Phylogenomic resolution of major tunicate relationships

Kevin M. Kocot[1], Michael G. Tassia[2], Kenneth M. Halanych[2], and Billie J. Swalla[3]

[1]Department of Biological Sciences and Alabama Museum of Natural History, University of Alabama. [2]Department of Biological Sciences, Auburn University. [3]Department of Biology and Friday Harbor Laboratories, University of Washington.

INTRODUCTION

Tunicata is a diverse clade of approximately 3,000 described species of marine, filter-feeding chordates. Tunicates are of great interest because 1) they are the closest living relatives of vertebrates, 2) there are numerous invasive species with economic consequences, and 3) some are edible. Despite their diversity and importance, relationships among major lineages of Tunicata are not completely resolved.

Traditional tunicate taxonomy is based on suites of gross morphological and life history characters (**Table 1**). Three classes are generally recognized: Ascidiacea (sea squirts), Thaliacea (pelagic salps, doliolids, and pyrosomes), and Appendicularia (larvaceans).

Despite recent studies (e.g., Swalla et al., 2000; Turon and López-Legentil, 2004; Tsagkogeorga et al., 2009; & Govindarajan et al., 2011) that have greatly advanced understanding of relationships within clades, many questions about tunicate higher-level phylogeny remain unanswered. Here, we supplemented data with transcriptomes that span the diversity of Tunicata and re-evaluated the higher-level evolutionary history using a phylogenomic approach.

Taxon	Benthic/ Pelagic	Solitary/ Colonial	Branchial sac	Gonad position
Appendicularia	Pelagic	Solitary	None	Dorsal
Phlebobranchia	Benthic	Usually solitary	Vascular	Enterogona
Aplousobranchia	Benthic	Colonial	Simple	Enterogona
Stolidobranchia	Benthic	Either	Folded	Pleurogona
Thaliacea	Pelagic	Either	Simple	Enterogona

Table 1. Gross morphology and life history features of major tunicate clades.

METHODS

We sampled all tunicate orders (with the exception of Doliolida, which was previously shown to be nested within the otherwise well-sampled taxon Thaliacea; Govindarajan et al., 2011) for genome or transcriptome data. RNA was extracted and purified using standard approaches and Illumina TruSeq RNA v2 libraries were prepared and sequenced using 2 X 100 bp paired-end sequencing on the Illumina HiSeq 2500 platform.

Dataset processing followed Kocot et al. (2017). Briefly, transcriptomes were assembled with Trinity 2.2.0 with the --trimmomatic and --normalize_reads flags and transcripts were translated with TransDecoder. For orthology inference, we used HaMStR 13 with the "model organisms" core-ortholog set. Redundant sequences that were identical were removed with UniqHaplo, leaving only unique sequences for each taxon. Each gene was aligned with MAFFT 7.273 and alignments were trimmed with Aliscore and Alicut to remove ambiguously aligned regions. Genes sampled for <14 of the 28 taxa after these steps were discarded. PhyloTreePruner was used to screen for paralogs and contamination. Concatenation of remaining sequences was performed using FASconCAT-G. To control for systematic artifacts, we calculated relative composition frequency variability (RCFV) and branch-length heterogeneity score (LB) for each gene and assembled data matrices corresponding to the best 50, 100, 200, and 500 genes according to RCFV and LB.

Maximum likelihood (ML) analyses were conducted in RAxML 8.2.8 using the "-f a" run modMatrices were partitioned by gene and the PROTGAMMAAUTO model was used for all partitions. Nodal support was assessed with rapid bootstrapping with the number of replicates determined by the autoMRE criterion. Internode certainty was calculated based on the complete dataset of all 798 genes.

Bayesian inference (BI) analyses were conducted in Phylobayes MPI 1.6j using the site-heterogeneous CAT+GTR+Γ4 model to account for site-specific rate heterogeneity (-cat -gtr -dgam 4). Because of the computationally intensive nature of analyses using this model, BI was only conducted for the RCFV_50 (i.e., best 50 genes according to RCFV) and LB_50 data sets. Phylobayes bpcomp maxdiff values (0.0039 for RCFV_50 and 0.0697 for LB_50) were used to assess convergence of chains. Only results form converged runs are reported.

RESULTS & DISCUSSION

We analyzed datasets of up to 798 genes sampled from 28 taxa (matrix occupancy = 77.43%; **Figure 1**). All analyses recovered Olfactores (Tunicata + Vertebrata) and Tunicata with maximal support. Within Tunicata, Appendicularia was recovered sister to the rest of Tunicata with maximal support. Consistent with some earlier studies (e.g., Swalla et al. 2000), all of our analyses recovered Thaliacea within Ascidiacea. Stolidobranchia was sister to a clade in which Thaliacea was sister to Phlebobranchia plus Aplousobranchia.

In most analyses, phlebobranch tunicates were recovered paraphyletic with respect to Aplousobranchia. Support for this varied but was strong in some cases. However, when only the 50 best genes based on compositional heterogeneity were analysed, Phlebobranchia and Aplousobranchia were recovered reciprocally monophyletic with strong support, consistent with traditional hypotheses. Although internode certainty values for the complete dataset were generally low, the node nesting Aplousobranchia within Phlebobranchia received zero support (data not shown). Relationships within major clades were consistent with earlier studies.

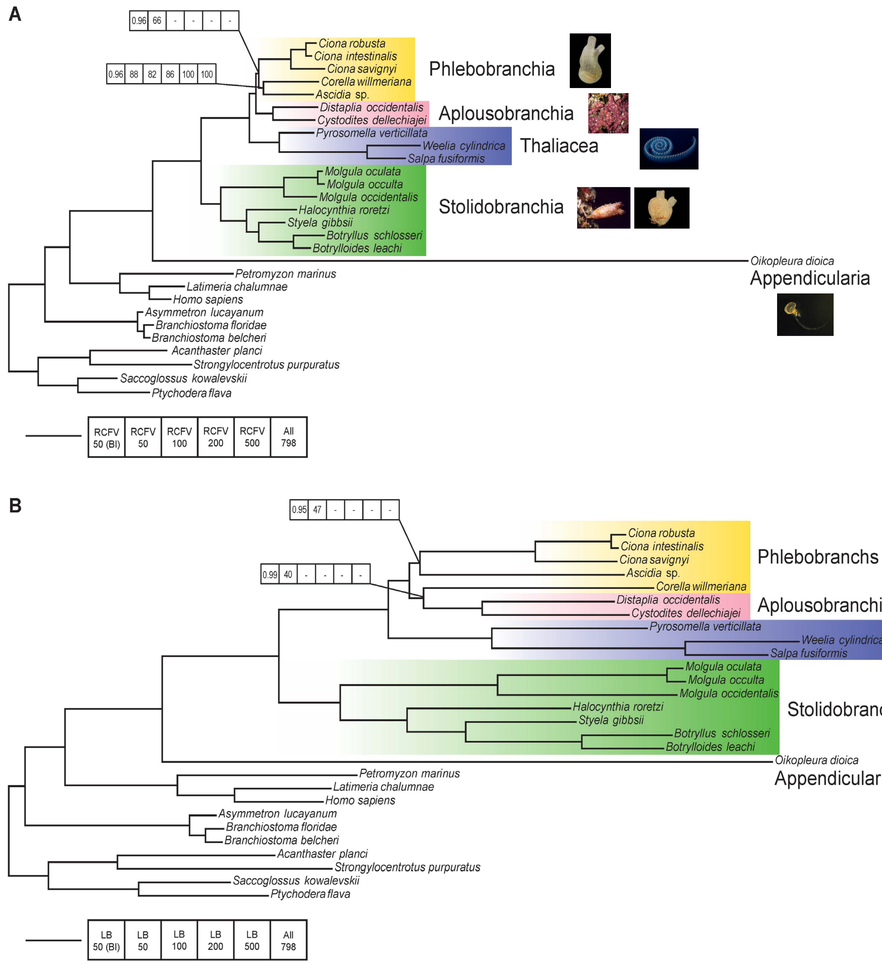

Figure 1. Phylogeny of Tunicata. **A.** Consensus phylogram from the Bayesian inference analysis of RCFV_50 with bootstrap support values from ML analyses of RCFV_50, RCFV_100, RCFV_200, RCFV_500, and the complete dataset shown. **B.** Consensus phylogram from the Bayesian inference analysis of LB_50 with bootstrap support values from ML analyses of LB_50, LB_100, LB_200, LB_500, and the complete dataset shown. Nodes without support matrices received maximal support in all BI and ML analyses. Scale bars represent 0.1 substitutions per site.

CONCLUSIONS

This study represents a first step toward resolving tunicate phylogeny using genomic data. Our results indicate that this approach holds promise to resolve tunicate evolutionary history at deep but also more shallow levels, which have also been challenging to address, particularly within Aplousobranchia.

PREPRINT

A pre-print describing this work in more detail is available online at bioRxiv.

THANKS!

This study was supported by UA start-up funds to KMK. We thank Alison Sweeney, James Townsend, Friday Harbor Labs, and the crew of the *R/V Nathaniel B. Palmer* for facilitating specimen collection. We thank Justin Cox and Emily Pabst for help with RNA extractions. We thank Deb Crocker and Robert Griffin for assistance with the University of Alabama High Performance Computing (UAHPC) system.

REFERENCES

Govindarajan, A.F., Bucklin, A., Madin, L.P., 2011. A molecular phylogeny of the Thaliacea. J. Plankton Res. 33, 843–853.

Kocot, K.M., Struck, T.H., Merkel, J., Waits, D.S., Todt, C., Brannock, P.M., Weese, D.A., Cannon, J.T., Moroz, L.L., Lieb, B., Halanych, K.M., 2017. Phylogenomics of lophotrochozoa with consideration of systematic error. Syst. Biol. 66, 256–282.

Swalla, B.J., Cameron, C.B., Corley, L.S., Garey, J.R., 2000. Urochordates are monophyletic within deuterostomes. Syst. Biol. 49, 52–64.Turon and López-Legentil, 2004;

Tsagkogeorga, G., Turon, X., Hopcroft, R.R., Tilak, M.-K., Feldstein, T., Shenkar, N., Loya, Y., Huchon, D., Douzery, E.J., Delsuc, F., 2009. An updated 18S rRNA phylogeny of tunicates based on mixture and secondary structure models. BMC Evol. Biol. 9, 187.

Turon, X., López-Legentil, S., 2004. Ascidian molecular phylogeny inferred from mtDNA data with emphasis on the Aplousobranchiata. Mol. Phylogenet. Evol. 33, 309–320.

THE UNIVERSITY OF ALABAMA®