PNAS
www.pnas.org

Supplementary Information for
The evolution of synaptic and cognitive capacity: Insights from the nervous system transcriptome of *Aplysia*

Joshua Orvis, Caroline B. Albertin, Pragya Shrestha, Shuangshuang Chen, Melanie Zheng, Cheyenne J. Rodriguez, Luke Tallon, Anup Mahurkar, Aleksey Zimin, Michelle Kim, Kelvin Liu, Eric R. Kandel, Claire M. Fraser, Wayne Sossin and Thomas W. Abrams

Corresponding author: Eric R. Kandel
erk5@columbia.edu

This PDF file includes:

Supplementary text
Figures S1 to S19 (not allowed for Brief Reports)
Tables S1 to S5
Legends for Datasets S1 to S2
SI References

Other supplementary materials for this manuscript include the following:

Datasets S1 to S2

Supplementary Information Text

Methods

RNA preparation. Trimmed Aplysia ganglia were frozen in liquid nitrogen and ground using a pre-cooled (at -80 ºC) mortar and pestle on a bed of dry ice. The pulverized sample was homogenized in a glass-Teflon homogenizer in 1 ml Trizol at 0 ºC, shaken on a rotator for 20 minutes at 0 ºC and centrifuged at 21,000g for 15 minutes at 4 ºC. The supernatant was collected and mixed well with 200 µl chloroform, incubated on ice for 5 minutes, and centrifuged. The upper chloroform phase was retained and combined with an additional 200 ul of chloroform, mixed well, centrifuged at 21,000g for 15 minutes, and the chloroform phase retained. This additional chloroform step eliminated any traces of Trizol from the solution. 40 µl of 3M sodium acetate (i.e., 10% of volume) and 1 ml of 100% ethanol (i.e., 2.5X volume) were added to the collected chloroform phase. The tube was centrifuged at 21,000g for 30 minutes at 4 ºC. The supernatant was carefully removed and discarded, the pellet was washed with 75% ethanol, and the tube centrifuged again. The air-dried pellet, taking care to avoid over-drying, was then dissolved in 50 µl nuclease-free water, and the RNA quantified using Nano-drop. PolyA RNA was sheared and cDNA libraries were prepared with the TruSeq RNA Sample Prep kit (Illumina, San Diego, CA). Adapters containing seven nucleotide indexes were ligated to the double-stranded cDNA. The DNA was purified between enzymatic reactions and the size selection of the library was performed with AMPure XP beads (Beckman Coulter Genomics, Danvers, MA). Libraries were sequenced on Illumina HiSeq 2000. The total reads from the CNS were 2.97 billion 100bp paired-end reads.

Assembly of reads. Paired-end reads were trimmed with Trimmomatic (1). Assembly was performed using three different approaches - *de novo* with Trinity (DN-Trinity) (2), genome-guided *de novo*, also with Trinity (GG-Trinity), and pure genome-guided assembly with StringTie (3) (see Table S1 and AplysiaTools.org for genome information). These were then merged using PASA (4). TransDecoder was used to generate CDS predictions; the longest CDSs called per transcript from both TransRate-filtered (5) PASA transcripts and TransRate-filtered *de novo* Trinity transcripts were selected and clustered using CD-HIT-EST (6, 7), with the following parameters:

-c 0.97 -G 0 -l 300 -aL 0.35 -AL max -aS 0.35 -AS max -A 0 -g 1 -r 1 -mask NX -M 8000 -T 3 -mismatch -4 -gap -12 -gap-ext -2 -d 50

A bespoke script filtered these transcripts to select the transcript within each cluster having the longest CDS. The longest transcript per cluster was then analyzed with TransDecoder to generate polypeptide predictions. These sequences were used as input for Trinotate to generate functional annotation, supplemented by manual annotation.

TransRate filtering of contigs. TransRate (5) generates a cumulative score representing assembly accuracy, which is the geometric mean of four "contig score components" and eliminates contigs likely to be incorrect (1). However, based on examining TOIs, we observed that valid transcripts were lost by this screening. Examination of the valid transcripts that were lost suggested that low scores for the sCseg component contributed to the erroneous assessment of assembly accuracy. s(Cseg) scores are based on the assumption that there will be equal reads across a transcript; we already had documented that fragmentation was due to poor coverage/assignment of reads across repeat regions, and thus, transcripts were prone to being eliminated by low Cseg scores. To ameliorate this problem, we scaled sCseg, so that it ranged from 0.5 to 1.0, using the following equation:
Scaled s(Cseg) = 0.5+ s(CSeg)*(1-0.5).
We then empirically chose 0.4 for the cutoff value for the geometric mean calculated using the scaled sCseg based on BUSCO and TOI coverage. This is an example of the difficulty of transcriptome assemblies in the absence of closely related assemblies and the benefit of TOIs in optimizing this somewhat arbitrary but important process.

Phylogenetic analysis. To identify key synaptic scaffold genes, we searched NCBI for candidate sequences in select animal clades. *Octopus bimaculoides* sequences were identified by tblastn and blastp from a Trinity-assembled transcriptome from RNAseq reads from supraesophageal brain, subesophageal brain, and optic lobes (8). In a number of cases, we were able to stitch together fragmented transcript sequences in other mollusks based on orthologous *Aplysia* sequences. Phylogenetic dendrograms were developed using alignments generated by MUSCLE. Maximum likelihood inference was generated with RAxMLGui (9, 10) with 500 replicates per dendrogram using the following options: ML+rapid bootstrap, VT amino acid substitution model and PROTGAMMAI model for rate substitution. Some alignments (indicated in legends) were trimmed prior to RAxML with TrimAI (11), with User Defined option and the following parameters: Minimum percentage of positions to conserve, 45%; Gap threshold, 0.19; Similarity threshold, 0; Window size, 1.0.. Dendrograms were visualized with Figtree (See Table S2 for software details).

Additional Postsynaptic Scaffold Proteins

Cornichons

Cornichons also serve as AMPA receptor regulatory proteins. There are four cornichons in vertebrates with no additional duplications in *Danio*. Two Cornichon-like sequences are found in the *Aplysia* assembly, one of which clusters with the vertebrate Cornichon 1-3 family and one of which clusters with vertebrate Cornichon 4 (Fig. S7). Unlike TARPS, Cornichons are evolutionarily ancient, and are present in plants and fungi. Cornichon 1-3 family members are linked to AMPA receptor regulation in both vertebrates and invertebrates (12, 13), whereas Cornichon 4 has been implicated in G protein-coupled receptor (GPCR) regulation and localization (14). In *Trichoplax*, which lacks neurons and synapses, there are distinct transcripts encoding a Cornichon 1 family member and a Cornichon 4 family member (Fig. S7). The earliest-branching phylum in which features of the Cornichon 1 and 4 families appear is Porifera. Ctenophores and choanoflagellates encode a single family member that does not clearly segregate with either of the two Cornichon families.

Intracellular Postsynaptic Density Scaffold Proteins

DLGs

There are a number of intracellular proteins important for anchoring these ligand-gated receptors and receptor-associated proteins at synapses. The most well-known postsynaptic scaffolding protein group is the discs-large family, DLG1-4, including PSD-95, SAP102 and two additional members in vertebrates. DLG1 also duplicated in *Danio*. These proteins conserve three PDZ domains followed by an SH3 domain and an inactive guanylate kinase domain (Fig. S5). The divergence of DLGs from other MAGUK (Membrane-Associated Guanylate Kinase) members occurred at the base of the metazoan lineage (15). There is only one member in all protostomes examined, including *Aplysia*, that conserve this domain structure (Fig. S8). Previous results suggesting multiple members of this family in octopuses (8) were an artifact due to fragmented assembly of a single transcript. As discussed in the main text, DLGs regulate AMPA receptors by binding to the C-terminal TTPV motif of TARPs, and this sequence is conserved in most non-insect members of the invertebrate TARP-A family.

Homer

Homers have two conserved domains, an N-terminal EVH1 domain and a C-terminal coiled-coil domain, and these domains are conserved in most metazoans (Figs. S5) and in choanoflagellates. While three isoforms are present in vertebrates (five in *Danio*) only one isoform is present in invertebrates including *Aplysia* (Fig. S9). Note, no Homer orthologue is present in *C. elegans* (16), which is an example of the loss of a synaptic scaffold gene. In the PSD, Homer is important for linking IP3 receptors to PSD proteins such as mGluR receptors and Shank (17, 18). Interestingly, whereas the IP3 receptor-Homer interaction and Homer oligomerization are present in the choanoflagellate orthologue (16), the Shank interaction is not, consistent with some specific Homer interactions appearing later in evolution, after these proteins were exapted to serve a neuronal role.

Shank

SH3 and multiple ankyrin repeat protein (Shank), similar to the other intracellular scaffold proteins, first originated at the base of the metazoan lineage (19), and is present in choanoflagellates. Shank contains multiple ankyrin repeats at the N-terminus, followed by an SH3 domain, a PDZ domain and a conserved SAM domain at the C-terminus (Fig. S5). The Shank SH3 domain has several changes that prevent binding to the classic proline-rich ligand of SH3 domains (20), and therefore it lacks known binding partners. Interestingly, this SH3 domain is poorly conserved in molluscs, particularly in *Aplysia*. There is one Shank orthologue present in all

non-bilaterians and in protostomes, including *Aplysia*, but three isoforms are present after the vertebrate genome duplications, with an additional two members in *Danio* (Fig. S10).

GRIP

The glutamate receptor-interacting protein (GRIP) contains 7 PDZ domains (Fig. S5) and has many ligands. Given that many proteins have multiple PDZ domains, it is somewhat difficult to trace phylogenetically. However, based on our reverse-BLAST test (21), in which the putative orthologous protein sequence identified in a non-model organism must retrieve the same protein when BLASTed against a well-annotated reference species (21), almost all non-bilaterian sequences identified by BLAST to GRIP failed, except for one sequence from the octocoral *Dendronephthya gigantea* (Fig. S11). Thus, it appears that the protein was present in the last common ancestor of cnidarians and bilaterians, but was lost in most cnidarians. There is one GRIP orthologue present in all protostomes, but two isoforms are present in vertebrates after the genome duplication, and an additional member is present in *Danio*.

In summary, the intracellular scaffold proteins generally have multiple isoforms in vertebrates but only a single member in most protostomes and invertebrate deuterostomes examined (Table S4). All of these proteins, except for GRIP, were present before synapses evolved and have orthologues in choanoflagellates. The roles of these proteins in organizing sub-domains of cells clearly predated their use in organizing post-synaptic densities. Although the increase in diversity of these proteins in vertebrates (Table S4) is consistent with the proposal that complexity of synapses is linked to an increase in synaptic diversity, there are no increases in the number of these scaffolding proteins between *Aplysia* and the more cognitively complex *Octopus* (Table S4). Moreover, there are considerably more isoforms in the cognitively simpler *Danio* than in mammals (Table S4).

Additional Presynaptic Scaffold Proteins

RIM-Binding Protein

RIM binding proteins (RIM-BPs) have a series of three SH3 domains, some of which bind to RIM and Ca channels, and one to three fibronectin type 3 domains situated between the first and second SH3 domain (Fig. S12). RIM-BP acts in concert with RIM in recruiting Ca^{2+} channels to the active zone, and together with RIM and Munc13, plays an important role in vesicle priming (22, 23). In *Drosophila* RIM-BP is involved in homeostatic plasticity (24). *Aplysia* has a single RIM Binding Protein (RIM-BP), whereas mammals have three RIM-BPs, two of which have redundant roles at central synapses. RIM-BP3 is expressed in peripheral tissues, and has a specialized role in spermiogenesis. Zebrafish also has three RIMBPs, but there are two RIM-BP2 genes, so all three may have roles in the CNS. RIM-BP has a distant homolog in *Salpingoeca*, a colonial choanoflagellate. Better conserved RIM-BP homologs are present in *Trichoplax* and in cnidarians. As with *Aplysia*, all bilaterian invertebrates have a single RIM-BP (Fig. S15).

CAST/ELKS

ELKS family members include ERC and CAST. CAST/ELKS proteins have a large, highly conserved CAST domain, which occupies most of the protein. ELKS contribute importantly, together with RIM, to active zone organization. ELKS have a broad range of interactions within additional proteins within active zones, and with Ca^{2+} channel subunits (25). Phosphorylation of CAST regulates vesicle release during high frequency firing, reducing synaptic depression (26). The *Aplysia* sequence is unusual, having a zinc finger GAT1 sequence at the N terminus, albeit poorly conserved (Fig. S12) (This GAT1 domain was not found in the other molluscan ELKS sequences analyzed.) There is a single ELKS form in all bilaterian invertebrates, whereas mammals have two forms, sometimes known as CAST1/ERC2 and ELKS. At the C terminus, mammalian ERC2 has a PDZ recognition motif, GIWA, which binds RIM1alpha (27); this same motif is found in the Saccoglossus and nematode ELKS, but not in molluscs. *Danio* has three CAST/ELKS genes. In *Drosophila*, there is a larger

orthologous protein, BruchPilot (28), with a large N terminal SMC domain, which is important for specific forms of memory (29). Bruchpilot is also found in at least some other arthropods (Fig. S16). CAST/ELKS is not found in *Trichoplax*, ctenophores or sponges, and first appears in the common ancestor of cnidarians and bilaterians (Fig. S18).

CASK

CASK is a presynaptic MAGUK, with a conserved CASK (CAlcium/calmodulin Serine Kinase) domain at its N terminus and a Guanylate Kinase domain at its C terminus. Interposed, there are two L27 domains, an SH3 domain and a PDZ domain. *Aplysia* CASK has a similar domain structure, but with a single L27 domain (Fig. S12) (Among molluscan CASKs, the L27 domains are quite variable: *Octopus* CASK has two L27 domains, as do *Pectin* and *Mizuhopecten* and also *Biomphalaria* CASKs; *Crassostrea* CASK has none; *Pomacea* is similar to *Aplysia*, with a single L27 domain). It should be noted that both in *Drosophila* and in mammals, CASK has postsynaptic, as well as presynaptic functions (30). CASK is not found in Porifera, but is present in *Trichoplax* and Cnidaria (Fig. S18). There is a single CASK gene in most species, including in mammals (Fig. S17). Again, *Danio* has an additional gene, as compared with mammals (Fig. S17).

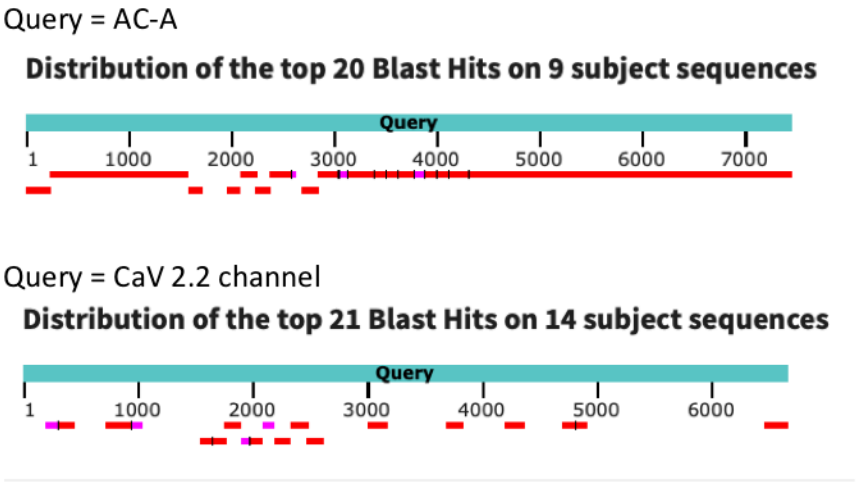

Figure S1. Examples of fragmentation for Reference Transcripts in 2013 *Aplysia* genome assembly from the Broad Institute (Table S1).

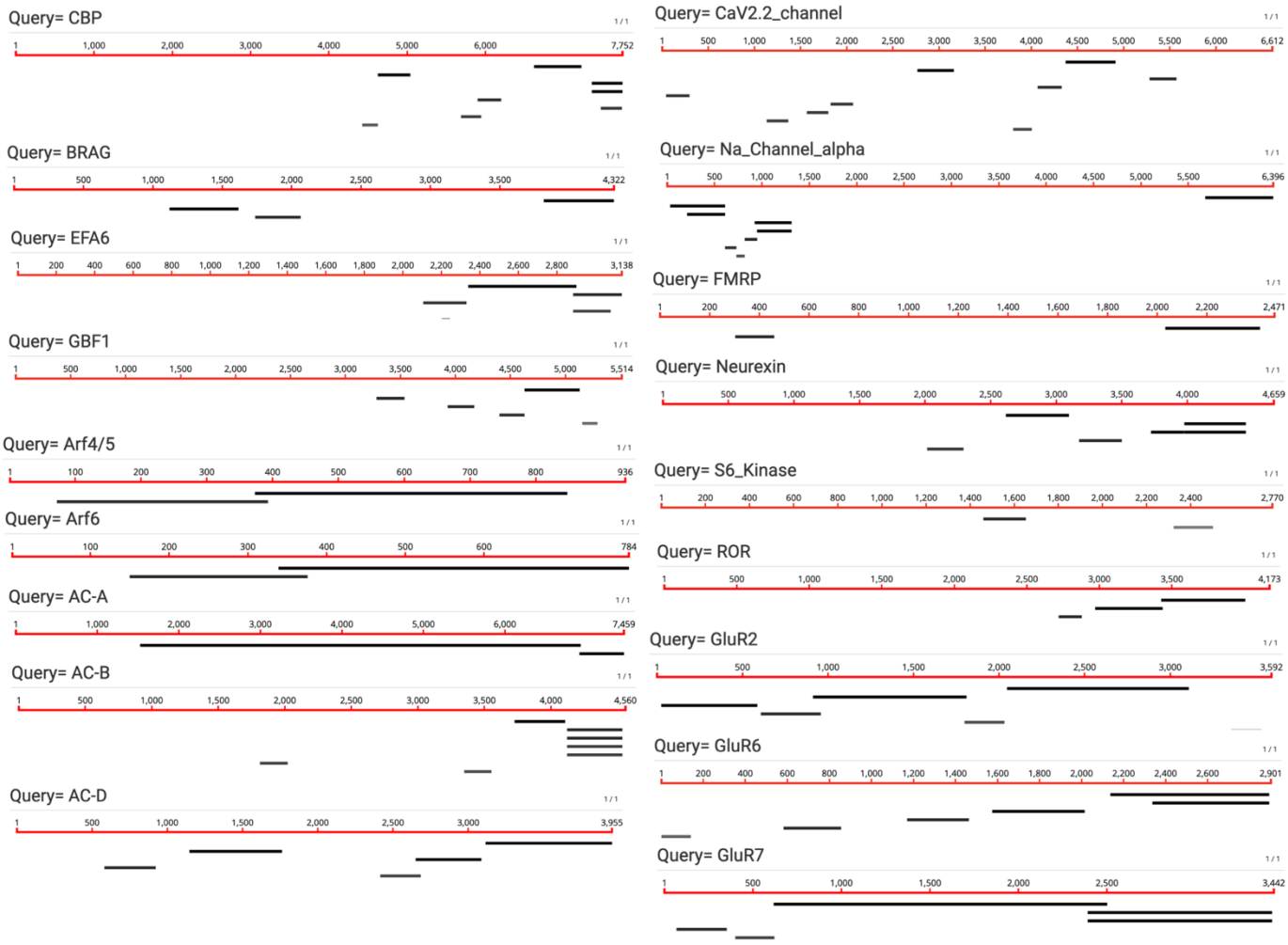

Figure S2. Examples of fragmentation for Reference Transcripts in the 2013 *Aplysia* Trinity transcriptome assembly from the Broad Institute (Table S1).

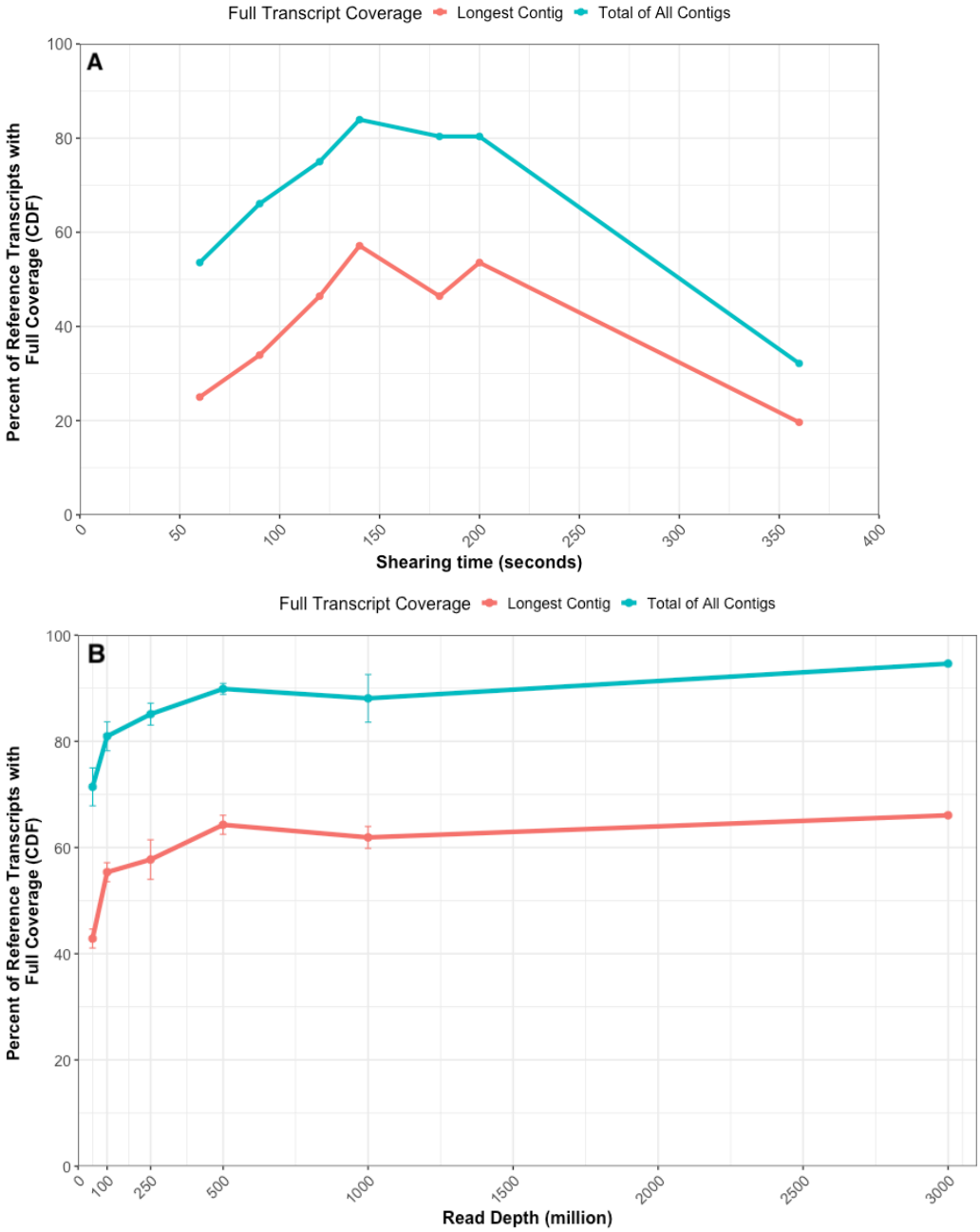

Figure S3. Effect of shearing time and read depth on completeness of transcripts assessed with Reference Transcripts.
A. Shearing time. RNA shearing time was reduced from the standard duration of 6 min to durations from 60 s to 200 s. Reads for each library were randomly downsampled so that an equal number of reads (248 thousand) were assembled for each shearing time. Plot shows percent of Reference Transcripts with full coverage by either longest single contig or the set of all contigs coding for each transcript.
B. Read depth. The full set of 2.97 billion Illumina reads was assembled with DN Trinity (2), or first downsampled to 50 million to 1 million reads prior to assembly. Plot shows percent of Reference Transcripts with full coverage (>97%) by either longest single contig or the set of all contigs coding for each transcript. Downsampling was repeated three times, as this step exhibited variability; points for all downsampled read sets are means +/- SEM.
Note, in both graphs, the substantial difference in coverage between longest contig and all of the contigs coding for the reference transcripts represents the residual fragmentation of Trinity transcripts for approximately one third of the Reference Transcripts, prior to further optimization using PASA and GG-Trinity and Stringtie contigs (Fig. 1A).

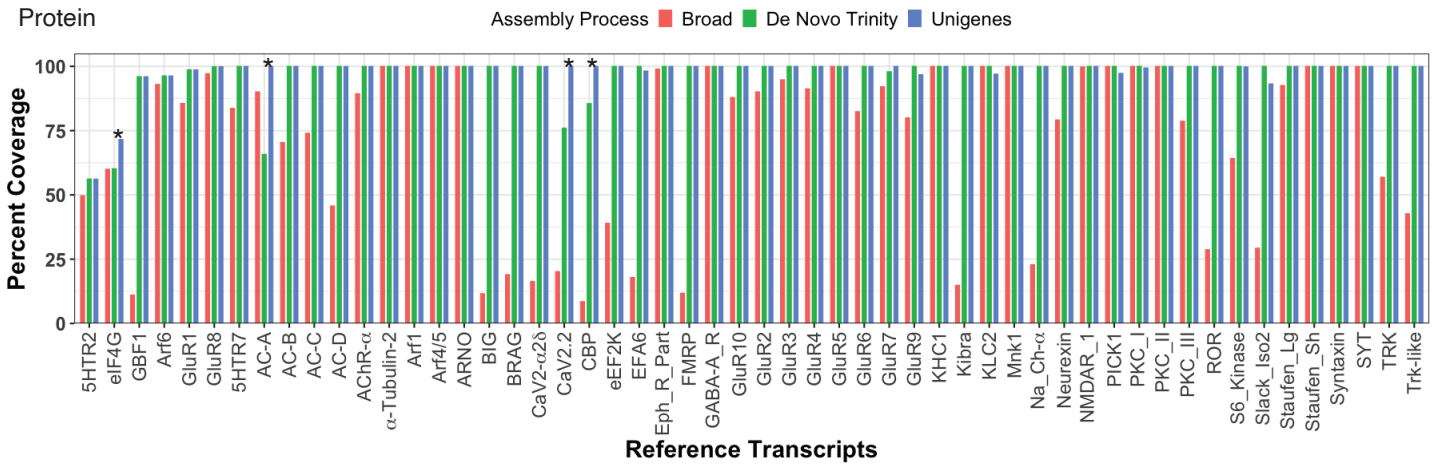

Figure S4. Coverage of protein sequences for reference transcripts by assemblies. The percent coverage of Reference Transcripts by predicted proteins from the Broad 2013 Trinity transcriptome assembly, the final *de novo* Trinity assembly, and the Unigene assembly. Asterisks highlight genes with substantial improvement of coverage with the PASA combined transcript set, which included the *de novo* Trinity, the genome-guided Trinity and the Stringtie contigs (i.e. the Unigene transcript set).

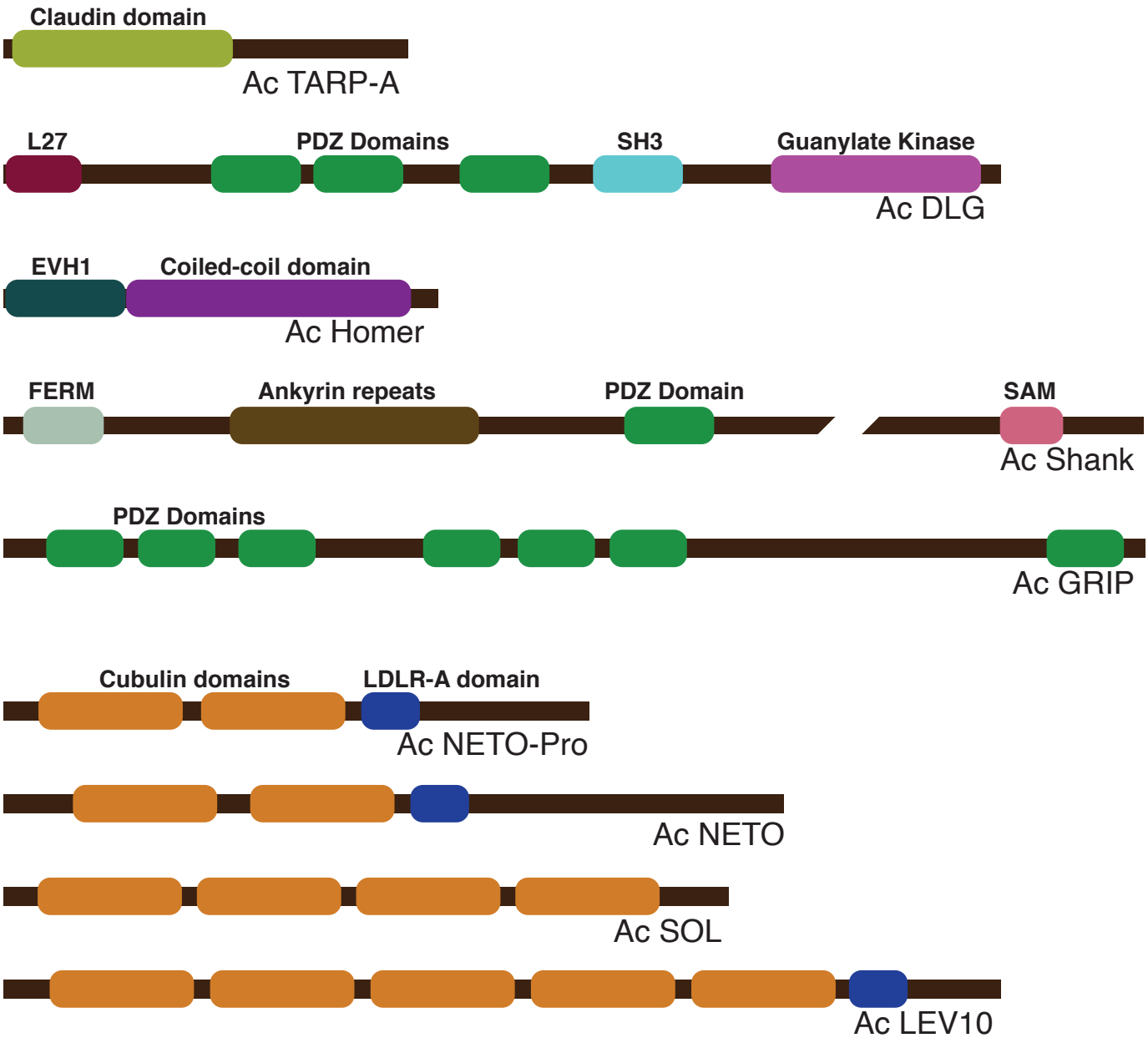

Figure S5. Structure of Aplysia postsynaptic scaffold proteins analyzed using NCBI conserved domains to identify domains. Ac TARP-B1 and Ac TARP-B2 have similar structures as Ac TARP-A. Lengths of scaffold proteins are shown to scale (except for Shank for which the C terminus is shown with a gap); lengths of proteins:TARP-A 350 AAs, Ac DLG 863 AA, Ac Homer 376 AA, Ac Shank 1939, Ac GRIP 1168, Ac NETO-Pro 477 AA, Ac NETO 620, Ac SOL 637, Ac LEV10 864 AA. The SH3 domain of SHANK has diverged sufficiently for it not to appear as a conserved domain, but residual homology does still exist in this region.

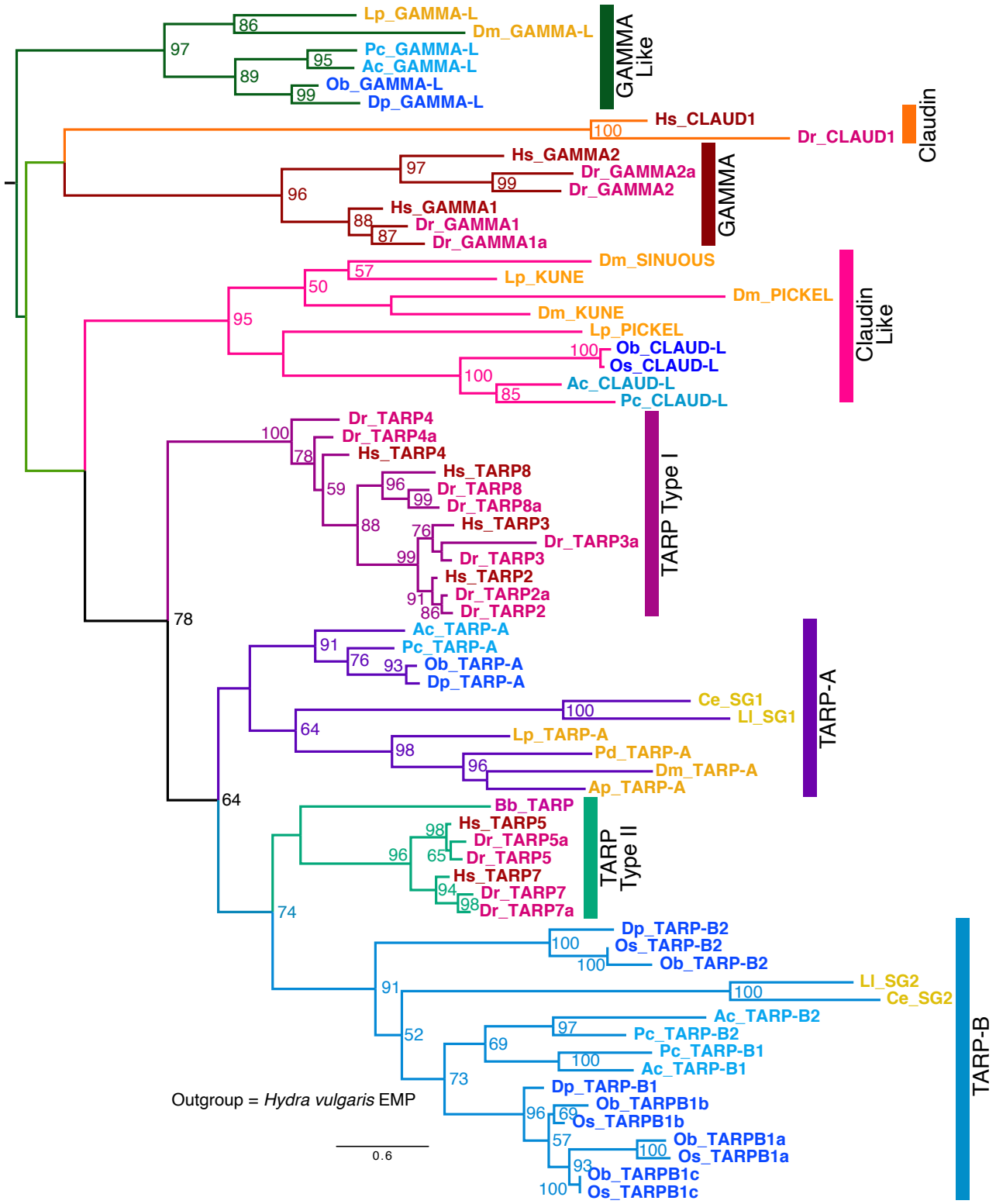

Figure S6. Extended TARP family dendrogram. Only one representative of the vertebrate claudins was used as there is a large expansion of this family in vertebrates. A prebilaterian tetraspanin from *Hydra vulgaris*, epithelial membrane protein (EMP), is the outgroup (not shown in figure). Kune, Sinuous and Pickel are Ecdysozoan proteins implicated in septate junctions previously called invertebrate claudins; the members of the Spiralia group that cluster with these are called Claudin-like (Claud-L).

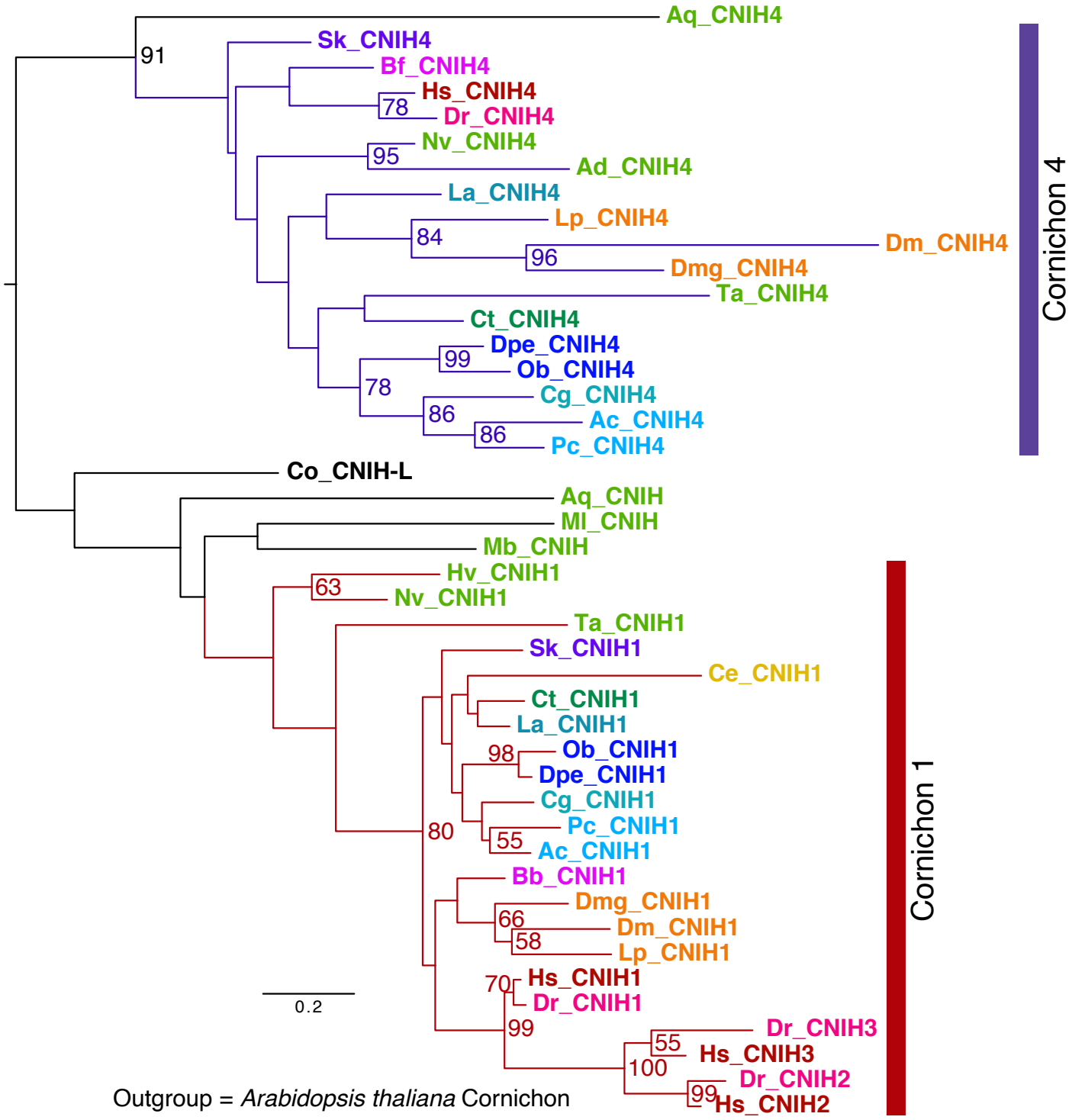

Figure S7. Cornichon dendrogram. Outgroup was the plant *Arabidopsis thaliana* cornichon (not shown in figure). Abbreviations for species not defined previously: Ad, *Acropora digitifera*, Co, *Capsaspora owczarzaki*; Mb *Monosiga brevicolus*; Ml, *Mnemiopsis leidyi*; TA, *Trichoplax adhaerens*. *Amphimedon queenslandica* CNIH and CNIH4 were found on a single transcript.

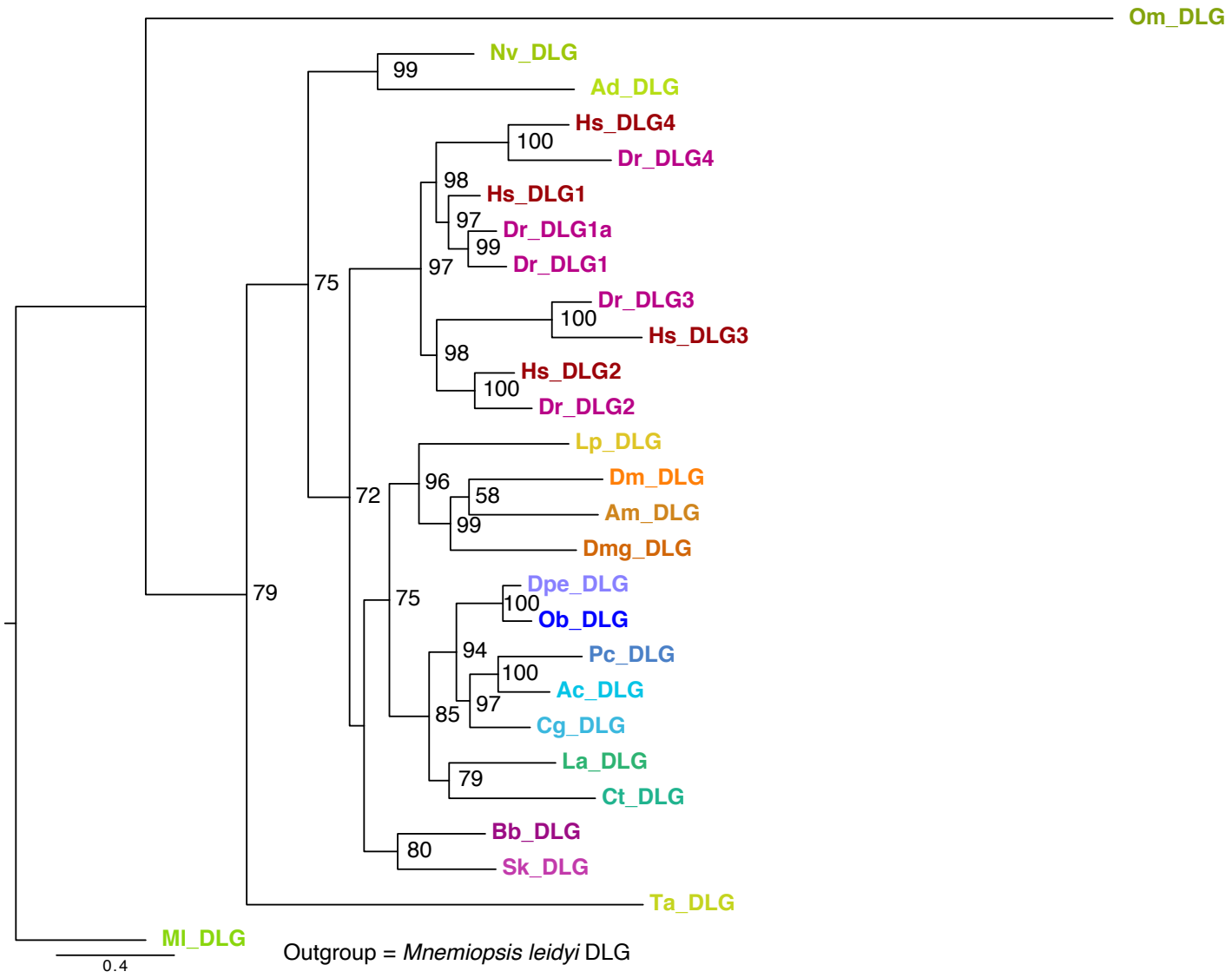

Figure S8. DLG dendrogram. The *Mnemiopsis leidyi* DLG protein was used as an outgroup. Abbreviations for species not defined previously; Om, *Oopsacas minuta* (Glass Sponge).

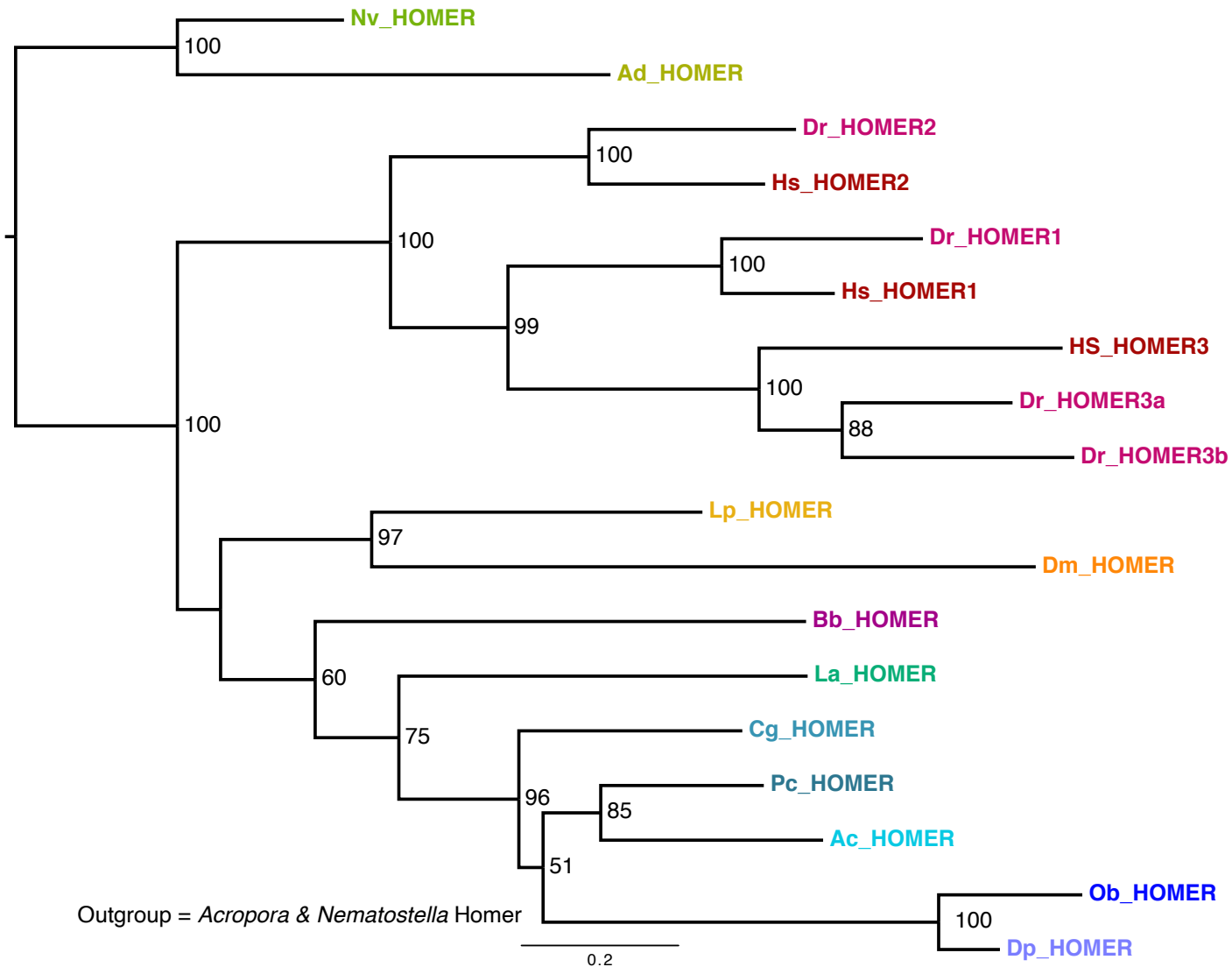

Figure S9. Homer dendrogram. Abbreviations for species not defined previously Sr, *Salpingoeca rosetta.* Outgroup is the two cnidarian Homer sequences, from *Acropora digitifera* and *Nematostella vectensis*.

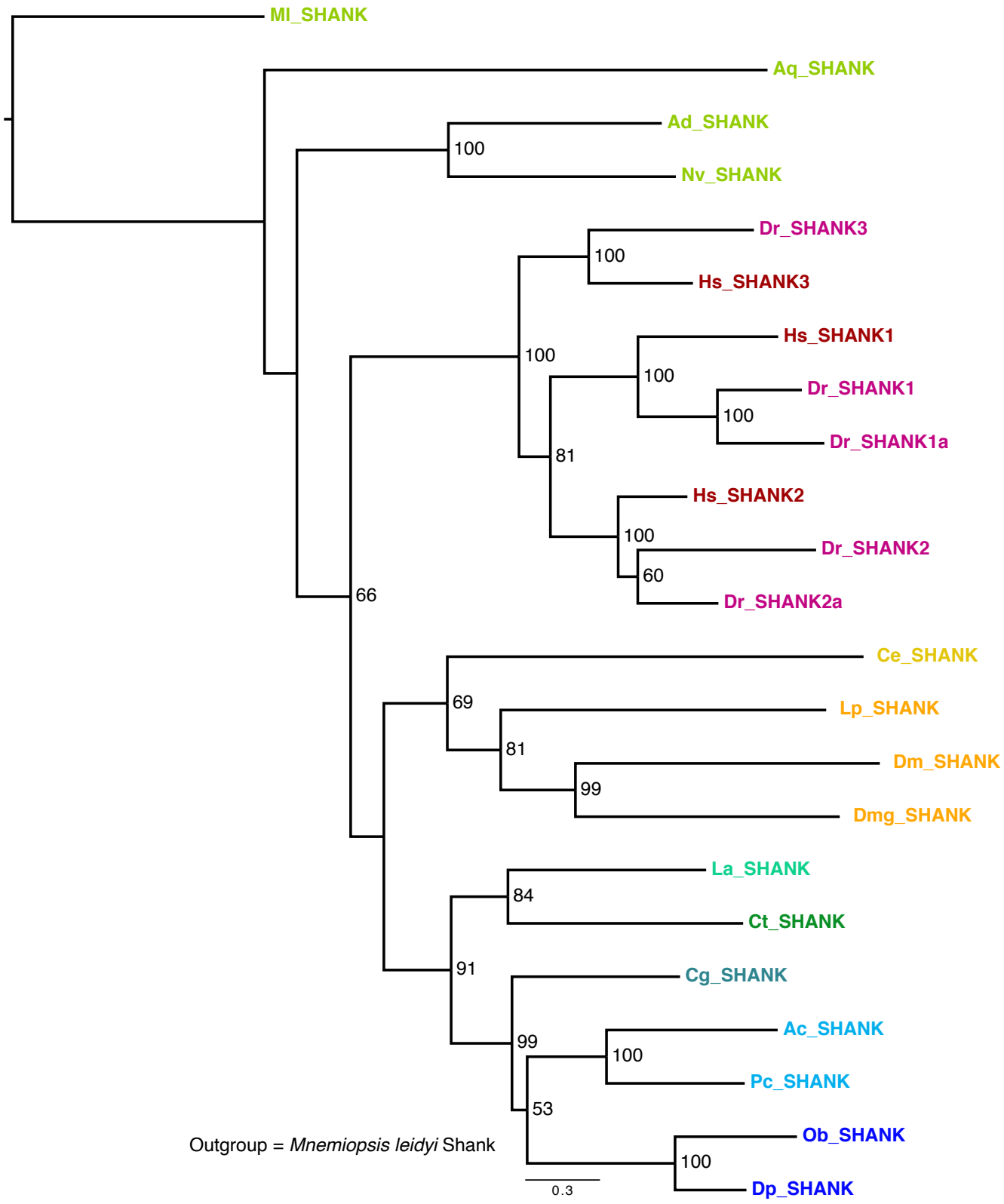

Figure S10. Shank Dendrogram. Abbreviations for species not defined previously: Dg, *Dendronephthya gigantea*; Tn, *Trichinella nativa*. Outgroup is *Mneiopsis leidyi* Shank.

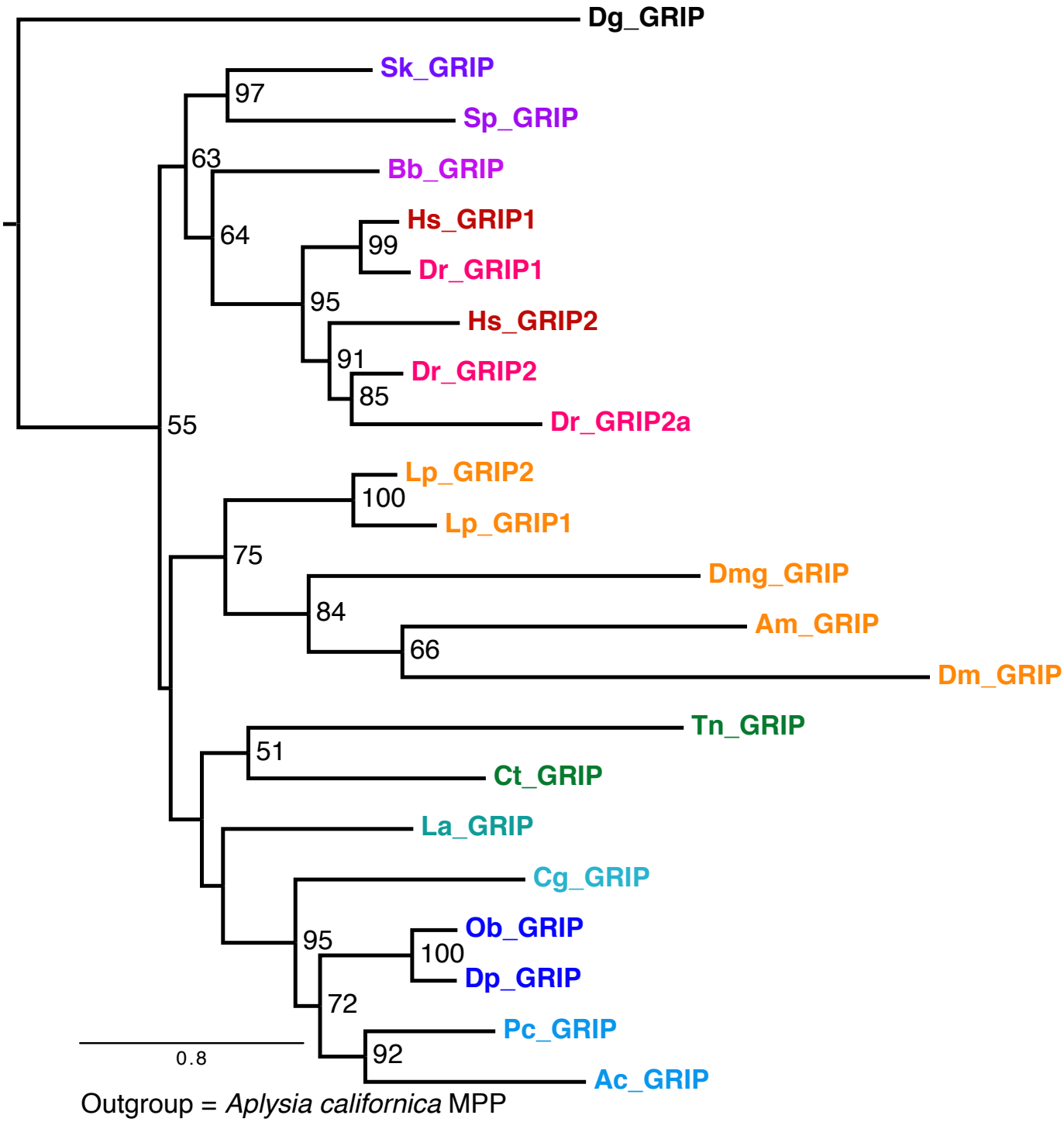

Figure S11. GRIP dendrogram. Outgroup was *Aplysia californica* MPP (not shown in figure).

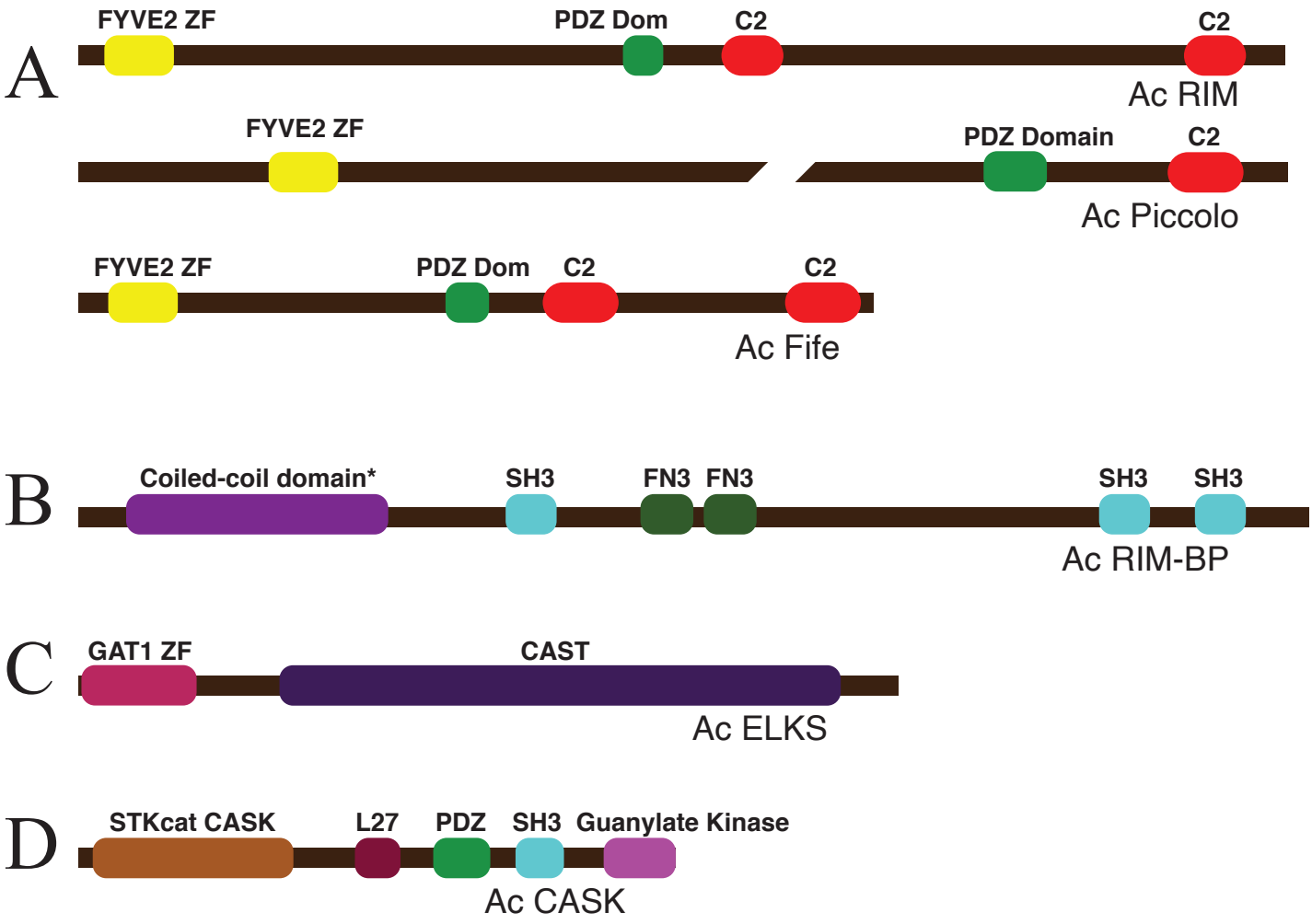

Figure S12. Presynaptic scaffold protein domains in *Aplysia*. **A.** *Aplysia* RIM Superfamily members. **B.** *Aplysia* RIM Binding Protein. **C.** *Aplysia* ELKS. **D.** *Aplysia* CASK. Schematics are to scale (except for Ac Piccolo, for which the C terminus is shown with a gap); lengths of proteins: Ac RIM 1717 AAs, Ac Piccolo 2850, Ac Fife 1132 AAs, Ac RIM-BP 1751 AAs, Ac ELKS 1167 AAs, Ac CASK 850 AAs.

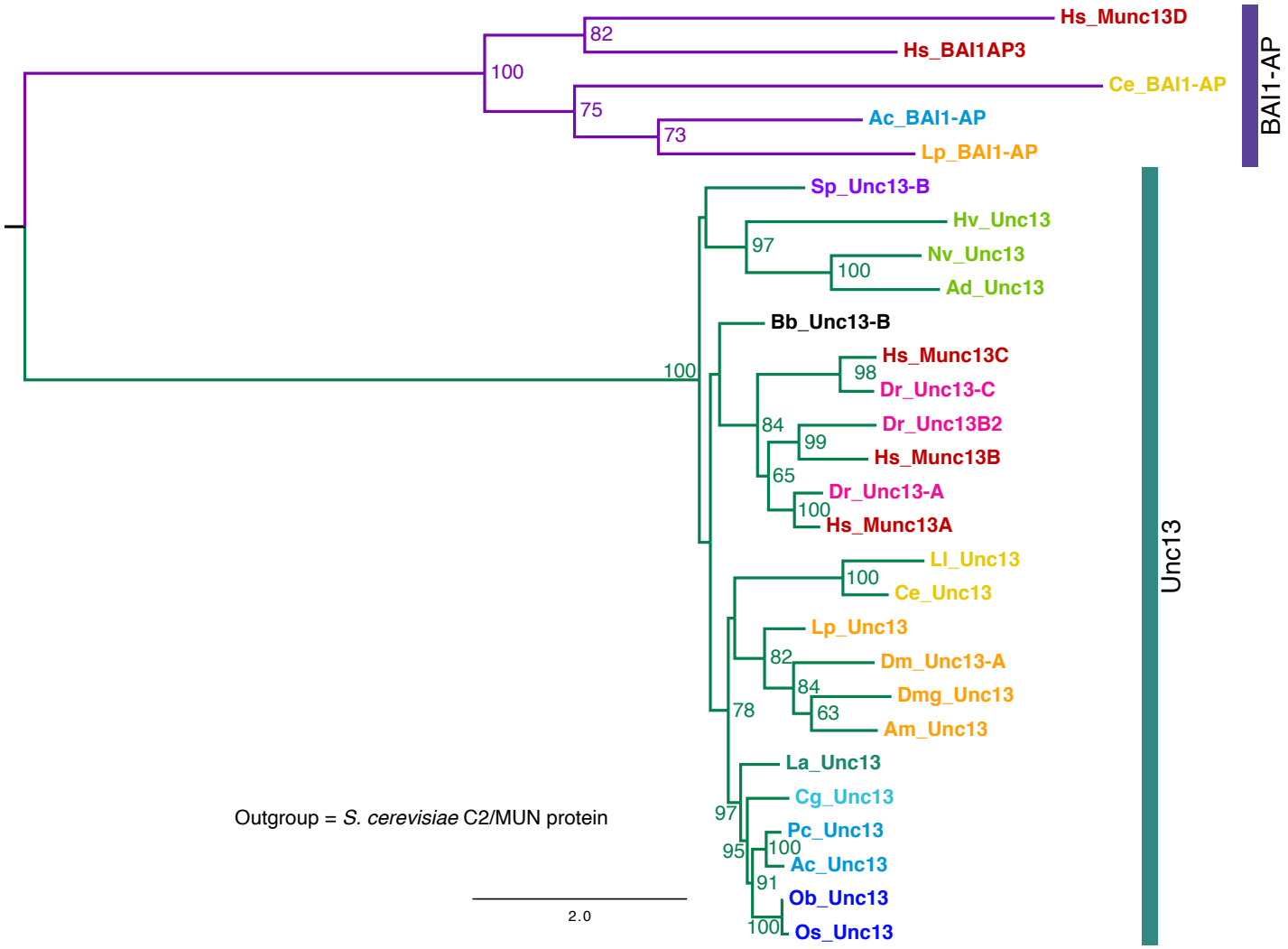

Figure S13. Unc13 dendrogram. Alignment trimmed with TrimAl prior to RAxML analysis (See Methods). Both BAI1-AP and Munc13D sequences [and also CAPS sequences (not included in dendrogram)] cluster separately from unc13 sequences.

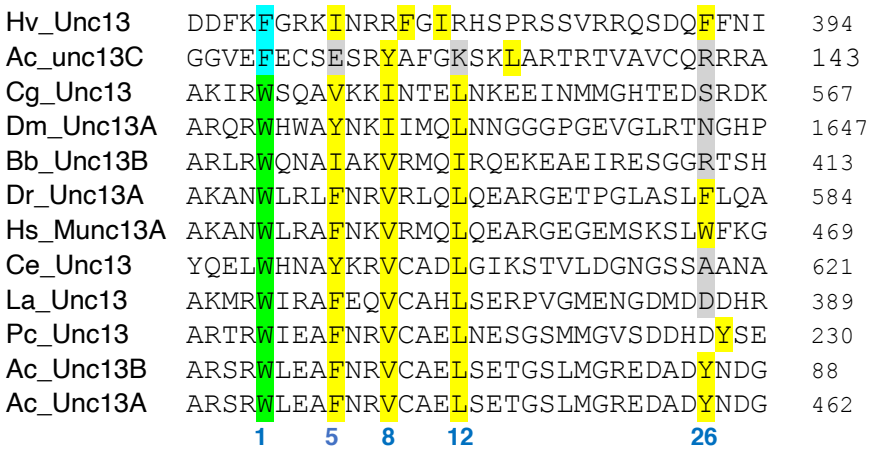

Figure S14. Putative calmodulin binding sites in *Aplysia* unc13 isoforms aligned with CaM-binding site in human Munc13 isoforms A. Initial hydrophobic anchor Tryp residues highlighted in green; additional hydrophobic residues highlighted in yellow; putative Tyr anchor residues highlighted in aqua. Note the characteristic pattern of 1-5-8-(12)-26 hydrophobic residues (31, 32) conserved across bilaterian phyla, with *Aplysia* unc13A and unc13B exhibiting this same pattern as *Danio* and Human Munc13-A. Some bilaterian species have lost the hydrophobic residue at postion 26; in *Pomacea*, this residue is shifted by one residue. *Aplysia* unc13-C displays a substantially 1-8-15 divergent pattern, which we speculate may similarly bind CaM, as it resembles another pattern of CaM-binding sequences (33). Human bMunc13-B has a more dramatically divergent CaM-binding site, with additional hydrophobic residues extending N terminally to the anchor residue (34, 35); because this bMunc13-B sequence is so distinct, it is not included in this alignment. [Note as indicated by the position of the N terminal residues shown by numbers at right, these sequences vary substantially in the length of the region N terminal to the CaM-binding region, located proximal to the C1-C2 module (e.g. the *Aplysia* isoforms illustrated in Fig. 5).

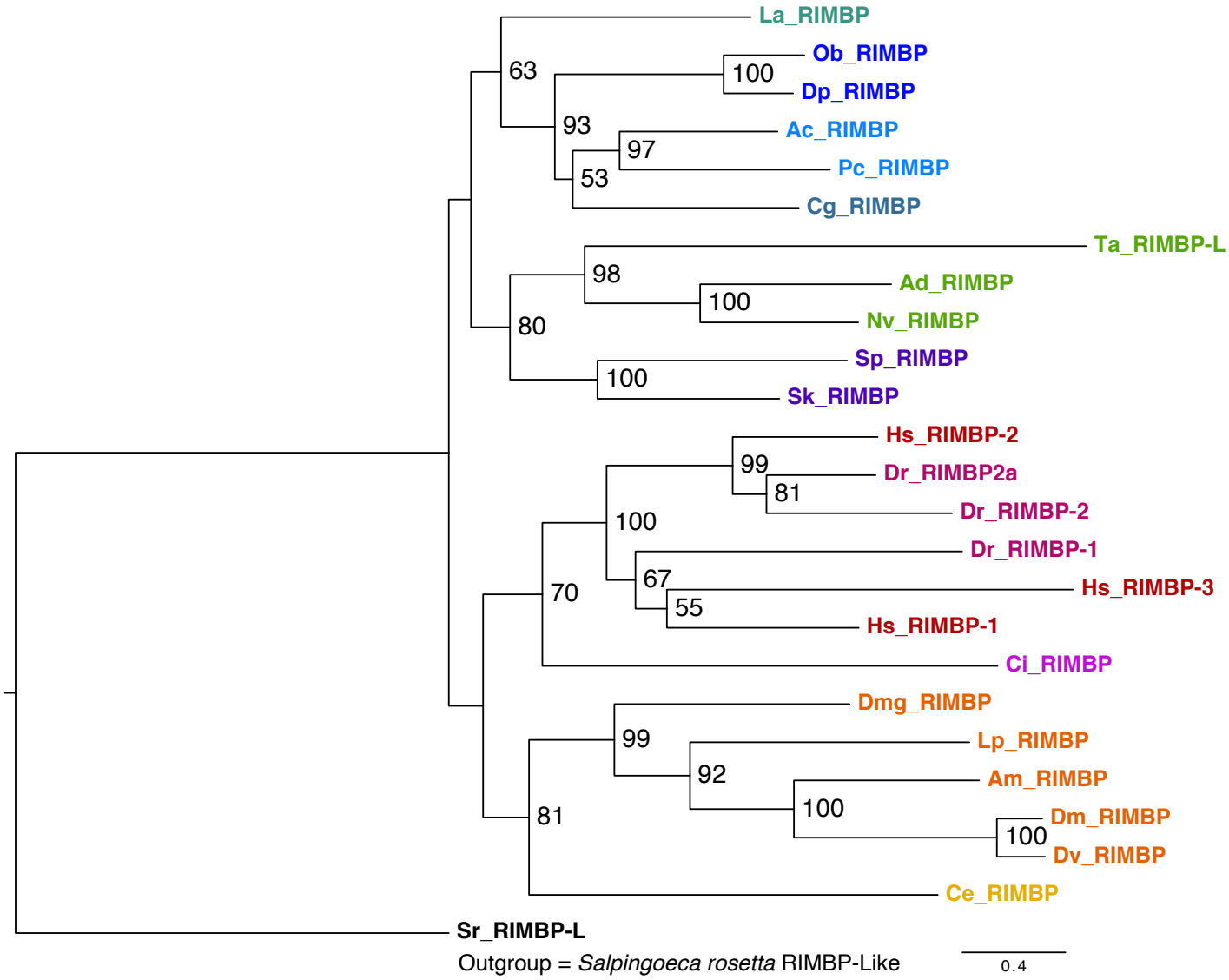

Fig. S15. RIM-BP dendrogram. Abbreviations for species not defined previously: Dv, *Drosophila virilis.* A distantly related gene in the choanoflagellate *Salpingoeca rosetta* (Sr) served as the outgroup.

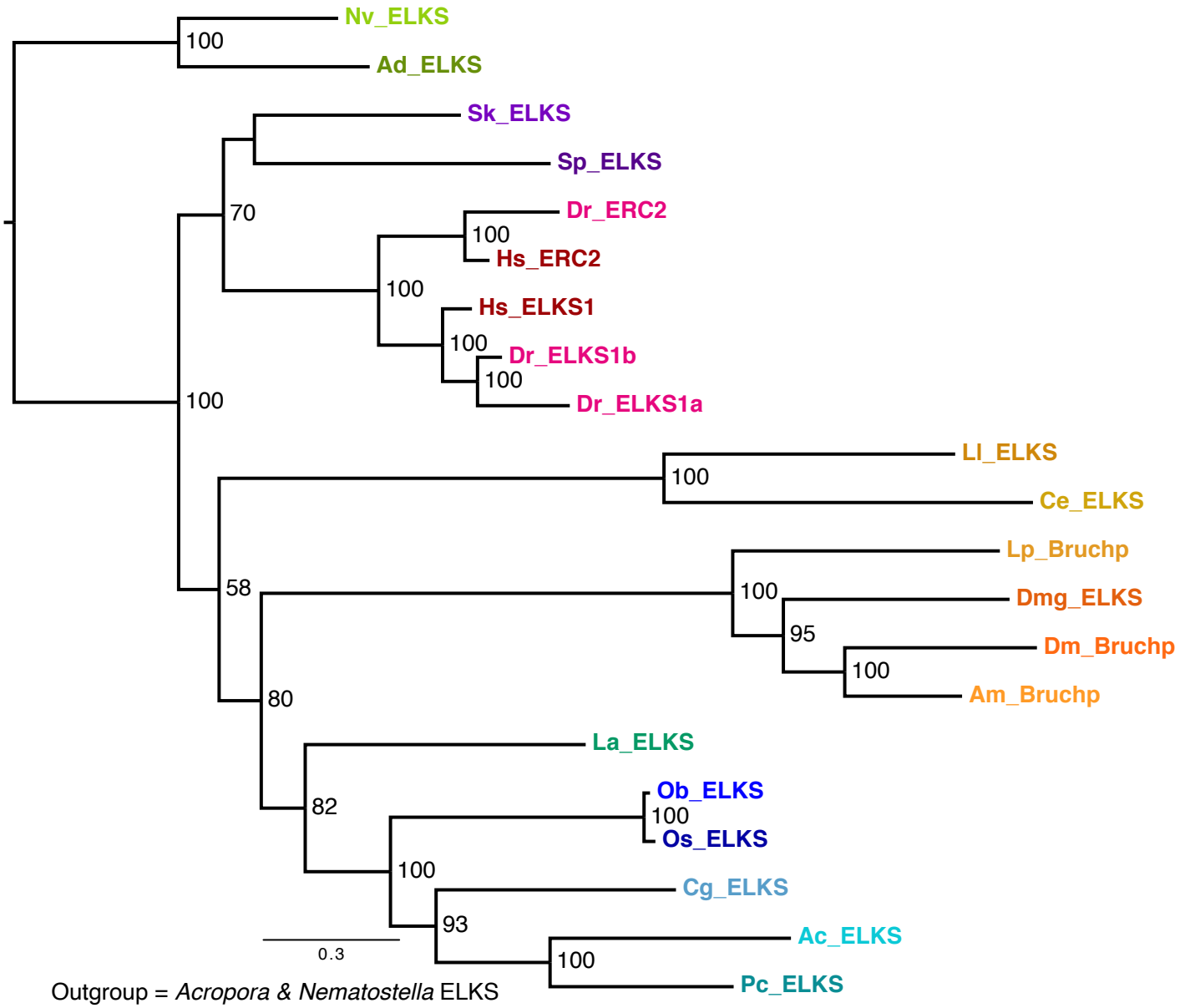

Fig. S16. CAST/ELKS dendrogram. Abbreviations as in previous dendrograms. The two cnidarian ELKS sequences from *Acropora* and *Nematostella* served as the outgroup.

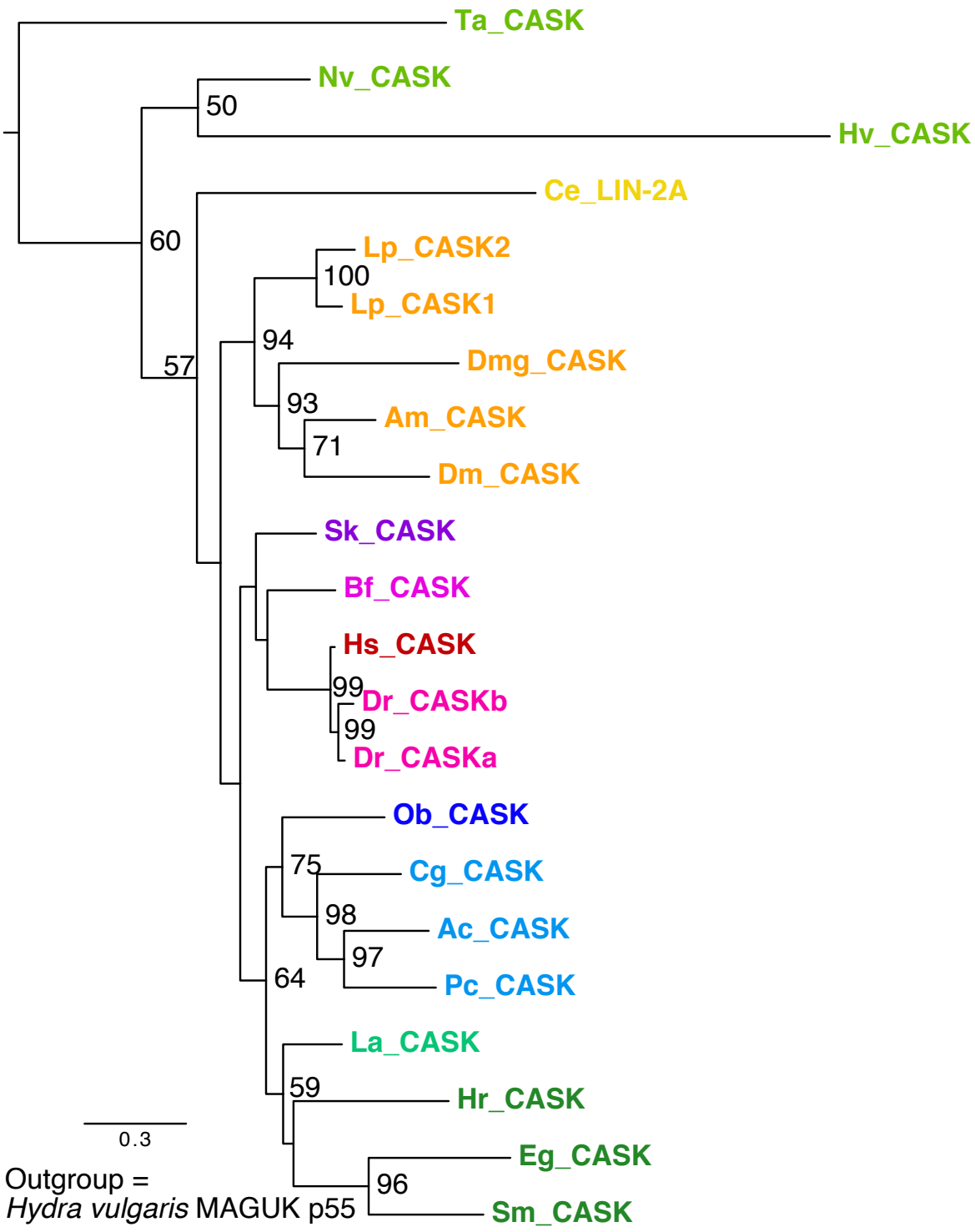

Fig. S17. CASK Dendrogram. Abbreviations as in previous dendrograms, with two additional abbreviations: Eg, *Echinococcus* granulosus; Sm, *Schmidtea mediterranea* (both platyhelminthes); Hr, *Helobdella robusta* (an annelid). *Hydra vulgaris* MAGUG P5 served as the outgroup (not shown in figure).

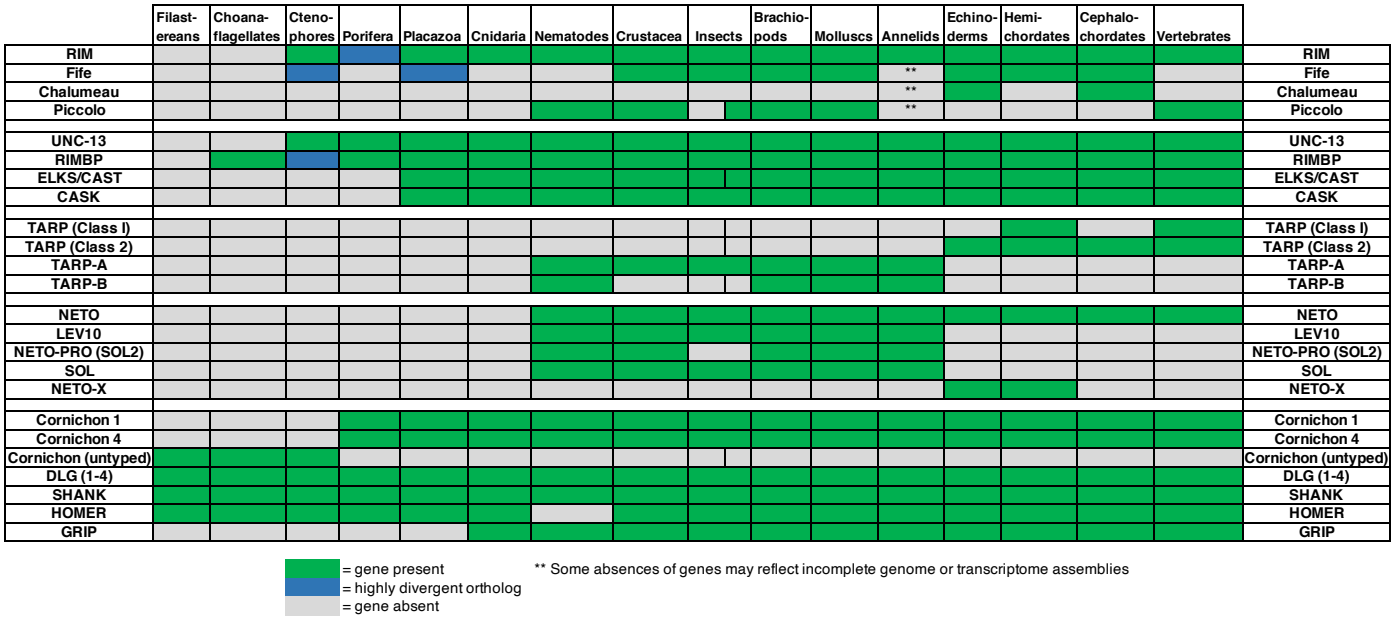

Figure S18. Representation of scaffold proteins in Holozoa. In general, the absence of a gene in a particular clade could reflect incomplete genome or transcriptome assemblies. However, where there are more well-studied species, such as in insects, an incomplete assembly is less of a likely explanation. We use Chalumeau to refer to the group of RIM superfamily genes found in echinoderms and lancelets (Fig. 4A).

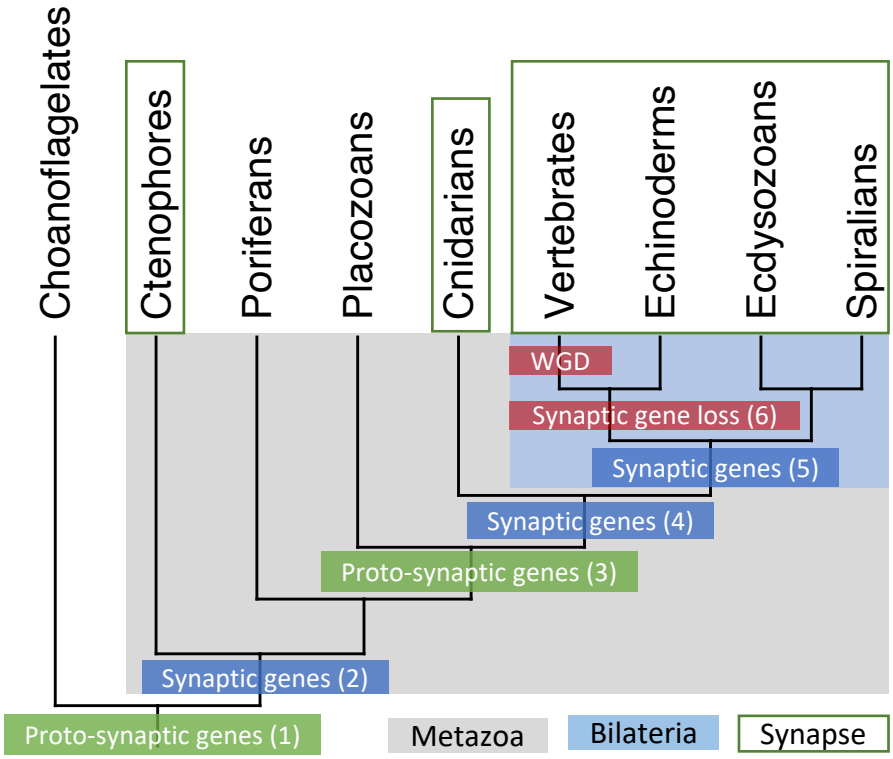

Figure S19. Gradual addition of synaptic genes through exaptation or appearance of novel sequences in early phyla with synapses. Numbers in parentheses represent refer to key nodes in the appearance of genes used in bilaterian synapses. Examples include: (1) Synaptic SNAREs; (2) RIM, CaV2; (3) vGlut, Synaptotagmin 1; ELKS; (4) GRIP, vGAT; (5) Piccolo & Fife; TARP, NETO; (6) genes lost, Fife, SOL. Green outline signifies clades with neurons and synapses. WGD, whole genome duplication. Grey shading signifies metazoan clades. Blue shading signifies bilaterian clades.

Assembly	Source	Bioproject
Aplysia Transcriptome	Broad Institute	NCBI Bioproject PRJNA77701
Aplysia Genome	Broad Institute	NCBI Bioproject PRJNA13635*
Aplysia Transcriptome	This paper	NCBI Bioproject PRJNA792581
Aplysia Genome	Zimin et al.	Aplysia Gene Tools http://www.aplysiatools.org

Table S1. Assemblies used. *Note, the assemblies on NCBI are continuing to evolve based on new sequences and new scaffolds; for example, the genomic sequences now available on BLAST include scaffolds and may be less fragmented than the original Broad genome that we analyzed, or than what is available on NCBI BLAST of WGS assemblies or available for download (currently the Broad genome contigs).

Software	Version	Reference or website
Trimmomatic	ver 0.38	Bolger et al. (2014) (1)
Trinity with Diginorm*	ver 2.9.1	Haas et al. (2013) (2)
Stringtie	ver 2.1.1	Pertea et al. (2015) (3)
PASA	ver 2.4.1;	Haas et al. (2008) (4)
TransRate	ver 1.0.3	Smith-Unna et al. 2016 (5)
Transdecoder	ver 5.5.0	https://github.com/TransDecoder/TransDecoder
CD-HIT-EST	ver 4.7	Li et al. (2006), Fu et al (2012) (6, 7)
Trinotate	ver 3.2.1	https://github.com/Trinotate/Trinotate
RAxML GUI	ver 2.0.5	Stamatakis (2014), Edler et al. (2020) (9, 10)
TrimAl	ver 1.3	Capella-Gutierrez et al. (2009) (11) http://phylemon2.bioinfo.cipf.es/utilities.html
FigTree	ver 1.4.4	http://tree.bio.ed.ac.uk/software/figtree/

Table S2. Software used.
*Diginorm step in Trinity provides read normalization.

Total No. of Contigs	71,104
Total bases	126,237,157
Average sequence length	1,775.4
N50	3,151
N90	699
Shortest sequence length	303
Longest sequence length	42,564
GC percentage	44.5

Table S3. Statistics for Unigene contigs. Transdecoder minimum ORF length: 300 bp (Fig. 1A).

Family	Aplysia	Octopus	Danio	Human
TARPs	3	5	12	6
Cubulin containing (Neto/Sol/LEV-10)	4	4	2	2
Cornichons	2	2	4	4
DLGs (PSD-95 and others)	1	1	5	4
GRIP	1	1	3	2
SHANK	1	1	5	3
HOMER	1	1	5	3
Postsynaptic Total	14	15	36	24
RIM/Fife/Piccolo/Bassoon	3	3	7	4
Unc 13	1	1	3	3
CASK	1	1	2	1
RIM-BP	1	1	3	2
ELKS	1	1	3	2
Presynaptic Total	7	7	19	12

Table S4. Synaptic scaffold proteins. Numbers of family members per species.

	Human	Danio	Human ÷ Danio		Octopus	Aplysia	Octopus ÷ Aplysia		Human ÷ Aplysia
GPCR and signaling	2296	4260	0.54		1273	1359	0.94		1.69
Scaffolding domains	3976	7879	0.5		2633	2335	1.13		1.7
Ion channels/Transporters	589	1298	0.45		818	707	1.16		0.83
RNA binding domains	589	491	1.2		283	247	1.14		2.38
Calcium signaling	381	842	0.45		220	165	1.33		2.31
Extracellular domains	931	2815	0.55		630	541	1.16		1.72
Synaptic vesicle/transmitter	61	110	0.49		48	32	1.51		1.91

Table S5. The number of proteins containing a PFAM domain using (PFAMA33.1) was determined using the Uniprot reference proteome sets from *Homo sapiens* (Hs) and *Danio rerio* (Dr), the predicted protein set from *Octopus bimaculoides* (Ob (8) or from the Unigenes for *Aplysia* (Ac) in this paper. The ratios (Hs/Dr; Ob/Ac and Hs/Ac) are calculated from the values shown. The underlying PFAMs used for each family is in Dataset 2.

Dataset 1 Legend. NCBI sequence IDs for proteins in dendrograms. There is one sheet for each dendrogram, other than for Figure 2 and Figure S6 which are both included in the TARP sheet.

Dataset 2 Legend. PFAM families of signaling, scaffolding and neural PFAMs. The individual PFAMs chosen to examine the numbers of proteins with specific signaling (i.e. kinases), scaffolding (I.e. PDZ domains) and neuron specific (I.e ion channel) domains are shown. The number of proteins containing a PFAM domain using (PFAMA33.1) was determined using the Uniprot reference proteome sets from *Homo sapiens* (Hs) and *Danio rerio* (Dr), the predicted protein set from *Octopus bimaculoides* (Ob (8)) or from the Unigenes for Aplysia (Ac) in this paper.

SI References

1. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).
2. B. J. Haas *et al.*, De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8**, 1494-1512 (2013).
3. M. Pertea *et al.*, StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**, 290-295 (2015).
4. B. J. Haas *et al.*, Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol* **9**, R7 (2008).
5. R. Smith-Unna, C. Boursnell, R. Patro, J. M. Hibberd, S. Kelly, TransRate: reference-free quality assessment of de novo transcriptome assemblies. *Genome Res* **26**, 1134-1144 (2016).
6. L. Fu, B. Niu, Z. Zhu, S. Wu, W. Li, CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150-3152 (2012).
7. W. Li, A. Godzik, Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658-1659 (2006).
8. C. B. Albertin *et al.*, The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature* **524**, 220-224 (2015).
9. D. Edler, J. Klein, A. Antonelli, D. Silvestro, raxmlGUI 2.0: A graphical interface and toolkit for phylogenetic analyses using RAxML. *Methods in Ecology and Evolution* **12**, 373-377 (2020).
10. A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313 (2014).
11. S. Capella-Gutierrez, J. M. Silla-Martinez, T. Gabaldon, trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972-1973 (2009).
12. P. J. Brockie *et al.*, Cornichons control ER export of AMPA receptors to regulate synaptic excitability. *Neuron* **80**, 129-142 (2013).
13. J. Schwenk *et al.*, High-resolution proteomics unravel architecture and molecular diversity of native AMPA receptor complexes. *Neuron* **74**, 621-633 (2012).
14. E. Sauvageau *et al.*, CNIH4 interacts with newly synthesized GPCR and controls their export from the endoplasmic reticulum. *Traffic* **15**, 383-400 (2014).
15. A. de Mendoza, H. Suga, I. Ruiz-Trillo, Evolution of the MAGUK protein gene family in premetazoan lineages. *BMC Evol Biol* **10**, 93 (2010).
16. P. Burkhardt *et al.*, Evolutionary insights into premetazoan functions of the neuronal protein homer. *Mol Biol Evol* **31**, 2342-2355 (2014).
17. J. C. Tu *et al.*, Coupling of mGluR/Homer and PSD-95 complexes by the Shank family of postsynaptic density proteins. *Neuron* **23**, 583-592 (1999).
18. J. C. Tu *et al.*, Homer binds a novel proline-rich motif and links group 1 metabotropic glutamate receptors with IP3 receptors. *Neuron* **21**, 717-726 (1998).
19. A. Alie, M. Manuel, The backbone of the post-synaptic density originated in a unicellular ancestor of choanoflagellates and metazoans. *BMC Evol Biol* **10**, 34 (2010).
20. S. K. Ponna, M. Myllykoski, T. M. Boeckers, P. Kursula, Structure of an unconventional SH3 domain from the postsynaptic density protein Shank3 at ultrahigh resolution. *Biochem Biophys Res Commun* **490**, 806-812 (2017).
21. T. W. Abrams, W. S. Sossin, "Invertebrate genomics provide insights into the origin of synaptic transmission." in Oxford Handbook of Invertebrate Neurobiology J. H. Byrne, Ed. (2018).
22. M. M. Brockmann *et al.*, A Trio of Active Zone Proteins Comprised of RIM-BPs, RIMs, and Munc13s Governs Neurotransmitter Release. *Cell Rep* **32**, 107960 (2020).

23. A. G. Petzoldt *et al.*, RIM-binding protein couples synaptic vesicle recruitment to release sites. *J Cell Biol* **219** (2020).
24. M. Muller, O. Genc, G. W. Davis, RIM-binding protein links synaptic homeostasis to the stabilization and replenishment of high release probability vesicles. *Neuron* **85**, 1056-1069 (2015).
25. R. G. Held, P. S. Kaeser, ELKS active zone proteins as multitasking scaffolds for secretion. *Open Biol* **8** (2018).
26. S. Hamada, T. Ohtsuka, CAST: Its molecular structure and phosphorylation-dependent regulation of presynaptic plasticity. *Neurosci Res* **127**, 25-32 (2018).
27. J. Lu, H. Li, Y. Wang, T. C. Sudhof, J. Rizo, Solution structure of the RIM1alpha PDZ domain in complex with an ELKS1b C-terminal peptide. *J Mol Biol* **352**, 455-466 (2005).
28. R. J. Kittel *et al.*, Bruchpilot promotes active zone assembly, Ca2+ channel clustering, and vesicle release. *Science* **312**, 1051-1054 (2006).
29. S. Knapek, S. Sigrist, H. Tanimoto, Bruchpilot, a synaptic active zone protein for anesthesia-resistant memory. *J Neurosci* **31**, 3453-3458 (2011).
30. L. E. LaConte *et al.*, CASK stabilizes neurexin and links it to liprin-alpha in a neuronal activity-dependent manner. *Cell Mol Life Sci* **73**, 3599-3621 (2016).
31. S. Herbst, N. Lipstein, O. Jahn, A. Sinz, Structural insights into calmodulin/Munc13 interaction. *Biol Chem* **395**, 763-768 (2014).
32. F. Rodriguez-Castaneda *et al.*, Modular architecture of Munc13/calmodulin complexes: dual regulation by Ca2+ and possible function in short-term synaptic plasticity. *EMBO J* **29**, 680-691 (2010).
33. C. Andrews, Y. Xu, M. Kirberger, J. J. Yang, Structural Aspects and Prediction of Calmodulin-Binding Proteins. *Int J Mol Sci* **22** (2020).
34. N. Lipstein *et al.*, Nonconserved Ca(2+)/calmodulin binding sites in Munc13s differentially control synaptic short-term plasticity. *Mol Cell Biol* **32**, 4628-4641 (2012).
35. C. Piotrowski *et al.*, Delineating the Molecular Basis of the CalmodulinbMunc13-2 Interaction by Cross-Linking/Mass Spectrometry-Evidence for a Novel CaM Binding Motif in bMunc13-2. *Cells* **9** (2020).