Siphonophore Phylogeny

Stefan Siebert[1,2], Felipe Zapata[1,3], Mark Howison[4], Cat Munro[1], Freya Goetz[1,5], Phil Pugh, Steven H.D. Haddock[4], Casey W. Dunn[1]*

[1] Department of Ecology and Evolutionary Biology, Brown University, Providence, RI, USA

[2] Current address: Department of Molecular & Cellular Biology, University of California at Davis, Davis, CA, USA

[3] Current address: Department of Ecology and Evolutionary Biology, University of California Los Angeles, Los Angeles, CA, USA

[4] Brown Data Science Practice, Brown University, Brown University, Providence, RI, USA

[5] Current address: Smithsonian, Washington DC, USA

[4] Monterey Bay Aquarium Research Institute, Moss Landing, CA, USA

* Corresponding author, casey_dunn@brown.edu

Abstract

Introduction

Siphonophores are…

Methods

This manuscript is an executable document computed directly from the data, providing an explicit and reproducible description of all findings. All scripts for the analyses are available in a git repository at https://github.com/caseywdunn/siphonophore_phylogeny_2017. The most recent commit at the time of the analysis presented here was a5eaa3d6c6614d766451f1702a3b67a396169ff9.

Collecting

Collection data on all examined specimens, a description of the tissue that was sampled from the colony, collection mode, sample processing details, mRNA extraction methods, sequencing library preparation methods and sequencing details are summarized in supplementary table 1. Monterey Bay and Gulf of California specimens were collected by remotely operated underwater vehicle (ROV) or during blue-water scuba dives. *Chelophyes appendiculata* and *Hippopodius hippopus* specimen were collected in the bay of Villefranche-sur-Mer, France, during a plankton trawl on 04/13/11. Available physical vouchers have been deposited at the Museum of Comparative Zoology (Harvard University), Cambridge, MA, and at the United States National Museum (Smithsonian Institution), Washington, DC. Accession numbers are given in supplementary table X. In cases where physical vouchers were unavailable we provide photographs to document species identity (table x).

mRNA isolation and RNAseq library preparation

When possible specimens were starved overnight in filtered seawater at temperatures close to ambient water temperatures at the time point of specimen collection (supplementary table 1). mRNA was extracted directly from tissue using Magnetic mRNA Isolation Kit (NEB, #S1550S), Invitrogen Dynabeads mRNA Direct Kit (Ambion, #61011) or from total RNA after Trizol (Ambion, #15596026) extraction and through purification using Dynabeads mRNA Purification Kit (Ambion, #61006). Extractions were performed according to the manufacturer's instruction. In case of very small total RNA quantities, only a single round of bead purification was performed (supplementary table 1). Resulting higher rRNA read counts were dealt with

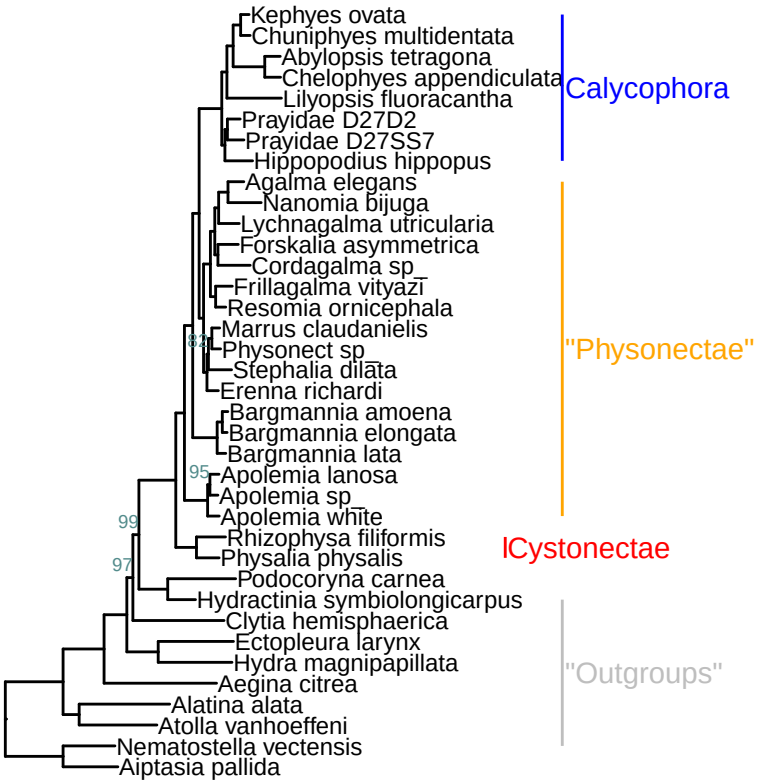

Figure 1: Phylogram of siphonophore relationships. Node labels indicate bootstrap support percent, unnumbered nodes have 100% support. The image was rendered with ggtree (Yu et al. 2016)

further downstream in the bioinformatics workflow. Libraries were prepped for sequencing using the Illumina TruSeq RNA Sample Prep Kit (Illumina, #FC-122-1001, #FC-122-1002), the Illumina TruSeq Stranded Library Prep Kit (Illumina, #RS-122-2101) or the NEBNext RNA Sample Prep. Master Mix Set (NEB, #E6110S). We collected long read paired end Illumina data for *de novo* transcriptome assembly. Summary statistics for expression libaries are given in Table 1.

Analysis

Sequence assembly, annotation, phylogenetic analysis, and derivation of expression counts were conducted with the tool Agalma (Dunn et al. 2013), v0.XXX. Subsequent analyses were conducted in R and integrated into this manuscript with the `knitr` package. See Supplementary Information for R package version numbers.

Results and Discussion

Sample collecting and sequencing

Table 1 (at the end of file): Summary statistics for libraries.

Species phylogeny

The phylogenetic relationships between siphonophores species presented here is based on XXX genes from XXX siphonophore species, including new data presented here for XXX species. These findings are entirely consistent with a previous analysis based on two genes (16S and 18S ribosomal RNA) (Dunn et al. 2005). Cystonectae is the sister group to the remaining siphonophores, and Calycophorae is nested within the

paraphyletic “Physonectae”. In addition, multiple nodes that were not resolved in the previous two-gene analysis do recieve strong support in this XXX-gene transcriptome analysis. These findings include XXX.

Conclusion

Acknowledgements

This work was supported by the National Science Foundation (DEB-1256695 and the Waterman Award). Sequencing at the Brown Genomics Core facility was supported in part by NIH P30RR031153 and NSF EPSCoR EPS-1004057. Data transfer was supported by NSF RII-C2 EPS-1005789. Analyses were conducted with computational resources and services at the Center for Computation and Visualization at Brown University, supported in part by the NSF EPSCoR EPS-1004057 and the State of Rhode Island. We also thank the MBARI crews and ROV pilots for collection of the specimens.

Supplementary Information

Supplementary Analyses

Software versions

This manuscript was computed on Sat Feb 11 08:48:32 2017 with the following R package versions.

```
R version 3.3.2 (2016-10-31)
Platform: x86_64-apple-darwin13.4.0 (64-bit)
Running under: macOS Sierra 10.12.3

locale:
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

attached base packages:
[1] grid      parallel  stats     graphics  grDevices utils     datasets
[8] methods   base

other attached packages:
 [1] phytools_0.5-64    picante_1.6-2      nlme_3.1-128
 [4] vegan_2.4-1        lattice_0.20-34    permute_0.9-4
 [7] agalmar_0.0.0.9000 ape_4.0            hutan_0.0.0.9000
[10] edgeR_3.14.0       limma_3.28.21      RGraphics_2.0-14
[13] gridExtra_2.2.1    seriation_1.2-1    fields_8.10
[16] maps_3.1.1         spam_1.4-0         GGally_1.3.0
[19] ggtree_1.7.7       treeio_0.99.9      cowplot_0.7.0
[22] xtable_1.8-2       jsonlite_1.2       knitr_1.15.1
[25] digest_0.6.11      magrittr_1.5       forcats_0.1.1
[28] stringr_1.1.0      dplyr_0.5.0        purrr_0.2.2
[31] readr_1.0.0        tidyr_0.6.1        tibble_1.2
[34] ggplot2_2.2.1      tidyverse_1.0.0    devtools_1.12.0

loaded via a namespace (and not attached):
 [1] colorspace_1.3-2          class_7.3-14
 [3] modeltools_0.2-21         mclust_5.2.1
 [5] rprojroot_1.1             htmlTable_1.8
 [7] XVector_0.12.1            GenomicRanges_1.24.3
 [9] base64enc_0.1-3           flexmix_2.3-13
```

```
[11] AnnotationDbi_1.34.4       mvtnorm_1.0-5
[13] codetools_0.2-15           splines_3.3.2
[15] mnormt_1.5-5               robustbase_0.92-7
[17] geneplotter_1.50.0         Formula_1.2-1
[19] annotate_1.50.1            cluster_2.0.5
[21] kernlab_0.9-25             backports_1.0.4
[23] assertthat_0.1             Matrix_1.2-7.1
[25] lazyeval_0.2.0             acepack_1.4.1
[27] htmltools_0.3.5            tools_3.3.2
[29] igraph_1.0.1               coda_0.19-1
[31] gtable_0.2.0               clusterGeneration_1.3.4
[33] fastmatch_1.0-4            Rcpp_0.12.9.1
[35] msm_1.6.4                  Biobase_2.32.0
[37] trimcluster_0.1-2          gdata_2.17.0
[39] iterators_1.0.8            fpc_2.1-10
[41] phangorn_2.1.1             gtools_3.5.0
[43] XML_3.98-1.5               dendextend_1.3.0
[45] DEoptimR_1.0-8             zlibbioc_1.18.0
[47] MASS_7.3-45                scales_0.4.1
[49] TSP_1.1-4                  SummarizedExperiment_1.2.3
[51] expm_0.999-0               animation_2.4
[53] RColorBrewer_1.1-2         yaml_2.1.14
[55] memoise_1.0.0              rpart_4.1-10
[57] reshape_0.8.6              latticeExtra_0.6-28
[59] stringi_1.1.2              RSQLite_1.1-1
[61] genefilter_1.54.2          gclus_1.3.1
[63] S4Vectors_0.10.3           plotrix_3.6-4
[65] foreach_1.4.3              checkmate_1.8.2
[67] caTools_1.17.1             BiocGenerics_0.18.0
[69] BiocParallel_1.6.6         GenomeInfoDb_1.8.7
[71] prabclus_2.2-6             bitops_1.0-6
[73] evaluate_0.10              plyr_1.8.4
[75] DESeq2_1.12.4              R6_2.2.0
[77] IRanges_2.6.1              gplots_3.0.1
[79] Hmisc_4.0-2                combinat_0.0-8
[81] DBI_0.5-1                  whisker_0.3-2
[83] foreign_0.8-67             withr_1.0.2
[85] mgcv_1.8-16                scatterplot3d_0.3-38
[87] survival_2.40-1            RCurl_1.95-4.8
[89] nnet_7.3-12                KernSmooth_2.23-15
[91] rmarkdown_1.3              locfit_1.5-9.1
[93] data.table_1.10.0          diptest_0.75-7
[95] numDeriv_2016.8-1          stats4_3.3.2
[97] munsell_0.4.3              registry_0.3
[99] quadprog_1.5-5
```

References

Dunn C., Pugh P., Haddock S. 2005. Molecular Phylogenetics of the Siphonophora (Cnidaria), with Implications for the Evolution of Functional Specialization. Systematic biology. 54:916–935.

Dunn C.W., Howison M., Zapata F. 2013. Agalma: an automated phylogenomics workflow. BMC Bioinfor-

matics. 14:330.

Yu G., Smith D.K., Zhu H., Guan Y., Lam T.T.-Y. 2016. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution.