Molgulid Ascidians Share a Unique Gene Complex

Peter X. Wu¹, Daniel W. Seufert², Billie J. Swalla¹,³*

¹ Box 351800; 24 Kincaid Hall; Biology Department, University of Washington, Seattle, WA 98195-1800
² Department of Zoology, Ohio Wesleyan University, Delaware, OH 43015
³ Station Biologique, BP° 74, 29682 Roscoff Cedex, France

Total number of text figures, graphs, and charts
Abbreviated title: Molgulid ascidians share a unique gene complex

Correspondence to:
Billie J. Swalla
Box 351800
24 Kincaid Hall
Biology Department
University of Washington
Seattle, WA 98195-1800

Phone: 206-616-9367
Fax: 206-616-2011
bjswalla@u.washington.edu

Supporting grant information: National Science Foundation Cooperative Agreement No. DBI-0939454 BEACON: Evolution in Action
ABSTRACT (250 words)

Typical chordate features found in ascidian tadpole larvae have been evolutionarily lost several times independently within the Molgulidae family. While tailed molgulids retain a tail with muscle cells, a notochord, and a dorsal neural tube, the notochord and muscle cells have been lost within the tail-less species. Of the ascidians, there are just two extant species with tail-less larvae other than the in the Molgulidae, which are found in the related Styelidae. A locus containing an unusual gene arrangement of the Bobcat gene within the first intron of the Manx gene has been shown to be essential for the development of chordate features in molgulid tadpole larvae. Sequencing and closer examination of ascidian genomes shows that there is a unique gene rearrangement of the NA-14 gene adjacent upstream to Manx and Bobcat within the Molgulidae also not found in styelid and cionid ascidians. Expression of these key genes could be affected by one another’s close proximity, disturbing normal larval development, specifically that of chordate features. We propose that the unique rearrangement that took place in the molgulid ancestor may be responsible for the numerous instances of the evolution of tail-lessness found in the Molgulidae.
INTRODUCTION

Recent phylogenomic data supports a chordate phylogeny that places the tunicates, or urochordates, as the sister group to the vertebrates (Bourlat and others, '06; Delsuc and others, '06; '08). Ascidians are a class of tunicates that are filter-feeding marine invertebrates and are sessile as adults. Most of the nearly 3,000 species of ascidians have a chordate tadpole larva, with a notochord, muscle cells, a dorsal neural tube, and sensory organs, an otolith and an ocellus (Swalla, '04). Molgulidae is a monophyletic family of ascidians that has evolved tailless larvae at least three times or more independently (Huber and others, '00). Few other extant ascidian species have evolved tail-less larva; only two tailless species are found in the Styelidae (Huber and others, '00). Therefore Molgulidae appear to have a propensity to independently evolve tailless species (Huber and others, '00) and we are interested in the molecular basis of this phenomenon. In previous studies, it has been shown that some tailed larval characters can be rescued in the tailless species Molgula occulta, by crossing with its tailed sister species Molgula oculata, suggesting loss-of-function mutation(s) as the cause for tailless development (Swalla and Jeffery, '90; Jeffery and Swalla, '92).

The Manx gene is a zinc-finger transcription factor vital for the development of chordate features in molgulid larva (Swalla and Jeffery, '96). Manx is present in the Ciona intestinalis genome, but it becomes increasingly difficult to establish homology outside of the tunicates. It was previously reported that in the tailed-larva species M. oculata, Bobcat, a p68 RNA helicase gene, is actually located within the first intron of Manx, with the two genes sharing two small exons (Swalla and others, '99). This gene complex was the first to be found in ascidians to show trans-splicing and also to be alternatively spliced to form multiple protein transcripts.
Subsequently, transplicing has been reported in both *Ciona* (Sierro and others, '09) and the distantly related *Oikopleura* (Ganot and others, '04) genome, so it is likely common in all tunicate genomes.

The expression of *Bobcat* in *M. oculata* and *M. occulta*, was found to be required for neural sensory organ, secondary muscle cell, and tail development (Swalla and others, '99). p68 (*DDX5*) and differentially spliced p72/p82 (*DDX17*) are paralogous DEAD-box RNA helicases found in vertebrates that are involved in the processing of mRNA and microRNA, both functions that were found to be essential for development in mice (Fukuda and others, '07).

The gene sequence of *Bobcat* in *M. oculata* and *M. occulta* is interesting because it lacks introns in its coding region (Swalla and others, '99), suggesting a previous retrotransposition event that inserted the processed RNA within the first intron of the *Manx* gene.

We have significantly expanded the mapping of this gene complex to include a third gene, *NA-14*, closely 5’ upstream of *Manx* and *Bobcat*. *NA-14*, also known as Sjogren syndrome nuclear antigen 1 (SSNA1), is a nuclear gene highly conserved within eukaryotes. This protein lacks a nuclear localization signal, but is thought that its small size of 14 kD would allow it to diffuse through nuclear pores, even though it remains in the nucleus (Ramos-Morales and others, '98). *NA-14* is localized to the centrosome and to sperm flagella (Pfannenschmid and others, '03) and has been shown to be involved in targeting ATPase *spastin* to microtubules in the centrosome (Errico and others, '04). Its close proximity upstream to *Bobcat* and *Manx* is a unique feature in the Molgulidae and could be affecting *NA-14, Bobcat*, and/or *Manx* expression. Here, we characterize this expanded gene complex in multiple molgulid species, and show that this unique rearrangement is likely to have occurred in the molgulid ancestor.
We hypothesize that this gene rearrangement contributes to the propensity for molgulid larvae to evolve a tailless phenotype.

**MATERIALS AND METHODS**

**PCR and Sequencing**

Initial PCR and sequencing was done by Dr. Dan Seufert at the Pennsylvania State University during his postdoctoral studies in the late 1990’s and were recently expanded by Peter X. Wu at the University of Washington. Table 1 summarizes the primer sequences used for PCR and sequencing different genes from different species. The *NA-14, Manx/Bobcat* complex was compared in *Molgula oculata, M. occulta, M. arenata, M. manhattensis, M. provisionalis, M. occidentalis, M. bleizi, M. pacifica, M. citrina, and Styela clava*, and *Ciona intestinalis* in order to examine which ascidian species share this unique gene arrangement.

**Gene Tree Construction**

Protein sequences of *NA-14, Bobcat*, and *Manx* homologues were found via blast searches on NCBI (http://www.ncbi.nlm.nih.gov/).

*NA-14*: Bf, __; Ci, __; Dr, __; Gg, __; Hs, __; Nv, __; Sk, __; Sp, __

*Bobcat*: Bf, XP_002599469.1; Ci, XP_002130131.1; Dr/DDX5, AAH67585.1; Dr/DDX17, XP_001923830.1; Gg/DDX5, NP_990158.1; Gg/DDX17, XP_416260.2; Hs/DDX5, NP_004387.1; Hs/DDX17, CAG30318.1; Nv, XP_001630124.1; Sk, XP_002731189.1; Sp, XP_780035.1. For *C. intestinalis* and *S. purpuratus* with multiple isoforms of the gene, the longest isoform was retrieved. *B. floridanae* has two hits of *p68*, the copy with 13 exons was chosen.
Phylogenetic programs were used for the construction of gene trees, using proteins obtained through database searches. The ClustalW alignment (Thompson and others, '02) found using the MacVector program (Rastogi, '00) was then run through MrBayes (Ronquist and Huelsenbeck, '03) to infer phylogeny and finally visualized using Treeview (Page, '02).

**NA-14 Expression**

This would be future work done.
RESULTS

NA-14/Manx/Bobcat Gene Complex

Figure 1 shows the expanded 10kb NA-14/Manx/Bobcat complex found in molgulid ascidians. Previously the 3’ part of this gene complex was reported in Swalla and others ‘99. The expanded molgulid complex shows a third gene, NA-14, less than 500bp 3’ to the trans-spliced 5’ UTR of Bobcat and the long Manx transcript. We found this arrangement in a number of molgulid ascidians and decided to test whether the expression of NA-14 is altered in tailless ascidians, since both Manx and Bobcat are silenced in the tailless M. occulta (Swalla and others, ’99).

We examined the relative arrangements of NA-14, Manx, and Bobcat’s homologues in other organisms to confirm the uniqueness of the gene complex in the Molgulidae. In humans, two homologues of Bobcat, p68 and p72/p82, can be found on chromosomes 17 and 22, respectively, and both genes contain introns that are not found in molgulid Bobcat. Manx could not be reliably detected in the human genome, but the homologous NA-14 gene is on chromosome 9, far apart from the Bobcat homologues. All the human genes can be said with certainty to be unlinked to each other in the genome. In the genome of the more related tunicate Ciona intestinalis, Manx and Bobcat homologues were found to be at separate loci instead of within a complex (Maglott and others, ’11), suggesting that these genes are also not linked in Ciona. However, C. intestinalis is distantly related to the molgulids, so here we report that the genes are not linked in Styela clava, a sister group to the molgulids (Zeng and others, ’06).
The genomic arrangement of Bobcat in chordates strongly suggests that it was retrotransposed into the manx locus in molgulid ascidians (Figure 2). Both human (Homo sapiens) DDX5 and tunicate (Ciona intestinalis) p68 have 13 exons that are spliced together to make a mature mRNA (Figure 2). In contrast, Bobcat in Molgula oculata, M. occulta, M. arenata, M. manhattensis, M. provisionalis, M. occidentalis, M. bleizi, M. pacifica, and M. citrina contains only two exons, as shown in green in Figure 2. The 3’ UTR is trans-spliced onto a single exon containing the entire p68 coding region. In spite of these dramatically different genomic organization, the coding region of p68 is highly conserved (Swalla and others, ’99).

DISCUSSION

The molgulid genome is highly compact compared to Ciona intestinalis, so it is not surprising to find that many genes have lost introns. Similar results have been reported for a distantly related tunicate, Oikopleura dioica, which has an even more compact genome (Seo and others, ’01). However, the fact that Bobcat completely lacks introns, but is flanked by two genes that have the same number of introns as found in C. intestinalis, very strongly suggests that it was originally retrotransposed from a Bobcat mRNA. The original Bobcat gene containing introns must have then been subsequently lost from the ancestral molgulid genome, because we have not been able to detect a second gene by PCR or southern blot in any of the ten molgulid genomes examined. Complete genome sequencing will be necessary to see if there is any evidence of the ancestral Bobcat gene in the extant molgulid genomes.
ACKNOWLEDGMENTS

We would like to thank Matt Sandel who worked on this project with Dr. Dan Seufert as an undergraduate at Pennsylvania State University. This material is based in part upon work supported by the National Science Foundation under Cooperative Agreement No. DBI-0939454. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.
LITERATURE CITED


### Table 1. Primers used for sequencing

<table>
<thead>
<tr>
<th>Species</th>
<th>Primer</th>
<th>DNA sequence</th>
<th>AA sequence</th>
<th>Target Region</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>S. clava</td>
<td>5’LIDF (328)</td>
<td>TTATTGACTTTCTGGAGTCAC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3’GDRHV (329)</td>
<td>CCDATICRKRTGNACRTARTCYTC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>U999</td>
<td>TGCCHGHHATTGTHCAYAT</td>
<td>LPAIVHI</td>
<td>Within p68</td>
<td>Seufert et al. 2000 JEZ</td>
</tr>
<tr>
<td></td>
<td>U937 (424)</td>
<td>GCGTGACTGGGCTTGAATG</td>
<td>RDVVLN region</td>
<td>Upstream of p68 intron</td>
<td>Iggo et al. 1991 MCB:11, p1326</td>
</tr>
<tr>
<td></td>
<td>D1043 (423)</td>
<td>TTATGACAAACTTGCATCGGAC</td>
<td>SDVKFI region</td>
<td>Downstream of p68 intron</td>
<td></td>
</tr>
<tr>
<td>M. ocullata</td>
<td>5-18MR684</td>
<td>CCGATTAGACTAAGTTCGCAG</td>
<td></td>
<td>Sequencing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5-18MF712</td>
<td>ATGGTATGGCATTCCAGG</td>
<td></td>
<td>Sequencing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M13F1307</td>
<td>CTTCAGAATGTGTGGGAG</td>
<td></td>
<td>Sequencing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NASMQEE</td>
<td>ATGCARGARGARGARAARCA</td>
<td>MQEE</td>
<td>Upstream in NA14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NA3IKMYA</td>
<td>ATYTTCATRTANGCTGAYTCNGTYTC</td>
<td>IKMYA</td>
<td>Downstream in NA14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NASTEKLLAKINE</td>
<td>ACIGARAARYTIGCNAARATHAAYGA</td>
<td>TEKLLAKINE</td>
<td>Upstream in NA14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P683Mocc</td>
<td>GGTCACTATATCCAGCATGTGA</td>
<td></td>
<td>Downstream in p68 exon 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NAS9ISQV</td>
<td>GATCTCYCARGTGCAGAGTGG</td>
<td>SISQV</td>
<td>Upstream in NA14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NAS5VESGK</td>
<td>GTGAAGATGAAAAACAAAAAGCAGA</td>
<td>VESGK</td>
<td>Upstream in NA14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P68EX1</td>
<td>TCAAGCGGSAACCAACTACC</td>
<td></td>
<td>In p68 exon 1</td>
<td></td>
</tr>
<tr>
<td>M. occulta</td>
<td>NAS5VESGK</td>
<td>GTCAAGATGGAAAAACAAAAAGCAGA</td>
<td>VESGK</td>
<td>Upstream in NA14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P68EX1</td>
<td>TCAAGCGGSAACCAACTACC</td>
<td></td>
<td>In p68 exon 1</td>
<td></td>
</tr>
<tr>
<td>M. pacifica</td>
<td>NAS5TEKLLAKINE</td>
<td>ACIGARAARYTIGCNAARATHAAYGA</td>
<td>TEKLLAKINE</td>
<td>Upstream in NA14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P683Mocc</td>
<td>GGTCACTATATCCAGCATGTGA</td>
<td></td>
<td>Downstream in p68 exon 1</td>
<td></td>
</tr>
<tr>
<td>M. citrina</td>
<td>NAS5VESGK</td>
<td>GTCGAAGATGGAAAAACAAAAAGCAGA</td>
<td>VESGK</td>
<td>Upstream in NA14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Prp683 (388)</td>
<td>GATTG*MTCATAAATGCGG</td>
<td></td>
<td>In p68</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P68EX1</td>
<td>TCAAGCGGSAACCAACTACC</td>
<td></td>
<td>In p68 exon 1</td>
<td></td>
</tr>
</tbody>
</table>
**Figure Legends**

Fig. 1  *NA-14/Bobcat/Manx* Gene Complex found in Molgulid Ascidians. *NA-14*, shown in orange, is upstream and contains three exons. The first exon of both *Bobcat* and the long minor *Manx* transcript share a 5’ UTR, which is indicated as an open green box. Minor *Manx* also shares its second exon in its 5’ UTR with the 3’ UTR of the *Bobcat* transcript. Noncoding UTRs are depicted as open boxes, coding exons as solid boxes. *NA-14* is orange, *Manx* is red, *Bobcat* is light blue, *Bobcat* and *Manx* shared exons are green.

Fig. 2  *Bobcat* gene structure and its homologues. Shown are the gene structures of *Bobcat* and its homologues in human, *Ciona intestinalis*, and *Styela clava*. The homologues in human, *C. intestinalis*, and *S. clava* are seen with 13 coding exons with introns separating them, contrasting with the long single coding exon found in *Bobcat* in *M. oculata*. Predicted retro-transposition would be responsible for the lack of introns in *Bobcat* in *M. oculata*. Noncoding UTRS are depicted as open boxes, coding exons as solid boxes.

Fig. 3  *NA-14* gene structure. Shown are the gene structures of *NA-14* in human, *C. intestinalis*, *S. clava*, and *M. oculata*. 

