Molgulid Ascidians Share a Unique Gene Complex

Peter X. Wu[1], Daniel W. Seufert[2], Billie J. Swalla[1,3]*

[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. floridae* 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 intenstinalis*, 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

Bourlat SJ, Juliusdottir T, Lowe CJ, Freeman R, Aronowicz J, Kirschner M, Lander ES, Thorndyke M, Nakano H, Kohn AB, Heyland A, Moroz LL, Copley RR, Telford MJ. 2006. Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. Nature 444(7115):85-88.

Delsuc F, Brinkmann H, Chourrout D, Philippe H. 2006. Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature 439(7079):965-968.

Delsuc F, Tsagkogeorga G, Lartillot N, Philippe H. 2008. Additional molecular support for the new chordate phylogeny. Genesis 46(11):592-604.

Errico A, Claudiani P, D'Addio M, Rugarli EI. 2004. Spastin interacts with the centrosomal protein NA14, and is enriched in the spindle pole, the midbody and the distal axon. Hum Mol Genet 13(18):2121-2132.

Flicek P, Amode MR, Barrell D, Beal K, Brent S, Chen Y, Clapham P, Coates G, Fairley S, Fitzgerald S, Gordon L, Hendrix M, Hourlier T, Johnson N, Kähäri A, Keefe D, Keenan S, Kinsella R, Kokocinski F, Kulesha E, Larsson P, Longden I, McLaren W, Overduin B, Pritchard B, Riat HS, Rios D, Ritchie GR, Ruffier M, Schuster M, Sobral D, Spudich G, Tang YA, Trevanion S, Vandrovcova J, Vilella AJ, White S, Wilder SP, Zadissa A, Zamora J, Aken BL, Birney E, Cunningham F, Dunham I, Durbin R, Fernández-Suarez XM, Herrero J, Hubbard TJ, Parker A, Proctor G, Vogel J, Searle SM. 2011. Ensembl 2011. Nucleic Acids Res 39(Database issue):D800-806.

Fukuda T, Yamagata K, Fujiyama S, Matsumoto T, Koshida I, Yoshimura K, Mihara M, Naitou M, Endoh H, Nakamura T, Akimoto C, Yamamoto Y, Katagiri T, Foulds C, Takezawa S, Kitagawa H, Takeyama K, O'Malley BW, Kato S. 2007. DEAD-box RNA helicase subunits of the Drosha complex are required for processing of rRNA and a subset of microRNAs. Nat Cell Biol 9(5):604-611.

Ganot P, Kallesøe T, Reinhardt R, Chourrout D, Thompson EM. 2004. Spliced-leader RNA trans splicing in a chordate, Oikopleura dioica, with a compact genome. Mol Cell Biol 24(17):7795-7805.

Huber JL, da Silva KB, Bates WR, Swalla BJ. 2000. The evolution of anural larvae in molgulid ascidians. Semin Cell Dev Biol 11(6):419-426.

Jeffery WR, Swalla BJ. 1992. Factors necessary for restoring an evolutionary change in an anural ascidian embryo. Dev Biol 153(2):194-205.

Maglott D, Ostell J, Pruitt KD, Tatusova T. 2011. Entrez Gene: gene-centered information at NCBI. Nucleic Acids Res 39(Database issue):D52-57.

Page RD. 2002. Visualizing phylogenetic trees using TreeView. Curr Protoc Bioinformatics Chapter 6:Unit 6.2.

Pfannenschmid F, Wimmer VC, Rios RM, Geimer S, Kröckel U, Leiherer A, Haller K, Nemcová Y, Mages W. 2003. Chlamydomonas DIP13 and human NA14: a new class of proteins associated with microtubule structures is involved in cell division. J Cell Sci 116(Pt 8):1449-1462.

Ramos-Morales F, Infante C, Fedriani C, Bornens M, Rios RM. 1998. NA14 is a novel nuclear autoantigen with a coiled-coil domain. J Biol Chem 273(3):1634-1639.

Rastogi PA. 2000. MacVector. Integrated sequence analysis for the Macintosh. Methods Mol Biol 132:47-69.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19(12):1572-1574.

Seo HC, Kube M, Edvardsen RB, Jensen MF, Beck A, Spriet E, Gorsky G, Thompson EM, Lehrach H, Reinhardt R, Chourrout D. 2001. Miniature genome in the marine chordate Oikopleura dioica. Science 294(5551):2506.

Sierro N, Li S, Suzuki Y, Yamashita R, Nakai K. 2009. Spatial and temporal preferences for trans-splicing in Ciona intestinalis revealed by EST-based gene expression analysis. Gene 430(1-2):44-49.

Swalla BJ. 2004. Procurement and culture of ascidian embryos. Methods Cell Biol 74:115-141.

Swalla BJ, Jeffery WR. 1990. Interspecific hybridization between an anural and urodele ascidian: differential expression of urodele features suggests multiple mechanisms control anural development. Dev Biol 142(2):319-334.

Swalla BJ, Jeffery WR. 1996. Requirement of the Manx gene for expression of chordate features in a tailless ascidian larva. Science 274(5290):1205-1208.

Swalla BJ, Just MA, Pederson EL, 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 126(8):1643-1653.

Thompson JD, Gibson TJ, Higgins DG. 2002. Multiple sequence alignment using ClustalW and ClustalX. Curr Protoc Bioinformatics Chapter 2:Unit 2.3.

Zeng L, Jacobs MW, Swalla BJ. 2006. Coloniality has evolved once in Stolidobranch Asicidians. Integrative and Comparative Biology 46(3):255-268.

TABLES

Table 1. Primers used for sequencing

Species	Primer	DNA sequence	AA sequence	Target Region	References
S. clava	5'LIDF (328)	TTATTGACTTTCTGGAGTCAC			
	3'GDRHV (329)	CCDATICKRTGNACRTARTCYTC		3' of p68	
	U999	TGCCHGGHATTGTHCAYAT	LPAIVHI	Within p68	Seufert et al. 2000 JEZ
	U937 (424)	GCGTGACTGGGTCTTGAATG	RDWVLN region	Upstream of p68 intron	Iggo et al. 1991 MCB:11, p1326
	D1043 (423)	TTATGACAAACTTGACATCGGAC	SDVKFVI region	Downstream of p68 intron	
M. oculata	5-18MR684	CCGATTAGACTAAGTTCGCAG		Sequencing	
	5-18MF712	ATGGTATGGCATTCCAGG		Sequencing	
	M13F1307	CTTTCAGAATGATGTGCGAG		Sequencing	
	NA5MQEE	ATGCARGARGARGARGARAARCA	MQEE	Upstream in NA14	
	NA3IKMYA	ATYTTCATRTANGCTGAYTCNGTYTC	IKMYA	Downstream in NA14	
	NA5TEKLLAKINE	ACIGARAARYTIGCNAARATHAAYGA	TEKLLAKINE	Upstream in NA14	
	P683Mocc	GGTCACTATATCCAGGCATTGTA		Downstream in p68 exon 1	
	NA5SISQV	GYATYTCYCARGTCGAAAGTGG	SISQV	Upstream in NA14	
	NA5VESGK	GTCGAAAGTGGAAAACAAAAAGCAGA	VESGK	Upstream in NA14	
	P68EX1	TCAAGCGGSAACCAACTACC		In p68 exon 1	
M. occulta	NA5VESGK	GTCGAAAGTGGAAAACAAAAAGCAGA	VESGK	Upstream in NA14	
	P68EX1	TCAAGCGGSAACCAACTACC		In p68 exon 1	
M. pacifica	NA5TEKLLAKINE	ACIGARAARYTIGCNAARATHAAYGA	TEKLLAKINE	Upstream in NA14	
	P683Mocc	GGTCACTATATCCAGGCATTGTA		Downstream in p68 exon 1	
M. citrina	NA5VESGK	GTCGAAAGTGGAAAACAAAAAGCAGA	VESGK	Upstream in NA14	
	Prp683 (388)	GATTGG*GTCATAAARATGGCG		In p68	
	P68EX1	TCAAGCGGSAACCAACTACC		In p68 exon 1	

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*.

FIGURES

Fig. 1

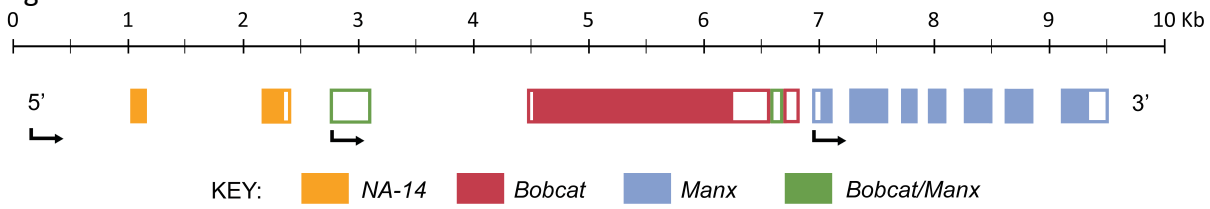

Fig. 2

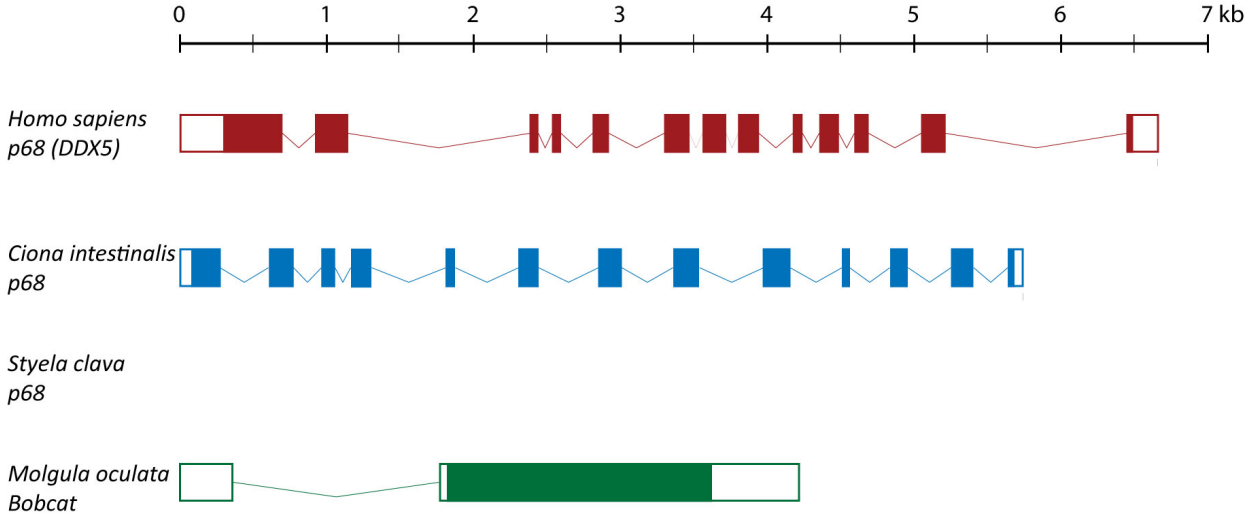

Fig. 3