Published in final edited form as:

Biol Bull. 2023 August ; 245(1): 1–8. doi:10.1086/730536.

Cymric, a maternal and zygotic HTK-16-like SHARK-family tyrosine kinase gene, is disrupted in *M. occulta*, a tailless ascidian

Kazuhiro W. Makabe[a,b], **Hannah I. Jensen**[c], **Alexander C. A. Fodor**[c], **William R. Jeffery**[d,e], **Noriyuki Satoh**[b,f], **Billie J. Swalla**[c,d,*]

[a]Graduate School of Technology, Industrial and Social Sciences, Tokushima University, Tokushima 770-8506, Japan

[b]Department of Zoology, Graduate School of Science, Kyoto University, Kyoto, 606-8502, Japan

[c]Biology Department and Friday Harbor Laboratories, University of Washington, Seattle, WA 98195

[d]Station Biologique, Roscoff, France

[e]Biology Department, University of Maryland, College Park, MD, 20742

[f]Marine Genomics Unit, Okinawa Institute of Science and Technology, Onna, Okinawa 904-0495, Japan

Abstract

We describe the cloning and expression of a nonreceptor tyrosine kinase, *cymric* (*Uro-1*), a *HTK16-like (HydraTyrosineKinase-16)* gene, identified in a subtractive screen for maternal ascidian cDNAs in *Molgula oculata*, an ascidian species with a tadpole larva. *Cymric* encodes a 4 kb mRNA expressed in gonads, eggs, and embryos in the tailed *M. oculata*, but is not detected in eggs or embryos of the closely related tailless species, *M. occulta*. There is a large insertion in *cymric* in the *M. occulta* genome, as shown by transcriptome and genome analyses, resulting in it becoming a pseudogene. The *cymric* amino acid sequence encodes a non-receptor tyrosine kinase (TK) with an N-terminal region containing two SH2 domains and five ankyrin repeats, similar to the HTK-16-like gene found in other ascidians. Thus, the ascidian *cymric* genes are members of the SHARK (Src-homology ankyrin-repeat containing tyrosine kinase) family of non-receptor TKs, which have been found throughout invertebrates. We show that *cymric* is lacking the tyrosine kinase domain in the tailless *M. occulta*, although the truncated mRNA is still expressed in transcriptome data. This maternal and zygotic HTK-16-like tyrosine kinase is another described pseudogene from *M. occulta* and appears to not be necessary for adult development.

Keywords

Evolution and development; pseudogenes; signaling; subtractive cloning; tyrosine kinases

*Address correspondence and proofs to: Dr. Billie J. Swalla, 620 University Road, Friday Harbor Laboratories, University of Washington, Friday Harbor, WA 98195, Fax: 206-616-2011, bjswalla@u.washington.edu, bjswalla@uw.edu.

1. Introduction

Ascidians are invertebrate chordates with a larval notochord, dorsal hollow nerve tube, muscle cells in their tails, and pharyngeal slits as adults (Swalla, 2004 a, b; Fodor *et al*. 2021 a). Ascidian larvae exhibit the simplest chordate body plan and gastrulate early in development, after the 110-cell stage (Swalla, 2004b). Cleavage is determinate and invariant with a known cell lineage in all solitary ascidians that have been studied, and there is great interest in understanding ascidian embryonic development for clues to the evolution of the chordate body plan (Swalla, 2004b; Zeng and Swalla, 2005) Two closely related ascidian species with completely different larvae were used for the studies described here – *Molgula oculata*, which has a tailed (urodele) larva exhibiting typical chordate features, and *Molgula occulta*, which develops into a tailless (anural) larva, lacking chordate features (Fig. 1) (Berrill, 1931; Swalla and Jeffery 1990). Hybrids can be made between the two species, if the tailed *M. oculata* egg is used then the resulting larva is always tailed, if the tailless *M. occulta* egg is fertilized with a tailed *M. oculata* sperm, then it results in a larva with a non-functional shortened tail (Swalla and Jeffery, 1990).

Subtractive hybridization between RNA isolated from gonads of the two species allowed the isolation of three RNAs expressed in the tailed species but not in the tailless species (Swalla *et al*. 1993; Swalla 1996). Two of those RNAs correspond to genes present in the genome of both species but were differentially expressed in the tailed *M. oculata* species (Swalla *et al*. 1993). One of these genes, *manx*, is found in a multigene complex with a differentially expressed RNA helicase, *bobcat, p68*, or *DDX5/17* (Swalla *et al*. 1999; Fodor *et al*. 2021a). Expression of *manx* and *bobcat* is necessary for the formation of the larval tail in *M. oculata*, as shown by antisense oligo treatment (Swalla and Jeffery 1996). In contrast, the gene reported here, a non-receptor tyrosine kinase, *cymric*, is altered in the tailless *M. occulta* genome. It was isolated in the original screen because the *M. occulta* transcript lacks the 3'noncoding region and the tyrosine kinase domain, which is present in *M. oculata*, but missing in *M. occulta's* disrupted kinase. The mutations to the *cymric* gene are not suspected to be the main cause of tail loss in *M. occulta* because a complete *cymric* gene is present in other ascidian species with tailless larvae (unpublished data). Instead, it is thought that *cymric* has become a pseudogene due lack of evolutionary time to complete purifying selection for larval genes. Since the loss of function of *cymric* in *M. occulta* does not seem to have affected the adult animal, we suspect that *cymric* is necessary for the development of the tadpole and has little, or a reduced role, in the adult.

Cymric (previously called *Uro-1*) belongs to the *HTK-16-like* SHARK family of non-receptor tyrosine kinases. These are unusual tyrosine kinases with five ankyrin repeats flanked by two SH2 domains at the 5'end (Ferrante *et al*. 1995). The first gene isolated from this family was described in hydra (Chan *et al*. 1994) and later also found in *Drosophila melanogaster* (Ferrante *et al*. 1995). In *Drosophila*, SHARK is present on the apical surface of ectodermal cells, and genetic evidence suggests that it may be involved in signaling by crumbs, but the exact function is still unknown (Ferrante *et al*. 1995). SHARK tyrosine kinase has been shown to be critical in signaling the migration of leading-edge cells to form the dorsal closure in *Drosophila* (Fernandez *et al*. 2000). Leading-edge cells express *dpp*, a *TGFβ* superfamily member, and this leads to the migration of the lateral ectoderm

toward the dorsal midline. *Drosophila* SHARK is located upstream of *dpp* and downstream of *Bullwinkle* (*bwk*) in the Jun-kinase signaling pathway, regulating embryonic epidermal morphogenesis (Fernandez *et al.* 2000, Tran and Berg 2003). In *Drosophila* embryos, *bwk* regulates cell migration during the formation of dorsal appendages (Tran and Berg 2003). It has been shown in *Drosophila* that *SHARK* will restore *bwk* mutants to the wild phenotype, but the precise interaction at the protein level is not yet known (Tran and Berg 2003).

SHARK is also important in *Drosophila* for the recruitment of inflammatory macrophages via Src42A-Draper-Shark signaling in response to damage induced by H_2O_2. In *Drosophila*, *Src42A* is a homologue of zebrafish *lyn*. Src42A detects H_2O_2 in response to a wound and interacts with Draper to recruit SHARK, possibly equivalent to the SFK-ITAM-Syk pathway used by vertebrates in adaptive immune response (Evans *et al.* 2015). It has been demonstrated that *SHARK* mutant embryos fail to recruit macrophage cells and re-expression of *SHARK* in the mutants rescues the phenotype (Evans *et al.* 2015). *SHARK* has also been shown to be necessary in signaling *Draper* to initiate glial phagocytosis in *C. elegans* and axon-injury-induced changes. Knocking down *SHARK* in the glia complex of *C. elegans* causes a failure of phagocytosis and axonal healing (Stanley *et al.* 2008).

Ankyrin repeats have been identified in diverse proteins and are important in protein-protein interactions (Michaely and Bennett 1992). Ankyrin is a critical component of the ascidian egg myoplasm, which is necessary for axis determination and formation (Swalla 1992) and is frequently missing or altered in the tailless species (Jeffery and Swalla 1992; Jeffery and Swalla 1993; Swalla *et al.* 1993; Jeffery *et al.* 1999). It is intriguing that a specific tyrosine kinase might be anchored in the egg cortex of tailed species and necessary for the initial events of axis formation to occur during the first cell cycle, which will in turn lead to the proper movements during gastrulation (Swalla 1992, 2004b). In this paper, we describe the cloning and expression of *cymric* RNA in *M. oculata* and M. *occulta* ascidians. We present genomic analyses that *cymric* is intact in *M. oculata* and *M. occidentalis*, two tailed Molgula species, but is disrupted in the tailless *M. occulta*.

2. Materials and Methods

2.1 Biological materials

Molgula oculata and *Molgula occulta* were collected by dredging sand flats at Point de Bloscon near Roscoff, France. Gonads were dissected from gravid adults to obtain sperm and eggs. Procedures for obtaining gametes, fertilization, and culturing embryos have been described previously (Swalla and Jeffery 1990).

2.2 Isolation of DNA and RNA

Genomic DNA was prepared by homogenizing isolated testes in 1 mg/ml Proteinase K, incubating the mixture at 65°C for 15 min and then at 37°C overnight, followed by phenol:chloroform extraction (Davis *et al.* 1986). RNA was prepared from gonads or embryos by the guanidine isothiocyanate method (March *et al.* 1985).

2.3 Isolation of cDNA clones by subtractive hybridization

DNA was prepared from *M. oculata* gonad poly $(A)^+$ RNA and a subtracted probe was prepared using *M. occulta* cDNA as described previously (Swalla *et al.* 1993). The subtracted cDNA was amplified by PCR and used to screen an *M. oculata* gonad cDNA library for positive clones. Probes made from the positive clones were then used to screen an *M. occulta* cDNA library. Clones that screened positive in the *M. oculata* cDNA library and negative in the *M. occulta* cDNA library were retained. The resulting *M. oculata* clones were then subjected to RNA slot blot hybridization (Swalla *et al.* 1993) to distinguish those encoding differentially expressed RNAs of the tailed and tailless species from the background. A complete description of the subtraction procedure is available elsewhere (Swalla 1996). One of the cDNA clones isolated in the subtractive hybridization was Urodele 1 (Uro 1), which encodes the 3' region of the *cymric* mRNA (Fig. 2A). A probe made from the Uro-1 insert was used to screen the *M. oculata* cDNA library for longer clones. This screen resulted in the isolation of the overlapping clones, from which the *cymric* cDNA sequence of *M. oculata* was derived.

We have since sequenced the transcriptomes of *M. oculata*, *M. occulta* and the hybrid embryos (made with *M. occulta* eggs, fertilized by *M. oculata* sperm) at three different time points: 3 hours after fertilization (F+3; gastrulation), 4 hours after fertilization (F+4; neurulation), and 5 hours after fertilization (F+5; early tailbud) (Lowe *et al.* 2014). De novo assembly of the transcriptome of these three stages yielded about 16,000 genes in *M. oculata* and *M. occulta*. Next, we sequenced the genomes of both species of molgulid ascidians, with a third species, the tailed *Molgula occidentalis*, that is available all year from Gulf Specimen Marine Laboratory Inc. in Panacea, Florida, USA. Genomes were sequenced in collaboration with the lab of Lionel Christiaen at New York University (Stolfi *et al.* 2014). The genomic sequences for *M. oculata*, *M. occulta*, and *M. occidentalis* were deposited into ANISEED (Ascidian Network for In Situ Expression and Embryological Database) (Brozovic *et al.* 2017; Tassy *et al.* 2010) to facilitate sharing of these unique genomic resources.

2.4 Sequencing and computer analysis

The original Uro-1 cDNA clones were Sanger sequenced (old school) on both strands using [^{35}S]-dATP (800 Ci/mmole; New England Nuclear, Boston, MA) by the dideoxy chain termination method (Sanger *et al.* 1977) with Sequenase (United States Biochemical Corp., Cleveland, OH). Oligonucleotide primers were made on a Pharmacia LKB Gene Assembler Plus (Pharmacia Biosystems, Inc.; Piscataway, NJ) to create overlapping sequence information. Sequences were read and compared with the MacVector, Inc. Program (MacVector, Inc.; Apex, NC). The genomic sequences for *cymric* in *M. oculata* and *M. occulta* as well as the other ascidian species were retrieved from ANISEED by BLAST searching, using the sequenced cDNA clones as a query. The resulting sequences were then used to search the ascidian transcriptomes. The deduced protein sequences were compared with MacVector, Inc. along with sequences present in the National Biomedical Research Foundation protein database and the on-line GenBank protein sequences available through the National Library of Medicine. The accession number for *M. oculata cymric* is OQ445879, *M. occulta* (OQ589864), *M. occidentalis* (OQ589865),

C. intestinalis (XP_002121112), *C. savignyi* (OQ589867), *H. roretzi* (OQ589866), *P. mammillata* (CAB3267847) and *S. clava* (XP_039270686).

2.5 In situ hybridizations

Whole mounts of eggs and embryos were subjected to in situ hybridization with DIG labeled probes as described by Swalla *et al.* (1994). Embryos were dechorionated manually after fixation or by treatment with Sodium thioglycolate and Pronase E (Sigma Chemical Corp., St. Louis, MO) before fixation. Proteinase K treament was carried out at a concentration of 10μg/ml for 30min at 37°C. The hybridization and washing temperatures were 47°C and the final wash was done in 0.5XSSC, 50 % formamide at 47°C. The whole mounts were cleared in 1:2 benzyl alcohol: benzyl benzoate and photographed.

2.6 Gene Tree Analyses

Protein sequences were deduced from *M. oculata* and *M. occulta* transcriptome data using MacVector. Other species' protein sequences were found by Blast searching using *M. oculata* protein sequences as a query in NCBI and ANISEED (Brozovic *et al.* 2017). Proteins were then aligned in MacVector using a Muscle alignment.

3. Results

3.1 Isolation and characterization of *cymric* cDNA clones

The *cymric* cDNA clone was obtained in a subtractive screen designed to identify maternal genes expressed in the tailed ascidian species, *Molgula oculata*, but not expressed (or down regulated) in the tailless ascidian species, *Molgula occulta* (Fig. 1) (Swalla *et al.* 1993). Three different clones, *Uro-1*, *Uro-2*, and *Uro-11*, were isolated in this screen. The full-length *Uro-2* (*lynx*) and *Uro-11* (*manx*) cDNAs encode leucine zipper and zinc finger proteins respectively, which are expressed preferentially in the tailed species (Swalla *et al.* 1993). The initial *cymric* clone was incomplete, although it contained an open reading frame corresponding to the C-terminal region of a protein tyrosine kinase (Figure 2; (Swalla 1996). To obtain longer cDNA clones, the *M. oculata* cDNA library was re screened with the *cymric* insert, and five additional overlapping clones were isolated and sequenced. Further screening failed to produce longer cDNA clones. The longest clone, *Uro-1–2*, consists of a 2073 nucleotide (nt) coding region and a 1215 nt 3' UTR, which terminates in a polyadenylation site and a poly (A) tail (Figure 2). The gene corresponding to the *Uro-1* cDNA compiled sequence was designated *cymric*, consistent with the practice of naming ascidian *Uro* genes after tailless cats (Swalla 1996). The translated amino acid sequence shows that *cymric* encodes a HTK-16-like tyrosine kinase of the SHARK family (Figure 2). The Genbank accession number for *M. oculata cymric* is OQ445879.

3.2 The cymric gene encodes a SHARK family protein tyrosine kinase

There was high sequence identity between Cymric and proteins associated with the HTK-16 and SHARK families of tyrosine kinases (Chan *et al.* 1994, Ferrante *et al.* 1995). Figure 3 shows an alignment of tailed and tailless Cymric to other metazoans including *M. occidentalis* (OQ589865), *Ciona intestinalis* (XP_002121112.2), *Strongylocentrotus purpuratus* (XP_001180232.2), *Aplysia californica* (XP_005102068.1), *Hydra vulgaris*

(NP_00129668.1) (Chan *et al*. 1994), *Amphimedon queenslandica* (BAA81720) (Suga *et al*. 1999), and *Drosophila melanogaster* (NP_524743.2) (Ferrante *et al*. 1995). All metazoan HTK-16 genes share 2 SH2 domains (Fig. 3 pink) flanking a series of ankyrin repeats (Fig. 3 blue), and all except for the tailless ascidian, *Molgula occulta*, have a very conserved tyrosine kinase domain (Fig. 3 green). The SHARK domains share a high similarity between sponges and the distantly related ascidians, indicating that this protein was present in the metazoan ancestor and has been highly conserved. Blast searches of ascidian SHARK proteins did not yield any vertebrate homologues, indicating that this gene was lost in the vertebrate lineage prior to whole genome duplications.

The Cymric protein also exhibits a long proline-rich region linking the N-terminal SH2 domain and ankyrin repeats with the C-terminal catalytic domain (Ferrante *et al*. 1995). The HTK-16, sponge SHARK and ascidian Cymric proteins have truncated proline-rich regions linking their N- and C-terminal domains. Cymric is a new member of the SHARK family of protein tyrosine kinases; and the fact that SHARK proteins are found in many invertebrate metazoan groups, including cnidarians, bugs, beetles and bivalves suggests that it is a very ancient and conserved protein in the evolution of metazoans. However, there were no *cymric* gene sequences reported for vertebrates, which suggests that it was lost in a long ago vertebrate ancestor.

3.3 The single-copy *cymric* gene is disrupted in the tailless *M. occulta*

A single copy of intact *Cymric* was found in the genomes of the two tailed species *M. occidentalis* and *M. oculata*, and a partial copy, lacking the tyrosine kinase domain, was found in the tailless species, *M. occulta* (Fig. 2A). When looking at the genomic organization, *M. oculata cymric* contains 14 exons which code for the entire protein. However, *M. occulta cymric* has 10 exons, the first exon in the 5' noncoding region is unique and not expressed in the tailed species; the next 9 coding exons share high homology with the first 9 exons of *M. oculata* (Fig. 2A). There is an interruption after the 10th exon in the tailless *M. occulta* genome and the rest of the gene is truncated before the tyrosine kinase domain (Fig. 2B). All of the other ascidian species' genomes in ANISEED had a complete *cymric* that contained all of the critical domains whereas the *cymric* gene from *M. occulta* was truncated before reaching the tyrosine kinase domain (Fig. 2).

Transcriptomes for 3 different stages of development (gastrula, neurula, and tailbud) are available for three Molgulid species, *M. oculata*, *M. occulta*, and *M. occidentalis*. The transcripts in each stage were then combined into a "developmental transcriptome" which we used to search for the *cymric* transcript. The transcriptome data confirms the genomic findings. Both *M. oculata* and *M. occulta cymric* are found expressed in the transcriptome; the tailed *M. oculata* transcript encodes a full-length protein, while the tailless *M. occulta's* transcript is truncated, missing the tyrosine kinase domain (Fig. 2A). To summarize, the *M. occulta* Cymric protein lacks a tyrosine kinase domain, making it nonfunctional, and the *cymric* gene is now a pseudogene.

3.4 *Cymric* is expressed in oocytes and embryos of the tailed species, *M. oculata*

The accumulation of *cymric* transcripts during development of the tailed *M. oculata* was examined by *in situ* hybridization, since earlier studies showed that *cymric* is expressed in *M. oculata* gonads and not *M. occulta* gonads (Swalla *et al.* 1993). As shown in Figure 4, *cymric* mRNA was observed in the vegetal pole region of the oocytes of the tailed species (Fig. 4A). Expression remains in the vegetal and later the posterior poles of the embryo where the larval muscle cells are determined (Fig. 4B–C), then was seen posteriorly in the muscle cell precursors during gastrulation and neurulation (Fig. 4D–E), and expression was lost in the tailbud stage (Fig. 4F).

4. Discussion

Subtractive hybridization between two molgulid ascidian species, *Molgula occulta* and *Molgula oculata*, with different developmental modes (Swalla *et al.* 1993), allowed the isolation of *cymric*, an *HTK-16 SHARK* family non-receptor tyrosine kinase. *Lynx* (*Uro-2*) and *manx* (*Uro-11*), two other genes from the same screen, have been previously published (Swalla *et al.* 1993). From the results shown here of the genome and transcriptome analyses, we conclude that the full length *cymric* transcript is present in *M. occidentalis* and *M. oculata*, two molgulid species that make tadpole larvae, but the tailless *M. occulta* transcript is missing the tyrosine kinase domain, making *cymric* non-functional at phosphorylation of other proteins.

The role of *cymric* in ascidian development is still not clear, but it is intriguing that a protein containing ankyrin repeats has been implicated in protein-protein interactions. Ascidian eggs contain a special cortical cytoskeletal domain, the myoplasm, which is known to be important in later tail muscle cell development (Swalla 1992, 1993, 2004b). The cortex contains ankyrin (Jeffery and Swalla 1993) and p58 (Swalla *et al.* 1991) in all species that develop into tailed tadpole larvae. Cymric may be anchored in the cortical egg myoplasm and be important in signaling during muscle cell specification in ascidian embryos.

The myoplasm, an egg cytoplasmic region that is segregated to muscle lineages in ascidian embryos (Swalla 1992), is modified in the eggs of anural species (Swalla *et al.* 1991; Jeffery and Swalla 1992b. The modification occurs during oogenesis, after vitellogenesis begins (Swalla *et al.* 1991). Later, this cortical cytoplasm is inherited by tail muscle cells, which then undergo myogenesis (Swalla 1992, 2004b). In contrast, *Molgula occulta* and other tailless species' eggs contained p58, but failed to initiate and maintain cortical localization of the protein (Swalla *et al.* 1991), which was correlated with the development of embryos lacking muscle cells (Swalla and Jeffery 1990; Jeffery and Swalla 1992a). The mechanism(s) of the changes in oogenesis are not fully understood, but it has also been shown that all of the investigated tailless species in the family Molgulidae lack an ankyrin-like protein in their eggs and embryos (Jeffery and Swalla 1993). This raises the interesting possibility that the Molgulid ancestor modified the ankyrin-like protein, possibly being a pre-adaptation to the further changes during oogenesis seen in all the tailless Molgulid species. This indicates that although the truncated *cymric* transcript is present in the tailless *M. occulta*, if translated from the mRNA, it would be a truncated nonfuctional protein, lacking the tyrosine kinase. Further research will focus on how changes in oogenesis

and gene expression result in the morphological changes seen in the tadpole larvae in other closely related Molgula species.

There have been several genes critical for larval development that have become pseudogenes in the tailless *Molgula occulta*, including the larval muscle actin genes (Kusakabe *et al*. 1996) and tyrosinase, a gene vital for forming the ocellus (a pigmented light sensing organ) which is absent in *M. occulta* (Racioppi *et al*. 2017). *Cymric* has become a pseudogene in the tailless *Molgula occulta* as well, where the mRNA with the truncated tyrosine kinase domain is still expressed in the transcriptome, but it no longer produces a protein with a tyrosine kinase domain. This indicates that the tail loss is a fairly recent development in the evolutionary history of *M. occulta* and that these pseudogenizations are likely a secondary effect of other mutations that are the primary cause of the tail loss.

Cumulative evidence suggests that gene pseudogenization is common in the tailless Molgulids. Not only have *cymric, tyrosinase*, and the *muscle actin* genes all become pseudogenes in *M. occulta* (Kusakabe *et al*. 1996, Racioppi *et al*. 2017), but muscle actin has also become a pseudogene in both of the tailless species *Molgula bleizi* and *Molgula tectiformis* (Gyoja *et al*. 2007, Kusakabe *et al*. 1996). Recent studies have suggested that pseudogenes may play a role in regulation and development, and they can arise in various ways in genomes (Li *et al* 2013). The Molgulid family offers a unique system to study what some of those effects may be while we investigate upstream factors for the primary change causing the tailless phenotype in *Molgula occulta*.

Acknowledgments:

The *Molgula oculata*, *M. occulta* and *M. occidentalis* genomes are available on Aniseed. *Cymric* mRNAs in *M. occulta* and *M. oculata* are available in Genbank as well. We thank Barbara Baldwin, Ruth Tower, Margaret Just, Jana Machula, David Martasian, and Jennifer Reardon who worked on this project in W.R. Jeffery's laboratory at Bodega Marine Lab at the University of California at Davis in the early 1990's. Kristen Suling and Allison Hatmaker worked on this project as undergraduates doing independent study in B.J. Swalla's lab at Vanderbilt University from 1994–1997. We are also indebted to Nicole Sanseau, Dr. Andre Toulmond, Dr. Laurent Meijer, and the collection staff at Station Biologique, Roscoff for their support of this work.

Data Availability Statement:

Sequences for the Cymric gene were retrieved from ANISEED (http://www.aniseed.cnrs.fr/), a public database containing the genomes of various tunicate species. These sequences are also available in greater detail in the NCIB (https://www.ncbi.nlm.nih.gov/). The accession numbers are as follows: *Styela clava* XP_039270686.1, *Phallusia mammillata* CAB3267847.1, *Ciona intestinalis* XP_002121112.2, *Molgula oculata* OQ445879, *Molgula occulta* OQ445879, *Molgula occidentalis* OQ589865, *Halocynthia rorezi* OQ589866, *Ciona savigny* OQ589867.

Literature Cited

Berrill NJ 1931. Studies in tunicate development, Part II. Abbreviation of development in the Molgulidae. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences. 219: 281–346.

Brozovic M, Dantec C, Dardaillon J, Dauga D, Faure E, Gineste M, Louis A, Naville M, Nitta KR, Piette J, Reeves W, Scornavacca C, Simion P, Vincentelli R, Bellec M, Ben Aicha S, Fagotto M, Guéroult-Bellone M, Haeussler M, … Lamaire P 2017. ANISEED.

Chan TA, Chu CA, Rauen KA, Kroiher M, Tatarewicz SM, and Steele RE 1994. Identification of a gene encoding a novel protein-tyrosine kinase containing SH2 domains and ankyrin-like repeats. Oncogene. 9(4): 1253–1259. [PubMed: 8134129]

Davis LG 1986. Basic methods in molecular biology. Elsevier.

Evans IR, Rodrigues FS, Armitage EL, and Wood W 2015. Draper/CED-1 Mediates an Ancient Damage Response to Control Inflammatory Blood Cell Migration In Vivo. Current Biology. 25(12): 1606–1612. 10.1016/j.cub.2015.04.037 [PubMed: 26028435]

Fernandez R, Takahashi F, Liu Z, Steward R, Stein D, and Stanley ER 2000. The Drosophila shark tyrosine kinase is required for embryonic dorsal closure. Genes and Development. 14(5): 604–614. 10.1101/gad.14.5.604 [PubMed: 10716948]

Ferrante AW, Reinke R, and Stanley ER 1995. Shark, a Src homology 2, ankyrin repeat, tyrosine kinase, is expressed on the apical surfaces of ectodermal epithelia. Proceedings of the National Academy of Sciences – PNAS. 92(6): 1911–1915. 10.1073/pnas.92.6.1911

Fodor ACA, Powers MM, Andrykovich K, Liu J, Lowe EK, Brown CT, Di Gregorio A, Stolfi A, and Swalla BJ 2021. The Degenerate Tale of Ascidian Tails. Integrative and Comparative Biology. 61:(2), 358–369. 10.1093/icb/icab022 [PubMed: 33881514]

Fodor A, Liu J, Turner L, and Swalla BJ 2021. Transitional chordates and vertebrate origins: Tunicates. Current Topics in Developmental Biology. 141: 149–171. 10.1016/bs.ctdb.2020.10.001 [PubMed: 33602487]

Gyoja F, Satou Y, Shin-i T, Kohara Y, Swalla BJ, and Satoh N 2007. Analysis of large scale expression sequenced tags (ESTs) from the anural ascidian, Molgula tectiformis. Developmental Biology. 307:(2), 460–482. 10.1016/j.ydbio.2007.03.035 [PubMed: 17540363]

Higgins DG, and Sharp PM 1988. CLUSTAL: A package for performing multiple sequence alignment on a microcomputer. Gene. 73(1): 237–244. 10.1016/0378-1119(88)90330-7 [PubMed: 3243435]

Jeffery WR, and Swalla BJ 1992a. Evolution of alternate modes of development in ascidians. BioEssays. 14(4): 219–226. 10.1002/bies.950140404 [PubMed: 1596271]

Jeffery WR, and Swalla BJ 1992b. Factors necessary for restoring an evolutionary change in an anural ascidian embryo. Developmental Biology. 153(2): 194–205. 10.1016/0012-1606(92)90105-P [PubMed: 1397677]

Jeffery WR, and Swalla BJ 1993. An ankryin-like protein in ascidian eggs and its role in the evolution of direct development. Zygote (Cambridge). 1(3): 197–208. 10.1017/S0967199400001477

Jeffery WR, Swalla BJ, Ewing N, and Kusakabe T 1999. Evolution of the ascidian anural larva: Evidence from embryos and molecules. Molecular Biology and Evolution. 16(5): 646–654. 10.1093/oxfordjournals.molbev.a026147 [PubMed: 10335658]

Kusakabe T, Swalla BJ, Satoh N, and Jeffery WR 1996. Mechanism of an Evolutionary Change in Muscle Cell Differentiation in Ascidians with Different Modes of Development. Developmental Biology. 174(2): 379–392. 10.1006/dbio.1996.0082 [PubMed: 8631509]

Li W, Yang W, and Wang X-J 2013. Pseudogenes: Pseudo or Real Functional Elements. Journal of Genetics and Genomics. 40(4): 171–177. 10.1016/j.jgg.2013.03.003 [PubMed: 23618400]

Lowe EK, Swalla BJ, and Brown C Titus. 2014. Evaluating a lightweight transcriptome assembly pipeline on two closely related ascidian species. PeerJ Preprints. 10.7287/peerj.preprints.505v1

March CJ, Mosley B, Larsen A, Cerretti DP, Braedt G, Price V, Gillis S, Henney CS, Kronheim SR, Grabstein K, Conlon PJ, Hopp TP, and Cosman D 1985. Cloning, sequence and expression of two distinct human interleukin-1 complementary DNAs. Nature (London). 315(6021): 41–647. 10.1038/315641a0

Michaely P, and Bennett V 1992. The ANK repeat: A ubiquitous motif involved in macromolecular recognition. Trends in Cell Biology. 2(5): 127–129. 10.1016/0962-8924(92)90084-Z [PubMed: 14731966]

Racioppi C, Valoroso MC, Coppola U, Lowe EK, Brown CT, Swalla BJ, Christiaen L, Stolfi A, and Ristoratore F 2017. Evolutionary loss of melanogenesis in the tunicate Molgula occulta. EvoDevo. 8(1): 11–11. 10.1186/s13227-017-0074-x [PubMed: 28729899]

Sanger F, Nicklen S, and Coulson AR 1977. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences – PNAS. 74:(12), 5463–5467. 10.1073/pnas.74.12.5463

Stanley ER, Freeman MR, Ziegenfuss JS, Biswas R, Avery MA, Hong K, Sheehan AE, and Yeung Y-G 2008. Draper-dependent glial phagocytic activity is mediated by Src and Syk family kinase signalling. Nature. 453(7197): 935–939. 10.1038/nature06901 [PubMed: 18432193]

Stolfi A, Lowe EK, Racioppi C, Ristoratore F, Brown CT, Swalla BJ, and Christiaen L 2014. Divergent mechanisms regulate conserved cardiopharyngeal development and gene expression in distantly related ascidians. ELife. 3, e03728–e03728. 10.7554/eLife.03728 [PubMed: 25209999]

Suga H, Koyanagi M, Hoshiyama D, Ono K, Iwabe N, Kuma K, and Miyata T 1999. Extensive gene duplication in the early evolution of animals before the parazoan-eumetazoan split demonstrated by G proteins and protein tyrosine kinases from sponge and hydra. Journal of Molecular Evolution. 48(6): 646–653. 10.1007/PL00006508 [PubMed: 10229568]

Swalla B 1996. Strategies for cloning developmental genes using closely related species. Pp. 197–208 in Molecular Zoology: Advances, Strategies and Protocols, Ferraris JD and Palumbi SR. Wiley-Liss, John Wiley and Sons, Inc.

Swalla B 2004a. Procurement and Culture of Ascidian Embryos. Pp. 115–141 in Methods in Cell Biology: Experimental Analysis of the Development of Sea Urchins and Other Non-Vertebrate Deuterostomes, Ettenson C, Wessell G, and Wray G. Elsevier Science/Academic Press,.

Swalla B 2004b. Protochordate Gastrulation: Lancelets and Ascidians. Pp. 139–149 in C. Stern, Gastrulation. Cold Spring Harbor Press.

Swalla BJ 1992. The role of maternal factors in ascidian larval muscle development. Seminars in Developmental Biology. 3(4): 287–295.

Swalla BJ 1993. Mechanisms of gastrulation and tail formation in ascidians. Microscopy Research and Technique. 26(4): 274–284. 10.1002/jemt.1070260403 [PubMed: 8305720]

Swalla BJ, Badgett MR, and Jeffery WR 1991. Identification of a cytoskeletal protein localized in the myoplasm of ascidian eggs: Localization is modified during anural development. Development (Cambridge). 111(2): 425–436. 10.1242/dev.111.2.425

Swalla BJ, and Jeffery WR. 1990. Interspecific hybridization between an anural and urodele ascidian: Differential expression of urodele features suggests multiple mechanisms control anural development. Developmental Biology. 142(2): 319–334. 10.1016/0012-1606(90)90353-K [PubMed: 2124192]

Swalla BJ, and Jeffery WR. 1996. Requirement of the Manx Gene for Expression of Chordate Features in a Tailless Ascidian Larva. Science (American Association for the Advancement of Science). 274: 1205–1208. 10.1126/science.274.5290.1205

Swalla BJ, Just MA, Pederson EL, and Jeffery WR 1999. A multigene locus containing the Manx and bobcat genes is required for development of chordate features in the ascidian tadpole larva. Development (Cambridge). 126: 1643–1653. 10.1242/dev.126.8.1643

Swalla BJ, Makabe KW, Satoh N, and Jeffery WR 1993. Novel genes expressed differentially in ascidians with alternate modes of development. Development (Cambridge). 119(2): 307–318. 10.1242/dev.119.2.307

Swalla BJ, White ME, Zhou J, and Jeffery WR 1994. Heterochronic expression of an adult muscle actin gene during ascidian larval development. Developmental Genetics. 15(1): 51–63. 10.1002/dvg.1020150107 [PubMed: 8187350]

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S 2011. MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution. 28(10): 2731–2739. 10.1093/molbev/msr121 [PubMed: 21546353]

Tassy O, Dauga D, Daian F, Sobral D, Robin F, Khoueiry P, Salgado D, Fox V, Caillol D, Schiappa R, Laporte B, Rios A, Luxardi G, Kusakabe T, Joly J-S, Darras S, Christiaen L, Contensin M, Auger H, … Lemaire P 2010. The ANISEED database: Digital representation, formalization, and elucidation of a chordate developmental program. Pp. 1459–1468 in Genome research 20(10). Cold Spring Harbor Laboratory Press. 10.1101/gr.108175.110 [PubMed: 20647237]

Tran DH, and Berg CA 2003. Bullwinkle and shark regulate dorsal-appendage morphogenesis in Drosophila oogenesis. Development (Cambridge). 130(25): 6273–6282. 10.1242/dev.00854

Zeng L, and Swalla BJ 2005. Molecular phylogeny of the protochordates: Chordate evolution. Canadian Journal of Zoology. 83(1): 24–33. 10.1139/z05-010

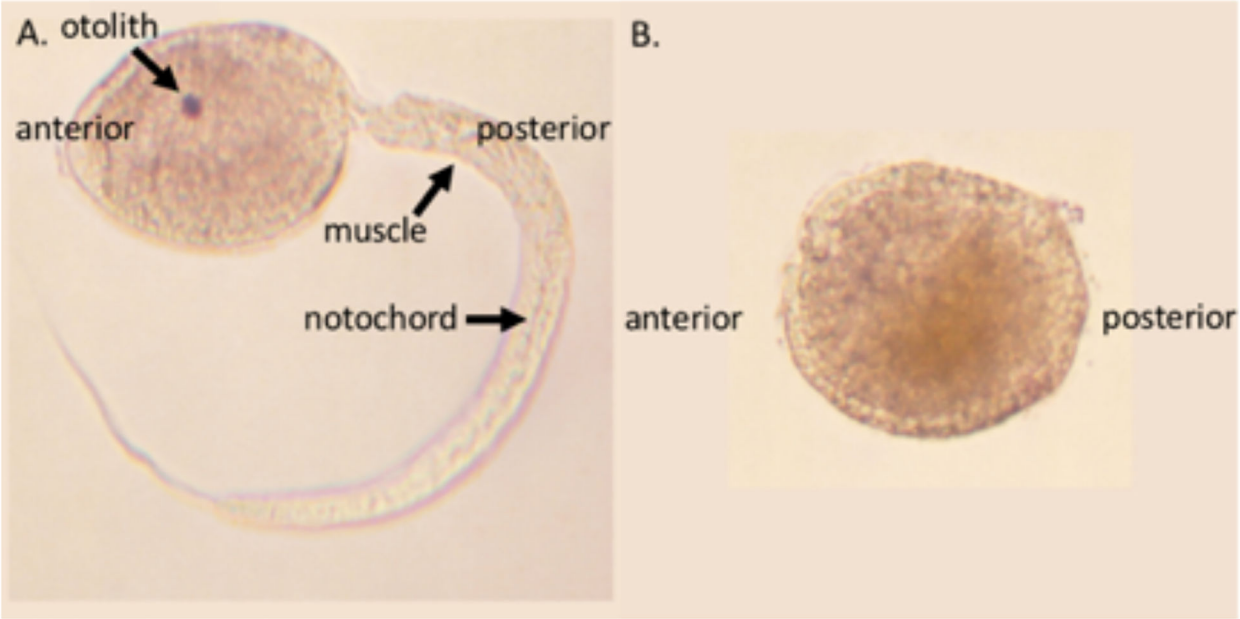

Figure 1.
Photographs of A. *Molgula oculata* and B. *Molgula occulta* larvae. *M. oculata* is a tailed ascidian larva with arrows pointing to notochord and muscle cells in the tail, and a pigmented sensory organ (the otolith) in the head. In contrast, *M. occulta* is tailless, lacking a sensory otolith, notochord, and muscle cells (Swalla and Jeffery 1990; Berrill, 1931).

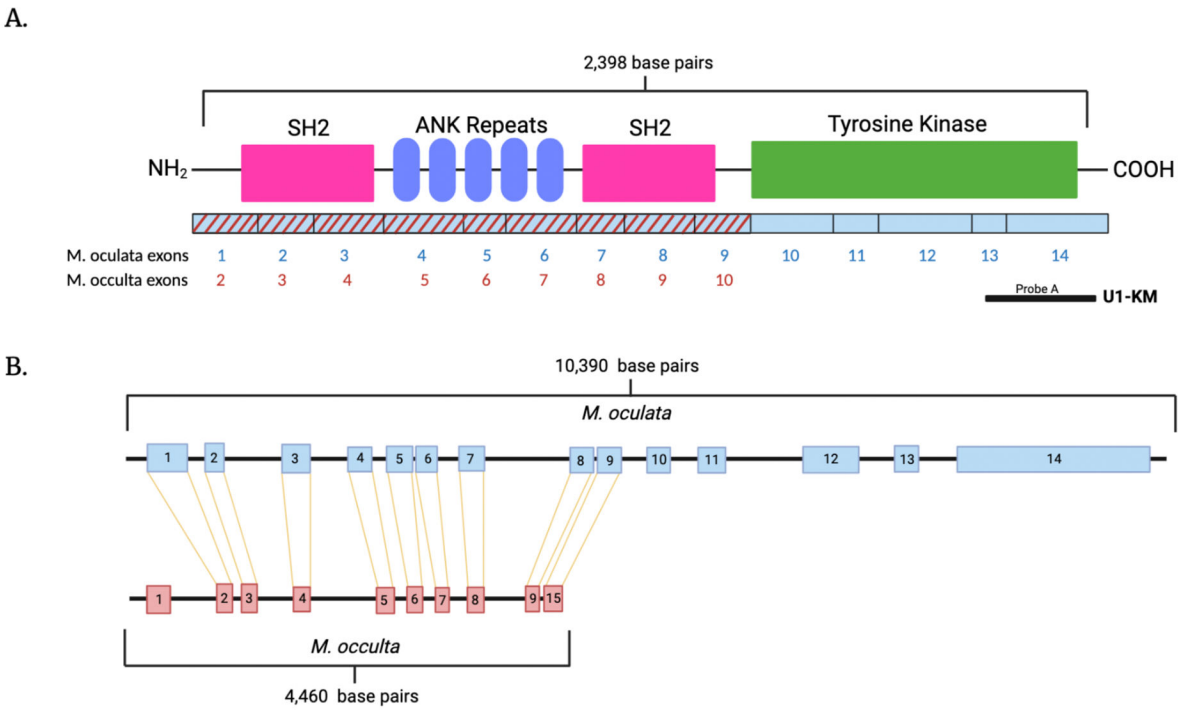

Figure 2.

A. A schematic diagram of the putative *M. oculata cymric* protein. The amino and carboxy termini are labeled, the locations of SH2 domains are in pink, ankyrin repeats are in blue, and the tyrosine kinase catalytic domain is in green. Below the protein diagram are the corresponding exons from *M. oculata* in blue and *M. occulta* in red that encode the cymric protein. Shown below is the position of cDNA clone (probe A) used for identifying the cymric amino acid sequence and as probes for *in situ* hybridizations. B. Intron and exon organization of *M. oculata* and *M. occulta SHARK tyrosine kinase* transcripts compared to the *M. oculata* and *M. occulta* genomes. The *M. occulta* genome is missing the last 5 exons transcribed from the *M. oculata* genome (exons 10–14) and has an extra exon at the 5' end of the transcript (not shown). The *M. occulta cymric* transcript is truncated missing the tyrosine kinase domain.

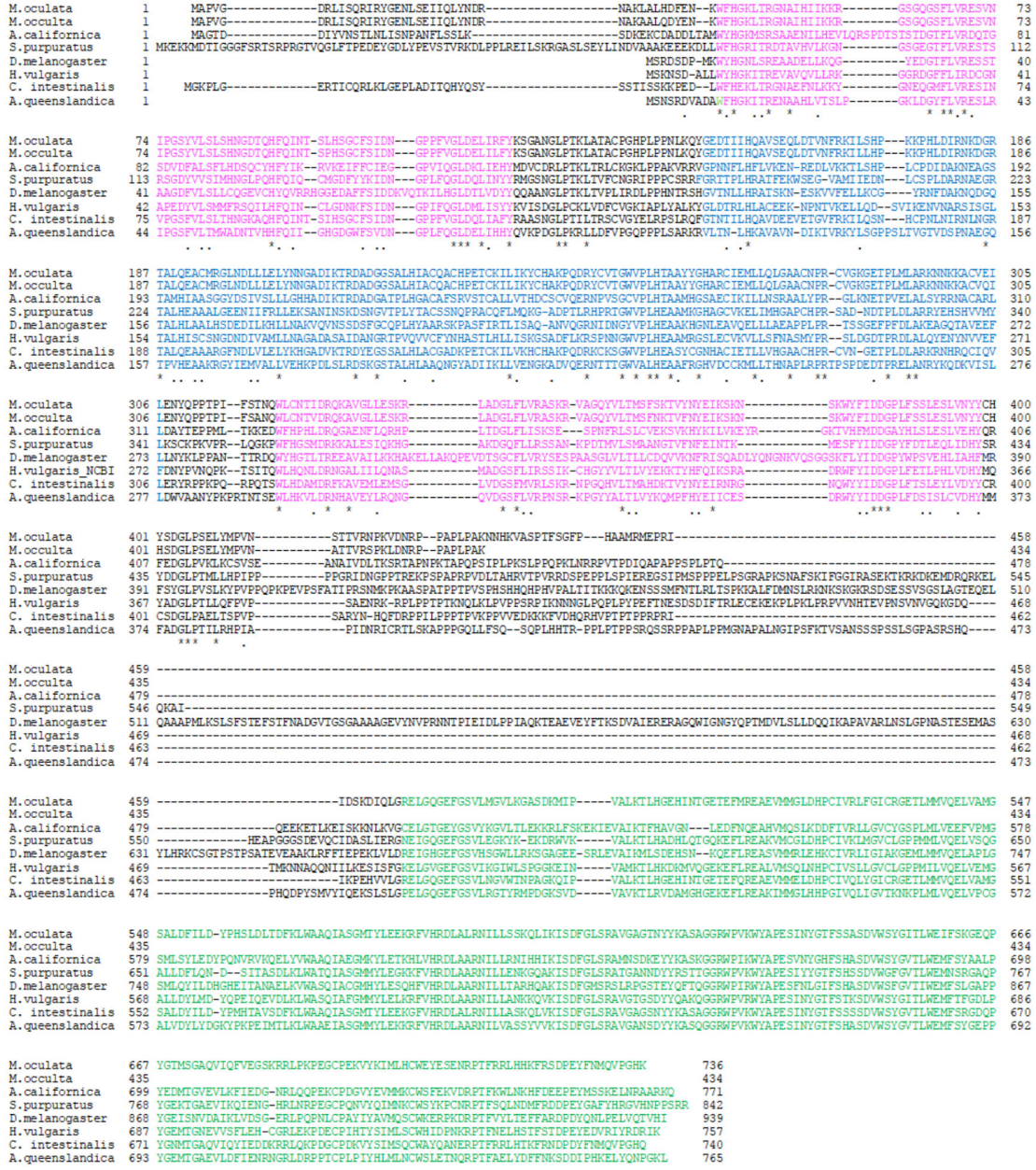

Figure 3.

Alignment of SHARK tyrosine kinases including ascidian (*M. oculata, M. occulta, M. occidentalis*, and *C. intestinalis* (XP_002121112.2)) cymric, *H. vulgaris* HTK16 (NP_00129668.1) (Chan et al., 1994), *A. queenslandica* (sponge) SHARK (BAA81720) (Suga et al. 1999), *D. melanogaster* SHARK (NP_524743.2) (Ferrante et al., 1995), *S. purpuratus* SHARK (XP_001180232.2), and *A. californica* SHARK (XP_005102068.1) tyrosine kinases. The amino acid positions are numbered at the beginning and end of each row. The gaps in the alignment are indicated by dashes. The SH2 domains are in pink, the ankyrin repeats are blue, and the tyrosine kinase domain is green. There is very high sequence similarity across metazoans, suggesting that *cymric* was present in the metazoan common ancestor and has been very conserved.

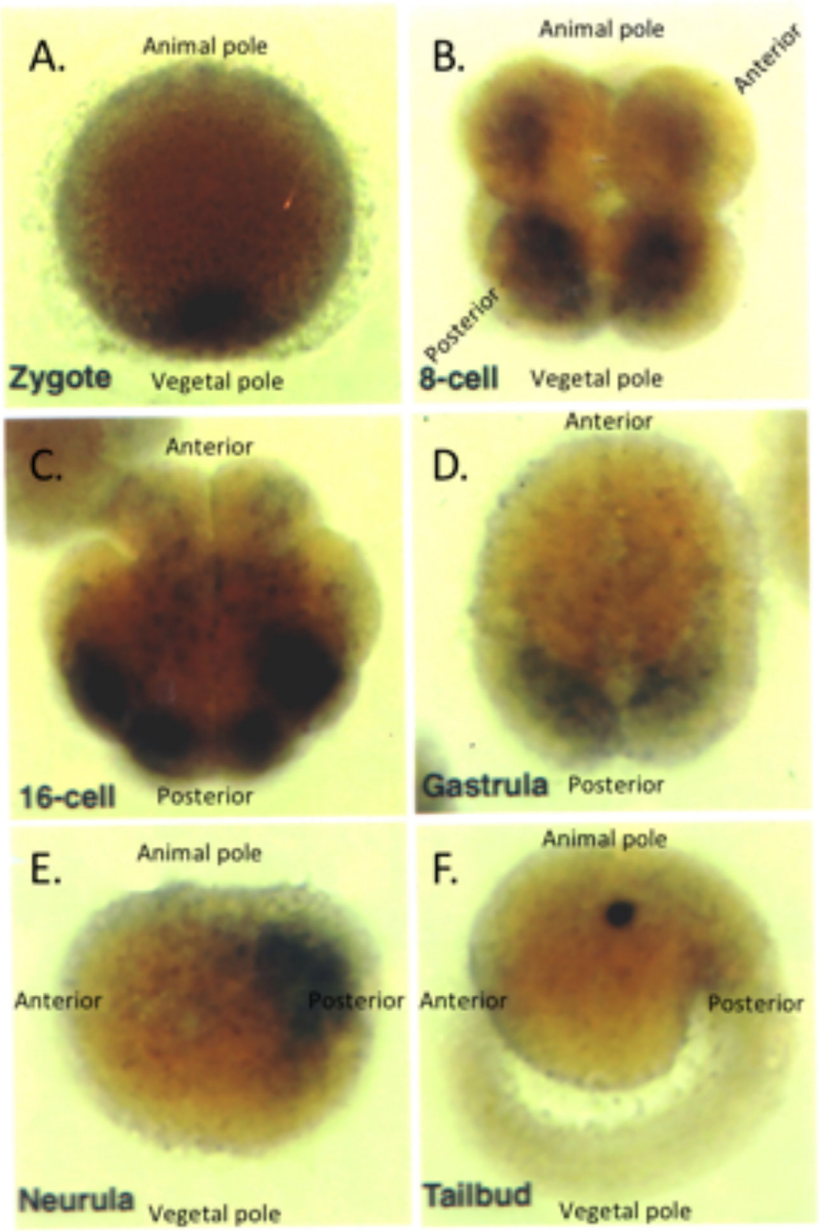

Figure 4.

Distribution of *cymric* transcripts throughout the tailed *Molgula oculata* development as determined by *in situ* hybridization. A. Oocyte at first ooplasmic segregation with animal pole up and vegetal pole down, the *cymric* transcript is expressed in the vegetal pole. B. 8-cell with animal pole up and vegetal pole down. *Cymric* expression accumulates in the blastomeres on the vegetal and posterior poles. C. 16-cell embryo in a vegetal view. *Cymric* transcripts are sequestered into the 4 tail muscle cells at the posterior of the embryo. D. A vegetal view of a gastrula showing *cymric* accumulation in all the larval tail muscle cells. E. Neurula with the animal pole up, the vegetal pole down, the anterior left and posterior right; *cymric* transcripts still remain expressed in the tail muscle cells. G. Tailbud stage with

animal pole up, vegetal down, anterior left, and posterior right. *Cymric* expression is no longer detectable in the larva.