green anole lizard. *PLoS One*. 4:7614.

Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. 2008. NCBI BLAST: a better web interface. *Nucleic Acids Res*. 36:W5–W9.

Kall L, Krogh A, Sonnhammer EL. 2007. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res*. 35:W429–W432.

Kuhn RM, Karolchik D, Zweig AS, et al. (22 co-authors). 2009. The UCSC genome browser database: update 2009. *Nucleic Acids Res*. 37:D755–D761.

Larkin MA, Blackshields G, Brown NP, et al. (13 co-authors). 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947–2948.

Lu G, Moriyama EN. 2004. Vector NTI, a balanced all-in-one sequence analysis suite. *Brief Bioinform*. 5:378–388.

Marchler-Bauer A, Anderson JB, Chitsaz F, et al. (28 co-authors). 2009. CDD: specific functional annotation with the conserved domain database. *Nucleic Acids Res*. 37:D205–D210.

McCrea PD, Gu D. 2010. The catenin family at a glance. *J Cell Sci*. 123:637–642.

Morishita H, Yagi T. 2007. Protocadherin family: diversity, structure, and function. *Curr Opin Cell Biol*. 19:584–592.

Nelson WJ, Fuchs E. 2010. Cell-cell junctions (ISBN 978-087969884-3). Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.

Nelson WJ, Nusse R. 2004. Convergence of Wnt, beta-catenin, and cadherin pathways. *Science* 303:1483–1487.

Noda T, Satoh N. 2008. A comprehensive survey of cadherin superfamily gene expression patterns in Ciona intestinalis. *Gene Expr Patterns*. 8:349–356.

Nollet F, Kools P, van Roy F. 2000. Phylogenetic analysis of the cadherin superfamily allows identification of six major subfamilies besides several solitary members. *J Mol Biol*. 299:551–572.

Noonan JP, Grimwood J, Schmutz J, Dickson M, Myers RM. 2004. Gene conversion and the evolution of protocadherin gene cluster diversity. *Genome Res*. 14:354–366.

Oda H, Tagawa K, 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.

Oda H, Tsukita S. 1999. Nonchordate classic cadherins have a structurally and functionally unique domain that is absent from chordate classic cadherins. *Dev Biol*. 216:406–422.

Pettitt J. 2005. The cadherin superfamily. In: Moerman DG, Kramer JM, editors. WormBook. The *C. elegans* research community. p. 1–9 [cited 2005 December 29]. Available from: http://www.wormbook.org.

Putnam NH, Butts T, Ferrier DEK, et al. (37 co-authors). 2008. The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453:1064–1071.

Putnam NH, Srivastava M, Hellsten U, et al. (19 co-authors). 2007. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* 317:86–94.

Rokas A. 2008. The origins of multicellularity and the early history of the genetic toolkit for animal development. *Annu Rev Genet*. 42:235–251.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.

Ryan TJ, Grant SG. 2009. The origin and evolution of synapses. *Nat Rev Neurosci*. 10:701–712.

Saburi S, McNeill H. 2005. Organising cells into tissues: new roles for cell adhesion molecules in planar cell polarity. *Curr Opin Cell Biol*. 17:482–488.

Simons M, Mlodzik M. 2008. Planar cell polarity signaling: from fly development to human disease. *Annu Rev Genet*. 42:517–540.

Srivastava M, Begovic E, Chapman J, et al. (21 co-authors). 2008. The *Trichoplax* genome and the nature of placozoans. *Nature* 454:955–960.

Takeichi M. 1990. Cadherins: a molecular family important in selective cell-cell adhesion. *Annu Rev Biochem*. 59:237–252.

Takeichi M. 1995. Morphogenetic roles of classic cadherins. *Curr Opin Cell Biol*. 7:619–627.

Takeichi M. 2007. The cadherin superfamily in neuronal connections and interactions. *Nat Rev Neurosci*. 8:11–20.

Tanabe K, Takeichi M, Nakagawa S. 2004. Identification of a nonchordate-type classic cadherin in vertebrates: chicken Hz-cadherin is expressed in horizontal cells of the neural retina and contains a nonchordate-specific domain complex. *Dev Dyn*. 229:899–906.

Tanihara H, Kido M, Obata S, Heimark RL, Davidson M, St John T, Suzuki S. 1994. Characterization of cadherin-4 and cadherin-5 reveals new aspects of cadherins. *J Cell Sci*. 107:1697–1704.

Thornton JW, Need E, Crews D. 2003. Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling. *Science* 301:1714–1717.

van Roy F, Berx G. 2008. The cell-cell adhesion molecule E-cadherin. *Cell Mol Life Sci*. 65:3756–3788.

Vanhalst K, Kools P, Staes K, van Roy F, Redies C. 2005. Delta-protocadherins: a gene family expressed differentially in the mouse brain. *Cell Mol Life Sci*. 62:1247–1259.

Warren WC, Hillier LW, Marshall Graves JA, et al. (102 co-authors). 2008. Genome analysis of the platypus reveals unique signatures of evolution. *Nature* 453:175–183.

Weiner JA, Wang XZ, Tapia JC, Sanes JR. 2005. Gamma protocadherins are required for synaptic development in the spinal cord. *Proc Natl Acad Sci U S A*. 102:8–14.

Whittaker CA, Bergeron KF, Whittle J, Brandhorst BP, Burke RD, Hynes RO. 2006. The echinoderm adhesome. *Dev Biol*. 300:252–266.

Yagi T. 2008. Clustered protocadherin family. *Dev Growth Differ*. 50:S131–S140.