

# Integrating Embryonic Development and Evolutionary History to Characterize Tentacle-Specific Cell Types in a Ctenophore

Leslie S. Babonis,<sup>\*1</sup> Melissa B. DeBiasse,<sup>1</sup> Warren R. Francis,<sup>2</sup> Lynne M. Christianson,<sup>2</sup> Anthony G. Moss,<sup>3</sup> Steven H.D. Haddock,<sup>2</sup> Mark Q. Martindale,<sup>1</sup> and Joseph F. Ryan<sup>\*1</sup>

<sup>1</sup>Whitney Laboratory for Marine Bioscience, University of Florida, St. Augustine, FL

<sup>2</sup>Monterey Bay Aquarium Research Institute (MBARI), Moss Landing, CA

<sup>3</sup>Department of Biological Sciences, Auburn University, Auburn, AL

**\*Corresponding authors:** E-mails: babonis@whitney.ufl.edu; joseph.ryan@whitney.ufl.edu.

**Associate editor:** Iñaki Ruiz-Trillo

## Abstract

**The origin of novel traits can promote expansion into new niches and drive speciation. Ctenophores (comb jellies) are unified by their possession of a novel cell type: the colloblast, an adhesive cell found only in the tentacles. Although colloblast-laden tentacles are fundamental for prey capture among ctenophores, some species have tentacles lacking colloblasts and others have lost their tentacles completely. We used transcriptomes from 36 ctenophore species to identify gene losses that occurred specifically in lineages lacking colloblasts and tentacles. We cross-referenced these colloblast- and tentacle-specific candidate genes with temporal RNA-Seq during embryogenesis in *Mnemiopsis leidyi* and found that both sets of candidates are preferentially expressed during tentacle morphogenesis. We also demonstrate significant upregulation of candidates from both data sets in the tentacle bulb of adults. Both sets of candidates were enriched for an N-terminal signal peptide and protein domains associated with secretion; among tentacle candidates we also identified orthologs of cnidarian toxin proteins, presenting tantalizing evidence that ctenophore tentacles may secrete toxins along with their adhesive. Finally, using cell lineage tracing, we demonstrate that colloblasts and neurons share a common progenitor, suggesting the evolution of colloblasts involved co-option of a neurosecretory gene regulatory network. Together these data offer an initial glimpse into the genetic architecture underlying ctenophore cell-type diversity.**

**Key words:** trait loss, comparative transcriptomics, ctenophore, novelty, development, biological adhesive.

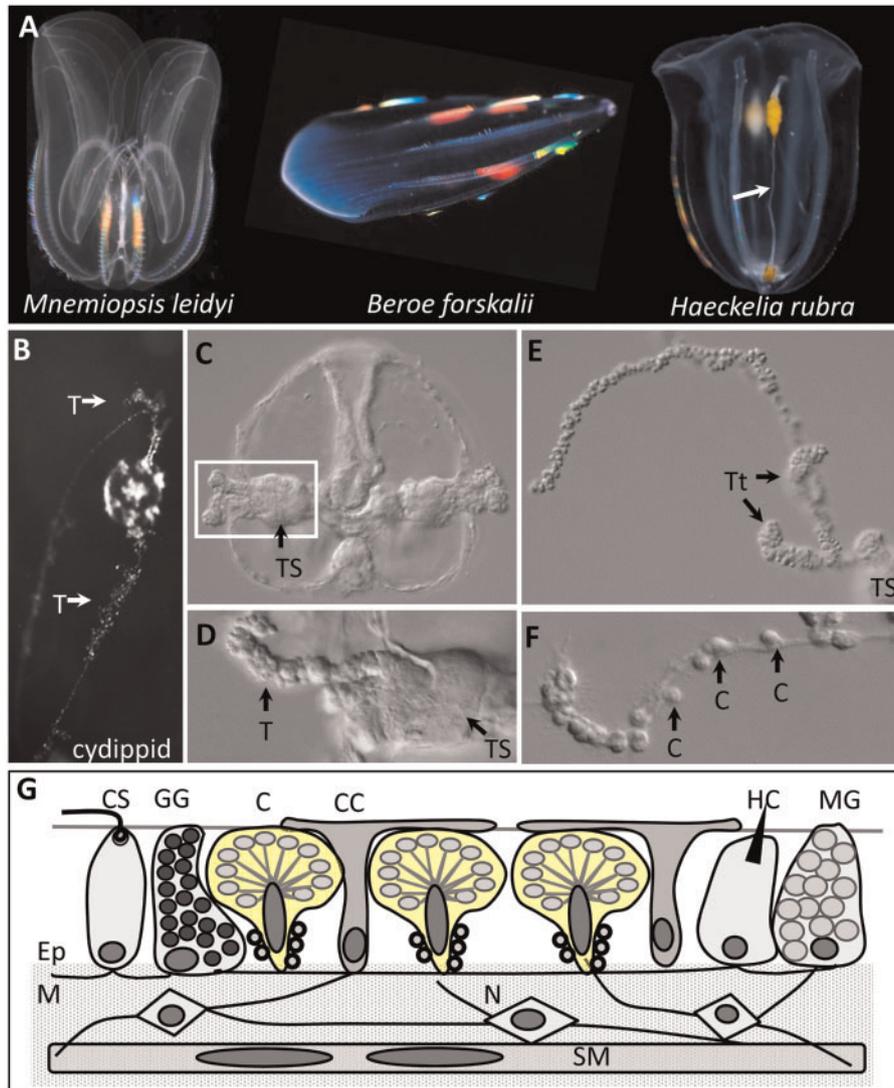
## Background

Insight into how novelty is generated is important for understanding the origin and diversification of multicellular life. An outstanding challenge, however, is finding a model for which the direction of evolutionary change is known and the novelty of interest is easy to characterize. Ctenophores (comb jellies) are gelatinous marine invertebrates that diverged from the rest of animals over 800 Ma (Dohrmann and Worheide 2017); although their phylogenetic position remains contentious (Dunn et al. 2008; Ryan et al. 2013; Moroz et al. 2014; Borowiec et al. 2015; Simion et al. 2017; Whelan et al. 2017), they are clearly among the first lineages to diverge from the rest of animals. While they share several anatomical features in common with bilaterian animals (e.g., neurons and muscle cells), ctenophores are defined by two novel traits: parallel rows of cilia organized into “combs,” and colloblasts, the adhesive cells used to capture prey (fig. 1). Found exclusively in the tentacles, colloblasts are typified by a crown of adhesive-filled secretory vesicles and an extensible basal apparatus (Eckhaut et al. 1997). Upon contact with prey, the apical membrane of the colloblast ruptures, releasing the adhesive (Franc 1978). Their association with tentacles and their specialized role in prey capture have led some to propose that

colloblasts are the functional analogs of the cnidarian cnidocyte (stinging cell; Alie et al. 2011; Borisenko and Ereskovsky 2013).

The tentacles of ctenophores are composed of a central axis of muscle and nerve fibers embedded in a gelatinous extracellular layer (the mesoglea) surrounded by a monolayer of epidermal cells. In many (but not all) species of ctenophore, the tentacles are adorned by numerous side branches (tentilla) and in some ctenophores (e.g., *Euplokamis*), these side branches are extensible and prehensile (Mackie et al. 1988). During feeding, the tentacles and tentilla (when present) are extended or uncoiled into the water column to ensnare passing prey (Mackie et al. 1988; Emson and Whitfield 1991). While colloblasts have been described as the predominant cell type of the tentacle/tentillum epidermis, several other cell types are known to populate these tissues (fig. 1G): covering cells, also known as cap cells or support cells, two types of sensory neurons (ciliated sensory cells and hoplocytes/peg cells), and two types of gland cells (mucus-secreting and granular gland cells; Horridge 1965; Emson and Whitfield 1991; Eckhaut et al. 1997; Borisenko and Ereskovsky 2013; Carre and Carre 1989).

The feeding behaviors of ctenophores are diverse but typically involve entangling prey in the extended tentacles or



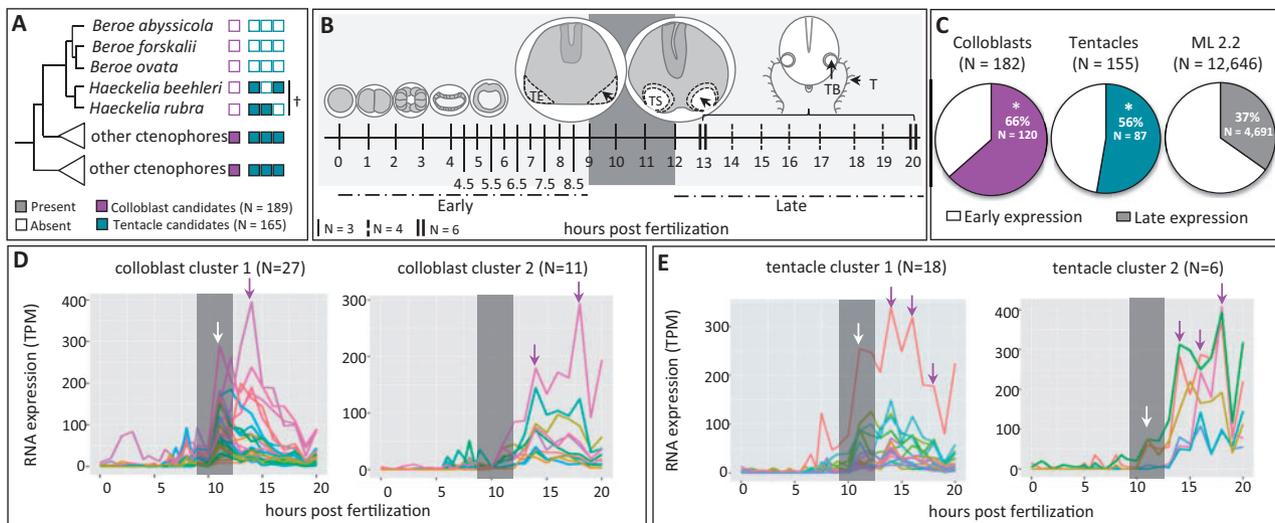
**FIG. 1.** Tentacle morphology varies across ctenophores. (A) *Mnemiopsis leidyi* has reduced tentacles as an adult, *Beroe forskalii* lacks tentacles completely, and *Haeckelia rubra* has long unbranched tentacles (white arrow). (B) The cydippid (larval) stage of *M. leidyi* has long branched tentacles (T). (C) DIC micrograph of a cydippid with tentacles retracted into the tentacle sheath (TS). (D) Higher magnification of the boxed area in C showing a branched tentacle (T) emerging from the tentacle sheath. (E) Emerging tentacle showing partially contracted side branches (tentilla, Tt). (F) Fully extended tentacle showing colloblast islets (C) along the tentacle. (G) Tentacle cell types—neurons (N) and smooth muscle cells (SM) are in the mesogleal core (M); colloblasts (C; yellow), covering cells (CC), ciliated sensory cells (CS), granular gland cells (GG), hoplocytes (HC), and mucous gland cells (MG) are in the epidermis (Ep). Images in panel A courtesy of Bruno Vellutini (*M. leidyi*) and Steve Haddock (*B. forskalii*, and *H. rubra*).

trapping prey with the oral lobes (Haddock 2007). One group of ctenophores (genus *Haeckelia*) has tentacles devoid of colloblasts; instead, their tentacles are populated by cnidocytes sequestered from their cnidarian prey (Carre and Carre 1980; Mills and Miller 1984). Lacking tentacles completely, ctenophores in the genus *Beroe* (the sister group to *Haeckelia*; Podar et al. 2001; Simion et al. 2015) engulf their prey (other ctenophores) with expanded lips and remove chunks of tissue using “teeth” made from modified cilia (Tamm 1983; Haddock 2007). Many species of lobate ctenophore (e.g., *Mnemiopsis leidyi*, *Bolinopsis infundibulum*) undergo ontogenetic change in their behavior, relying on the use of tentacles in the juvenile stage and oral lobes as adults. In these taxa, the adult tentacles are short and become

restricted to an oral fringe following metamorphosis. In contrast, the adult *Beroe* develops directly from an atentaculate larva. Thus, whereas the gene regulatory network underlying the development of tentacles may be downregulated in the adult stage of many lobate species, this network may not function at any stage in beroids.

Although they are a clear example of an evolutionary novelty, little is known about the origin of colloblasts. In this study, we leveraged the evolutionary history of ctenophores (including phylogeny, genes loss, and trait loss) to identify genes specific to this novel cell type. We hypothesized that some of the genes associated with colloblast development would have been lost during the diversification of *Beroe* and *Haeckelia* from their colloblast-bearing ancestor. Likewise, we





**Fig. 3.** Identifying colloblast and tentacle candidate genes. (A) Summarized 18S tree. Colloblast candidates ( $N = 189$ ) were present in  $\geq 70\%$  of the sampled taxa but absent from *Beroe* and *Haeckelia* (magenta boxes). Tentacle candidates ( $N = 165$ ) were present  $\geq 70\%$  of the sampled taxa, including at least one species of *Haeckelia* (dagger), but were absent from *Beroe* (teal boxes). (B) Developmental transcriptome sampling for *Mnemiopsis leidyi*. Early cleavage occurs during the first 8 h post fertilization (hpf). A thickened tentacle epithelium (TE) is visible at 9 hpf and invagination of the epithelium to form the tentacle sheath (TS) is complete by 12 hpf. Larval tentacles (T) grow continuously from the tentacle bulb (TB). (C) Candidate data sets are enriched for genes expressed during tentacle morphogenesis, relative to gene models from *M. leidyi* (ML2.2) ( $P < 0.0001$  for both). (D, E) The two largest clusters of gene expression profiles identified from candidate genes; colored lines represent different genes. Grey bars denote the 9–12 hpf window highlighted in panel B. White arrows indicate peaks in expression during tentacle bulb invagination, magenta arrows indicate peaks during tentacle morphogenesis. TPM—transcripts per million mapped reads.

exhibited higher expression during tentacle morphogenesis. Using a random sampling approach (see Materials and Methods), we found that both sets of candidate genes were significantly enriched for late-expressed genes ( $P < 0.0001$  for each).

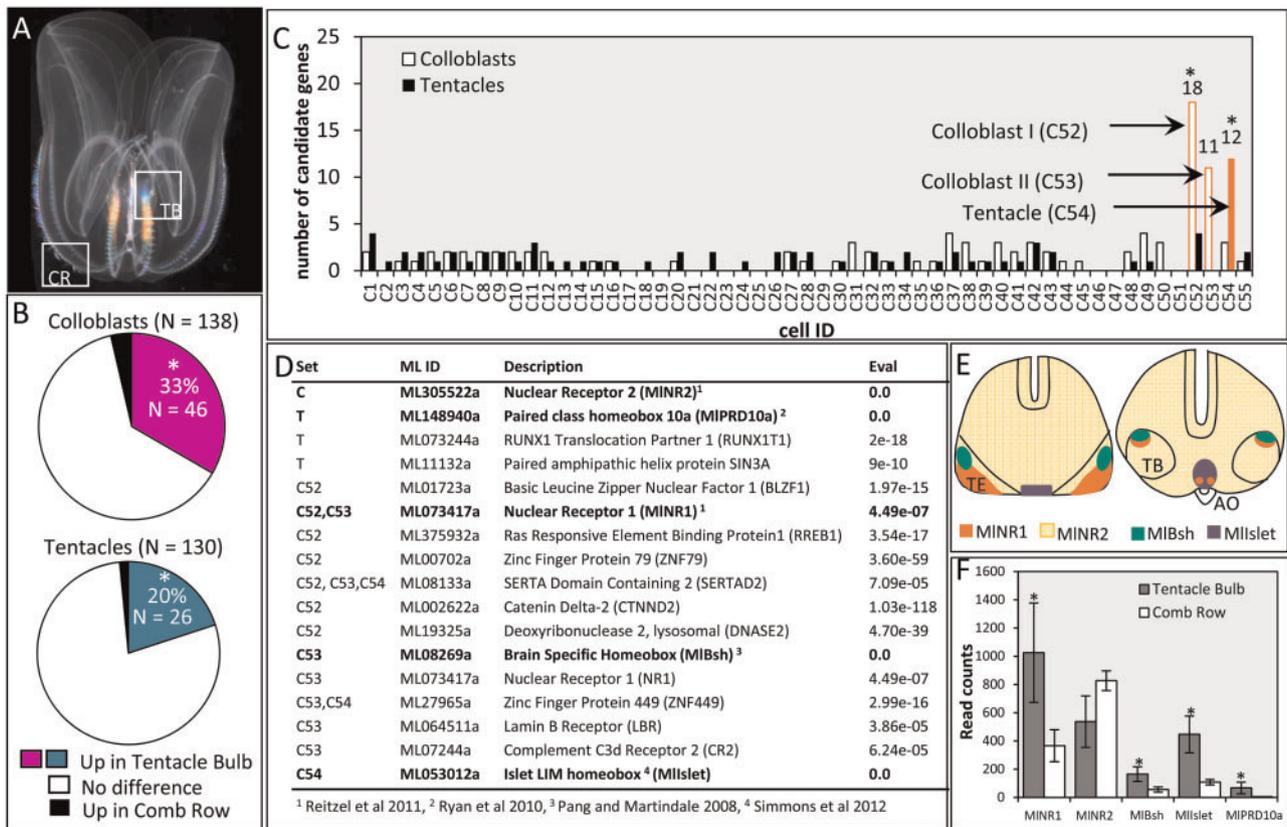
Next, we used quality threshold (QT) clustering (Heyer et al. 1999) to group candidate genes with similar expression patterns. Among colloblast candidates, the two largest clusters consisted of 27 and 11 genes (fig. 3D). The cluster containing 27 genes was characterized by a peak in expression at 11 hpf followed by a second peak at 14 hpf whereas genes in the cluster containing 11 genes first peaked at 14 hpf with a second peak at 18 hpf. The two largest clusters of tentacle candidates consisted of 18 and 6 genes (fig. 3E). Both clusters exhibited an early peak at 11 hpf followed by peaks at 14 hpf, 16 hpf, and 18 hpf. (Accession numbers for clustered genes are provided in supplementary file 1, Supplementary Material online.)

We further validated the colloblast and tentacle candidates by examining their expression in two adult tissues: tentacle bulbs and comb rows (fig. 4A). Over 70% of the candidate genes were also expressed in the adult tissues we sampled ( $N = 138/189$  colloblast candidates,  $N = 130/165$  tentacle candidates). Using differential expression analysis, we found that 33% of the colloblast candidates ( $N = 46/138$ ) and 20% of the tentacle candidates ( $N = 26/130$ ) were significantly upregulated in the tentacle bulb compared with the comb row (fig. 4B). Both sets of candidates were significantly enriched for tentacle bulb expression, compared with randomly selected data sets ( $P < 0.0001$  for both).

Using a published data set reporting differential expression of genes across individual cell types in *M. leidyi* (Sebe-Pedros,

Chomsky, et al. 2018), we found significant clustering of colloblast candidate genes ( $N = 18$ ) in a single cell (C52;  $P < 0.0001$ ) and another large cluster ( $N = 11$ ) of colloblast candidates in a second cell (C53; fig. 4C). We also found significant clustering of tentacle candidates ( $N = 12$ ) in a third cell (C54,  $P = 0.0015$ ). The remaining expressed candidate genes were distributed across the other cell types, none of which had a cluster of more than four candidate genes. Cells C52, C53, and C54 were undescribed by Sebe-Pedros, Chomsky, et al. (2018); however, based on the significant overrepresentation of candidate genes in these cells, we suggest that C52 and C53 are colloblasts and C54 is another tentacle-specific cell type.

To characterize these putative colloblast and tentacle cell types further, we first searched both sets of candidates and all genes expressed in cells C52, C53, and C54 for transcription factors that have been previously characterized in *M. leidyi* (Pang and Martindale 2008; Jackson et al. 2010; Pang et al. 2010; Yamada et al. 2010; Pang et al. 2011; Reitzel et al. 2011; Schnitzler et al. 2012; Simmons et al. 2012; Schnitzler et al. 2014; Reitzel et al. 2016). Additionally, we performed reciprocal BLAST of these data sets against the human proteome and annotated the results of both searches using Gene Ontology (GO). To identify transcription factors, we searched specifically for the following GO terms: GO: 0003677—DNA binding; GO: 0003700—DNA binding, transcription factor activity; GO: 0006351—transcription, DNA templated; and GO: 0006355—regulation of transcription, DNA templated. These combined approaches led to the discovery of seventeen transcription factors (fig. 4D), five of which have been previously studied in *M. leidyi*: Nuclear Receptor 2 (MINR2,



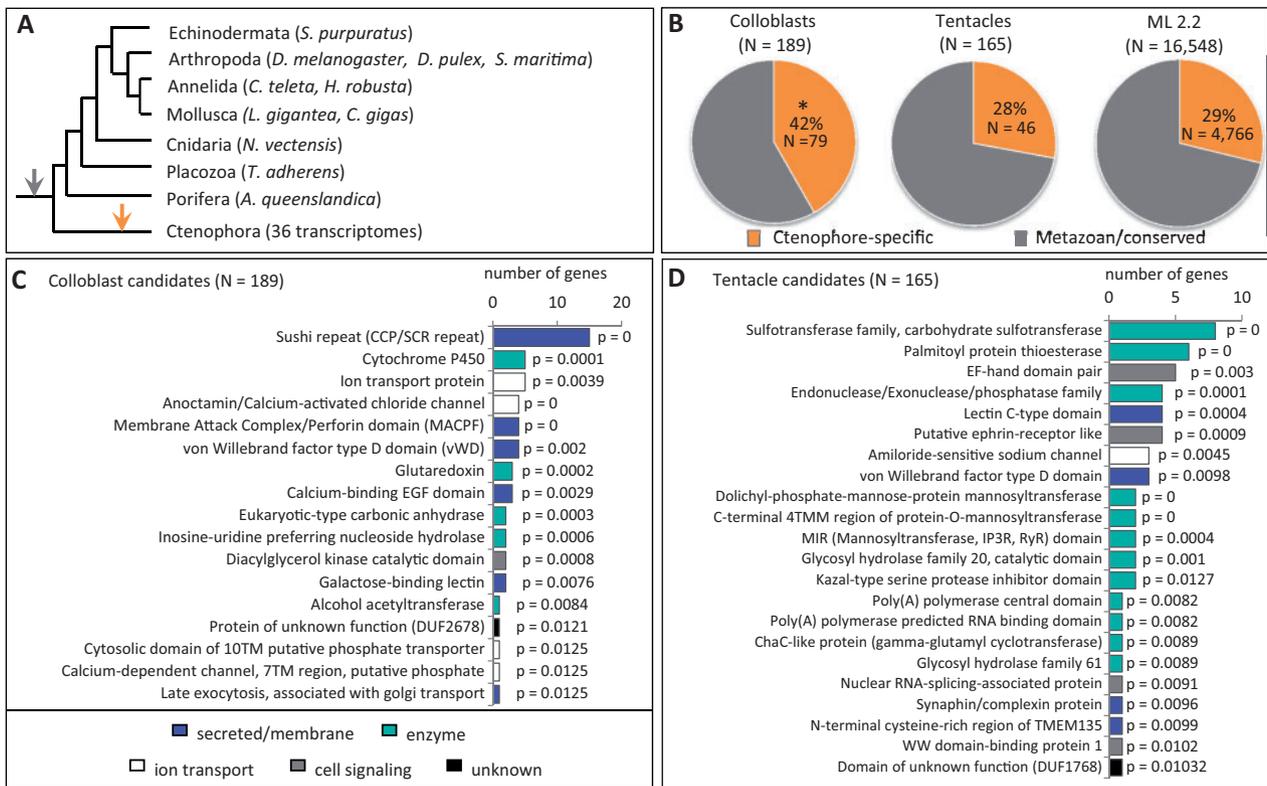
**Fig. 4.** Colloblast and tentacle candidates are expressed together in adult *Mnemiopsis leidyi*. (A) Tissue sampling protocol for tentacle bulb (TB) and comb row (CR) transcriptome sequencing. (B) Over 70% of the colloblast and tentacle candidate genes were expressed in adult tissues ( $N = 138/189$  colloblast candidates,  $N = 130/165$  tentacle candidates). In both data sets, a significant proportion of the expressed genes ( $P < 0.0001$  for both colloblast and tentacle candidates) were upregulated in the tentacle bulb, relative to the comb row ( $\geq 2 \log_2$ -fold change,  $\text{padj} < 0.05$ ). (C) There was a significant cluster of colloblast candidates in cell C52 ( $P < 0.0001$ ), with a second cluster in cell C53. Tentacle candidates clustered significantly in cell C54 ( $P = 0.0015$ ). Cell IDs refer to single-cell sequencing results reported in Sebe-Pedros, Chomsky, et al. (2018). (D) Seventeen transcription factors were identified from colloblast candidates (C), tentacle candidates (T), or cells C52, C53, or C54 using GO terms GO: 0003677, GO: 0003700, GO: 0006351, and GO: 0006355. E-values represent the reciprocal best BLAST hit in the human genome, except for genes already identified from *M. leidyi* (Eval = 0.0). Only significant hits ( $E \leq 1e-03$ ) are shown. Genes previously characterized by in situ hybridization are indicated in bold. (E) Summarized expression of *MINR1*, *MINR2*, *MIBsh*, and *MIslet* in the presumptive tentacle epithelium (TE), the invaginated tentacle bulb (TB), and near the apical organ (AO) during embryonic development. (F) Read counts for previously characterized transcription factors in tissues from adult *M. leidyi*. *MINR1*, *MIBsh*, *MIslet*, and *MIPRD10A* are significantly upregulated ( $\geq 2 \log_2$ -fold change) in the tentacle bulb (*MINR1*  $P = 0.004$ , *MIBsh*  $P = 0.010$ , *MIslet*  $P < 0.001$ , *MIPRD10A*  $P = 0.026$ ). *MINR2* is not differentially expressed ( $P = 0.893$ ).

ML305522a) from the colloblast candidate data set, Paired Class Homeobox 10a (*MIPRD10a*, ML148940a) from the tentacle candidate data set, Nuclear Receptor 1 (*MINR1*, ML073417a) from cells C52 and C53, Brain Specific Homeobox (*MIBsh*, ML08269a) from cell C53, and the LIM Homeobox gene *Islet* (*MIslet*, ML053012a) from cell C54 (Pang and Martindale 2008; Ryan et al. 2010; Reitzel et al. 2011; Simmons et al. 2012). Four of these (excluding *MIPRD10a*) have been previously characterized during embryonic development in *M. leidyi* using in situ hybridization. Whereas *MINR2* is expressed ubiquitously throughout development, *MINR1* is expressed in the tentacle bulb and apical organ, *MIBsh* is restricted to the tentacle bulb, and *MIslet* is restricted to the apical organ (fig. 4E). Using tissue-specific transcriptomes from adults, we confirmed that *MINR1*, *MIBsh*, *MIslet*, and *MIPRD10a* are all upregulated in the tentacle bulb, relative to the comb rows (fig. 4F). *MINR2* was expressed in both tentacle bulbs and comb rows but was not differentially expressed.

### Characterizing Candidate Genes

Consistent with other studies of metazoan novelties (Johnson and Tsutsui 2011; Babonis et al. 2016), we hypothesized that the set of colloblast candidates would be enriched for novel (ctenophore-specific) genes. To test this, we used a reciprocal BLAST strategy to search candidate genes against a database of animal genomes (fig. 5A); we considered genes that lacked significant hits outside of Ctenophora ( $E \geq 1e-02$ ) to be ctenophore-specific. Over 40% (79/189) of the colloblast candidates were ctenophore-specific, whereas only 28% (46/165) of the tentacle candidates and 29% (4,766/16,548) of all protein models (ML2.2) were ctenophore-specific (fig. 5B). Random sampling confirmed that colloblast candidates were significantly enriched for novel genes ( $P < 0.0001$ ) whereas tentacle candidates were not ( $P = 0.643$ ).

To evaluate their putative function, we annotated both sets of candidate genes against the InterPro Consortium database using Interproscan (Jones et al. 2014). We found that



**Fig. 5.** Ontology of candidate genes. (A) Cladogram of the taxa used to identify ctenophore-specific genes. Genes with significant BLAST hits ( $E \leq 1e-02$ ) to taxa other than Ctenophora are assumed to have been present in the common ancestor of all animals (grey arrow). Genes lacking significant hits to other metazoan taxa are assumed to have originated after ctenophores diverged from the rest of animals or to have been lost in the stem lineage giving rise to the rest of animals (orange arrow). (B) Ctenophore-specific genes (orange) are significantly enriched in the set of late-expressed colloblast candidates ( $P = 0.0007$ ), but comprise similar proportions of the tentacle candidate genes and complete set of protein models from *Mnemiopsis leidyi* (ML 2.2). (C, D) Overrepresented annotations from colloblast and tentacle candidates, relative to ML2.2 protein model annotations.

61% (116/189) of the colloblast candidates and 76% (125/165) of the tentacle candidates were annotated (supplementary file 1, Supplementary Material online). We further identified 17 GO terms that were significantly overrepresented in colloblast candidates (fig. 5C) and 22 GO terms significantly overrepresented in tentacle candidates (fig. 5D). Among colloblast candidates, overrepresented categories were largely associated with secretion/cell membrane recognition (e.g., sushi/SCR, MACPF, vWD, Ca-EGF, lectin, golgi transport) and enzymes involved in cellular metabolism (e.g., cytochrome P450, glutaredoxin, carbonic anhydrase, nucleoside hydrolase, acetyltransferase). Among tentacle candidates, the largest overrepresented category consisted of enzymes involved in posttranslational modification (i.e., sulfotransferase, thioesterase, phosphatase, mannosyltransferase, glycosyl hydrolase, cyclotransferase).

### Searching for Ctenophore Adhesive Proteins

We used BLAST to search candidate genes against a set of known adhesive proteins from other invertebrates (Hennebert et al. 2015). (Sequences provided in supplementary file 2, Supplementary Material online.) Five colloblast candidates and four tentacle candidates had significant hits to adhesive proteins ( $E \leq 1e-03$ ), yet each of these genes had

better hits to other proteins in the Uniprot database (www.uniprot.org; table 1). Next, we compared protein family (Pfam) domains from the known adhesives to domains identified from candidate genes using Interproscan. From the 48 confirmed adhesive proteins, we identified 17 Pfam domains. One domain was shared among all three data sets (von Willebrand factor type D domain, PF00094), one was shared by adhesives and colloblasts only (EGF-like calcium-binding domain, PF07645), and one was shared by adhesives and tentacles (Chitin binding domain, PF01607). Consistent with our BLAST results, neither colloblast- nor tentacle candidates exhibited significant overlap with Pfam domains from known adhesives (colloblasts  $P = 0.5008$ , tentacles  $P = 0.4459$ ).

Given that we did not find strong BLAST support for the homology of candidate genes and proteins from other biological adhesives, we searched instead for features known to be enriched among described adhesive proteins, including: secretion signal peptide, single-pass transmembrane domains, and regions of low sequence complexity (Waite et al. 2005; Endrizzi and Stewart 2009). Using SignalP (Petersen et al. 2011) and TMHMM (Krogh et al. 2001), we found that 28% (53/189) of the colloblast candidates encoded a signal peptide and 36% (68/189) encoded one or more transmembrane

**Table 1.** Putative Adhesive Genes from Colloblast (C) and Tentacle (T) Data Sets. Only Significant Hits ( $\leq 1E-03$ ) are Shown.

Set	ML Gene ID	Top hit—Adhesives	E-Value	Top Hit—Uniprot	E-Value
C	ML020113a <sup>a</sup>	ABR68008.1 matrilin-like 85 kDa protein ( <i>Ambigolimax valentianus</i> )	6.16E-11	O89103 C1QR1_MOUSE Complement component C1q receptor ( <i>Mus musculus</i> )	1.37e-20
C	ML056914a	ABR68008.1 matrilin-like 85 kDa protein ( <i>Ambigolimax valentianus</i> )	3.12 E-10	Q14246 AGRE1_HUMAN Adhesion G protein-coupled receptor E1 ( <i>Homo sapiens</i> )	5.55e-25
C	ML50011a	ABR68008.1 matrilin-like 85 kDa protein ( <i>Ambigolimax valentianus</i> )	2.35 E-08	O08999 LTBP2_MOUSE Latent-transforming growth factor beta-binding protein 2 ( <i>Mus musculus</i> )	7.16e-24
C	ML223525a <sup>a</sup>	AHN92641.1 sea star footprint protein 1 ( <i>Asterias rubens</i> )	3.04 E-06	Q9HC84 MUC5B_HUMAN Mucin-5B ( <i>Homo sapiens</i> )	1.91e-15
C	ML32223a <sup>a</sup>	AHN92641.1 sea star footprint protein 1 ( <i>Asterias rubens</i> )	1.19 E-06	Q9Y493 ZAN_HUMAN Zonadhesin ( <i>Homo sapiens</i> )	8.27e-16
T	ML14246a	AHN92641.1 sea star footprint protein 1 ( <i>Asterias rubens</i> )	1.15 E-12	Q02817 MUC2_HUMAN Mucin-2 ( <i>Homo sapiens</i> )	1.22e-17
T	ML14247a <sup>a</sup>	AHN92641.1 sea star footprint protein 1 ( <i>Asterias rubens</i> )	3.07 E-12	Q9HC84 MUC5B_HUMAN Mucin-5B ( <i>Homo sapiens</i> )	3.79e-17
T	ML154123a <sup>a</sup>	AHN92641.1 sea star footprint protein 1 ( <i>Asterias rubens</i> )	8.48 E-09	Q02817 MUC2_HUMAN Mucin-2 ( <i>Homo sapiens</i> )	3.51e-13
T	ML056959a	AFP57565.1 aggregate gland silk factor 1, partial ( <i>Latrodectus hesperus</i> )	1.34 E-10	A2VD00 EIF3A_XENLA Eukaryotic translation initiation factor 3 subunit A ( <i>Xenopus laevis</i> )	2.72e-22

NOTE.—Candidate genes were searched using BLAST against a database of empirically derived adhesive genes concatenated with the complete Uniprot database.

<sup>a</sup>Indicates presence of signal peptide. ML gene IDs refer to ML2.2 (<https://research.nhgri.nih.gov/mnemiopsis/>).

domains (fig. 6A and B). By comparison, 19% (31/165) of tentacle candidates encoded a signal peptide and 36% (59/165) encoded transmembrane domains. Both sets of candidate genes were significantly enriched for signal peptides and transmembrane domains, relative to random sets drawn from ML2.2 ( $P < 0.0001$ , for both). The number of transmembrane domains varied from 1 to 9 among colloblast candidates and from 1 to 13 among tentacle candidates (fig. 6C), although the number of transmembrane domains in these candidate gene sets did not differ significantly from samples drawn randomly from ML2.2 (colloblasts  $P = 0.5961$ , tentacles  $P = 0.4756$ ). Therefore, neither set of candidate genes was enriched for single-pass transmembrane domains. Finally, we assessed sequences in both data sets for regions of low complexity using the program Segmasker (Wootton 1994). Contrary to our expectations based on other biological adhesives, colloblast and tentacle candidates were not enriched for regions of low-complexity (colloblasts  $P = 0.427$ , tentacles  $P = 0.98$ ).

### Searching for Ctenophore Toxin Proteins

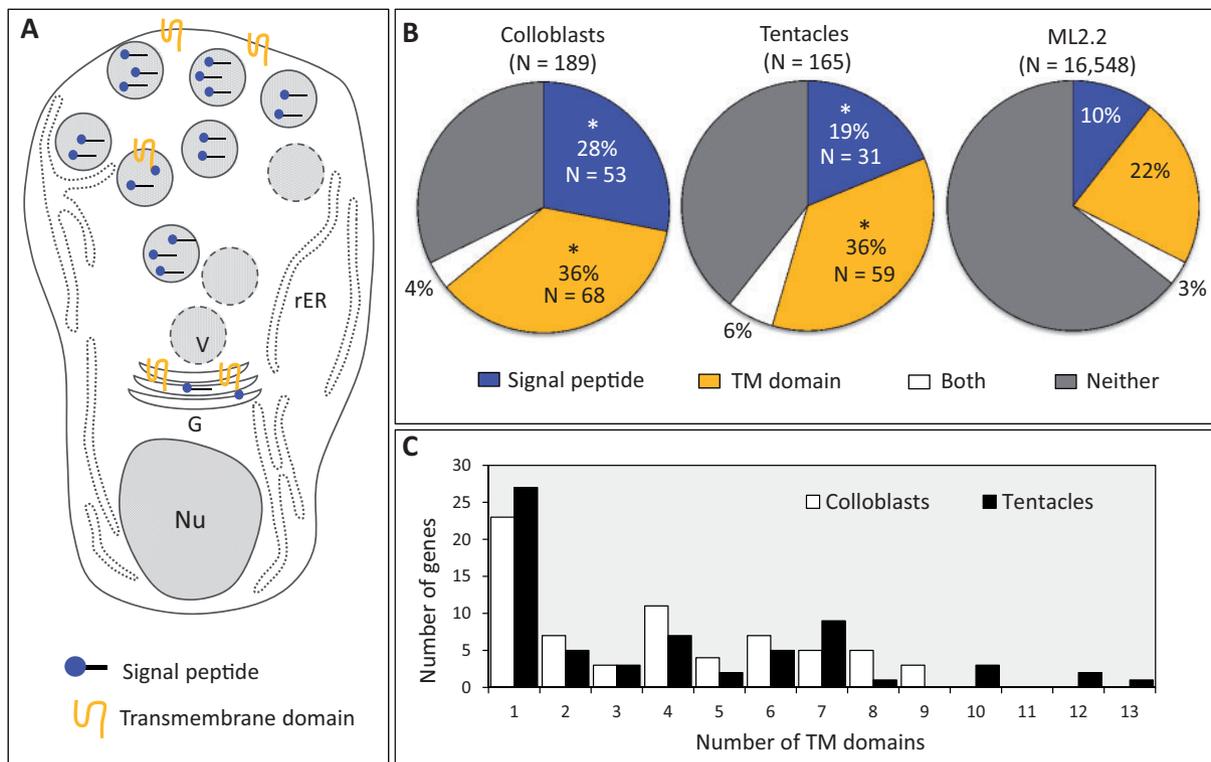
Moss et al. (2001) suggested the possibility that colloblasts or other secretory cells in ctenophores may secrete a toxin. To test this hypothesis, we used BLAST to search both sets of candidate genes against a database of known animal venoms/toxins, referred to hereafter as “ToxProt” (Jungo et al. 2012). From the colloblast candidates, we identified a single gene (ML263512a) with a significant match in the ToxProt database ( $E \leq 1e-03$ ); however, this gene had better hits to Uniprot proteins outside of the ToxProt database (table 2). Among tentacle candidates, we identified 12 sequences with significant hits in the ToxProt database, only one of which (ML435831a) had an equivalent/better hit to a protein in the ToxProt database than to any nontoxin proteins in the Uniprot database.

We then compared Pfam domains from the ToxProt database with domains from the colloblast and tentacle data sets. From the 6,665 genes in ToxProt, we identified 174 Pfam domains. Four domains were shared among all three data sets: Ankyrin repeat-containing domain (PF12796), Membrane attack complex component/perforin (MACPF) domain (PF01823), Thrombospondin type-1 (TSP1) repeat (PF00090), and Thyroglobulin type-1 (PF00086); three domains were shared between ToxProt and colloblasts only: Beta-propeller repeat TECPR (PF06462), DNA/RNA nonspecific endonuclease (PF01223), Immunoglobulin domain (PF13927); and five domains were shared between ToxProt and tentacles only: Ankyrin repeat (PF00023, PF13637), C-type lectin-like (PF00059), EF-hand domain (PF13499, PF13833), Kazal domain (PF00050, PF07648), and ShK domain (PF01549). Contrary to our expectations, neither set of candidates exhibited significant overlap with ToxProt Pfam domains (colloblasts  $P = 0.7987$ , tentacles  $P = 0.0873$ ).

Finally, we searched the suite of genes identified from cells C52, C53, and C54 (Sebe-Pedros, Chomsky, et al. 2018) against the ToxProt database to determine if there was additional support for the secretion of toxins in these putative tentacle cell types. We identified five cells with significant clusters of ToxProt genes: C17 ( $P = 0.016$ ), C21 ( $P = 0.036$ ), C25 ( $P = 0.037$ ), C47 ( $P = 0.029$ ), and C54 ( $P = 0.020$ ; supplementary fig. 1, Supplementary Material online). Cell C54 was identified by the authors only by the presence of a protein with an ShK domain, a domain originally identified from sea anemone toxins.

### A Common Origin for Colloblasts and Neurons

Although cell fate has been fairly well characterized in *M. leidyi* (Martindale and Henry 1997; Martindale and Henry 1999; Henry and Martindale 2001), previous studies of cell fate have been performed only up to the 60-cell (pregastrula)



**Fig. 6.** Identification of features associated with subcellular localization. (A) Diagram of an immature colloblast after Storch and Lehnert-Moritz (1974) showing developing secretory vesicles. (B) Both colloblast and tentacle candidates were significantly enriched for genes encoding signal peptides (blue) and transmembrane domains (yellow) ( $P < 0.0001$  for all). (C) Number of genes encoding single- and multi-pass transmembrane domains from colloblast and tentacle candidates.

stage; as such, little is known about the fate of cells differentiating at later stages of development. To characterize the developmental origin of the colloblasts in *M. leidy*, we randomly labeled single cells in the vicinity of the presumptive tentacle epithelium in late gastrula stage embryos with a fluorescent dye (DiI; fig. 7A–D) and allowed embryos to develop to the cydippid stage, as previously described. From 28 embryos with individually labeled cells, we recovered seven cydippids (25%) with labeled colloblasts on the side corresponding to the injected micromere. Surprisingly, all seven of these cydippids also exhibited DiI-labeled neurons, either in the floor of the apical organ (fig. 7E–G) or in the peripheral nerve net (fig. 7H), suggesting that colloblasts and neurons differentiate from a common progenitor that acquires its identity after gastrulation.

### No Common Origin for Colloblasts and Cnidocytes

We tested the hypothesis that colloblasts and cnidocytes share a common evolutionary origin by searching for orthologous genes in these two cell types. Using OrthoFinder, we generated orthology groups using protein models from *M. leidy* and the sea anemone *Nematostella vectensis*. From this analysis, we identified four groups containing at least one candidate gene (colloblast or tentacle) and at least one of the proteins identified as cnidocyte-specific in a recent study using single-cell sequencing from *N. vectensis* (Sebe-Pedros, Saudemont, et al. 2018; table 3).

Two groups included colloblast candidates and cnidocyte genes; the first group contains orthologs of fibrillin, a glycoprotein component of the extracellular matrix, and the second contains orthologs of retinoic acid receptors (RxRs). Upon closer inspection, the colloblast candidate in this latter group turned out to be the previously studied nuclear receptor *MINR2* (Reitzel et al. 2011). We used BLAST to search *MINR2* against *B. ovata* and confirmed that this gene is missing from both the transcriptome and the genome of *B. ovata*. The other two groups contained tentacle candidates and cnidocyte genes. The first of these contained orthologs of protein-O-mannosyl transferase 2, an important regulator of protein glycosylation. Genes in the second group share homology with DELTA-alicotoxin, a pore-forming toxin from sea anemones. Compared with data sets sampled randomly from ML2.2, orthology groups containing cnidocyte genes were not significantly enriched for colloblast ( $P = 0.8519$ ) or tentacle ( $P = 0.7869$ ) candidates.

### Discussion

First described nearly 200 years ago (Eschscholtz 1829), ctenophores remain a poorly understood group of animals. By combining phylogeny, natural variation in morphology, analyses of embryonic and adult gene expression, and detailed sequence annotations, we have identified and characterized genes associated with tentacle-specific cell types. While we recognize the possibility that our data sets may include genes

**Table 2.** Putative toxin Genes from Colloblast (C) and Tentacle (T) Data Sets. Only Significant Hits ( $E \leq 1e-03$ ) are Shown.

Set	ML Gene ID	Top Hit—ToxProt	E-Value	Top Hit—Uniprot	E-Value
C	ML263512a	Q25338 LITD_LATTR Delta-latroinsectotoxin-Lt1a <i>Latrodectus tredecimguttatus</i> (Mediterranean black widow spider)	5.37E-07	G5E8K5 ANK3_MOUSE Ankyrin-3 <i>Mus musculus</i> (mouse)	3.32e-12
T	ML01511a	Q3SB11 CALGL_TROCA Calglandulin <i>Tropidechis carinatus</i> (Australian rough-scaled snake)	3.68E-12	P02595 CALM_PATSP Calmodulin <i>Patinopecten</i> sp. (scallop)	1.19e-24
T	ML01571a	Q3SB11 CALGL_TROCA Calglandulin <i>Tropidechis carinatus</i> (Australian rough-scaled snake)	3.52E-27	P24044 CALM_PLAFA Calmodulin <i>Plasmodium falciparum</i>	7.35e-39
T	ML01786a	Q66S03 LECG_THANI Galactose-specific lectin natterctin <i>Thalassophryne nattereri</i> (Copper Joe toadfish)	5.93E-07	P82596 PLC_HALLA Perluin <i>Haliotis laevigata</i> (Smooth Australian abalone)	2.17e-09
T	ML056959a	Q66S03 LECG_THANI Galactose-specific lectin natterctin <i>Thalassophryne nattereri</i> (Copper Joe toadfish)	6.35E-05	A2VD00 EIF3A_XENLA Eukaryotic translation initiation factor 3 subunit A <i>Xenopus laevis</i> (African clawed frog)	2.73e-22
T	ML056965a	A3FM55 LECM1_HYDHA C-type lectin 1 <i>Hydrophis hardwickii</i> (Hardwick's spine-bellied seasnake)	2.97E-05	P21328 RTJK_DROME RNA-directed DNA polymerase from mobile element jockey <i>Drosophila melanogaster</i> (fruit fly)	1.98e-16
T	ML070216a	Q3SB11 CALGL_TROCA Calglandulin <i>Tropidechis carinatus</i> (Australian rough-scaled snake)	2.87E-13	P25071 CML12_ARATH Calmodulin-like protein 12 <i>Arabidopsis thaliana</i> (Mouse-ear cress)	5.02e-30
T	ML075218a	Q02989 LITA_LATTR Alpha-latroinsectotoxin-Lt1a <i>Latrodectus tredecimguttatus</i> (Mediterranean black widow spider)	5.69E-06	P16157 ANK1_HUMAN Ankyrin-1 <i>Homo sapiens</i> (human)	3.50e-10
T	ML14246a	P0DKM9 TU11_LOPAL Turriptide OL11-like <i>Lophiotoma albina</i> (Sea snail)	2.72E-04	Q02817 MUC2_HUMAN Mucin-2 <i>Homo sapiens</i> (human)	1.22e-17
T	ML273210a	Q8AY75 CALGL_BOTIN Calglandulin <i>Bothrops insularis</i> (Golden lancehead snake)	5.69E-25	P27164 CALM3_PETHY Calmodulin-related protein <i>Petunia hybrida</i> (Petunia)	3.94e-56
T	ML35385a	G0LXV8 LATA_LATHA Alpha-latrotoxin-Lh1a <i>Latrodectus hasseltii</i> (Redback spider)	1.14E-06	Q8Q0U0 Y045_METMA Putative ankyrin repeat protein MM_0045 <i>Methanosarcina mazei</i> (anaerobic archaeobacter)	4.38e-14
T	ML435831a <sup>a,b</sup>	P58912 TX60B_PHYSE DELTA-alicitoxin-Pse2b <i>Phyllodiscus semoni</i> (Night anemone)	1.20E-20	P58912 TX60B_PHYSE DELTA-alicitoxin-Pse2b <i>Phyllodiscus semoni</i> (Night anemone)	1.20e-20
T	ML45397a	Q9XZC0 LCTA_LATTR Alpha-latrocrustotoxin-Lt1a <i>Latrodectus tredecimguttatus</i> (Mediterranean black widow spider)	1.37E-10	Q5ZLC8 ANR52_CHICK Serine/threonine-protein phosphatase 6 regulatory ankyrin repeat subunit C <i>Gallus gallus</i> (chicken)	1.90e-14

NOTE.—Candidate genes were searched using BLAST against a database of known animal venom/toxin genes (ToxProt) concatenated with the complete Uniprot database.

<sup>a</sup>Indicates presence of signal peptide.

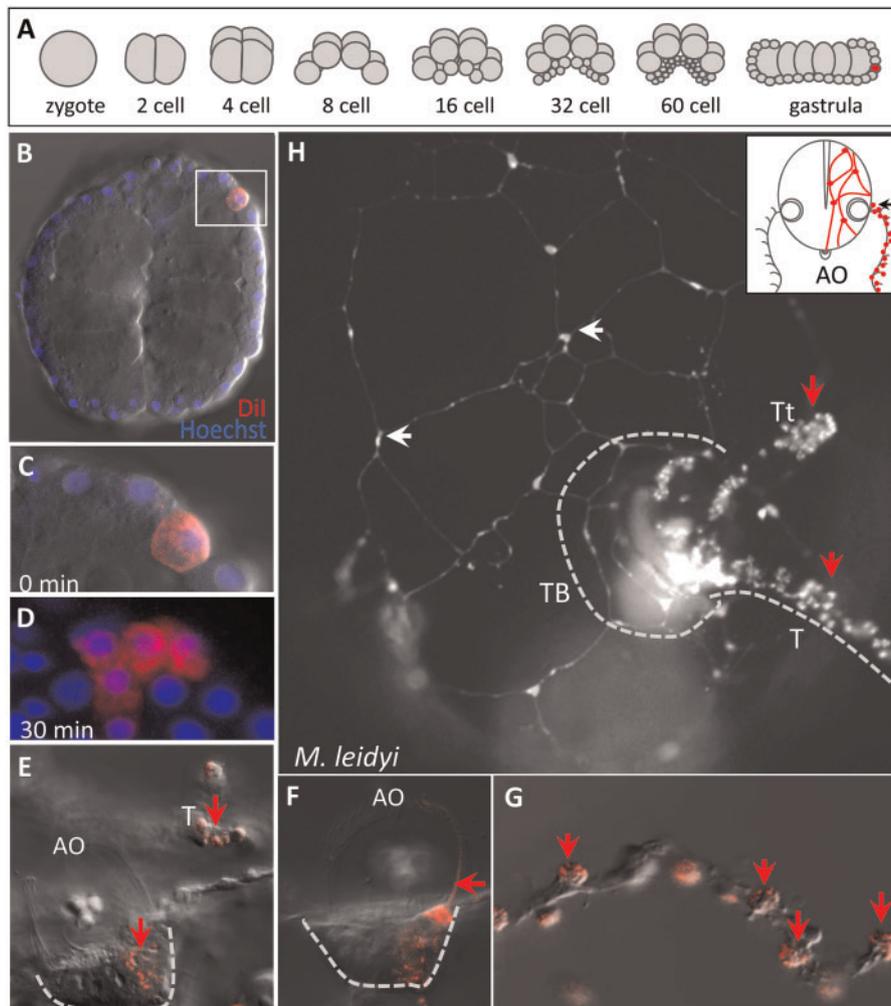
<sup>b</sup>Indicates an equivalent or better hit in the ToxProt database. ML gene IDs refer to ML2.2 (<https://research.nhgri.nih.gov/mnemiopsis/>).

not associated with these traits, the overrepresentation of late-expressed genes (fig. 3C) with numerous peaks in expression (fig. 3D and E), significant upregulation in the adult tentacle bulb (fig. 4B), and significant clustering in specific isolated cell types (fig. 4C), suggests this approach was effective for identifying genes associated with tentacle cell identity.

As part of this work, we have likely uncovered novel components of an undescribed biological adhesive. Consistent with other adhesives, colloblast candidate genes were enriched for domains associated with secretion, membrane recognition, and subcellular protein trafficking (fig. 5, supplementary file 1, Supplementary Material online). Furthermore, colloblast candidates were enriched for a hydrophobic N-terminal signal peptide (fig. 6). Signal peptides are important for directing proteins to the vesicles in numerous secretory cell types including cnidocytes (Anderluh et al. 2000), cells from venom glands (Jones et al. 1992), and adhesive-secreting cells from other animal groups (Hennebert et al. 2015). Thus, the genes we identified as colloblast candidates are consistent with the genes expected to be expressed in a cell undergoing synthesis, packaging, and storage of secreted proteins.

Surprisingly, we found no BLAST support for the homology of colloblast candidates with other biological adhesive proteins (table 1) and, unlike other biological adhesives, colloblast candidates were not enriched for regions of low-complexity. Combined with the overrepresentation of ctenophore-specific genes among colloblast candidates (fig. 5), our results suggest that the origin of the colloblast adhesive was largely independent from the evolution of adhesives in other biological systems. Unlike other animal adhesives (e.g., sea star foot protein, mussel byssal threads), the colloblast adhesive must be fast-acting (“instantaneous”) but need not be permanent (Flammang et al. 2009) and these constraints may have facilitated the origin of an adhesive with unique properties in the stem ctenophore. Indeed, we suggest that rapid evolution of existing genes (Martin-Duran et al. 2017), resulting in *de novo* acquisitions of novel peptide motifs may have promoted the origin of the colloblast adhesive.

We further leveraged the secondary loss of tentacles in the genus *Beroe* to identify compelling candidate genes for future studies of other tentacle specific cell types in ctenophores. Tentacle candidates were enriched for signal peptides as well



**Fig. 7.** Neurons and colloblasts share a common progenitor. (A) Embryonic cleavage stages in *Mnemiopsis leidyi*. Random micromeres (red) were injected at the late gastrula stage. (B) Optical section of a live gastrula immediately following injection of Dil (red); nuclei are labeled with Hoechst. (C) High magnification image of the boxed area in B showing a single injected micromere. (D) High magnification image of a clone of labeled cells 30 min after Dil injection. (E–G) A live cydippid larva with two labeled populations of cells: in the floor plate of the apical organ (AO) and the colloblasts of the tentacle (T). (E) Both cell populations are depicted in a single focal plane. (F) A different focal plane showing labeling of cells and dome cilia (arrow) on one side of the apical organ. (G) High magnification image of the extended tentacle; colloblasts (arrows) are the only labeled cells in the tentacle. (H) Cydippid larva with Dil-labeled neurons of the subepidermal nerve net in the body wall (white arrows) and colloblasts (red arrows); the animal is viewed from the tentacular plane (black arrow) as summarized in the inset. TB—tentacle bulb, T—tentacle, Tt—tentilla.

as enzymes involved in posttranslational protein modification (figs. 5 and 6, [supplementary file 1, Supplementary Material](#) online). One intriguing interpretation is that these enzyme-rich tentacle secretory cells are some type of gland cell engaged in the production and secretion of a ctenophore toxin. In support of this, we identified one gene from the tentacle candidates (ML435831a) that encodes both a signal peptide and a MACPF domain, and appears to be an ortholog of actinoporin, a pore-forming DELTA-alcitoxin found in sea anemone cnidocytes (table 3; Oshiro et al. 2004; Rachamim et al. 2015). Further supporting the potential role of this tentacle cell type in producing a toxin, we demonstrate significant clustering of tentacle candidates (fig. 4F) in a single cell (C54) that also expresses the largest number of genes with significant hits in the ToxProt database ([supplementary fig. 1, Supplementary Material](#) online). While empirical

observations are essential for evaluating the function of this cell type, these results suggest that ctenophores may incapacitate their prey by secretion of pore-forming toxins from a tentacle specific gland cell. A toxin-secreting cell may have provided many ecological benefits, even among taxa lacking colloblasts, which could explain why this cell type may have been retained in *Haeckelia*.

Notably, both data sets (colloblast and tentacle candidate genes) were largely devoid of transcription factors. Essential for activating and/or repressing the expression of effector genes (e.g., secreted or structural products), transcription factors are known to be highly pleiotropic, regulating gene expression in numerous regulatory networks. Sox genes, for example, are likely involved in tentacle morphogenesis based on their expression in the tentacle bulb of both *M. leidyi* and *Pleurobrachia bachei* (Jager et al. 2008; Schnitzler et al. 2014);

**Table 3.** Orthology Groups Containing Colloblast (C) or Tentacle (T) Candidates from *M. leidy* and Cnidocyte Genes from *N. vectensis*.

Set	ML Gene ID	NV Gene ID	Others in Group	Description
C	ML020113a, ML056914a, ML50011a	NVJ_2203	ML282520a, NVJ_108241, NVJ_113453, NVJ_117150, NVJ_117297, NVJ_119340, NVJ_123710, NVJ_129169, NVJ_137797, NVJ_142234, NVJ_146869, NVJ_154796, NVJ_157742, NVJ_16432, NVJ_198567, NVJ_202189, NVJ_208146, NVJ_209642, NVJ_210066, NVJ_223762, NVJ_224641, NVJ_22881, NVJ_2483, NVJ_3250, NVJ_32913, NVJ_37776, NVJ_46752, NVJ_48353, NVJ_61301, NVJ_67572, NVJ_6789, NVJ_70073, NVJ_79239, NVJ_79524, NVJ_80132, NVJ_80370, NVJ_83827, NVJ_84687, NVJ_87211, NVJ_87454, NVJ_89626, NVJ_9760, NVJ_99210	Fibrillin; Latent-transforming growth factor beta-binding protein 4
C	ML305522a	NVJ_165424	NVJ_101676, NVJ_108851, NVJ_114090, NVJ_132075, NVJ_134436, NVJ_167880, NVJ_169225, NVJ_183874, NVJ_189134, NVJ_203423, NVJ_209681, NVJ_242271, NVJ_89471, NVJ_93844, NVJ_94673, NVJ_99425	Nuclear receptor; Retinoic acid receptor RXR-gamma-B
T	ML00965a	NVJ_99284, NVJ_175881		Protein O-mannosyltransferase 2
T	ML10468a, ML435831a	NVJ_200058	ML41821a, ML020060a, NVJ_109596, NVJ_1099, NVJ_1115, NVJ_1173, NVJ_166322, NVJ_196985, NVJ_205444, NVJ_211816	DELTA-alicitoxin/sea anemone venom protein

NOTE.—Cnidocyte genes were extracted from Sebe-Pedros, Saudemont, et al. 2018. ML gene IDs refer to ML2.2 (<https://research.nhgri.nih.gov/mnemiopsis/>). NV gene ID refers to the *N. vectensis* genome (<https://genome.jgi.doe.gov/Nemve1/Nemve1.home.html>, last accessed September 10, 2018).

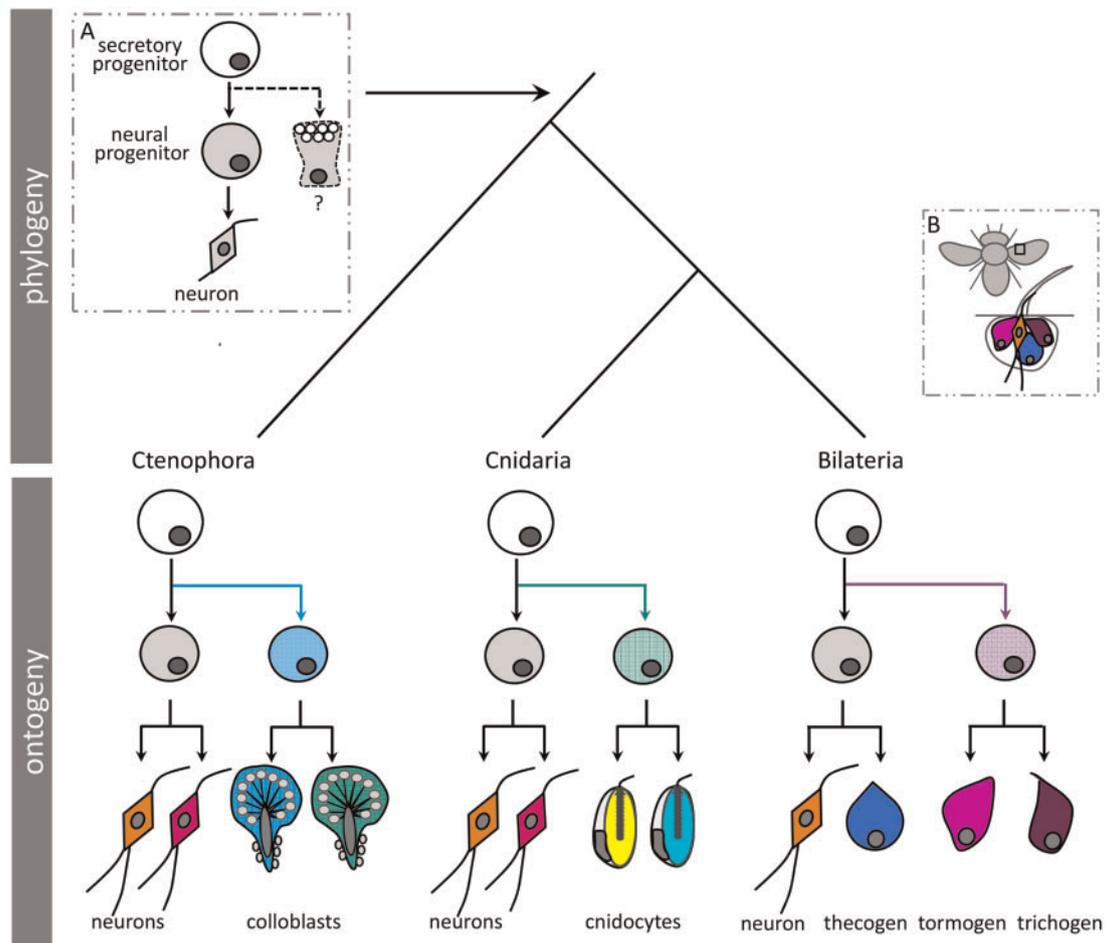
however, the fact that each *Sox* gene is expressed in additional domains outside of the tentacle bulbs in both species suggests these genes play many roles in the development of ctenophores. Consistent with this, *Sox* genes were not identified among the colloblast or tentacle candidates and the transcriptome of *B. ovata* encodes complete orthologs of all six ctenophore *Sox* genes (supplementary file 3, Supplementary Material online).

Annotation of both the candidate gene data sets and the putative colloblast (C52, C53) and tentacle (C54) cell types published previously (Sebe-Pedros, Chomsky, et al. 2018) enabled us to identify seventeen putative transcription factors that may play a role in patterning tentacle-specific cell types in *M. leidy* (fig. 4D–F). Possible colloblast transcription factors (*MINR1*, *MINR2*, and *MIBsh*) are all known to be expressed in the tentacle bulb during tentacle morphogenesis (Pang and Martindale 2008; Reitzel et al. 2011) and we demonstrate significant upregulation of *MINR1* and *MIBsh* in the adult tentacle bulb as well. The role of the putative toxin cell transcription factors (*Mllslet* and *MIPRD10a*) is not as clear. While both are upregulated in the adult tentacle bulb, *Mllslet* does not appear to be expressed in the tentacle primordia during embryonic development (Simmons et al. 2012) and the spatial expression of *MIPRD10a* has not been characterized (Ryan et al. 2010). Intriguingly, the genome of *B. ovata* encodes clear orthologs of *MINR1*, *MIBsh*, and *Mllslet*, but lacks orthologs of *MINR2* and *MIPRD10a*. Given that *MINR2* and *MIPRD10a* were identified as candidate genes from our phylogenetic analysis, we propose that knockdown of these genes in *M. leidy* should result in loss of colloblasts and other tentacle-specific secretory cells.

Surprisingly, our data suggest a common embryological origin for colloblasts and neurons, as both cell lineages appear to be the descendants of a single micromere labeled in the late gastrula stage in *M. leidy* (fig. 7). Previous cell lineage studies performed at earlier stages of development found

that neural and epidermal cells arose from a common precursor (Martindale and Henry 1997; Martindale and Henry 1999). Our results extend these observations, showing that epidermal cells differentiate from this common lineage before the separation of the neuronal and colloblast identities, as only the latter two cell types arose from micromeres labeled at later stages of development. This confirms a closer embryological relationship of the latter two cell types. Assuming that neurons are homologous across ctenophores (Hernandez-Nicaise 1973), these results imply that the loss of colloblasts resulted from disruption of the colloblast-specific branch of this lineage, independent of the segregation of neurons. Considering that ctenophores in the genus *Haeckelia* have tentacle bulbs and tentacles but lack colloblasts, we further suggest that the loss of colloblasts was independent of the development of the tentacle bulb. Additional studies of cell fate during embryogenesis in ctenophores with and without tentacle bulbs would shed much needed light on the evolution of morphological diversity in this group.

The shared embryological origin of colloblasts and neurons underscores one striking commonality between colloblasts and cnidocytes, as both cell types differentiate from a progenitor cell that also gives rise to neurons (Richards and Rentzsch 2014; Flici et al. 2017). Importantly, however, we found no additional evidence of a shared origin for these two cell types. Indeed, we found that colloblasts and cnidocytes express largely unique suites of genes as only four orthology groups were identified from among the hundreds of colloblast and cnidocyte candidates (table 3). Thus, rather than inferring the origin of some ancestral colloblast/cnidocyte prototype, we suggest that these novel secretory cell types arose independently in each lineage by co-option of a progenitor cell that already had the capacity for regulated cell secretion (fig. 8). Assuming nervous systems are homologous across animals (Jekely et al. 2015; Ryan and Chiodin 2015), it is likely that this progenitor cell already



**Fig. 8.** A possible scenario for the origin of secretory cell diversity in animals. (A) The common ancestor to ctenophores, cnidarians, and bilaterians may have already had a secretory cell progenitor (white) giving rise to neurons and some other unspecialized secretory cell. This pathway may have been independently co-opted (colored lines) to give rise to novel secretory cell types in ctenophores, cnidarians, and bilaterians. (B) The common origin of the neural and secretory cells responsible for making the sensory apparatus in flies supports this hypothesis in bilaterians. Sponges and placozoans have been excluded for simplicity.

gave rise to neurons and possibly other secretory cell types in the ancestor to the lineage encompassing ctenophores, cnidarians, and bilaterians (inset A, [fig. 8](#)). Studies characterizing the development of the epidermal sensory organs (sensilla) in flies support this explanation for the origin of novel secretory cells in bilaterians as well, since both the neural and secretory cells (thecogen, tormogen, and trichogen cells) underlying the sensilla also differentiate from a common progenitor (inset B; [Hartenstein and Posakony 1989](#)).

The relationship of specialized animal secretory cells to neurons suggests that there may be some underlying property of “neural” progenitor cells that makes them more likely to give rise to novel cell types. Because of their critical role in cell–cell communication, neurons have a phenotype that enables the packaging, storage, and delayed secretion of their products. It is possible that this pathway is easy to co-opt for other secretory functions, which could explain why multiple independent lineages of novel cell types seem to have evolved from a progenitor giving rise to neurons. Alternatively, cells that secrete a novel product may simply be easy to positively identify as novel cell types, artificially inflating the relationship

of neurons to novelty. Considering *Sox* genes are expressed in the common progenitor of neurons and cnidocytes in cnidarians ([Richards and Rentzsch 2014](#)) and in the tentacle bulb of ctenophores ([Jager et al. 2008](#); [Schnitzler et al. 2014](#)), we suggest that *Sox* genes may be good candidates for conferring general secretory cell identity across metazoans. Understanding the origin of other types of secretory cells (e.g., gland cells) in ctenophores and cnidarians and characterizing their developmental relationship to colloblasts/cnidocytes and neurons will be important for further assessing the ubiquity of this relationship between *Sox* gene expression and secretory cell phenotype.

The candidate genes described here now form the basis of future investigations into the origin, differentiation, and development of colloblasts and other tentacle-specific cell types in ctenophores. Future studies aimed at constructing the regulatory networks underlying ctenophore secretory cells (including neurons, colloblasts, and gland cells) will provide a unique opportunity to simultaneously characterize the poorly understood nervous system of ctenophores and probe the process by which novel secretory cells evolve. Cells with novel functions can be important for facilitating expansion

into new ecological niches, ultimately promoting speciation and diversification. Over evolutionary time, *Beroë* and *Haeckelia* have transitioned to prey types (other ctenophores and cnidarians, respectively) that are atypical for ctenophores, suggesting trophic specialization and evolutionary loss of cell types may have facilitated diversification in Ctenophora.

## Materials and Methods

### Animal Collection, Tissue Processing, and Transcriptome Assembly

Most specimens were collected during blue-water dives or using remotely operated-underwater vehicles from a region of the Eastern Central Pacific near the Monterey Bay Aquarium Research Institute (Moss Landing, CA), as described previously (Francis et al. 2015). These samples were snap frozen in liquid nitrogen and sequenced using a paired-end sequencing protocol at the University of Utah on an Illumina HiSeq 2000 platform with 100 amplification cycles. Briefly, read order was randomized and low-quality reads, adapters, and repeats were removed. For efficiency, subsets of reads were used to assemble transcriptomes. Assembly was performed with both Velvet/Oases v1.2.09/0.2.08 (Zerbino and Birney 2008; Schulz et al. 2012) and Trinity r2012-10-05 (Grabherr et al. 2011). Transcripts from both assemblers were combined and redundant sequences were removed using the sequniq utility in the GenomeTools package (Gremme et al. 2013).

For the developmental transcriptome series, adult *M. leidy* were collected from the estuary behind the University of Florida's Whitney Laboratory for Marine Bioscience (St. Augustine, FL) and maintained in the dark for 8 h to induce spawning. Zygotes were collected before first cleavage (time 0) and embryos were collected every 30–60 min for the first 20 h of development. Embryos were collected individually ( $N = 3–6$  embryos per time) and snap frozen on dry ice. Samples were prepared and sequenced on an Illumina HiSeq 2500, as described previously (Levin et al. 2016). *B. ovata* was collected from a public boat ramp on the Intracoastal Waterway in Port Orange, FL. Adults were spawned in the lab following the protocol for *M. leidy* and embryos were collected individually ( $N = 4$  per collection time) at 0, 6, 10, and 20 h post fertilization. RNA was extracted from all 16 embryos and from 4 adults and sent to the Genomic Sequencing and Analysis Facility at the University of Texas, Austin, for library preparation and sequencing on an Illumina HiSeq 2500.

For validation of putative colloblast and tentacle genes, we assembled tissue-specific transcriptomes from adult *M. leidy* collected from the estuary behind the Whitney Lab. Tissues (tentacle bulbs and comb rows) were freshly isolated from wild caught animals and snap frozen on dry ice. RNA extraction, library preparation, and sequencing were performed by the Interdisciplinary Center for Biotechnology Research at the University of Florida. Three independent replicates of each tissue were sequenced on a single lane of a HiSeq 3000 using a paired-end protocol. Differential expression analysis was performed using DESeq2 v1.20.0 (Love et al. 2014) in R v3.5.0 (R Core Development Team, 2008). Transcripts with  $\geq 2 \log_2$ -

fold change and an adjusted  $P$ -value  $\leq 0.05$  were considered differentially expressed. Raw sequence data have been deposited in the European Nucleotide Archive (accession PRJEB28334).

### Phylogenetics/18S Tree

We aligned 18S sequences from 36 ctenophores using MAFFT with default parameters (Nakamura et al. 2018). Trees were generated using three approaches: IQ-TREE v1.5.5 (Nguyen et al. 2015), RAxML (Stamatakis 2014) with 10 maximum parsimony starting trees, and RAxML with 10 random starting trees. A likelihood value for each tree was generated using RAxML; the tree produced by IQ-TREE had the highest likelihood value. We used AfterPhylo v0.9.1 (<https://github.com/qiyunzhu/AfterPhylo>; last accessed September 10, 2018) to create a tree that collapsed all branches with  $< 50\%$  bootstrap support.

### Identification and Annotation of Candidate Genes

To identify genes that had been lost in the lineage of ctenophores lacking colloblasts (*Beroë* + *Haeckelia*), we first created orthologous gene groups for the complete transcriptomes of all 36 species of ctenophore using OrthoFinder v1.1.8 (Emms and Kelly 2015). Colloblast candidates were genes present in orthology groups containing  $\geq 70\%$  of the taxa (including *M. leidy*) but were missing from all three species of *Beroë* (*B. ovata*, *B. forskalii*, and *B. abyssicola*) and from both species of *Haeckelia* (*H. rubra* and *H. beehleri*). Requiring these genes to be present in at least 70% of the transcriptomes (rather than 100%) allowed us to account for stochasticity in gene expression (i.e., genes not expressed at the time the animal was collected) and for gene losses that did not affect the maintenance of colloblasts. The number of candidate genes we recover varies considerably when we allow this cutoff to range from 50% to 100%, but there was no clear choice for the single best proportion to use (supplementary fig. 2, Supplementary Material online). We arbitrarily chose 70% but FASTA files of candidate genes recovered for all other cutoffs are provided in the GitHub Repository for this publication: [https://github.com/josephryan/2018-Babonis\\_et\\_al\\_Ryan](https://github.com/josephryan/2018-Babonis_et_al_Ryan). Tentacle candidates were present in orthology groups containing  $\geq 70\%$  of the taxa (including *M. leidy* and at least one species of *Haeckelia*) but lacking any species of *Beroë*. Three character states were possible for genes expressed in *Haeckelia* (see dagger, fig. 3A): present in both species, present in *H. rubra* only, or present in *H. beehleri* only. In each case, *Haeckelia* was counted toward the 70% total required to constitute ubiquitous expression across ctenophores.

### Temporal Expression of Candidate Genes in *M. leidy* Embryos

We examined the expression of candidate genes during embryonic development in *M. leidy* using stage-specific RNA-Seq data (NCBI GEO accessions GSE60478 and GSE111748). First, we removed colloblast and tentacle candidate genes with no expression ( $N = 7/189$  colloblast candidates and  $N = 10/165$  tentacle candidates with TPM = 0 at all time points). We also removed sequences with no expression

from the ML2.2 (<https://research.nhgri.nih.gov/mnemiopsis/>) protein models ( $N = 3,902/16,548$  genes with TPM = 0 at all time points). We then compared ratios of late gene expression (after the onset of tentacle morphogenesis; 12–20 hpf) to early gene expression (0–9 hpf). Late-expressed genes were those with a ratio  $>1$ . We searched the set of gene models (ML2.2) using the same approach. We used QT clustering (Heyer et al. 1999) to cluster candidate genes with similar expression patterns.

### Cell Specific Expression of Candidate Genes

We examined the distribution of candidate genes across individual cells from adult *M. leidy* isolated for single cell sequencing by Sebe-Pedros, Chomsky, et al. (2018). Significant clustering of candidate genes in individual cells was assessed with 10,000 random draws of similarly sized data sets.

### Identifying Ctenophore-Specific Genes

To test the hypothesis that phylum-specific cell types are enriched in novel proteins (encoded by phylum-specific genes), we first examined the colloblast and tentacle data sets for ctenophore-specific genes using *alien\_index* ([https://github.com/josephryan/alien\\_index](https://github.com/josephryan/alien_index)). In brief, this method uses a reciprocal BLAST strategy to identify taxon-specific genes as those which have sufficiently poor matches ( $E > 1e-02$ ) to taxa outside Ctenophora. Genomes for non-ctenophore metazoan taxa examined in this study were downloaded from EnsemblMetazoa on January 21, 2016 and consist of the following: *Amphimedon queenslandica* (Porifera), *Capitella teleta* (Annelida), *Crassostrea gigas* (Mollusca), *Daphnia pulex* (Arthropoda), *Drosophila melanogaster* (Arthropoda), *Helobdella robusta* (Annelida), *Lottia gigantea* (Mollusca), *N. vectensis* (Cnidaria), *Strigamia maritima* (Arthropoda), *Strongylocentrotus purpuratus* (Echinodermata), *Trichoplax adhaerens* (Placozoa). Genes that lacked significant hits in this database of animal taxa were considered ctenophore-specific. We also used Interproscan v5.26-65.0 (Jones et al. 2014) to annotate candidate genes against the InterPro Consortium database using the default settings. These analyses were used to evaluate the percentage of candidate genes with annotations. GO annotations had previously been assigned to *M. leidy* gene models (ML2.2) using Trinotate (Levin et al. 2016).

### Presence of Signal Peptides and Transmembrane Domains in Target Genes

We searched candidate genes for signal peptides using SignalP v4.1 (Petersen et al. 2011) and for transmembrane domains using TMHMM v2.0 (Krogh et al. 2001). Generally, the genome of *M. leidy* encodes fewer signal peptides and transmembrane domains than does the human genome (supplementary fig. 3, Supplementary Material online); however, this may simply reflect the fact that ctenophore sequences were not included in the training set for the SignalP and TMHMM algorithms. To test if the number of signal peptides and transmembrane domains identified by SignalP and TMHMM in our candidate gene data sets was greater than random chance, we built 10,000 randomly assembled size-

matched data sets from ML2.2 ( $N = 189$  for colloblast candidates and  $N = 165$  for tentacle candidates). We then ran SignalP on these random sets to determine how many searches produced more signal peptides and transmembrane domains than our initial search.

### Amino Acid Composition and Low Complexity Sequences

We determined the composition of amino acids in the colloblast candidate and tentacle candidate data set and compared them to 10,000 randomly assembled size-matched data sets. To determine if these candidate data sets had high numbers of low-complexity sequence stretches, we used Segmasker v1.0.0 (Wootton 1994) to identify regions of low complexity in these data sets as well as the random data sets.

### Sequence Similarity to Known Adhesive- and Toxin-Related Proteins

To identify putative adhesive genes, we used BLASTP v2.5.0 (Altschul et al. 1990) to search candidates against the Uniprot database concatenated with the 48 adhesive proteins reported previously (Hennebert et al. 2015). To identify venoms/toxins, we searched candidates against the ToxProt database, a Uniprot database annotated for known venom/toxin genes using the Animal Toxin Annotation Project ([www.uniprot.org/program/Toxins](http://www.uniprot.org/program/Toxins)). The Uniprot database was downloaded on October 13, 2017 and the ToxProt database was downloaded on September 5, 2017. Hits with  $E \leq 1e-03$  were considered significant.

### Domain Similarity between Candidates and Known Adhesive- and Toxin-Related Proteins

We used the Interproscan results to test whether our sets of candidate genes disproportionately shared Pfam domains with proteins in the Adhesives and ToxProt databases. Towards this, we compared the number of domains shared by the candidate genes and 10,000 randomly assembled size-matched data sets drawn from each database.

### Cnidocyte Orthology Analysis

Using OrthoFinder (as above) we grouped ML2.2 with the complete set of protein models from *N. vectensis* downloaded from JGI (<https://genome.jgi.doe.gov/Nemve1/Nemve1.home.html>), and a subset of proteins identified by Sebe-Pedros, Saudemont, et al. (2018) as cnidocyte specific but not found in JGI ([www.cnidariangenomes.org/download/nve.gene\\_models.vie130208](http://www.cnidariangenomes.org/download/nve.gene_models.vie130208)). We then searched for orthology groups containing at least one candidate (colloblast or tentacle) from *M. leidy* and at least one cnidocyte-specific protein. We assessed significance by searching for shared orthology groups in 10,000 randomly assembled size-matched groups drawn from ML2.2 and the NVJ database augmented with additional cnidocyte-specific sequences.

### Cell Lineage Tracing in *M. leidy*

Experiments were performed as described previously (Martindale and Henry 1997; Martindale and Henry 1999).

Briefly, individual micromeres of gastrula stage embryos were microinjected with saturated Dil (DiI18(3); Molecular Probes, OR, USA) prepared in soybean oil. Embryos were either imaged immediately or reared to the cydippid stage in 0.2  $\mu\text{m}$  filtered seawater at room temperature before imaging.

### Statistics and Code Availability

We used a Monte Carlo approach to assess the significance of our observations. Briefly, we randomly selected 10,000 data sets each of size  $N = 189$  genes or size  $N = 165$  genes from *M. leidyi* gene models (ML2.2) and compared the distribution of these random draws to colloblast and tentacle candidates, respectively. This approach was used to detect enrichment of late-expressed genes, ctenophore-specific genes, GO annotations, shared Pfam domains, signal peptides, transmembrane domains, regions of low sequence complexity, clustering in individual cells, and clustering with cnidocyte orthologs. Scripts and files for these analyses are in the GitHub Repository for this publication: [https://github.com/josephryan/2018-Babonis\\_et\\_al\\_Ryan](https://github.com/josephryan/2018-Babonis_et_al_Ryan).

### Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

### Acknowledgments

The developmental transcriptomes for *M. leidyi* were prepared, sequenced, and assembled by the laboratory of Dr Itai Yanai. We are grateful to Drs Christy Schnitzler and Sergey Koren for their help with the *Beroe ovata* genome and to Dr Francois Michonneau for his translation and interpretation of the results presented in Carre and Carre (1980, 1989).

### Authors' Contributions

Study design/concept: L.S.B., M.Q.M., J.F.R.; animal/tissue collection: W.R.F., S.H.D.H., L.M.C., A.G.M.; transcriptome assembly: W.R.F., S.H.D.H., L.M.C., M.B.D.; phylogenetics: M.B.D., L.M.C., S.H.D.H., J.F.R.; microinjection: M.Q.M.; other analyses: L.S.B., M.B.D., J.F.R.; writing: L.S.B., J.F.R.; editing: M.Q.M., M.B.D., S.H.D.H., W.R.F., L.M.C., A.G.M. All authors read and approved the final manuscript.

### Funding

This work was supported by the Alabama Established Program to Stimulate Competitive Research (grant 0447675 to A.G.M.), the David and Lucile Packard Foundation; the National Institutes of Health (grant 5-R01-GM087198 to S.H.D.H.); the National Aeronautics and Space Administration (grant NNX14AG70G to M.Q.M.); and the National Science Foundation (grants 1542679 to S.H.D.H. and 1542597 to J.F.R.).

### References

Alie A, Leclere L, Jager M, Dayraud C, Chang PR, Le Guyader H, Queinnee E, Manuel M. 2011. Somatic stem cells express *Piwi* and *Vasa* genes in an adult ctenophore: ancient association of "germline genes" with stemness. *Dev Biol.* 350(1):183–197.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215(3):403–410.

Anderluh G, Podlessek Z, Macek P. 2000. A common motif in parts of cnidarian toxins and nematocyst collagens and its putative role. *Biochim Biophys Acta-Protein Struct Molec Enzymol.* 1476(2):372–376.

Babonis LS, Martindale MQ, Ryan JF. 2016. Do novel genes drive morphological novelty? An investigation of the nematosomes in the sea anemone *Nematostella vectensis*. *BMC Evol Biol.* 16(1):1–22.

Borisenko I, Ereskovsky AV. 2013. Tentacular apparatus ultrastructure in the larva of *Bolinopsis infundibulum* (Lobata: ctenophora). *Acta Zool.* 94(2):193–202.

Borowiec ML, Lee EK, Chiu JC, Plachetzki DC. 2015. Extracting phylogenetic signal and accounting for bias in whole-genome data sets supports the Ctenophora as sister to remaining Metazoa. *BMC Genomics.* 16:987.

Carre C, Carre D. 1980. Les cnidocystes du ctenophore *Euchlora rubra* (Koluker 1853). *Cahiers De Biol Mar.* 21:221–226.

Carre D, Carre C. 1989. Acquisition de cnidocystes et differentiation de pseudocolloblastes chez les larves et les adultes de deux ctenophores du genre *Haeckelia* Carus, 1863. *Canadienne De Zoologie* 67:2169–2179.

Dohrmann M, Worheide G. 2017. Dating early animal evolution using phylogenomic data. *Sci Rep.* 7(1):3599.

Dunn CW, Hejnol A, Matus DQ, Pang K, Browne WE, Smith SA, Seaver E, Rouse GW, Obst M, Edgecombe GD. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* 452(7188):745–749.

Eckhaut I, Flammang P, Lo Bue C, Jangoux M. 1997. Functional morphology of the tentacles and tentilla of *Coeloplana bannworthi* (Ctenophora, Platyctenida), an ectosymbiont of *Diadema setosum* (Echinodermata, Echinoida). *Zoomorphology* 117(3):165–174.

Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16:157.

Emsen RH, Whitfield PJ. 1991. Behavioral and ultrastructural studies on the sedentary platyctenean ctenophore *Vallicula multififormis*. *Hydrobiologia* 216–217(1):27–33.

Endrizzi BJ, Stewart RJ. 2009. Glueomics: an expression survey of the adhesive gland of the sandcastle worm. *J Adhes.* 85(8):546–559.

Eschscholtz F. 1829. System der Acalephen. Eine ausführliche Beschreibung aller medusenartigen Strahltiere. Berlin (Germany): Ferdinand Dümmler.

Flammang P, Lambert A, Bailly P, Hennebert E. 2009. Polyphosphoprotein-containing marine adhesives. *J Adhes.* 85(8):447–464.

Flici H, Schnitzler CE, Millane RC, Govinden G, Houlihan A, Boomkamp SD, Shen SB, Baxevanis AD, Frank U. 2017. An Evolutionarily conserved SoxB-Hdac2 crosstalk regulates neurogenesis in a cnidarian. *Cell Reports* 18:1395–1409.

Franc J-M. 1978. Organization and function of ctenophore colloblasts: an ultrastructural study. *Biol Bull.* 155(3):527–541.

Francis WR, Shaner NC, Christianson LM, Powers ML, Haddock SHD. 2015. Occurrence of isopenicillin-N-synthase homologs in bioluminescent ctenophores and implications for coelenterazine biosynthesis. *PLoS ONE.* 10(6):e0128742.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 29(7):644–652.

Gremme G, Steinbiss S, Kurtz S. 2013. GenomeTools: a comprehensive software library for efficient processing of structured genome annotations. *IEEE-ACM Trans Comput Biol Bioinf.* 10(3):645–656.

Haddock SHD. 2007. Comparative feeding behavior of planktonic ctenophores. *Integr Comp Biol.* 47(6):847–853.

Hartenstein V, Posakony JW. 1989. Development of adult sensilla on the wing and notum of *Drosophila melanogaster*. *Development* 107(2):389–405.

- Hennebert E, Maldonado B, Ladurner P, Flammang P, Santos R. 2015. Experimental strategies for the identification and characterization of adhesive proteins in animals: a review. *Interface Focus* 5(1):20140064.
- Henry JQ, Martindale MQ. 2001. Multiple inductive signals are involved in the development of the ctenophore *Mnemiopsis leidyi*. *Dev Biol* 238(1):40–46.
- Hernandez-Nicaise ML. 1973. Le système nerveux des ctenaires. I. Structure et ultrastructure des réseaux épithéliaux. *Z Zellforsch* 137(2):223–250.
- Heyer LJ, Kruglyak S, Yooshep S. 1999. Exploring expression data: identification and analysis of coexpressed genes. *Genome Res* 9(11):1106–1115.
- Horridge GA. 1965. Non-motile sensory cilia and neuromuscular junctions in a ctenophore independent effector organ. *Proc R Soc Lond B Biol Sci* 162(988):333–350.
- Jackson DJ, Meyer NP, Seaver E, Pang K, McDougall C, Moy VN, Gordon K, Degnan BM, Martindale MQ, Burke RD, et al. 2010. Developmental expression of *COE* across the Metazoa supports a conserved role in neuronal cell-type specification and mesodermal development. *Dev Genes Evol* 220(7–8):221–234.
- Jager M, Queinnec E, Chiori R, Le Guyader H, Manuel M. 2008. Insights into the early evolution of *Sox* genes from expression analyses in a ctenophore. *J Exp Zool B Mol Dev Evol* 310(8):650–667.
- Jekely G, Paps J, Nielsen C. 2015. The phylogenetic position of ctenophores and the origin(s) of nervous systems. *Evodevo* 6:1.
- Johnson BR, Tsutsui ND. 2011. Taxonomically restricted genes are associated with the evolution of sociality in the honey bee. *BMC Genomics* 12:164.
- Jones D, Sawicki G, Wozniak M. 1992. Sequence, structure, and expression of a wasp venom protein with a negatively charged signal peptide and a novel repeating internal structure. *J Biol Chem* 267:14871–14878.
- Jones P, Binns D, Chang HY, Fraser M, Li WZ, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30(9):1236–1240.
- Jungo F, Bougueleret L, Xenarios I, Poux S. 2012. The UniProtKB/Swiss-Prot Tox-Prot program: a central hub of integrated venom protein data. *Toxicon* 60(4):551–557.
- Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305(3):567–580.
- Levin M, Anavy L, Cole AG, Winter E, Mostov N, Khair S, Senderovich N, Kovalev E, Silver DH, Feder M, et al. 2016. The mid-developmental transition and the evolution of animal body plans. *Nature* 531(7596):637–641.
- Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15(12):550–571.
- Mackie GO, Mills CE, Singla CL. 1988. Structure and function of the prehensile tentilla of *Euplokamis* (Ctenophora, Cydippida). *Zoomorphology* 107(6):319–337.
- Martin-Duran JM, Ryan JF, Vellutini BC, Pang K, Hejnal A. 2017. Increased taxon sampling reveals thousands of hidden orthologs in flatworms. *Genome Res* 27(7):1263–1272.
- Martindale MQ. 1986. The ontogeny and maintenance of adult symmetry properties in the ctenophore, *Mnemiopsis mccradyi*. *Dev Biol* 118(2):556–576.
- Martindale MQ, Henry JJ. 1997. Experimental analysis of tentacle formation in the ctenophore *Mnemiopsis leidyi*. *Biological Bulletin* 193(2):245–247.
- Martindale MQ, Henry JQ. 1999. Intracellular fate mapping in a basal metazoan, the ctenophore *Mnemiopsis leidyi*, reveals the origins of mesoderm and the existence of indeterminate cell lineages. *Dev Biol* 214(2):243–257.
- Mills CE, Miller RL. 1984. Ingestion of a medusa (*Aegina citrea*) by the nematocyst-containing ctenophore *Haeckelia rubra* (formerly *Euchlora rubra*) – phylogenetic implications. *Marine Biology* 78(2):215–221.
- Moroz LL, Kocot KM, Citarella MR, Dosung S, Norekian TP, Povolotskaya IS, Grigorenko AP, Dailey C, Berezikov E, Buckley KM, et al. 2014. The ctenophore genome and the evolutionary origins of neural systems. *Nature* 510(7503):109–114.
- Moss AG, Rapoza RC, Muellner L. 2001. A novel cilia-based feature within the food grooves of the ctenophore *Mnemiopsis mccradyi* Mayer. *Hydrobiologia* 451:287–294.
- Nakamura T, Yamada KD, Tomii K, Katoh K. 2018. Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics* 34(14):2490–2492.
- Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32(1):268–274.
- Oshiro N, Kobayashi C, Iwanaga S, Nozaki M, Namikoshi M, Spring J, Nagai H. 2004. A new membrane-attack complex/perforin (MACPF) domain lethal toxin from the nematocyst venom of the Okinawan sea anemone *Actinaria villosa*. *Toxicon* 43(2):225–228.
- Pang K, Martindale MQ. 2008. Developmental expression of homeobox genes in the ctenophore *Mnemiopsis leidyi*. *Dev Genes Evol* 218(6):307–319.
- Pang K, Ryan JF, Baxevas AD, Martindale MQ. 2011. Evolution of the *TGF-beta* signaling pathway and its potential role in the ctenophore, *Mnemiopsis leidyi*. *PLoS ONE* 6(9):e24152.
- Pang K, Ryan JF, Program NCS, Mullikin JC, Baxevas AD, Martindale MQ. 2010. Genomic insights into *Wnt* signaling in an early diverging metazoan, the ctenophore *Mnemiopsis leidyi*. *Evodevo* 1(1):10.
- Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8(10):785–786.
- Podar M, Haddock SHD, Sogin ML, Harbison GR. 2001. A molecular phylogenetic framework for the phylum Ctenophora using 18S rRNA genes. *Mol Phylogenet Evol* 21(2):218–230.
- Rachamim T, Morgenstern D, Aharonovich D, Brekhan V, Lotan T, Sher D. 2015. The dynamically evolving nematocyst content of an anthozoan, a scyphozoan, and a hydrozoan. *Mol Biol Evol* 32(3):740–753.
- Reitzel AM, Pang K, Martindale MQ. 2016. Developmental expression of “germline”- and “sex determination”-related genes in the ctenophore *Mnemiopsis leidyi*. *Evodevo* 7:17.
- Reitzel AM, Pang K, Ryan JF, Mullikin JC, Martindale MQ, Baxevas AD, Tarrant AM. 2011. Nuclear receptors from the ctenophore *Mnemiopsis leidyi* lack a zinc-finger DNA-binding domain: lineage-specific loss or ancestral condition in the emergence of the nuclear receptor superfamily?. *Evodevo* 2(1):3.
- Richards GS, Rentzsch F. 2014. Transgenic analysis of a *SoxB* gene reveals neural progenitor cells in the cnidarian *Nematostella vectensis*. *Development* 141(24):4681–4689.
- Ryan JF, Chiodin M. 2015. Where is my mind? How sponges and placozoans may have lost neural cell types. *Philos Trans R Soc Lond B Biol Sci* 370(1684):20150059.
- Ryan JF, Pang K, Mullikin JC, Martindale MQ, Baxevas AD, Progra NCS. 2010. The homeodomain complement of the ctenophore *Mnemiopsis leidyi* suggests that Ctenophora and Porifera diverged prior to the ParaHoxozoa. *Evodevo* 1(1):9.
- Ryan JF, Pang K, Schnitzler CE, Nguyen A-D, Moreland RT, Simmons DK, Koch BJ, Francis WR, Havlak P, Smith SA, et al. 2013. The genome of the ctenophore *Mnemiopsis leidyi* and its implications for cell type evolution. *Science* 342(6164):1242592.
- Schnitzler CE, Pang K, Powers ML, Reitzel AM, Ryan JF, Simmons D, Tada T, Park M, Gupta J, Brooks SY, et al. 2012. Genomic organization, evolution, and expression of photoprotein and opsin genes in *Mnemiopsis leidyi*: a new view of ctenophore photocytes. *BMC Biol* 10:107.
- Schnitzler CE, Simmons DK, Pang K, Martindale MQ, Baxevas AD. 2014. Expression of multiple *Sox* genes through embryonic development in the ctenophore *Mnemiopsis leidyi* is spatially restricted to zones of cell proliferation. *Evodevo* 5:15.

- Schulz MH, Zerbino DR, Vingron M, Birney E. 2012. Oases: robust *de novo* RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* 28(8):1086–1092.
- Sebe-Pedros A, Chomsky E, et al. 2018. Early metazoan cell type diversity and the evolution of multicellular gene regulation. *Nat Ecol Evol* 2(7):1176–1188.
- Sebe-Pedros A, Saudemont B, et al. 2018. Cnidarian cell type diversity and regulation revealed by whole-organism single-cell RNA-Seq. *Cell* 173:1520–1534.
- Simion P, Bekkouche N, Jager M, Queinnec E, Manuel M. 2015. Exploring the potential of small RNA subunit and ITS sequences for resolving phylogenetic relationships within the phylum Ctenophora. *Zoology (Jena)* 118(2):102–114.
- Simion P, Philippe H, Baurain D, Jager M, Richter DJ, Di Franco A, Roure B, Satoh N, Queinnec E, Ereskovsky A, et al. 2017. A large and consistent phylogenomic dataset supports sponges as the sister group to all other animals. *Curr Biol* 27(7):958–967.
- Simmons DK, Pang K, Martindale MQ. 2012. *Lim* homeobox genes in the ctenophore *Mnemiopsis leidyi*: the evolution of neural cell type specification. *Evodevo* 3(1):2.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.
- Storch V, Lehnert-Moritz K. 1974. Development of colloblasts of *Pleurobrachia pileus* Ctenophora. *Mar Biol* 28(3):215–219.
- Tamm SL. 1983. Motility and mechanosensitivity of macrocilia in the ctenophore *Beroe*. *Nature* 305(5933):430–433.
- R Core Development Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Available from: <http://www.R-project.org>.
- Waite JH, Andersen NH, Jewhurst S, Sun C. 2005. Mussel adhesion: finding the tricks worth mimicking. *J Adhes* 81(3–4):297–317.
- Whelan NV, Kocot KM, Moroz TP, Mukherjee K, Williams P, Paulay G, Moroz LL, Halanych KM. 2017. Ctenophore relationships and their placement as the sister group to all other animals. *Nat Ecol Evol* 1(11):1737–1746.
- Wootton JC. 1994. Nonglobular domains in protein sequences – automated segmentation using complexity-measures. *Comput Chem* 18(3):269–285.
- Yamada A, Martindale MQ, Fukui A, Tochinai S. 2010. Highly conserved functions of the Brachyury gene on morphogenetic movements: insight from the early-diverging phylum Ctenophora. *Dev Biol* 339(1):212–222.
- Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res* 18(5):821–829.