Evolution, morphology, and gene expression of functionally specialized zooids in Siphonophora

Catriona Munro

A dissertation submitted to the Graduate School of Brown University.

In partial fulfillment of the requirements for the degree of Doctor of Philosophy

Providence, RI

May 2019

Copyright

This dissertation by Catriona Munro is accepted in its present form by the Department of Ecology and Evolutionary Biology as satisfying the dissertation requirement for the degree of Doctor of Philosophy.

_____   _____

Casey W. Dunn, Advisor   Date

Recommended to the Graduate Council

_____   _____

Erika Edwards, Reader   Date

_____   _____

David Rand, Reader   Date

_____   _____

James Kellner, Reader   Date

Approved by the Graduate Council

_____   _____

Andrew G. Campbell, Dean of the Graduate School   Date

Curriculum vitae

Catriona Munro

Education

2013–2019	PhD Ecology and Evolutionary Biology, Brown University
	Advisor: Casey W. Dunn
2012–2013	MRes Ocean Science, University of Southampton
	Advisors: Sven Thatje, Chris Hauton
2008–2011	BSc Biology, First Class Honours, University College London

Peer-Reviewed Publications

Munro C, Siebert S, Zapata F, Howison M, Damian-Serrano A, Church SH, Goetz FE, Pugh PR, Haddock SHD, Dunn CW (2018) Improved phylogenetic resolution within Siphonophora (Cnidaria) with implications for trait evolution. Molecular Phylogenetics and Evolution. 127: 823-833. doi: 10.1016/j.ympev.2018.06.030, bioRxiv preprint: doi:10.1101/251116

Brown, A, Thatje S, Oliphant A, **Munro C**, Smith KE (2018) Temperature effects on larval development in the lithodid crab *Lithodes maja*. Journal of Sea Research. 139:73-84. doi: 10.1016/j.seares.2018.06.009

Dunn CW, Zapata F, **Munro C**, Siebert S, Hejnol A (2018). Pairwise comparisons are problematic when analyzing functional genomic data across species. Proceedings of the National Academy of Sciences. 115(3):E409-E417 doi:10.1073/pnas.1707515115. bioRxiv preprint: doi:10.1101/107177

Dunn CW, **Munro C** (2016) Comparative genomics and the diversity of life. Zoologica Scripta. 45:5-13. doi:10.1111/zsc.12211

Munro C, Morris JP, Brown A, Hauton C, Thatje S (2015). The role of ontogeny in physiological tolerance: decreasing hydrostatic pressure tolerance with development in the northern stone crab *Lithodes maja*. Proceedings of the Royal Society of London B. 282(1809): 20150577 doi: 10.1098/rspb.2015.0577

Shank TM, Baker ET, Embley RW, Hammond S, Holden JF, White S, Walker SL, Calderon M, Herrera S, Lin TJ, **Munro C**, Heyl T, Stewart L, Malik M, Lobecker M, Potter J (2012) GALREX 2011: Exploration of the Deep-Water Galapagos Region. Oceanography 25 Suppl.: 50-51. URL: http://tos.org/oceanography/archive/25-1_supplement.pdf

Academic Awards and Scholarships

June 2017– May 2018	NSF Doctoral Dissertation Improvement Grant (award no. 1701272) $21,028
May–June 2017	EMBRC-France funding (main PI: Casey Dunn)
April–May 2016	EMBRC-France funding (main PI: Casey Dunn)
March 2016	MBL Embryology Post-Course research $1484
2014–2015	RI-EPSCoR Graduate Student Fellowship (full tuition and stipend)
May 2014	Travel Award, Evolution of the First Nervous Systems II Meeting
2012–2013	Society for Underwater Technology, Educational Support Fund (full Masters' degree tuition)
2011	Dean's List, UCL
	Lilian Clarke Prize, UCL
2008–2011	Harold and Olga Fox Scholarship, UCL
2009	Darwin Prize, UCL

Research Positions

2012	**Research Assistant** Shank Lab, Woods Hole Oceanographic Institution
Summer 2011–end 2011	**Summer Student Fellow, then Guest Student** Shank Lab, Woods Hole Oceanographic Institution
Summer 2010	**Volunteer** Shank Lab, Woods Hole Oceanographic Institution

Conference Presentations

†Oral, ‡Poster, <u>presenting author</u>

†**<u>Munro C</u>**, S Siebert, F Zapata , CW Dunn (2018) Siphonophore Differential Gene Expression Patterns Analyzed within a Phylogenetic Context. SICB Meeting, San Francisco CA

†‡**<u>Munro C</u>**, S Siebert, M Howison , F Zapata , CW Dunn (2016) Gene expression patterns in siphonophore zooids, Hydroidfest, Bodega CA

‡**<u>Munro C</u>**, S Siebert, M Howison , F Zapata , CW Dunn (2016) Exploring the evolution of functional specialization in siphonophores using RNAseq, SICB Meeting, Portland OR

†**<u>Munro C</u>**, GW Luther III, RA Lutz, C Vetriani, TS Moore, S Herrera, TM Shank (2012) Temporal and spatial patterns of in situ community structure using time-lapse camera systems at a vent field on the East Pacific Rise, 13th International Deep-Sea Biology Symposium, Wellington New Zealand

Teaching

Graduate Teaching Assistant Brown University
Invertebrate Zoology (Undergraduate level): Fall semester, 2014
Invertebrate Zoology (Undergraduate level): Fall semester, 2013
Guest Lecture, University of Rhode Island, BIO460 Pelagic Ecology, 6 August 2014

Mentoring

Fall 2015 - Nicola Malakooti, undergraduate

Scientific Expeditions

R/V Western Flyer, ROV Doc Ricketts, Gulf of California, Mexico, March 8-16, 2015. Chief scientist: Steven Haddock

R/V Western Flyer, ROV Doc Ricketts, Monterey Bay, U.S.A, September 16-22, 2014. Chief scientist: Steven Haddock

R/V Endeavor, North East Atlantic, U.S.A, August 13-18, 2014. Chief scientist: Brad Seibel

R/V Western Flyer, ROV Doc Ricketts, Monterey Bay, U.S.A, May 17-23, 2014. Chief scientist: Steven Haddock

R/V Western Flyer, ROV Doc Ricketts, Monterey Bay, U.S.A, November 19-24, 2013. Chief scientist: Steven Haddock

R/V Falkor, ROV Global Explorer MK3, Deep-Sea Coral Shakedown cruise, Gulf of Mexico U.S.A, August 26 - September 6, 2012. Chief scientist: Peter Etnoyer

M/V Holiday Chouest, ROV UHD 34, HC3 Leg 1 and 2. October 2-26, 2011. Gulf of Mexico, U.S.A. Chief scientists: Charles Fisher and Erik Cordes

Science Outreach

Volunteering

Feb 2016 – Feb 2018, Hennessy Elementary School, Brown Junior Researchers Program (After-school science class, 5th grade students)

Science Writing

Munro, C. (2015) A day in the life of a siphonophore lab. http://thenode.biologists.com/a-day-in-the-life-of-a-siphonophore-lab/lablife/

Munro, C., S. Herrera, T. Heyl Muric (2011) Exploring the Paramount Seamounts. http://oceanexplorer.noaa.gov/okeanos/explorations/ex1103/logs/july14/july14.html

Munro, C. (2010) Volunteer Gets an Oceanful of Experience. Oceanus. http://www.whoi.edu/oceanus/viewArticle.do?id=84128

Courses and Workshops

Jun 7- Jul 18 2015	MBL Embryology, Woods Hole MA

Dedication

To my parents – for seeing a 9 year old who wanted to identify all the animals on the reef, and pointing this curiosity and passion towards science.

And to Larry – for your love, support, and much needed sense of perspective: reminding me of my achievements when I am dwelling my failures; and reminding me in my moments of extreme focus that there are other important things in life.

Acknowledgements

This work would not be possible without the support of so many people.

My family has always been a constant source of support and love. I'm grateful to my parents for moving so much, and for exposing the three of us to new people, cultures, and countries. We were always free to explore and make mistakes. My parents encouraged and respected my strong sense of intellectual independence. I always let them know about my achievements, challenges, and failures on my own terms – and when I did, they were always there to support my dreams and soothe my fears. My brothers grew up to become accomplished, kind, and generous men, and I love seeing what wonderful fathers and husbands they have become; Tilly and Danielle are both amazing sisters-in-law. To me, my brothers will always be the kids cooking up some scheme to access the top floors of the Liang Ma so that we can race remote controlled cars, often to their own doom down a set of stairs. I followed my older brothers everywhere, even when it wasn't the best idea – Lachlan, always the leader, and Iain, ever questioning authority. I can always count on them to bring me down to earth.

Larry has been with me throughout the journey of this PhD. He dropped me off outside Walter Hall for my in-person interview with Casey, and picked me up at the end of the day, full of hope about how it went. His love and support has ranged from apples secretly stashed in coat pockets ("just in case you get hungry") to agreeing to leave his comfort zone and family, and move to a new country for my next position as a postdoc. I am so lucky and grateful. We adopted Odin during the second year of my PhD, and he made our lives richer. He helped too, mostly by insisting that I take a break and go for a run or walk in the woods. I also want to thank Larry's family and friends for fully embracing me as one of their own and cheering me on throughout this process.

There are so many friends that have been there for me, good times and bad, too many to name them all. Henry Stanley and Charlotte Ballard have been there for me since UCL, and always make time to see me (even when "close by" simply means the same continent). Sarah Ye, my oldest friend, who I don't see nearly enough, but who I have been lucky to see more of over the past few years. My grad student cohort in EEB (David Sleboda, David Boerma, Robbie Lamb, Kim Neil, Brooke Osborne, Priya Nakka, Adam Spierer, and Kara Pellowe), have all grown as scientists alongside me and continue to inspire me. In addition to those in my cohort, KC Cushman, Lindsay McCulloch, Nikole Bonacorsi, Jillian Oliver, and Morgan Moeglein, are all brilliant women scientists and good friends. The EEB astronauts, led by Adam, got me out of bed at 6 am on Tuesdays and Thursdays for exercise, and more importantly, good company. Book club provided a literary diet of modern, diverse authors and also all of the best gossip. I've been lucky to be in such a friendly, supportive environment for five years.

My lab and office mates have helped me so much. Rebecca Helm paved the way as the first Dunn lab grad student, and continues to provide support and encouragement. She inspires me to be a better writer. August Guang patiently answered so many of my math/CS questions, and I am grateful to them for their expertise and insight. Stefan Siebert always pushed me to be a better, more thorough scientist - I've missed having his dry sense of humour around the lab. Felipe Zapata is a passionate scientist, and even after leaving the lab and starting his own, he continues to provide prompt and helpful feedback on my drafts and ideas.

Abby Moore always listened to my stories and ideas without any judgement, and I've missed sharing an office with her. Morgan, who moved to Yale with the Dunn/Edwards labs, gets a second mention because I'm so grateful for how many times she and Matt put me up on their sofa/camp bed in New Haven. Alex Damian Serrano is a gelatinous zooplankton enthusiast who I can always count on to know the siphonophore literature inside out. Zack Lewis, who provided great company on so many drives to and from New Haven, gave valuable feedback on chapter drafts, and is a role model to me for being such a meticulous scientist.

I'd also like to thank everyone who made my specimen- and field- work possible. Particularly Steve Haddock – without his generosity and invitations to come out on the Western Flyer, the collection of most of the samples for this dissertation would not have been possible. His great eye for design made the figures of chapter 2 beautiful. I am very grateful for all the help from Lynne Christianson before, during, and after cruises – she is amazing. The crew of the Western Flyer and the ROV pilots made the collection of deep sea siphonophores possible. Sam Church, a former Brown undergrad and co-author, assisted me on a particularly productive cruise. Stefan and Casey came to sea with me on separate occasions and taught me so much about working with siphonophores. Although none of the work conducted in Villefranche-sur-mer will make it into this dissertation, I am grateful to Lucas Leclère for inviting me to work with siphonophores there. All of the fieldwork collecting Portuguese man of war in Texas was made possible by the help and generosity of Richard Behringer and Maria Pia Miglietta. I'd also like to thank citizen scientists and jellywatchers Lea, Millie, Dignan, and Noah for alerting me, via Jellywatch (http://jellywatch.org/), to the presence of large numbers of Portuguese man o' war along the coast in 2016, and for their help collecting specimens. Last but not least, Phil Pugh provided invaluable taxonomic expertise and insight into siphonophore biology. I am particularly grateful for his thorough feedback on drafts, especially the *Physalia physalis* chapter.

I'd like to thank my committee, Erika Edwards, David Rand, and Jim Kellner, for providing their time and expertise and for all of their critical feedback. I am very grateful to David Rand for adopting me as his graduate student for the last year of my PhD. I'd also like to thank all the faculty and staff of Brown EEB for providing such a supportive community. Support staff in EEB, especially Shannon Silva, Lianne Mendonca, Jesse Marsh, and Irena Nedeljkovic Cunningham, who I interacted with most directly, were always so responsive, efficient, and capable, in addition to being wonderful people. Also everyone in BioMed facilities (Lenny, Gerry, Gilberto, to name but a few), who kept the building and the stockroom running, and always went above and beyond.

And of course, none of this would be possible without my advisor, Casey Dunn. His unwavering optimism and can-do attitude continues to inspire and surprise me. He would often take what I considered to be a colossal failure and re-frame it as a learning experience. As a mentor, he has guided me through the process of becoming a scientist with positivity and encouragement – his criticism is always constructive *and* kind. I have really enjoyed working together, and hope we continue to do so as colleagues.

Thank you.

Preface

"At 320 feet a lovely colony of siphonophores drifted past. At this level they appeared like spun glass. Others which I saw at far greater and blacker depths were illumined, but whether by their own or by reflected light I cannot say...Here in their own haunts they swept slowly along like an inverted spray of lilies-of-the-valley, alive and in constant motion. In our nets we find only the half-broken swimming bells, like cracked, crystal chalices, with all the wonderful loops and tendrils and animal flowers completely lost or contracted into a mass of tangled threads."

— William Beebe, *Half Mile Down*

Siphonophores have long been a subject of fascination, particularly in the 19th century, when voyages of scientific exploration brought planktonic siphonophores to the attention of zoologists and naturalists. Researchers at this time were fascinated by the paradoxical nature of these species. Siphonophores are colonial hydrozoans (Cnidaria) that are comprised of many "individuals" (zooids) that are either polyps or medusae, but that are completely physiologically integrated such that the colony acts as an individual. With the exception of researchers based in marine laboratories in Nice, Villefranche-sur-mer, and Messina, where upwellings brought some deep-sea species close to the surface, most researchers primarily gained insight into the morphology, colony organization, and systematics of Siphonophora from net-caught specimens. With the development of submersible technology in the 20th century, it soon became possible to observe and even collect intact deep-sea organisms, including siphonophores, in their natural habitat. As William Beebe describes from his pioneering dives in the Bathysphere in the 1930's, scientists were now able to learn so much more about the ecology, behavior, and morphology of these fragile siphonophore species.

In the 19th century, naturalists including Huxley, Haeckel, and Chun were fascinated by the functional specialization of zooids and studied comparative zooid morphology with an eye towards understanding division

of labor in siphonophores and understanding its relevance to division of labor in other biological systems. Several 20th century researchers built on these foundations and used descriptions of zooid morphology to make major contributions to siphonophore systematics and taxonomy (particularly A. K. Totton & P. R. Pugh), physiology and histology (such as G. O. Mackie), and development (e.g. C. & D. Carré). The 20th century also saw the rise of oceanography as a science, and with that, an increased focus on examining the distribution and ecology of siphonophores. For my graduate research, I return to questions about the evolution of functional specialization in siphonophores, and add a new character to these analyses of zooid diversity – gene expression.

Siphonophore colonies typically consist of a series of zooids, arranged linearly on a stem, that are each functionally specialized for a different task, including feeding, reproducing, swimming, and digesting. The whole colony originates from a single fertilized egg, which develops to give rise to a primary polyp (or protozooid) and growth zones, that in turn asexually produce each of the functionally specialized zooids within the colony (Carré & Carré, 1991, 1993). The first descriptions of the colony level development and order of appearance of these buds in different siphonophore species were given by Dunn & Wagner (2006). Within Codonophora, the clade that contains the greatest diversity of siphonophores, all of the zooids originate from a single bud within the growth zone that subdivides to give rise to all other zooids (Dunn *et al.*, 2005; Dunn & Wagner, 2006). We now know that the growth zone is the main source of interstitial stem cells (i-cells) and are a site of high cell proliferation (Siebert *et al.*, 2015). While the potency of these i-cells is unknown, cells differentiate and mature as buds are carried away from the growth zone by the elongating stem, and i-cell populations become restricted to a few sites within mature siphonophore zooids (Siebert *et al.*, 2015).

Functionally specialized zooids differ from one another not only in function, but also in form, and their cellular composition differs significantly among each of the different zooids (Mackie 1960; Carré, 1969;Church *et al.*, 2015). The diversity of functionally specialized zooids generated within a single siphonophore colony is greater than in any other colonial animal (Beklemishev, 1969), and yet, this diversity is generated from a single genome. 43 years ago, King and Wilson (1975) first described the very high genetic similarity between chimpanzees and humans, and proposed that the significant biological differences between these two

species may instead be driven by regulatory mutations and differences in gene expression. This observation is analogous to observations of high diversity among zooids within a single genetically-identical colony, as well as among zooids between closely related species. With the advent of high-throughput RNA sequencing technology, we are now in a position to investigate differences in gene expression across thousands of genes, and to apply these technologies to wild-caught species that cannot be grown in the laboratory. We are also able to use these technologies to investigate amino acid divergence, and to build phylogenies of rare and difficult to collect taxa, such as siphonophores. The first molecular phylogeny of Siphonophora was built by Dunn *et al.* (2005) using two ribosomal RNA genes, but many open questions remained about the deep relationships within Codonophora.

In my first chapter, I provide a description of the morphology and development of the Portuguese man of war, *Physalia physalis*. *Physalia physalis* was the first siphonophore species to be described, by Linnaeus in 1758, and it remains the best known of all siphonophore species. While it is the best known, it is also quite unlike any other siphonophore species, particularly with regards to its habitat, growth, and colony organization. In this chapter, I build on the extensive foundations set by Totton (1960) and Mackie (1960), and provide new photographic and 3D-microscopical images to describe colony growth, development, and organization. I also provide insights into the evolutionary origin of a functionally specialized zooid that is unique to *P. physalis.* This chapter provided an opportunity to advance our understanding of the alpha taxonomy, systematics, and morphology of this highly unusual pleustonic siphonophore.

My second chapter is a new molecular phylogeny of Siphonophora, which includes transcriptome data from 33 siphonophore species (29 were newly sequenced) and 10 outgroups. This new phylogeny, built using 1,423 genes, finds strong support for many of the key relationships identified by Dunn *et al.* (2005), but also resolves many deep relationships within the siphonophore phylogeny. This chapter also includes phylogenetic reconstructions of several traits that are central to siphonophore biology. The products of this chapter, including a species phylogeny and also thousands of gene trees, provide a key foundation for the subsequent two chapters.

In my third chapter, I investigate gene expression patterns across different zooids in seven siphonophore species. In this descriptive paper, I use gene expression patterns to build an understanding of the "molecular

anatomy" of these zooids and one specialized tissue, using expression patterns instead of histological methods to understand the molecular function and structure of these zooids. Gene expression patterns are used to provide an additional layer to existing anatomical descriptions, and are used to understand which genes show zooid-specific expression within species and how many of these genes are conserved across species. I then use this information to investigate the function of novel zooid types within particular species.

In my fourth and final chapter, I use phylogenetic methods to compare expression patterns across species. In this chapter, I adapt existing phylogenetic approaches in order to compare gene expression values of particular zooids and one specialized tissue across species. This approach moves beyond strict ortholog approaches, and considers complex evolutionary histories of speciation and duplication, enabling the mapping of expression values directly onto gene phylogenies with speciation and duplication events labelled at the nodes of the trees. By isolating key branches in gene trees that correspond to branches in the species tree, I was able to investigate evolutionary changes in expression. I not only learned about the global distribution of these changes across branches, but was able to investigate specific scenarios of change in particular zooids across specific gene tree branches. While the focus of this chapter is on siphonophores, the methods and results should be of broad interest to functional genomicists.

One of the exciting things about working with siphonophores is that there are still so many open questions about their biology. As a siphonophore researcher, I am not only accessing the most current literature on gene expression evolution and functional genomics (largely focused on Bilateria), as well as cnidarian and bilaterian development, but also delving back into the 19th and 20th century literature on siphonophore biology. Through each of these four chapters, I have been able to address aspects of old unanswered questions about siphonophore systematics, zooid identity, functional specialization, and evolution.

Works cited:

Beebe, W., 1951. *Half Mile Down.* 2nd Edition. Duell, Sloan and Pearce, New York.

Beklemishev, W.N., 1969. *Principles of Comparative Anatomy of Invertebrates. Volume I. Promorphology.* Oliver & Boyd, Edinburgh.

Carré, C., Carré, D., 1991. A complete life cycle of the calycophoran siphonophore Muggiaea kochi (Will) in the laboratory, under different temperature conditions: ecological implications. *Philosophical Transactions of the Royal Society B: Biological Sciences* 334, 27–32. https://doi.org/10.1098/rstb.1991.0095

Carré, C., Carré, D., 1993. Ordre des siphonophores, in: Grassé, P.-P. (Ed.), *Traité de Zoologie. Anatomie, Systématique, Biologie.* Paris:Masson, pp. 523–596.

Carré, D. 1969. Etude histologique du developpement de Nanomia bijuga (Chiaje, 1841), Siphonophore Physonecte, Agalmidae. *Cahiers de Biologie Marine*, 10, 325–341.

Church, S.H., Siebert, S., Bhattacharyya, P., Dunn, C.W., 2015. The histology of Nanomia bijuga (Hydrozoa: Siphonophora). *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* 324, 435–449. https://doi.org/10.1002/jez.b.22629

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. https://doi.org/10.1080/10635150500354837

Dunn, C.W., Wagner, G.P., 2006. The evolution of colony-level development in the Siphonophora (Cnidaria: Hydrozoa). *Dev. Genes Evol.* 216, 743–754. https://doi.org/10.1007/s00427-006-0101-8

King, M.C., Wilson, A.C. 1975. Evolution at two levels in humans and chimpanzees. *Science.* 188, 107–116

Mackie, G. O., 1960. Studies on Physalia physalis (L.). Part 2. Behavior and histology. *Discovery Reports,*30, 371–407.

Siebert, S., Goetz, F. E., Church, S. H., Bhattacharyya, P., Zapata, F., Haddock, S. H. D., Dunn, C. W., 2015. Stem cells in Nanomia bijuga (Siphonophora), a colonial animal with localized growth zones. *EvoDevo.* 6, 22. https://doi.org/10.1186/s13227-015-0018-2

Totton, A. K., 1960. Studies on Physalia physalis (L.). Part 1. Natural history and morphology. *Discovery Reports*, 30, 301–368.

Contents

List of Tables

List of Figures

Chapter 1

Morphology and development of the Portuguese man of war, *Physalia physalis*

Catriona Munro[1,‡], Zer Vue[2,3], Richard R. Behringer[2], Casey W. Dunn[4]

[1] Department of Ecology and Evolutionary Biology, Brown University, Providence, RI 02912, USA

[2] Department of Genetics, University of Texas MD Anderson Cancer Center, Houston, TX 77030 USA

[3] Graduate Program in Developmental Biology, Baylor College of Medicine, Houston, TX 77030, USA

[4] Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520, USA

‡ I conceived of this study along with CWD, conducted all sample collection and fixation, analysed and segmented images, and wrote the manuscript. ZV conducted the OPT imaging and along with RRB helped with sample collection.

1.1 Abstract

The Portuguese man of war, *Physalia physalis*, is a siphonophore that uses a gas-filled float as a sail to catch the wind. It is one of the most conspicuous, but poorly understood members of the pleuston, a community of organisms that occupy a habitat at the sea-air interface. The development, morphology, and colony organization of *P. physalis* is very different from all other siphonophores, in part because of adaptations to this unique lifestyle. Here, we propose a framework for homologizing the axes with other siphonophores, and also suggest that the tentacle bearing zooids should be called tentacular palpons. We also look at live and fixed larval and non-reproductively mature juvenile specimens, and use optical projection tomography to build on existing knowledge about the morphology and development of this species. Previous descriptions of *P. physalis* larvae, especially descriptions of budding order, were often framed with the mature colony in mind. However, we use the simpler organization of larvae and the juvenile specimens to inform our understanding of the morphology, budding order, and colony organization in the mature specimen. Finally, we review what is known about the ecology and lifecyle of *P. physalis.*

1.2 Introduction

The pleuston is the floating community of ocean organisms that live at the interface between water and air. This community is exposed to a unique set of environmental conditions including prolonged exposure to intense ultraviolet light, desiccation risk, and rough sea and wave conditions (Zaitsev, 1997). Despite their tolerance for extreme environmental conditions and the very large size of this habitat, which makes up 71% of the Earth's surface and is nearly three times the area of all terrestrial habitats, very little is known about the organisms that make up this highly specialized polyphyletic community. One of the most conspicuous, yet poorly understood, members of the pleuston is the siphonophore *Physalia physalis*, commonly known as the Portuguese man of war. The Portuguese man of war is aptly named after a warship: it uses part of an enlarged float filled with carbon monoxide and air as a sail to travel by wind for thousands of miles, dragging behind long tentacles that deliver a deadly venomous sting to fish (Clark and Lane, 1961; Iosilevskii and Weihs, 2009). This sailing ability, combined with a painful venomous sting and a life cycle with seasonal

blooms, results in periodic mass beach strandings and occasional human envenomations, making *P. physalis* the most infamous siphonophore (Prieto et al., 2015).

Siphonophores are a relatively understudied group of colonial hydrozoans. Most species are planktonic, and are found at most depths from the deep sea to the surface of the ocean (Mapstone, 2014; Munro et al., 2018; Pugh, 1984). They are fragile and difficult to collect intact, and must be collected by submersible, remotely operated vehicle, by hand while blue-water diving, or in regions with localized upwellings (Dunn et al., 2005b; Mackie et al., 1987). However, *Physalia physalis* is the most accessible, conspicuous, and robust siphonophore, and as such, much has been written about this species, including the chemical composition of its float, venom (especially envenomations), occurrence, and distribution (e.g. Araya et al., 2016; Copeland, 1968; Herring, 1971; Lane, 1960b; Larimer and Ashby, 1962; Prieto et al., 2015; Totton and Mackie, 1956; Wilson, 1947; Wittenberg, 1960; Wittenberg et al., 1962; Woodcock, 1956; Yanagihara et al., 2002). Fewer studies, however, have taken a detailed look at *P. physalis* structure, including development, histology of major zooids, and broader descriptions of colony arrangement (Bardi and Marques, 2007; Mackie, 1960; Okada, 1932, 1935; Steche, 1910; Totton, 1960). These studies provide an important foundation for understanding the morphology, cellular anatomy, and development of this pleustonic species. The morphology, growth, and development of *P. physalis* is difficult to understand, because they are so different from all other siphonophores it can be difficult to understand their morphology within the context of siphonophore diversity, and also because the colony consists of highly 3-dimensional branching structures.

Here, we combine what is already known about morphology and development with new microscopical techniques, including the use of optical projection tomography, and recent phylogenetic and histological knowledge from related siphonophore species, to add new perspectives on the morphology and development of *P. physalis*. As the colony organization is so distinct from other siphonophores, an important first step is to homologize the anatomical axes in developing and mature specimens with other siphonophores. It is then possible to describe the order, pattern, and directionality of budding, and place this within a broader phylogenetic context. There are also still open questions about the homology and origin of some of the unique zooids in *P. physalis*, including the gastrozooid and the tentacle bearing zooid (tentacular palpon). Additionally, understanding the complex 3D structure of *P. physalis* from written text and hand drawn diagrams

can be challenging for a reader that has not spent many hours looking at specimens under a microscope – 3D images and videos can help clarify the complex morphology and arrangement. Finally, we also review what is known about the ecology and lifecycle of this pleustonic species.

1.3 Methods

1.3.1 Collecting and fixing

Juvenile specimens, defined as colonies with float length 8-10cm and immature developing gonodendra, were collected from locations along the exposed Gulf coast of Galveston Island, TX in March 2016 and February 2017, from East Beach (Lat. Lon. 29.328090, -94.737542) to east of Galveston Island State park (Lat. Lon. 29.195358, -94.948335). Information on when to collect large numbers of *P. physalis* was obtained from sightings submitted to the citizen science website Jelly Watch (www.jellywatch.org). Juvenile specimens were collected fresh from the surf, and transferred directly to the lab for examination and fixation in 4% formalin in seawater after relaxation. Physical vouchers are deposited at the Peabody Museum of Natural History (Yale University), New Haven, CT. Developing specimens were obtained from the collections of Philip R. Pugh, and are now deposited at the Peabody Museum of Natural History (Yale University). These specimens were collected in various locations in the Atlantic Ocean during research expeditions in 1972, 1973 and 1983. Two additional containers of developing specimens were kindly provided by Dr. Pugh, however no collection information is available. Details of the collected specimens are provided in table 1.1.

1.3.2 Image capture and processing

Optical Projection Tomography (OPT) was used as a tool to collect serial images for three-dimensional reconstruction of fixed *P. physalis* tissue. Before imaging, formalin fixed specimens were washed 2x quickly in cold PBS, 3x PBS (5 min) and placed in 1:1000 DAPI and PBS overnight. Following DAPI staining, specimens were dehydrated into methanol (25% MeOH/H$_2$O, 50% MeOH/H$_2$O, 75% MeOH/H$_2$O for 4hs each step, and 100% MeOH overnight). Specimens were optically cleared using BABB, which is one part

Table 1.1: Collection information for larval developing Portuguese man of war specimens used in this study

Date of collection	Latitude, Longitude (Decimal Degrees)	Depth (m)	Expedition information
18 February 1972	22.80555556, -22.55277778	0	RRS Discovery 7800
22 February 1972	17.93888889, -24.88611111	0	RRS Discovery 7803 #13
1 March 1973	32.00277778, -34.38055556	0	RRS Discovery 8270
1 August 1983	13.31861111, -56.01861111	NA	BWP 1093-15 PPP106

benzyl alcohol and two parts benzyl benzoate (wash first in MeOH/BABB (1:1) 1h, then BABB overnight), before rehydration into PBS and embedding in agarose. Specimens were embedded in 2.5% Ultra Low Melting Point Agarose (Invitrogen, CA, 16520-050) within a syringe. Agarose cylinders, containing the embedded tissue, were then placed in MeOH for a full day, and washed in BABB for a full day. Specimens were imaged on a custom built optical projection tomography system in the Optical Imaging & Vital Microscopy Core, Baylor College of Medicine, Houston TX, with a camera pixel size of 6.7um, an image pixel size of 8.75um, and a round scanning trajectory. OPT images were reconstructed using NRecon software (Bruker microCT, v. 1.3). The files were subsequently resampled for segmentation and volume rendering by removing every other slice and also by scaling the images by half. The 3D reconstructions were created and segmented using Amira Software (ThermoScientific v. 5.3.3).

1.4 Results and Discussion

1.4.1 Axes, cormidia, and zooid types

Historically, there had been no consistent terminology to describe the axes of mature siphonophore colonies. Haddock et al. (2005) set up a standardized scheme to describe mature planktonic siphonophore colonies, with the anterior end of the colony as the pneumatophore and the posterior end of the colony as the oldest (Fig. 1.1A). The dorsal-ventral axis is perpendicular to this axis, with siphosomal zooids attached to the ventral side of the stem. Left and right are determined as perpendicular to the anterior-posterior and dorsal-ventral plane. Haddock *et al.* note that the oral end of the larva corresponds to the posterior of the mature colony. As *Physalia physalis* is a pleustonic species, with distinctive colony morphology and arrangement, it is important to homologize the axes with other siphonophores. Totton (1960) does not use
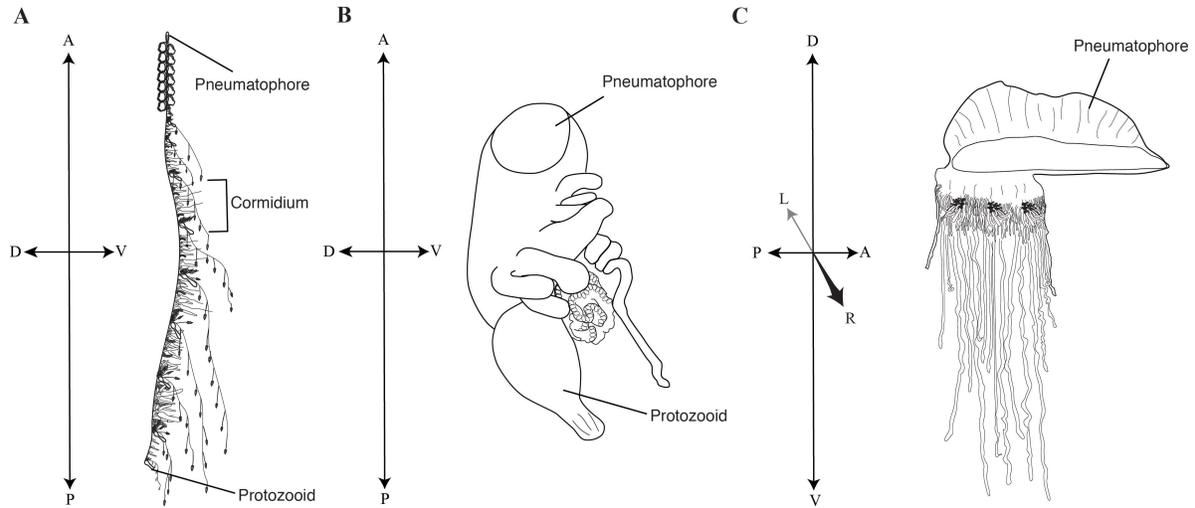
Figure 1.1: Colony orientation in siphonophores. A - anterior, P - posterior, D - dorsal, V - ventral, L - left, R - right. A. Schematic of a mature colony of the siphonophore *Nanomia bijuga*. Drawing by Freya Goetz, wikimedia commons. B. Schematic of a developing *Physalia physalis* larva. C. Schematic of a mature *Physalia physalis* colony.

the terms anterior-posterior, and defines an oral-aboral axis that corresponds directly to the larval axis, with the protozooid, the first feeding zooid (Fig. 1.1B), on the oral end and the apical pore (the pore is the site of invagination forming the pneumatophore) of the pneumatophore on the aboral end. The oral end of the colony thus corresponds to the posterior end as defined by Haddock et al. (2005). This corresponds directly with the anterior-posterior axis defined by other authors (Huxley, 1859; Steche, 1910), with the apical pore defined as the anterior of the colony. To keep terminology consistent across all siphonophores, we will follow this convention, with the anterior corresponding to the apical pore and the posterior corresponding to the protozooid (Fig. 1.1B). The dorsal-ventral axis is perpendicular to this plane, with the dorsal side towards the crest of the float and zooid attachment on the ventral side (Fig. 1.1C). We will follow the same left-right and proximal-distal axis conventions. While zooid attachment is on the ventral side, there are very clear left-right asymmetries in the placement and growth of zooids in this species, and colonies are either left-handed or right-handed.

Cormidia are typically defined as a group of zooids that are reiterated along the siphosomal stem in many siphonophore species (Fig. 1.1A) (Totton, 1965). Many authors refer to 'cormidia' in *Physalia physalis*. Cystonectae, the suborder to which *P. physalis* belongs, are sister to Codonophora (Dunn et al., 2005b;

Munro et al., 2018). Cystonects produce all zooids from single buds that arise along the stem, while probud subdivision (all zooids in a cormidium arise from a single bud) is a synapomorphy of Codonophora (Dunn and Wagner, 2006). Probud subdivision is associated with the origin of cormidia along the branch that leads to Codonophora (Dunn and Wagner, 2006). Due to this, and the fact that *P. physalis* has very distinct development and morphology, we will not apply the term 'cormidia' to describe *P. physalis* organization.

Siphonophores consist of a number of functionally specialized zooids that are homologous to free living polyps or medusae (Totton, 1965). Cystonects all lack a nectosome, a specialized region in the colony with a distinct growth zone that gives rise to nectophores (highly specialized medusae that play a key role in locomotion) (Totton, 1965). In long-stemmed cystonects (all cystonects except for *Physalia physalis*) gastrozooids (feeding polyps) arise as buds in the anterior of the colony and are carried to the posterior by an elongating stem, while gonodendra (reproductive structures) appear independently along the stem (Dunn and Wagner, 2006). In cystonects, the gonodendra are compound structures, containing gonophores (reduced medusae, bearing a gonad), palpons (derived gastrozooids, that lack tentacles in cystonects), and nectophores (Totton, 1965). *P. physalis* gonodendra have these zooids, as well as 'jelly polyps' that are reduced nectophores of unclear function (Mackie, 1960; Totton, 1960). Cystonects are dioecious, and all the gonodendra in individual colonies bear gonophores of only one sex.

As compared to other siphonophore species, including other cystonects, *Physalia physalis* is peculiar with regards to its colony organization. *P. physalis* is also the only species where the gastrozooid, the primary feeding zooid, does not have a tentacle for prey capture. The only exception is the protozooid, which is essentially a typical siphonophore gastrozooid, with a mouth, tentacle and small basigaster region (Church et al., 2015b; Totton, 1960). In *P. physalis*, the tentacle is borne on a separate zooid, that Totton (1960) called the ampulla. Other authors refer to either the zooid or the attached tentacle as a dactylozooid (Araya et al., 2016; Bardi and Marques, 2007; Jenkins, 1983; Lane, 1960b) - the term dactylozooid has historically been applied to palpons in other siphonophore species but is not currently used, and dactylozooids are specialized palpon-like defensive zooids in other hydrozoans (Cartwright et al., 1999; Cartwright and Nawrocki, 2010; Schuchert, 2003). To avoid confusion about the homology of this zooid, we suggest that the term dactylozooid should not be used, as we consider this zooid to be unique to *P. physalis* and not homologous to dactylozooids

in other hydrozoans. Additionally, the term ampulla is also commonly associated with the terminal vesicle of the tricornuate tentillum of agalmatids (Totton, 1965). We favor reviving Haeckel's 'tentacular palpon' to refer to this zooid (Haeckel, 1888), which not only has precedence, but also matches the likely hypothesized origin of this zooid (see below).

Haeckel (1888) outlined two possible hypotheses for the origin of tentacular palpons - the first hypothesis, promoted by Huxley (1859), is that they are not zooids, but are instead secondary diverticula at the base of the tentacle that function similarly to ampullae in Echinoderm tube feet. In the second hypothesis, modification and subfunctionalization of an ancestral gastrozooid gave rise to two separate zooids - a gastrozooid without a tentacle and a tentacular palpon with a tentacle. Totton (1960) proposed a modification of the first hypothesis, and suggested that the 'ampulla' is a hypertrophied basigaster (aboral region of a gastrozooid that plays an active role in nematogenesis) that has separated from the remainder of the gastrozooid. However, we favor the second hypothesis, based on observations of the gastrozooid and tentacular palpon (Figs. 1.2,1.3,1.8, 1.6). The gastrozooid and tentacular palpon are borne on separate peduncles (Figs. 1.8A, 1.6A, 1.6B), and develop from distinct, separate buds (Figs. 1.2A, 1.2B, 1.3). Thus, the tentacular palpon is a derived gastrozooid, unique to *Physalia physalis*, that has an enlarged tentacle, no mouth, and is functionally specialized for nematocyst production. The gastrozooids in *P. physalis* are also derived gastrozooids that have lost tentacles and are functionally specialized for feeding only. The subfunctionalized gastrozooid hypothesis is also more parsimonious than the other hypotheses, as the modification and subfunctionalization of zooids is common in siphonophores - palpons, for example, are considered to be derived, modified gastrozooids that typically have a reduced tentacle (Totton, 1965).

1.4.2 Larval development and morphology

Larval development has not been observed directly, and development has been described by comparing the morphology of fixed specimens (Okada, 1932, 1935; Totton, 1960). The smallest described larva is 2mm, and consisted of a pneumatophore and a developing protozooid with a tentacle (Totton, 1960). The pneumatophore forms in a manner similar to other siphonophores, with an invagination of the aboral end of planula forming the pneumatosaccus (Fig. 1.5) (Carré, 1969; Church et al., 2015b; Okada, 1935). Okada
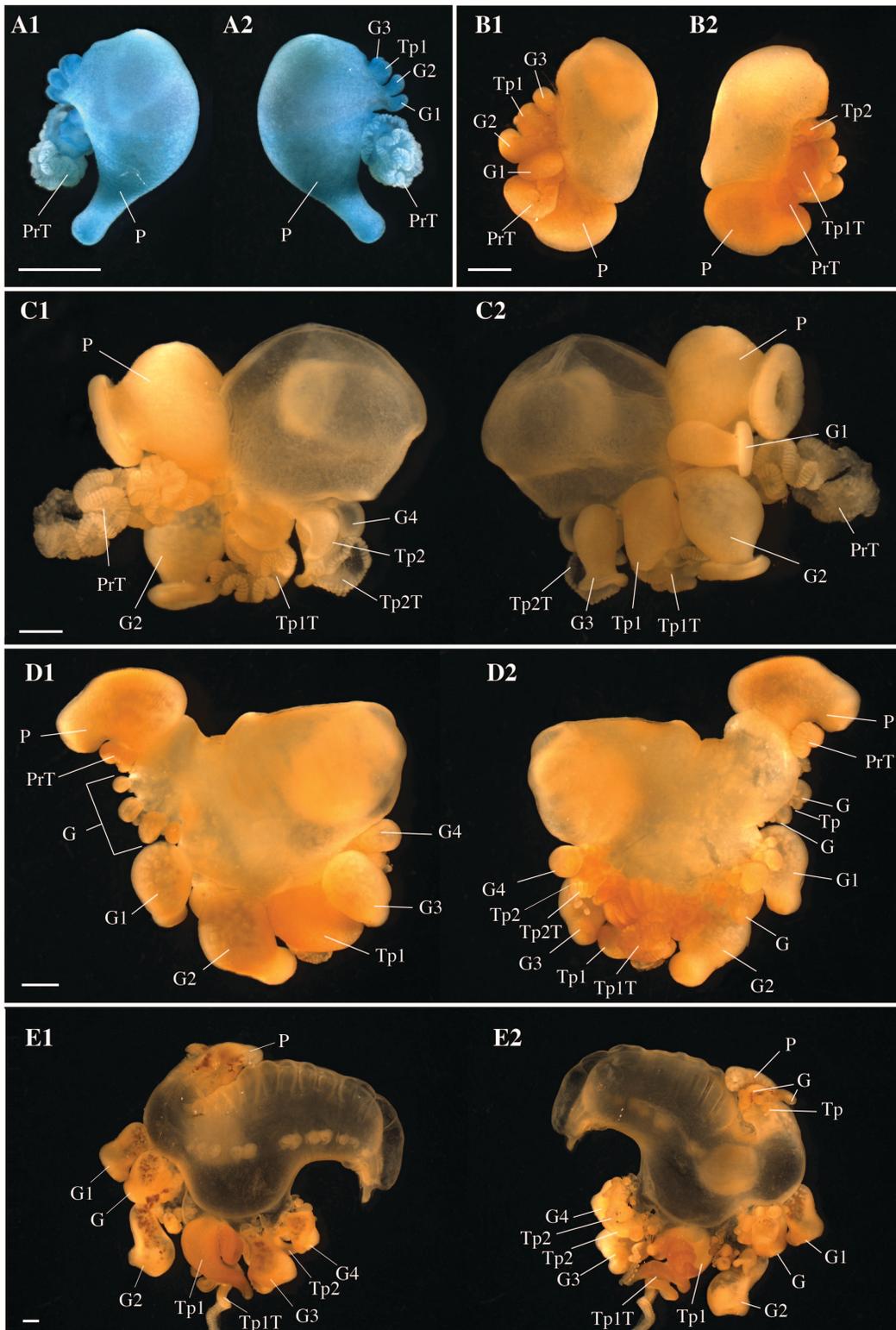
Figure 1.2: Photographs of formalin fixed developing *Physalia physalis*, all are different individuals. Photographs 1 and 2 represent left/right sides of the same specimen (specimens are a mix of left and right handed individuals). Scale bar is 1mm. Abbreviations: Tp: Tentacular palpon (number indicates order of appearance); G: Gastrozooid (number indicates hypothesized order of appearance) P: Protozooid; PrT - Tentacle of protozooid; TpT: tentacle of tentacular palpon (number indicates order of appearance).

(1932) suggests that the apical pore that is formed by this invagination is completely closed in larval *Physalia* (float length 2mm) and controlled gas release from the pneumatophore, as in some other siphonophore species, is no longer possible (Okada, 1932). However, Mackie (1960) suggests that the pore is not completely closed even in mature colonies, but the pore is so tightly constricted that gas release is unlikely to occur naturally. Other reports suggest that young *Physalia* may be able to release gas from the pore (Agassiz and Mayer, 1902). In the earliest stages, there is no separation between the gastric cavity of the protozooid and the main gastric cavity (Okada, 1935). The pneumatosaccus, that is formed via the invagination, protrudes into the main gastric cavity and is connected at the site of invagination (Okada, 1935). As the protozooid differentiates, a septum separates the gastric cavity of the protozooid from the main gastric cavity (Okada, 1932).

Anterior to the protozooid, three buds arise on the ventral side as three transverse folds (Totton, 1960). Based on our observations of the budding order and the relative size of the zooids, the posterior most of these three buds is a gastrozooid G1, followed by a second gastrozooid G2 and tentacular palpon (labelled Tp1) (Fig. 1.2A, B, 1.3). The third gastrozooid (G3) is hypothesized to appear anterior to gastrozooids G1, G2, and tentacular palpon Tp1. Totton (1960) numbers the buds based on the hypothesized 'cormidia' to which they belong in the mature colony, but not based on their order of appearance. Okada numbers the buds based on hypothesized order of appearance, which differs from ours only in that G2 is considered the first bud, perhaps based on size, and G1 is considered the second (Fig. 1.2A, B) (Okada, 1932, 1935). The gastrozooid labelled G2 here is larger in older specimens (Figs. 1.2C, 1.2D, 1.3), but not in the youngest developing specimen (Fig. 1.2A). Our numbering also follows Totton's observations that two gastrozooids (that he calls III and VII) appear first, followed by the first tentacular palpon (Tp1), and then the third gastrozooid (G3) (Totton, 1960).

Physalia physalis colonies can be either left or right handed, and the location of first tentacular palpon (Tp1) and the attachment point of the tentacle is the first indicator of left-right asymmetry (Okada, 1935; Totton, 1960). The tentacle of the tentacular palpon is placed either on the left or right side, depending on the handedness of the colony (Fig. 1.2, 1.3). The secondary series of buds always appear on the same side as the tentacular palpon tentacle. The attachment point of the tentacle of the protozooid may even be
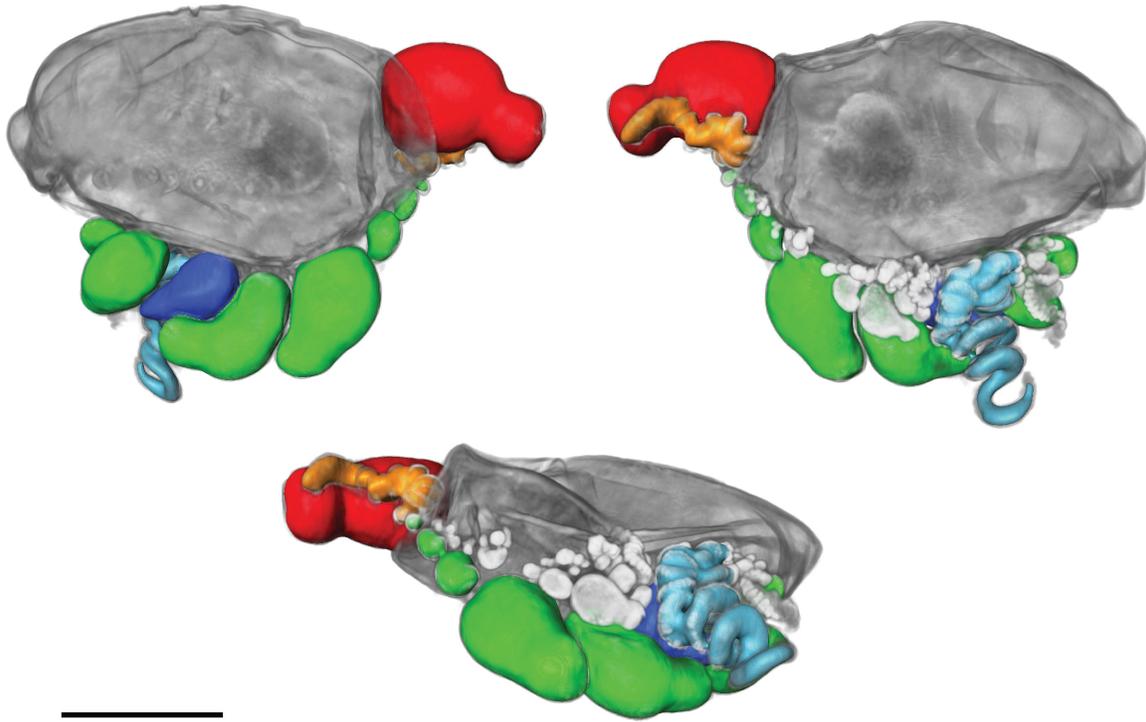
Figure 1.3: Images of formalin fixed larval *Physalia physalis*, images obtained by optical projection tomography. Images are different views of the same specimen. Scale bar is 2.5 mm. The 3D image was segmented and false-colored to highlight different morphological features. Green- gastrozooids; Red- Protozooid; Orange-tentacle associated with protozooid; Dark blue- Tentacular palpon; Light blue- tentacle associated with tentacular palpon. Gastrozooids and tentacular palpons forming at the base of the first set of gastrozooids and tentacular palpon are unlabelled and are light grey in color.

an earlier indication of left-right asymmetry (Fig. 1.2, 1.3). As live embryos are not available, it remains an open question as to whether left-right asymmetries are established via molecular mechanisms similar to those underlying left-right asymmetry in bilaterians (Levin, 2005).

As the organism grows and the pneumatosaccus expands anteriorly, new tentacular palpons grow at the base of the original gastrozoods (Fig. 1.2B2, 1.3). In larger specimens, new gastrozooid and tentacular palpon buds form anterior and posterior to the three gastrozooids (G1, G2, G3) and tentacular palpon (Tp1) (Fig. 1.2C1, D). A secondary series of buds also form at the base of the gastrozooids in line with the first tentacular palpon (either left or right, depending on the handedness of the colony) (Fig. 1.2D2, 1.3) (Okada, 1932). Additionally, in the expanding space between the protozooid and the primary series of gastrozooids, a series

of buds form (Fig. 1.2D, labelled "G"; Fig. 1.3, gastrozooids (in green) closest to protozooid). This region of growth directly anterior to the protozooid (Fig. 1.2D, labelled "G") is distinguished from the original region by Totton (1960) as the "oral zone", while the original series of buds (including G1, G2, G3, G4, Tp1, Tp2, and secondary buds) are the "main zone". To keep naming consistent with the axes, we propose calling the oral zone the "posterior zone". In older larvae the protozooid and posterior growth zone are physically separated from the main zone, due to elongation of the stem/float carrying the posterior growth zone away from the main growth zone (Fig. 1.2E).

As *P. physalis* continues to grow, new space along the ventral side in the main zone is occupied by new buds in line with the original series of gastrozooids (G1, G2, G3 etc.) and tentacular palpon. Additional secondary clusters of buds also continue to arise both in the posterior and main zone, either to the left or right according to the handedness of the colony (Fig. 1.3). A crest begins to become visible (Fig. 1.2D and 1.2E), and the float expands. Once the float is fully expanded, and the colony is floating on the ocean surface, branching and growth begins to occur in the dorsal-ventral plane (Fig. 1.1C). In fully mature specimens, zooids occupy the space between the posterior and main zones, and the gap (termed the basal internode) between the two zones of growth is not visible.

Superficially, the series of buds in the posterior zone resembles the growth zone of related species, such as *Nanomia bijuga* (Dunn, 2005; Siebert et al., 2015). We do not know the order of bud appearance, however based on the relative size of the gastrozooids (Fig. 1.3), new buds in the posterior zone appear to arise posterior-anterior along the ventral side in an inverse direction to other siphonophore species (Fig. 1.1B). This does not fit with the definition of axes as defined by Haddock et al. (2005), with buds arising in the anterior and being carried by elongation of the stem to the posterior. Patterns of growth are very different in *Physalia physalis*, however this may suggest that during early development growth patterns are inverted in this species. According to our numbering system, the original series of buds (G1, G2, Tp1) also arise posterior-anterior, although subsequent buds in the main zone arise both anterior and posterior to these zooids.

The patterns of growth that can be observed from fixed developing *Physalia physalis* specimens suggests that while there are many similarities between this species and other siphonophores, there are many differences

12
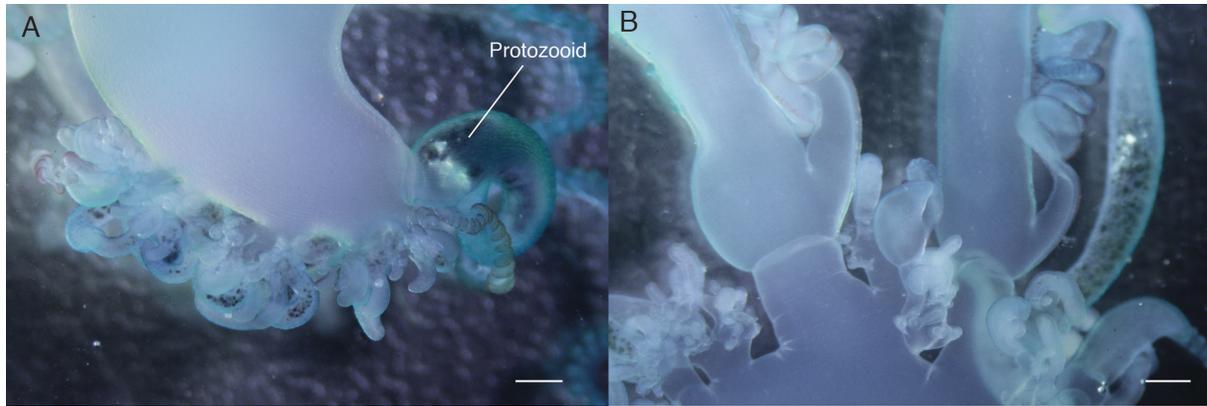
Figure 1.4: Regions of growth in juvenile *Physalia physalis* (float length 8-10 cm). A. Photograph of the posterior zone, with protozooid as the posterior most zooid. B. Photograph of mature tentacular palpons and gastrozooids, and developing tripartite groups forming proximally. Scale bar is 1 mm.

that are unique to this species. In other siphonophore species, ontogenetic series of zooids are arranged linearly along a stem with the oldest at the posterior and the youngest in the anterior (Dunn and Wagner, 2006; Siebert et al., 2015), although new zooids are observed to arise along the stem in some species (Siebert et al., 2013). In *P. physalis* there are three major axes of growth – along the ventral side, posterior-anterior in the posterior growth zone (Fig. 1.4A), as well as anterior and posterior of the main zone; secondary buds to left or right of the original series of buds along the ventral side, depending on the handedness of the individual; and finally in mature specimens, growth proceeds proximal-distal from the ventral side (Fig. 1.4B).

1.4.3 Morphology and zooid arrangement of mature *Physalia physalis*

Juvenile (sexually immature) and mature *Physalia physalis* float on the ocean surface with the pneumatophore, or float, above and on the surface of the water and all of the zooids are below the water surface. In juvenile *P. physalis* the pneumatophore will continue to grow in size, but it resembles the fully mature form. As in other siphonophores, the pneumatophore is a multi-layered structure, consisting of an outer codon, a pneumatosaccus, and a gas gland (Fig. 1.5) (Mackie, 1960). The outer codon consists of ectoderm, mesoglea, and endoderm (Mackie, 1960). Within the codon is the pneumatosaccus, formed by invagination, consisting of endoderm, mesoglea, ectoderm, a chitinous layer secreted by the ectoderm, and the gas space (Mackie, 1960). At one end of the pneumatosaccus is an expanded layer of ectodermal cells that form the gas
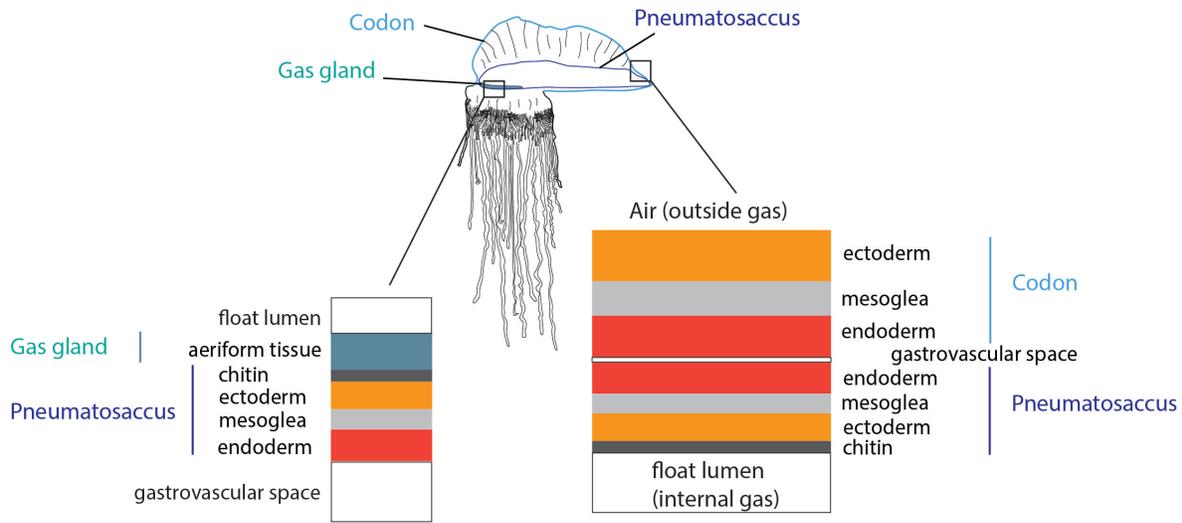
Figure 1.5: Schematic of the cell layers in the pneumatophore, showing the distinction between the codon, pneumatosaccus, and gas gland. Orange - ectoderm; dark grey - chitin; red - endoderm; light grey - mesoglea.

gland (Copeland, 1968; Mackie, 1960). Aeriform cells within the gas gland produce carbon monoxide to fill the float, however the percentage of carbon monoxide within the float is lower than other siphonophores due to diffusion and gas exchange (Larimer and Ashby, 1962; Pickwell et al., 1964; Wittenberg, 1960). Unlike other siphonophores, the pneumatophore is greatly expanded, and the pneumatosaccus is free within the gastric cavity and attached only to the site of invagination at the anterior of the colony (Mackie, 1960). Dorsal processes of the pneumatosaccus fit into pockets of the crest of the codon, and muscular contractions of the codon enable the pneumatosaccus to expand into this space and erect the sail – this 'pneumatic skeleton' is likened to a hydrostatic skeleton (Mackie, 1960). The zooids are all attached on the ventral side (displaced either to the left or right) and share this common gastric cavity – this region is likely homologous to the stem of other siphonophores (Chun, 1887).

In juvenile *Physalia physalis*, projections extend from the ventral 'stem', carrying zooids distally away from the float. The colony arrangement of *P. physalis* appears crowded and lacking in structure, particularly in fully mature specimens, however there is a distinct pattern of growth. The best descriptions of colony arrangement in mature specimens are given by Totton (1960), who suggested that growth occurs through the formation of tripartite groups (Fig. 1.8, 1.6). The tripartite groups consist of a tentacular palpon with an
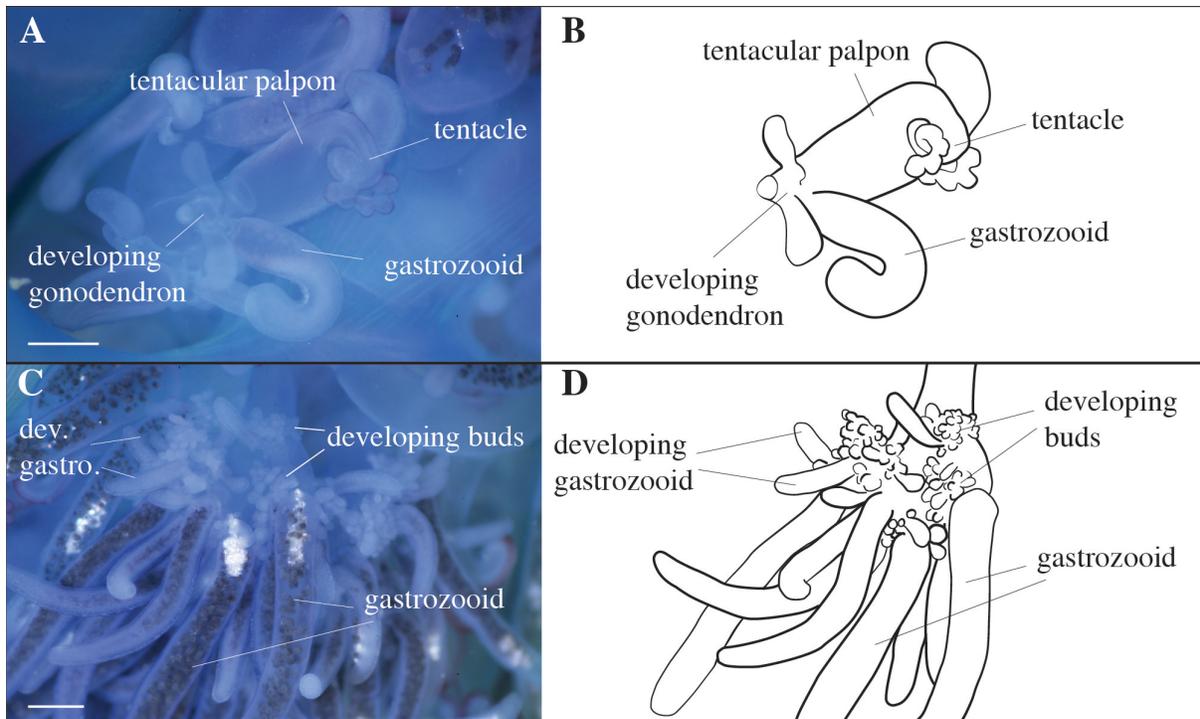
Figure 1.6: Photographs of live juvenile *Physalia physalis*. Scale bar is 1mm. A. Developing tripartite group, with gastrozooid, tentacular palpon and developing gonodendron. B. Schematic of the tripartite group. C. Developing gonodendron with mature gastrozooids and buds that will give rise to gonophores, nectophores, palpons. D. Schematic of the developing gonodendron.

associated tentacle, a gastrozoooid, and a gonodendron at the base of the gastrozooid (Totton, 1960). The morphology of *P. physalis* is clearest in juvenile specimens, where the gonodendron is not fully developed and developing tripartite groups are easily identifiable (Fig. 1.6, 1.8B). The gonodendron is a structure that consists of a number of different zooids, including gastrozooids, male or female gonophores (colonies are dioecious, and as such, colonies are either male or female), nectophores, jelly polyps, and also palpons.

Tripartite groups are carried down by elongated projections of the stem, with successive tripartite groups forming at the base of the older groups. In mature colonies, the oldest zooids are located distally, with developing zooids in tripartite groups forming proximally to the float (Fig. 1.4B). The exception to this appears to be the very oldest zooids that form during early development (Fig. 1.2, 1.3), that remain attached proximally to the stem via long peduncles. There are differences in the rate of growth and appearance of zooids in the tripartite groups: the tentacular palpon and gastrozooids both develop precociously, while the gonodendron develops and matures later (Totton, 1960). The developing and mature tentacles could be
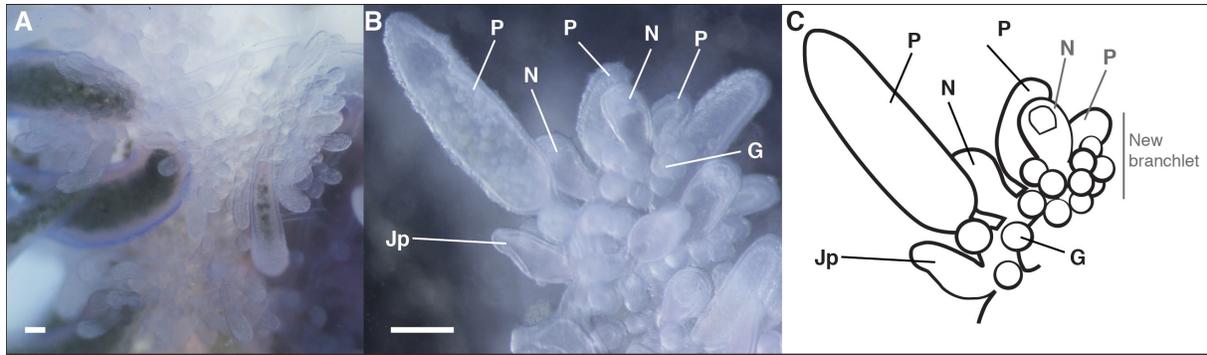
Figure 1.7: Photographs of branches within the developing gonodendra of live *Physalia physalis*. Scale bar is 500um. A. Overview of branch and branchlets with gastrozooids, nectophores, palpons, gonophores and jellypolyps. B. Close up of branchlet within the gonodendron, from proximal to distal: jelly polyp (Jp), palpon (P), nectophore (N), palpon, with gonophores (G) along the branchlet, additionally there is a nectophore, palpon and gonophores that are part of a new branchlet. C. Schematic of a close up of a branchlet within the gonodendron.

distinguished not only by size and length, but also by color – the tentacles of the mature tentacular palpons are a turquoise blue, while the buttons of the developing tentacle are a purple/pink color. The blue pigment of *P. physalis* is suggested to be a bilin-protein complex, and the green, purple, and pink coloration in other tissues are caused by unconjugated bile pigments, which are likely sourced from their diet (Herring, 1971).

The gonodendra are highly complex branching structures. We are not able to build much upon the description by Totton (1960) of the structure of the gonodendron, but we do attempt to simplify aspects of his description here, based on our observations. In the juvenile specimens, we were able to observe developing gonodendra with mature gastrozooids (what Totton (1960) calls 'gonozooids', or secondary gastrozooids) and clusters of buds at their base that will subdivide and give rise to all the other zooids within the gonodendron (Fig. 1.6B). The peduncles at the base of the gastrozooids form the major branches within the gonodendron (Totton, 1960). Branching can be observed at two levels: major branches formed by the peduncle of the gastrozooid (Fig. 1.6B); and branching structures at the base of the gastrozooids, that are formed by probuds (Fig. 1.6B "developing buds") that subdivide, branch and re-branch, and form a series of branchlets along which nectophores, jelly polyps, palpons, and gonophores form (Fig. 1.7A). The branchlets of the gonodendra typically consist of a series (proximal to distal) of a jelly polyp and more developed palpon, followed by a nectophore and palpon, with ~ 10 or more male or female gonophores (depending on the sex of the colony) forming along the branchlet (Fig. 1.7B,C). Totton (1960) refers to the section with the jelly polyp and

16

palpon as the terminal section of the branchlet, while the sub-terminal portion of the branchlet may become a palpon and nectophore (Fig. 1.7B,C), or continue dividing into a new terminal and subterminal portion. New probuds form in the region directly opposite the location of jelly polyps, giving rise to new branchlets, that in turn re-branch opposite the location of the jelly polyp (Totton, 1960). According to Totton (1960), sometimes a branchlet can consist only of a palpon and jelly polyp.

1.4.4 Ecology and lifecycle

Physalia physalis is a cosmopolitan species, found in tropical and subtropical regions of all oceans, as well as occasionally in temperate regions (Totton, 1960). Historically, a large number of *Physalia* species have been described on the basis of size, color, and location (Chun, 1887; Lamarck, 1801; Huxley, 1859; Totton, 1960), however, there is currently only one recognized species of *Physalia* – *P. physalis* (Totton, 1965). The different species that have been identified are suggested to be different developmental stages (Okada, 1932; Totton, 1960, 1965). However, nothing is known about genetic diversity among populations of *P. physalis* in the Atlantic or the Pacific/Indian Ocean. One local study has been conducted, using two genetic markers, that showed substantial genetic diversity among *Physalia* off the coast of New Zealand (Pontin and Cruickshank, 2012), however global studies using more markers would help clarify whether this reflects intra-specific genetic diversity or if there is cryptic diversity.

As larval development has not been observed directly, everything that is known about the early stages of this species is known from fixed specimens collected in trawl samples (Okada, 1932, 1935; Totton, 1960). Gonodendra are thought to be detatched by the colony once they are fully mature, and the nectophores may be used to propel the gonodendron through the water column (Steche, 1910; Totton, 1960). Released mature gonodendra have not been observed, and it is not clear what depth range they occupy (Steche, 1910; Totton, 1960). It is also not known how the gonodendra from different individuals occupy a similar space for fertilization, or if there is any seasonality or periodicity to sexual reproduction. Embryonic and larval development also occurs at an unknown depth below the ocean surface (Fig. 1.9) (Totton, 1960). After the float reaches a sufficient size, the juvenile *P. physalis* is able to float on the ocean surface.
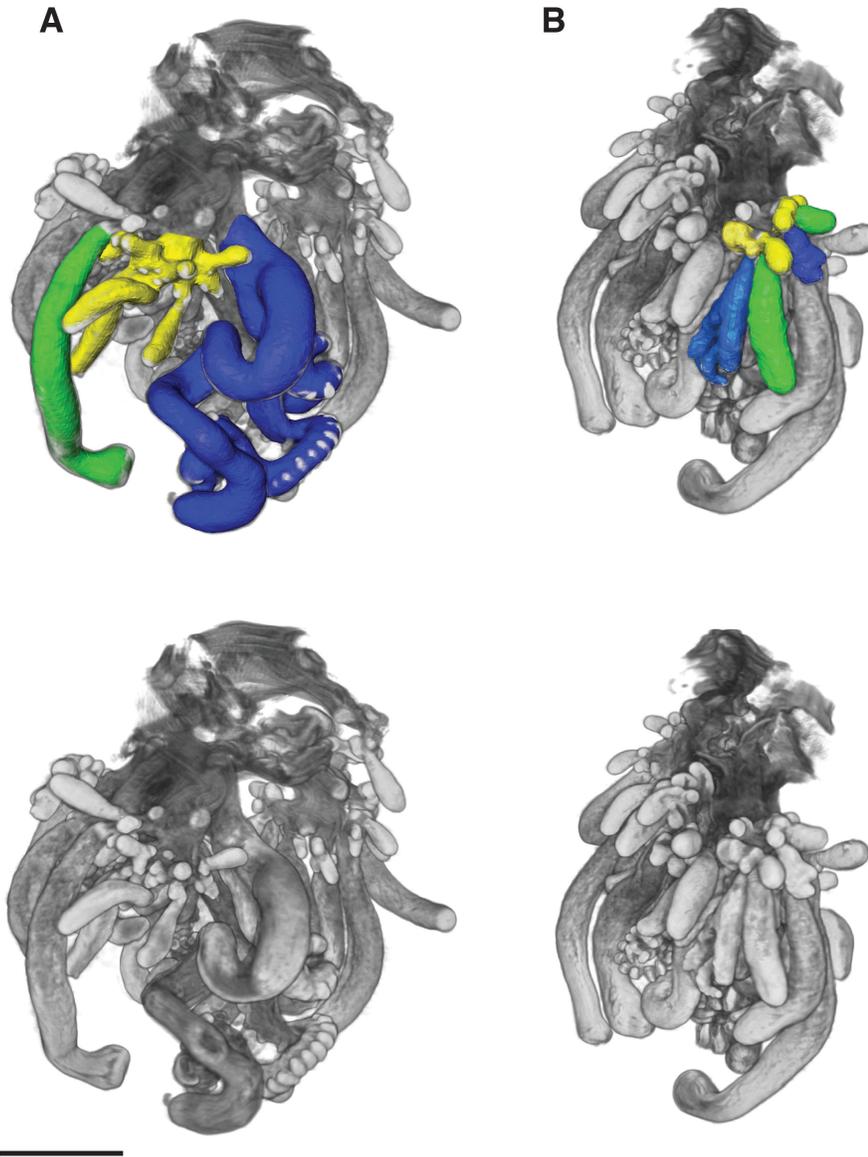
Figure 1.8: Images of formalin fixed juvenile *Physalia physalis* zooids, images obtained by optical projection tomography. Images are different views of the same specimen. Scale bar is 2mm. The 3D image was segmented and false-colored to highlight tripartite groups. The un-segmented image is shown below. Green-gastrozooid; Dark blue- tentacular palpon; Yellow- developing gonodendron. A. Tripartite group with developing tentacular palpon, gonodendron and gastrozooid. B. Two sets of developing tripartite groups at different developmental stages are highlighted, while others are visible but not segmented.
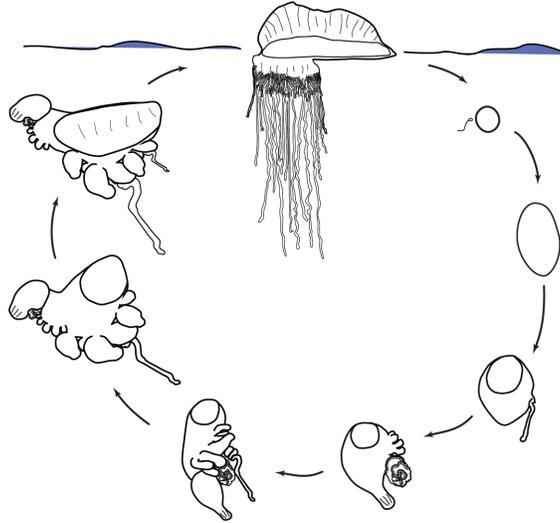
Figure 1.9: Schematic of the lifecycle of the Portuguese man of war. The mature *Physalia physalis* is pictured floating on the ocean surface, while early development is thought to occur at an unknown depth below the ocean surface. The egg and planula larva stage have not been observed. The egg and planula drawings are from a *Nanomia bijuga* lifecycle schematic drawn by Freya Goetz, wikimedia commons.

Mature *Physalia physalis* uses its sail to catch prevailing winds. Muscle contractions of the outer codon of the pneumatophore force increased pressure within the pneumatosaccus and enable the erection of the crest (Mackie, 1960). This is the only known active contribution to locomotion – *P. physalis* cannot change tack, and the nectophores within the gonophore are not thought to play any role in active propulsion of the colony (although they may play a role once the gonodendron is released) (Totton, 1960). The alignment of the sail relative to the wind (left-right handedness), is established during early development, and while it has been suggested that left-handed individuals are dominant in the Northern Hemisphere as a result of prevailing winds, and right-handed individuals are more prevalent in the Southern Hemisphere (Woodcock, 1944, 1956), there is no evidence to support this (Totton, 1960; Totton and Mackie, 1956). Wind fluctuations are likely to result in random distribution of both forms regardless of hemisphere, although strong sustained winds from the same direction do appear to result in the stranding of a particular type (Clark, 1970; Totton and Mackie, 1956). Totton (1960) suggests that left-right asymmetry is established by the prevailing wind on the first windy day, however this is unlikely, as the asymmetry is present early in developing specimens.

The tentacles of the Portuguese man of war can reach up to 30m in mature colonies, and are used as a fishing line to catch fish and fish larvae. Fish and fish larvae comprise 70-90% of their diet, and the

nematocyst batteries on the tentacles of *Physalia physalis* contain a single type of nematocyst that is only able to penetrate soft bodied prey (Purcell, 1981b, 1984). The nematocyst delivers a toxin that leads to hyperventilation, immobilization and, in high doses, death (Lane and Dodge, 1958). Once a tentacle comes into contact with its prey, the prey is carried up towards the gastrozooids near the base of the float. The gastrozooids respond immediately to the capture of prey, and begin writhing and opening their mouths (Lenhoff and Schneiderman, 1959). Many gastrozooids attach themselves to the prey – upwards of 50 gastrozooids have been observed to completely cover a 10cm fish with their mouths spread out across the surface of the fish (Wilson, 1947). The gastrozooids release proteolytic enzymes to digest the fish extracellularly, and are also responsible for intracellular digestion of particulate matter (Mackie, 1960; Mackie and Boag, 1963). The digested food products are released into the main gastric cavity for uptake by the rest of the colony (Mackie, 1960; Mackie and Boag, 1963). While *P. physalis* is a voracious predator of fish, it is predated upon by sea turtles (Babcock, 1938; Bingham and Albertson, 1974), and *Glaucus atlanticus* and *Glaucus marginatus*, species of nudibranch that store intact *Physalia* nematocysts and redeploy them for their own defense (Bieri, 1966; Thompson and Bennett, 1969; Valdés and Campillo, 2004). A number of juvenile fish live commensally with *Physalia* and are found near the gastrozooids and gonodendra, however one species, *Nomeus gronovii*, has been observed to swim among and feed upon the tentacles (Jenkins, 1983; Kato, 1933). *Nomeus gronovii* is significantly more tolerant of *Physalia* venom than other species, but can nevertheless be killed by *P. physalis* (Lane, 1960a; Totton, 1960).

1.5   Conclusions

Physalia physalis differs significantly from all other siphonophores in terms of its habitat, development, body plan, and colonial organization. The radical modification of the colony body plan is likely associated with a transition from a planktonic to pleustonic lifestyle. Using photographs, specimens and new volumetric imaging methods to create 3D reconstructions, we were able to clarify aspects of *P. physalis* colony organization in juvenile specimens, and also early development in larval specimens. The study underscores the value of fixed specimen collections – all of the developing specimens used in this study were collected in

the 1970s and 1980s, and it was still nevertheless possible to 3D image these individuals using fluorescent stains. Optical projection tomography is particularly useful for imaging these complex, highly branching structures, and we are able to use these images to build upon the existing knowledge about the development, morphology and colony organization of this species. In particular, larval and juvenile specimens were key for this work, because growth and secondary budding in mature specimens makes it significantly more difficult to understand the order and pattern of growth.

Many open questions remain about this species, however. While Totton (1960) was able to observe gonodendra that are more mature than those examined in this study, fully mature gonodendra with mature eggs or sperm have not been described yet. Mature gonodendra are hypothesized to be released into the water column, however there is no data on the depth ranges that the gonodendra occupy. Additionally, there is also no information about the depth at which any of the early developmental stages can be found, nor their ecology. While there is abundant data on the occurrence and location of *P. physalis*, particularly beached specimens, there is frequently little recorded information about the size of the colony, and it is not clear if there is seasonality to their reproduction. Most of our experiences of the Portuguese man of war are close to shore, where news stories warn of purple flags, vicious stings, and ruined beach days, however we still know almost nothing about their behavior, ecology, and lifecycle out in the open ocean.

Chapter 2

Improved phylogenetic resolution within Siphonophora (Cnidaria) with implications for trait evolution

Catriona Munro[1*,‡], Stefan Siebert[1,2*], Felipe Zapata[3], Mark Howison[4,5], Alejandro Damian Serrano[1,10], Samuel H. Church[1,6], Freya E. Goetz[1,7], Philip R. Pugh[8], Steven H.D. Haddock[9], Casey W. Dunn[10]

[1] Department of Ecology and Evolutionary Biology, Brown University, Providence, RI 02912, USA

[2] Current address: Department of Molecular & Cellular Biology, University of California Davis, Davis, CA 95616, USA

[3] Department of Ecology and Evolutionary Biology, University of California Los Angeles, Los Angeles, CA 90095, USA

[4] Brown Data Science Practice, Brown University, Brown University, Providence, RI 02912, USA

[5] Current address: Watson Institute for International and Public Affairs, Brown University, Providence, RI 02912, USA

[6] Current address: Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA

[7] Current address: Smithsonian Institution, National Museum of Natural History, Washington, DC 20560, USA

[8] National Oceanography Centre, Southampton, SO14 3ZH, UK

[9] Monterey Bay Aquarium Research Institute, Moss Landing, CA 95039, USA

[10] Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520, USA

* Authors contributed equally

2.1 Abstract

Siphonophores are a diverse group of hydrozoans (Cnidaria) that are found at most depths of the ocean - from the surface, like the familiar Portuguese man of war, to the deep sea. They play important roles in ocean ecosystems, and are among the most abundant gelatinous predators. A previous phylogenetic study based on two ribosomal RNA genes provided insight into the internal relationships between major siphonophore groups. There was, however, little support for many deep relationships within the clade Codonophora. Here, we present a new siphonophore phylogeny based on new transcriptome data from 29 siphonophore species analyzed in combination with 14 publicly available genomic and transcriptomic datasets. We use this new phylogeny to reconstruct several traits that are central to siphonophore biology, including sexual system (monoecy vs. dioecy), gain and loss of zooid types, life history traits, and habitat. The phylogenetic relationships in this study are largely consistent with the previous phylogeny, but we find strong support for new clades within Codonophora that were previously unresolved. These results have important implications for trait evolution within Siphonophora, including favoring the hypothesis that monoecy arose at least twice.

2.2 Introduction

Siphonophores (Fig. 2.1 and 2.2) are among the most abundant gelatinous predators in the open ocean, and have a large impact on ocean ecosystems (Choy et al., 2017; Pagès et al., 2001; Pugh, 1984; Pugh et al., 1997; Purcell, 1981a; Williams and Conway, 1981). Siphonophores, which belong to Hydrozoa (Cnidaria), are found at most depths in the ocean, with the deepest recorded species found around 4300m (Lindsay, 2005). The most familiar species is the Portuguese man of war *Physalia physalis*, which floats at the surface and can wash up conspicuously onto beaches (Totton, 1960). Most species are planktonic, living in the water column, where some grow to be more than 30 meters in length (Mackie et al., 1987). There is also a small clade of benthic siphonophores, Rhodaliidae, that are tethered to the bottom for part of their lives (Pugh, 1983). There are currently 187 valid described siphonophore species (Schuchert, 2018).

Siphonophores remain poorly known, in large part because they are fragile and difficult to collect. They

have, however, been of great interest for more than 150 years due to their unique structure and development (Mackie et al., 1987; Mapstone, 2014). Like many other cnidarians, they are colonial: they grow by incomplete asexual reproduction. Each colony arises from a single embryo that forms the protozooid, the first body. One or two growth zones (Fig. 2.2) then arise and asexually produce other genetically identical zooids that remain attached (Carré, 1967, 1969; Carré and Carré, 1991, 1993). In some species additional zooids are added outside the growth zone along the siphosomal stem (Siebert et al., 2013). These zooids are each homologous to a solitary animal, but are physiologically integrated (Totton, 1965; Mackie et al., 1987; Dunn and Wagner, 2006). Siphonophores differ significantly from other colonial animals in their colony structure and development – their zooids are highly functionally specialized and arranged in precise, repeating, species-specific patterns (Beklemishev, 1969; Cartwright and Nawrocki, 2010). Zooids are specialized for a range of functions, including feeding, reproduction, or swimming (Fig. 2.2) (Dunn and Wagner, 2006).

Understanding the unique ecology, morphology, and development of siphonophores requires a well-resolved phylogeny of the group. The relationship of siphonophores to other hydrozoans has been difficult to determine (Cartwright and Nawrocki, 2010; Cartwright et al., 2008; Kayal et al., 2013, 2015, 2018; Zapata et al., 2015), but there has been progress on their internal relationships. A phylogeny (Dunn et al., 2005a) based on two genes (16S, 18S) from 52 siphonophore taxa addressed several long standing questions about siphonophore biology, including the relationships of the three historically recognized groups, Cystonectae, Physonectae, and Calycophorae. Cystonectae was found to be sister to all other siphonophores, while Calycophorae were nested within "Physonectae". The name Codonophora was given to this clade of "Physonectae" and Calycophorae (Dunn et al., 2005a).

Major questions remained after this early work, though. In particular, there was little support for important deep relationships within Codonophora. Understanding these relationships is key to resolving the evolution of several traits of importance, including sexual systems (monoecy versus dioecy) and the gain and loss of particular zooids, such as palpons (Fig. 2.2). Here we present a broadly sampled phylogenetic analysis of Siphonophora that considers transcriptomic data from 33 siphonophore species and 10 outgroup species (2 outgroups were subsequently excluded due to poor sampling). Using 1,423 genes, we find strong support for many relationships found in the earlier phylogeny (Dunn et al., 2005a), and also provide new resolution
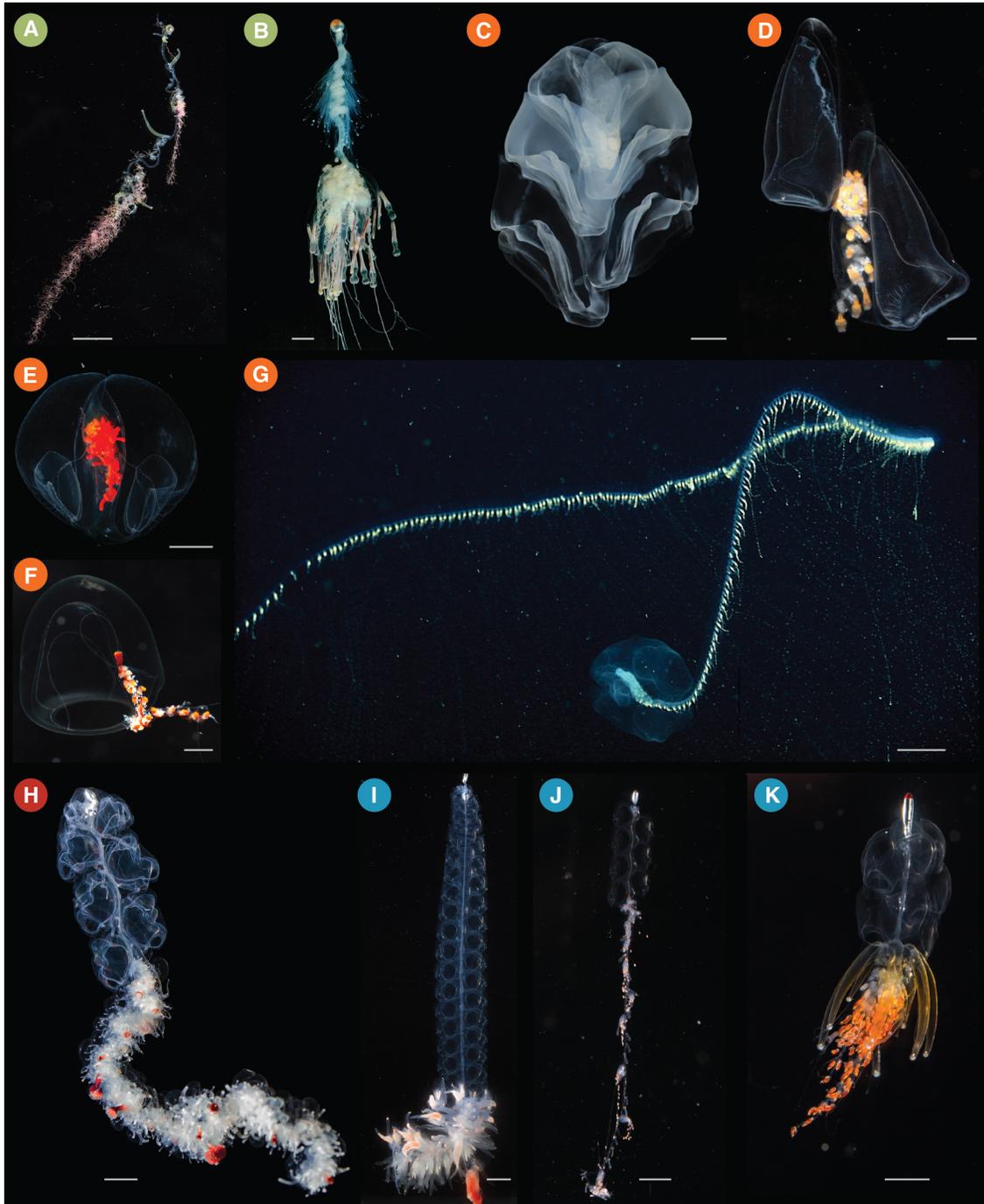
Figure 2.1: Photographs of living siphonophores. Colored circles correspond to the clades shown in Figure 3 as follows: Cystonectae (A-B), Calycophorae (C-G), Apolemiidae (H), and Clade A within Euphysonectae (I-K). (A) *Rhizophysa eysenhardtii*, scale bar = 1 cm. (B) *Bathyphysa conifera*, scale bar = 2cm. (C) *Hippopodius hippopus*, scale bar = 5 mm. (D) *Kephyes hiulcus*, scale bar = 2 mm. (E) *Desmophyes haematogaster*, scale bar = 5 mm. (F) *Sphaeronectes christiansonae*, scale bar = 2 mm. (G) *Praya dubia*, scale bar = 4 cm . (H) *Apolemia* sp., scale bar = 1 cm. (I) *Lychnagalma utricularia*, scale bar = 1 cm. (J) *Nanomia* sp., scale bar = 1 cm. (K) *Physophora hydrostatica*, scale bar = 5 mm. Photo credits: S. Siebert (C,H,I,K), S. Haddock (A,D,E,F), R. Sherlock (B), MBARI (G), C. Munro (J)
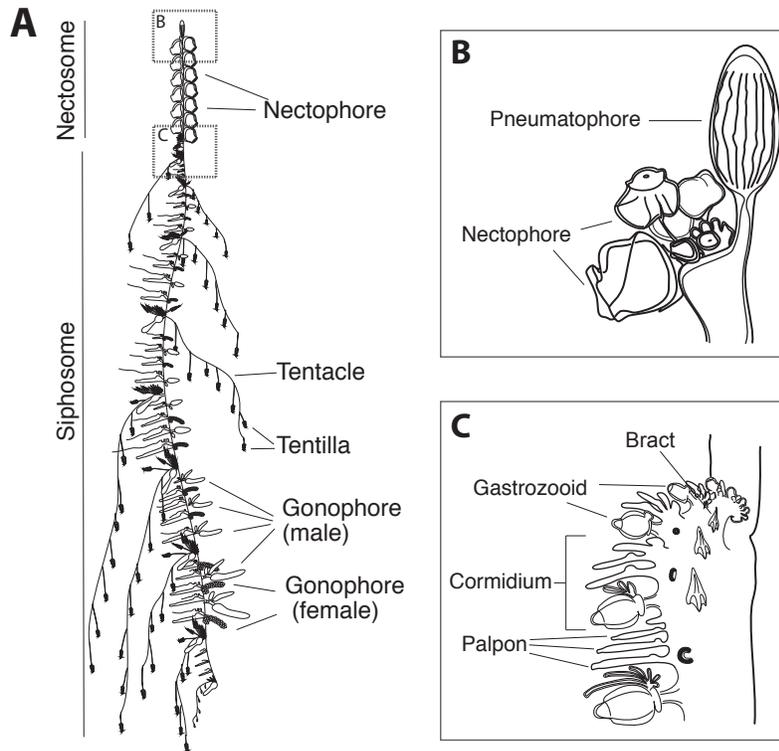
Figure 2.2: Schematic of the siphonophore *Nanomia bijuga*, oriented with the anterior of the colony at the top of the page, and the ventral side to the left. Adapted from http://commons.wikimedia.org/wiki/ File:Nanomia_bijuga_whole_animal_and_growth_zones.svg, drawn by Freya Goetz. (A) Overview of the whole mature colony. (B) Inset of the pneumatophore and nectosomal growth zone. A series of buds give rise to nectophores. (C) Inset of the siphosomal growth zone. Probuds subdivide to give rise to zooids in repeating units (cormidia). The gastrozooid (specialized feeding polyp) is the posterior-most zooid within each cormidium.

for key relationships that were unresolved in that previous study. Using this phylogeny, we reconstruct the

evolutionary history of characters central to the unique biology of siphonophores, including zooid type, life

history traits, and vertical habitat use.

2.3   Material and methods

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 1501118c with tag "paper_v2".

2.3.1 Collecting

Specimens were collected in the north-eastern Pacific Ocean, Mediterranean, and the Gulf of California (see table 1). 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 the file Supplementary data 1 (also found in the git repository) [see publication]. 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* (Fig. 2.1C) specimens were collected in the bay of Villefranche-sur-Mer, France, during a plankton trawl on 13 April 2011. Available physical vouchers have been deposited at the Museum of Comparative Zoology (Harvard University), Cambridge, MA, the Peabody Museum of Natural History (Yale University), New Haven, CT, or had been previously deposited at the Smithsonian National Museum of Natural History, Washington, DC. Accession numbers are given in Supplementary data 1 [see publication]. In cases where physical vouchers were unavailable we provide photographs to document species identity (see git repository: https://github.com/caseywdunn/siphonophore_phylogeny_2017/tree/master/supplementary_info/photographic_vouchers).

2.3.2 Sequencing

When possible, specimens were starved overnight in filtered seawater at temperatures close to ambient water temperatures at the time of specimen collection. Subsequently, mRNA was extracted directly from tissue using a variety of methods (Supplementary data 1 [see publication]): Magnetic mRNA Isolation Kit (NEB, #S1550S), Invitrogen Dynabeads mRNA Direct Kit (Ambion, #61011), Zymo Quick RNA MicroPrep (Zymo #R1050), or from total RNA after Trizol (Ambion, #15596026) extraction and through purification using Dynabeads mRNA Purification Kit (Ambion, #61006). In case of small total RNA quantities, only a single round of bead purification was performed. Extractions were performed according to the manufacturer's instruction. All samples were DNase treated (TURBO DNA-free, Invitrogen #AM1907; or on column DNase treatment with Zymo Quick RNA MicroPrep). Libraries were prepared 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. In the case of large tissue inputs, libraries were sequenced separately for each tissue, subsequently subsampled and pooled *in silico.* Libraries were sequenced on the HiSeq 2000, 2500, and 3000 sequencing platforms. Summary statistics for each library are given in the file Supplementary data 2 [see publication]. All sequence data have been deposited in the NCBI sequence read archive (SRA) with Bioproject accession number PRJNA255132.

2.3.3 Analysis

New data were analysed in conjunction with 14 publicly available datasets (Chapman et al., 2010; Dunn et al., 2013a; Fidler et al., 2014; Lehnert et al., 2012; Philippe et al., 2009; Putnam et al., 2007; Sanders et al., 2014; Sanders and Cartwright, 2015; Zapata et al., 2015), with a total number of 43 species. Sequence assembly, annotation, homology evaluation, gene tree construction, parsing of genes trees to isolate orthologous sequences, and supermatrix construction were conducted with Agalma (commit 6bd9988, running BioLite commit 784edc6) (Dunn et al., 2013a; Guang et al., 2017). This workflow integrates a variety of existing tools (Grabherr et al., 2011; Altschul et al., 1990; Enright et al., 2002; Katoh and Standley, 2013; Langmead and Salzberg, 2012; Li et al., 2009b; Li and Dewey, 2011; Sukumaran and Holder, 2010; Talavera and Castresana, 2007) and new methods.

Two outgroup species, *Atolla vanhoeffeni* and *Aegina citrea*, were removed from the final supermatrix due to low gene occupancy (gene sampling of 17.0% and 17.3% respectively in a 60% occupancy matrix with 3,379 genes). The final analyses presented here consider 33 siphonophore species and 8 outgroup species. This includes new data for 30 species. In the final analyses, we sampled 1,423 genes to generate a supermatrix with 80% occupancy and a length of 395,699 amino acids (Fig. S2.1).

We used ModelFinder (Kalyaanamoorthy et al., 2017), as implemented in IQTree v1.6.3 (Nguyen et al., 2015), to assess relative model fit. ModelFinder selected JTT + Empirically counted frequencies from alignment + FreeRate model with 7 categories based on the Bayesian Information Criterion. To assess the robustness

of our results, we conducted phylogenetic analyses using multiple software programs, methods (Maximum likelihood (ML) and Bayesian Inference (BI)), and models (including the model selected by ModelFinder and several other commonly used models). Maximum likelihood (ML) analyses were conducted with RAxML v8.2.0 (Stamatakis, 2006) and IQTree v1.6.3 (Nguyen et al., 2015; Hoang et al., 2018). Bayesian Inference (BI) were conducted with Phylobayes v. 1.7a-mpi (Lartillot et al., 2009). Sequence alignments, sampled and consensus trees, and voucher information are available in the git repository. Tree figures were rendered with ggtree (Yu et al., 2017).

RAxML ML analyses were conducted on the unpartitioned supermatrix using the WAG+Γ model of amino acid substitution (Fig. 2.3A). RAxML bootstrap values were estimated using 1000 replicates. IQTree ML analyses were run under JTT + Empirically counted frequencies from alignment + FreeRate model with 7 categories, the best model identified by ModelFinder, and the commonly used models GTR + Optimized base frequencies + Free rate model with 6 categories and WAG + Optimized base frequencies + Free rate model with 6 categories (Fig. S2.10).

BI was conducted in phylobayes using two different models: fixed-state WAG+Γ (Fig. S2.12) and CAT-Poisson (Fig. S2.11). Eight chains were run under the CAT-Poisson model. Four chains were run under WAG+Γ, these runs did not converge (maxdiff=1, meandiff=0.0130273). The CAT-Poisson runs did not converge (maxdiff=1, meandiff=0.0565898). Closer inspection revealed that chain1 and chain3 were stuck in local maxima with low likelihood relative to other chains after 1405 and 4695 iterations. These two chains were excluded from the analyses, and the results presented here are based on the remaining 6 CAT-Poisson chains (maxdiff=1, meandiff=0.0185032). Visual inspection of the traces indicated that a burn in of 400 trees was sufficient for all CAT-Poisson runs. This left 17893 trees in the CAT-Poisson posterior.

We used the Swofford-Olsen-Waddell-Hillis (SOWH) test (Swofford et al., 1996) to evaluate two hypotheses (Fig. 2.3C, S2): (i) "Physonectae" is monophyletic (Totton, 1965); (ii) monoecious species are monophyletic (Dunn et al., 2005a). The sexual mode of *Rudjakovia* is undescribed, but preliminary observations suggest that they are monoecious, so we include *Rudjakovia* as a monoecious species in this test. We used SOWHAT (Church et al., 2015a) dev. version 0.39 (commit fd68ef57) to carry out the SOWH tests in parallel, using the default options and an initial sample size of 100 (analysis code can be found in the git repository). For

each hypothesis we defined a topology with a single constrained node that was inconsistent with the most likely topology (Fig. 2.3). We used a threshold for significance of 0.05 and following the initial 100 samples, we evaluated the confidence interval around the p-value to determine if more samples were necessary.

Morphological character data used in trait mapping were obtained from the literature or direct observation of available voucher material. Depth distribution data was queried from the MBARI VARS database (http://www.mbari.org/products/research-software/video-annotation-and-reference-system-vars/) (Schlining and Stout, 2006). We used stochastic character mapping to infer the most probable evolution of traits on the tree in R using the `phytools` package (Huelsenbeck et al., 2003; Revell, 2012). For continuous character traits, model fit was tested using `fitContinuous` in the `geiger` R package. Subsequent analyses were conducted in R and integrated into this manuscript with the `knitr` package. See Supplementary Information for R package version numbers [see publication].

2.4 Results and Discussion

2.4.1 Species phylogeny and hypothesis testing

The phylogenetic relationships recovered in this study received strong support across analysis methods (Fig. 2.3A), with a couple of localized exceptions (Fig. 2.3B and S2.11). All of the ML analyses were congruent with each other, regardless of model and software used (Fig. S2.10). These ML results were also congruent with the Phylobayes BI WAG+Γ analyses (Fig. S2.12). The Phylobayes BI CAT-Poisson result (Fig. S2.11), however, had a strongly supported topology that differed (Fig. 2.3B) from the ML topology in localized regions as described below. The fact that the Phylobayes BI WAG+Γ is consistent with the WAG (and other) ML analyses suggests that the different topology recovered in the Phylobayes BI CAT-Poisson analyses is due to the different model rather than different software or methods. Here we take the conservative approach of considering relationships that differ between the Phylobayes BI CAT-Poisson analyses and other analyses to be unresolved.

Most clades are consistent with those found in a previous study based on two genes (16S and 18S ribosomal

New data	Species	Depth (m)	Lat Lon	SRA Number
Y&N	Nanomia bijuga	414/387	36.60 N 122.15 W	SRR1548376;SRR1548377;SRR871527
Y	Bargmannia elongata	412/805/636/818	36.12 N 122.67 W	SRR1548343–47
Y	Frillagalma vityazi	407	36.69 N 122.05 W	SRR1548362;SRR1548363;SRR1548364
Y	Apolemia rubriversa	767	36.70 N 122.05 W	SRR1548342
Y	Chelophyes appendiculata	3–20	43.696 N, 7.308 E	SRR1548354
Y	Chuniphyes multidentata	327	36.79 N 122.00 W	SRR1548355
Y	Cordagalma sp	252	36.70 N 122.06 W	SRR1548356
Y	Erenna richardi	1044	36.61 N 122.38 W	SRR1548360
Y	Forskalia asymmetrica	253	36.80 N 122.00 W	SRR1548361
Y	Hippopodius hippopus	3–20	43.69 N 7.315 E	SRR1548371
Y	Kephyes ovata	452	36.36 N 122.81 W	SRR1548372
Y	Lilyopsis fluoracantha	320	36.69 N 122.04 W	SRR1548373
Y	Lychnagalma utricularia	431	36.69 N 122.04 W	SRR1548374
Y	Marrus claudanielis	1427	36.07 N 122.29 W	SRR1548375
Y	Undescribed Sp. L	1463	36.70 N 122.57 W	SRR1548381
Y	Desmophyes sp.	1363	35.48 N 123.64 W	SRR1548358
Y	Resomia ornicephala	322	35.48 N 123.86 W	SRR1548382
Y	Rhizophysa filiformis	10	27.23 N 110.46 W	SRR1548383
Y	Stephalia dilata	3074	35.62 N 122.67 W	SRR1548384
Y	Apolemia lanosa	1073	36.70 N 122.08 W	SRR6512857
Y	Apolemia sp	461	36.60 N 122.15 W	SRR6512854
Y	Bargmannia amoena	1251	36.70 N 122.08 W	SRR6512862
Y	Bargmannia lata	1158	36.067 N 122.30 W	SRR6512863
Y	Rudjakovia sp	334	36.00 N 122.42 W	SRR6512851
Y	Stephalia sp	3255	36.39 N 122.67 W	SRR6512855
Y	Physophora gilmeri	242	36.36 N 122.40 W	SRR6512853
Y	Halistemma rubrum	313	24.68 N 109.90W	SRR6512852
Y	Athorybia rosacea	3–20	22.92 N 108.36 W	SRR6512856
Y	Diphyes dispar	3–20	35.93 N 122.93 W	SRR6512850;SRR6512858–61;SRR6512864;SRR6512867–68
	Agalma elegans	3–20	35.56 N 122.55 W	SRR6512865;SRR6512866
	Physalia physalis	0	13.831 N 129.943 W	SRR871528
	Abylopsis tetragona	3–20	43.696 N, 7.308 E	SRR871525
	Aegina citrea		36.697177 N 122.054095 W	SRS893439
	Aiptasia pallida			SRR6967; SRR6967; SRR6967
	Alatina alata		12.151891 N 68.278002 W	SRR1952741
	Atolla vanhoeffeni		36.707311 N 122.061062 W	SRR1952729
	Clytia hemisphaerica		43.696 N, 7.308 E	N/A
	Ectopleura larynx			SRR923510
	Hydra magnipapillata			N/A
	Hydractinia symbiolongicarpus			SRX474878
	Nematostella vectensis			N/A
	Podocoryna carnea			SRR1266262
	Craseoa lathetica	1530		SRR871529

Table 2.1: A complete list of specimens used in this work, information from already published datasets added where available. New data indicated by Y, blank fields indicate that data were already published. For the species not on SRA, a link to the data is included in supplementary data 1.

RNA) (Dunn et al., 2005a). Relationships that receive strong support in both include the placement of Cystonectae as sister to Codonophora (the clade that includes all other siphonophores), the placement of Apolemiidae as sister to all other codonophorans, and the placement of Calycophorae within the paraphyletic "Physonectae". Multiple nodes that were not resolved in the previous two-gene analysis receive strong support in the present 1,423-gene transcriptome analyses. There is strong support for Pyrostephidae as sister to all other non-apolemiid codonophorans. We provisionally refer here to Pyrostephidae as the clade including *Rudjakovia* sp., although sampling of *Pyrostephos vanhoeffeni* is needed in order to determine if *Rudjakovia* sp. falls within Pyrostephidae or is sister to it. Within the clade that is sister to Pyrostephidae, we find two main clades, Calycophorae and a clade we here name Euphysonectae (Fig. 2.3A). It includes the remaining

non-apolemiid, non-pyrostephid "Physonectae". We define Euphysonectae as the clade consisting of *Agalma elegans* and all taxa that are more closely related to it than to *Diphyes dispar*.

In ML analyses and BI WAG analyses, Euphysonectae consists of two reciprocally monophyletic groups that we here provisionally refer to as Clade A and Group B (Fig. 2.3A). In BI CAT-Poisson analyses, Group B is paraphyletic (Fig. 2.3B). The presence of an involucrum, a fold around the base of the cnidoband (Totton, 1965), is a potential synapomorphy for Clade A. Species of Clade A also have a descending mantle canal within the nectophores (Fig. S2.6, S2.18), a structure that is also present in some calycophorans. Members of Clade A are also monoecious (Fig. 2.5). There is not a clear synapomorphy for Group B. Within Group B there is high support for the placement of *Erenna richardi* in ML analyses and BI WAG (Fig. 2.3 and S2.12), but it is placed as sister to Clade A in BI CAT-Poisson analyses (Fig. 2.3B). More taxon sampling will be required to determine the relationship of species within this group.

Within Clade A, *Physophora gilmeri* along with *Lychnagalma utricularia* (Fig. 2.1I) (both not included in the previous phylogeny) are sister to Agalmatidae, a clade restricted to *Agalma*, *Athorybia*, *Melophysa*, *Halistemma* and *Nanomia* (Dunn et al., 2005a; Pugh, 2006). In the rDNA study, *P. hydrostatica* (the presumed sister species to *P. gilmeri*) was sister to Forskaliidae with low support. The position of *Cordagalma cordiforme* (= *C. ordinatum*) (Pugh, 2016) was previously unresolved, while in this analysis *Cordagalma* sp. is in a clade with *Forskalia asymmetrica*, falling outside of Agalmatidae. Placement of *Cordagalma* outside Agalmatidae is consistent with previous analyses of molecular and morphological data (Dunn et al., 2005a; Pugh, 2006).

Within Calycophorae, taxon sampling is less comprehensive here than in the previous study. The calycophoran relationships that can be investigated, however, are in broad agreement with the previous analysis. Calycophorans have in the past been split into two groups, prayomorphs and diphyomorphs, based on morphology after Mackie et al. (1987). As in the previous study, the results presented here indicate that the prayomorphs are paraphyletic with respect to the diphyomorphs. In the previous study, the relationship between *C. lathetica* and the clade including *H. hippopus* was unresolved. In this study, *Craseoa lathetica* and *Desmophyes* sp. are sister to *Hippopodius hippopus* in ML and BI-WAG analyses with high support, while in BI CAT-Poisson analyses, *H. hippopus* is sister to *Lilyopsis fluoracantha* and the diphyomorphs (Fig.

34

2.3B and S2.11).

Using the Swofford-Olsen-Waddell-Hillis (SOWH) test (Swofford et al., 1996), we evaluated the following two alternative phylogenetic hypotheses against the most likely tree topology (Fig. 2.3C): (i) monophyletic Physonectae, (ii) monophyletic monoecious siphonophores. In both tests the alternative hypothesis was rejected (p-value <0.01, confidence interval: <0.001 - 0.03, Fig. S2.2).

The broad taxon sampling and more extensive gene sampling of this phylogeny provide new evidence for the relationships between major siphonophore clades within Codonophora, specifically between Pyrostephidae, Calycophorae, and the newly named Euphysonectae. This opens up new questions about key relationships within both Calycophorae and Euphysonectae – where future transcriptome sampling efforts should be focused. Within Euphysonectae, two clades (Clade A and Group B) are hypothesized, although there is weaker support for Group B (Fig. 2.3A, 3B). Expanding sampling of species that probably fall in Group B, including other *Erenna* species, rhodaliids, and relatives of Undescribed sp L, will greatly expand our understanding of these two groups and perhaps provide evidence of Group B synapomorphies. Similarly, within Calycophorae, increased taxon sampling is needed. This study, and the previous phylogenetic study (Dunn et al., 2005a), suggest that the prayomorphs are paraphyletic, but for slightly different reasons given the different sampling of the analyses. In Dunn et al. (2005a), a clade of prayomorphs including *Praya dubia* (Fig. 2.1G), *Nectadamas diomedeae*, and *Nectopyramis natans* (not included in this study) were found to be sister to all other calycophorans, while in this study, the prayomorph *Lilyopsis fluoracantha* (not included in the previous study) is found in a clade including diphyomorph calycophorans that is sister to all other prayomorphs. Expanded transcriptome sequencing, particularly *P. dubia* or a nectopyramid, but also extensive sampling across the major prayomorph and diphyomorph groups, will expand our understanding of relationships within Calycophorae. This will be especially important for understanding trait evolution within Calycophorae, for example, the release of eudoxids (Fig. 2.4), or the arrangement of male and female zooids along the stem (see section 2.4.2 below).

Figure 2.3: (A) Maximum likelihood (ML) phylogram with bipartition frequencies from the ML bootstraps and the Bayesian posterior distribution of trees. Unlabeled nodes have support >0.99 for both bootstraps and posteriors. The numbers of valid described species estimated to be in each clade based on taxonomy are shown below each clade name on the right. (B) The topologies found in the posterior distribution of trees that conflict with the ML tree. (C) The topologies evaluated by the SOWH tests. For more details on the SOWH topologies refer to Fig. S2.2.

2.4.2   Evolution of Monoecy

In all siphonophores, each gonophore (sexual medusa that produces gametes) is either male or female. Within each siphonophore species, colonies are either monoecious (male and female gonophores are on the same colony) or dioecious (male and female gonophores are on different colonies). Previous analysis suggested that the common ancestor of siphonophores was dioecious, and was consistent with a single gain of monoecy within Codonophora and no secondary losses (Dunn et al., 2005a). The better-resolved tree in the current analyses indicates that the evolution of monoecy is more complicated than this. The two clades of monoecious siphonophores, Calycophorae and Clade A (Fig. 2.3A), do not form a monophyletic group. This is because Group B, which contains dioecious species, is also descended from their most recent common ancestor. The SOWH test strongly rejects the placement of the monecious clades Calycophorae and Clade A as a group that excludes Group B (Figs. 3C and S2). The positions of the only two taxa from Group B that were included in the previous analysis (Dunn et al., 2005a), *Erenna* and *Stephalia*, were unresolved in that study. This difference in conclusions regarding trait evolution, therefore, does not reflect a contradiction between alternative strongly supported results, but the resolution of earlier polytomies in a way that indicates there has been homoplasy in the evolution of monoecy.

The distribution of monoecy is consistent with two potential scenarios (Fig. 2.4). In the first, there is a single shift from dioecy to monoecy along the branch that gave rise to the most recent common ancestor of Calycophorae and Euphysonectae, followed by a shift back to dioecy along the branch that gave rise to Group B. In the second, monoecy arose twice: once along the branch that gave rise to Clade A and again along the branch that gave rise to Calycophorae.

Ancestral character state reconstructions favor the hypothesis that monoecy arose twice (Fig. 2.5A and S2.13), once in Calycophorae and once in Clade A. This is consistent with differences in the arrangements of male and female gonophores in the two clades. In Clade A, male and female zooids are found within the same cormidium (a single reiterated sequence of zooids along the stem, see Fig. 2.2). In these species, the male and female zooids are placed at different but well defined locations within the cormidium. Meanwhile in calycophorans, each cormidium bears either male or female gonophores. In this form of monoecy, the

male and female cormidia can either occur in an alternating pattern, or there can be several male or female cormidia in a row. In either case, male and female zooids are found at the same corresponding locations within the cormidia. One known exception to this can be found in abylid calycophorans, where both male or female gonophores may be found within the same eudoxid (Carré, 1967). In sum, homoplasy in sexual-system evolution along with variation in the spatial arrangement of gonophores within a colony suggest that siphonophores have evolved different ways to be monoecious. The sexual system and cormidial arrangement of *Rudjakovia* is undescribed, although preliminary observations suggest that this species may be monoecious and that monoecy arose a third time in the Pyrostephidae. A detailed redescription of *Rudjakovia* would help clarify this.

Both Calycophorae and Clade A have a large proportion of shallow water species (see section 2.4.6), suggesting that there may be an association between habitat depth and sexual mode. Similar independent transitions from gonochorism (separate sex) to hermaphroditism (both sexes in the same individual) have been identified in shallow-water scleractinian corals (Anthozoa, Cnidaria) (Kerr et al., 2011). To test this hypothesis, a more extensive taxon sampling of the Calycophorae is needed.

Within Calycophorae there are additional variations of the sexual mode: in *Sulculeolaria* (not included in this phylogeny) colonies appear to present a single sex at a time. However they are monoecious and protandrous, with female gonophores developing after the release of male gonophores (Carré, 1979). Environmental influences may also play a role in determining the expressed sex. Colonies of the calycophoran *Chelophyes appendiculata* collected in the field always bear both male and female gonophores, whereas when kept in culture only gonophores of one sex are maintained (Carré and Carré, 2000). This suggests a high plasticity of the sexual state in some calycophoran taxa and underlines the need for caution when evaluating the state of this character in rarely collected species.

2.4.3 The Evolution of Zooid Types

One of the most striking aspects of siphonophore biology is their diversity of unique zooid types (Beklemishev, 1969; Cartwright and Nawrocki, 2010). For example, *Forskalia* and other physonects have at least 5 basic

zooid types (nectophore, gastrozooid, palpon, bract, and gonophore), and in some species, there can be nine zooid subtypes (4 types of bract, male & female gonophores)(Pugh, 2003). Here we reconstruct the evolutionary origins of several zooid types on the present transcriptome-based tree (Fig. 2.4).

Nectophores (Fig. 2.2) are non-reproductive propulsive medusae. In Codonophora, the nectophores are localized to a region known as the nectosome (Fig. 2.2B), which has its own growth zone, and they are used for coordinated colony-level swimming. Planktonic cystonects like *Bathyphysa sibogae* and *Rhizophysa filiformis* (Fig. 2.1A) instead move through the water column using repeated contraction and relaxation of the stem, and in the case of *B. sibogae*, use modified flattened gastrozooids with wings (called ptera) to increase surface area and prevent colony sinking (Biggs and Harbison, 1976). Nectophores are also present within the gonodendra (reproductive structures) of cystonects, and are thought to propel the gonodendra when they detach from the colony (Totton, 1965, 1960). It is not clear whether the nectophores found within the siphosome of the cystonects are homologous to the nectophores borne on the nectosome of codonophorans. Similarly, the homology of the special nectophore associated with gonophores of the calycophoran *Stephanophyes superba* is also unclear (Chun, 1891). In this study, we only consider the evolution of the nectosome, and not the presence/absence of nectophores. The present analyses, as well as the analyses of Dunn et al. (2005a), are consistent with a single origin of the nectosome (Fig. S2.5, S2.17).

Within the nectosome, the nectophores can be attached along the dorsal or ventral side of the stem, following the orientation framework of Haddock et al. (2005). The apparent placement of the nectophores on opposite sides of the nectosome occurs through twisting of the stem during development. Our ancestral reconstructions for this character (Fig. S2.7, S2.19) suggest that ventral attachment of nectophores was the ancestral state in Codonophora, and that dorsal attachment has independently evolved twice – once along the stem of Agalmatidae and once along the stem of Pyrostephidae. The functional implication of dorsal vs. ventral attachment is not clear.

Bracts are greatly reduced zooids unique to siphonophores, where they are only present in Codonophora (Fig. 2.4). Bracts are functional for protection of the delicate zooids and to help maintain neutral buoyancy (Jacobs, 1937). Some calycophorans are able to actively exclude sulfate ions in their bracts to adjust their buoyancy along the colony (Bidigare and Biggs, 1980). Bracts were lost in Hippopodiidae, some clausophyids,
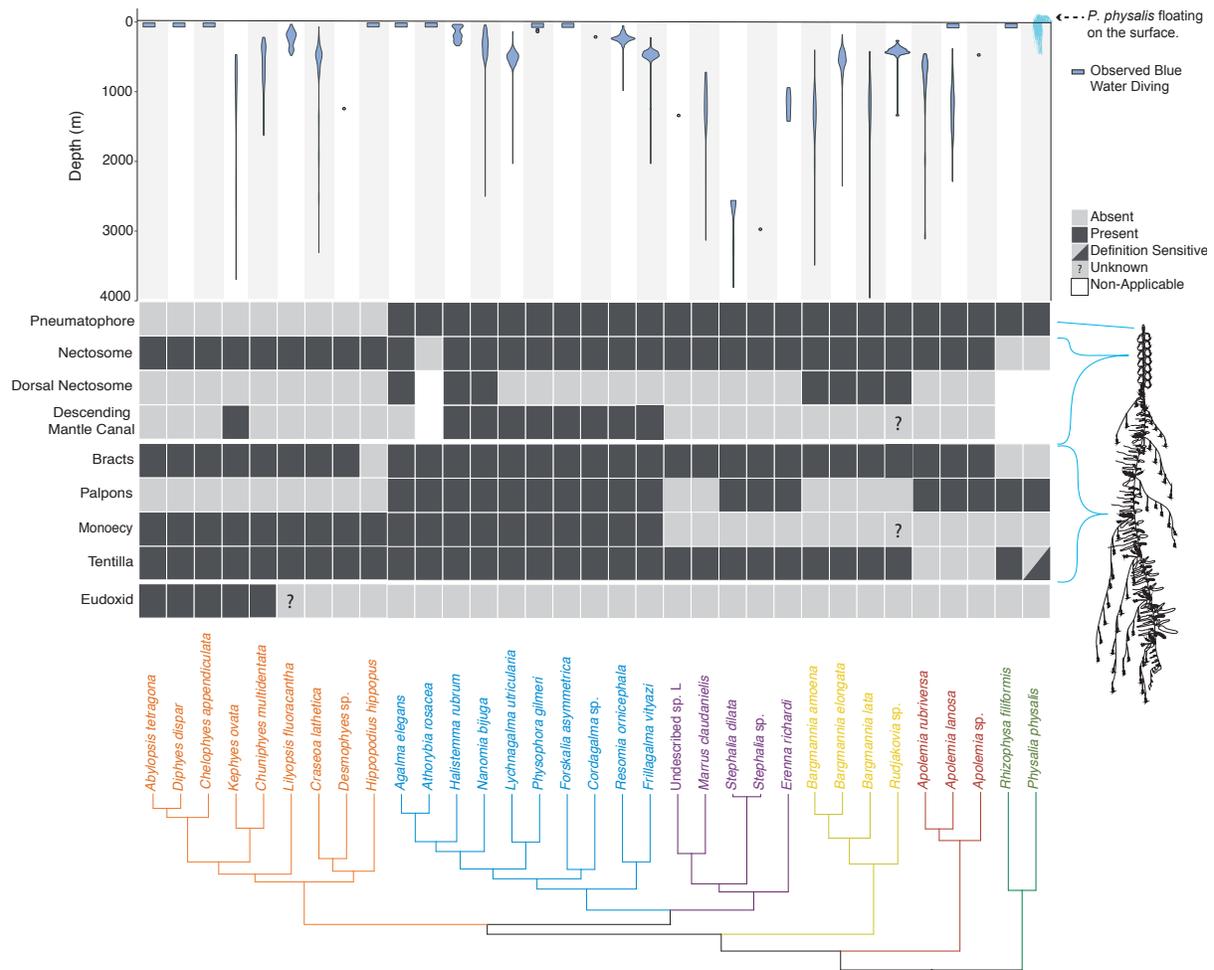
Figure 2.4: Siphonophore ML phylogeny showing the distribution of the main anatomical characters and the bathymetric distributions of the different species. Bottom: siphonophore ML phylogeny, colored by clade. Middle panel: diagram showing the presence/absence of traits across Siphonophora, with the physical location of the trait shown on a schematic of *Nanomia bijuga* (schematic by Freya Goetz). Top: Bathymetric distribution of siphonophore species. *Physalia* illustration by Noah Schlottman, taken from http://phylopic.org/

Physophora hydrostatica (Fig. 2.1K), and in *Gymnopraia lapislazula.* These patterns of loss are not captured in this study, as most of these species are not included in the present phylogeny. In species without bracts, other zooids appear to fulfill the roles of buoyancy control and protection. In *P. hydrostatica*, enlarged palpons surround all other siphosomal zooids and move in a coordinated manner to inflict a powerful sting (Totton, 1965). While in *Hippopodius hippopus* the nectophores play a role in maintaining neutral buoyancy and possibly also in defense, by bioluminescing and blanching in response to stimuli (Fig. 2.1C shows the blanching of nectophores)(Bassot et al., 1978).

Palpons are typically defined as modified reduced gastrozooids (Mackie et al., 1987). In many species palpons are thought to play a role in digestion and circulation of the gastrovascular fluid, while other species may use them for defense (e.g *Physophora*) or sensory functions (Totton, 1965). Palpons are subcategorised based on their location - palpons that are associated with gonodendra are termed gonopalpons (typically with a reduced tentacle, called a palpacle); palpons found along the stem of the siphosome are termed palpons (typically have a palpacle); and palpons found along the stem of the nectosome are termed nectosomal palpons (as in *Apolemia*) (Siebert et al., 2013; Totton, 1965). It is not clear how structure and function differs among different palpon subtypes, and a detailed histological investigation of palpons found at different locations within species is needed. For this reason, here we only assess the presence or absence of palpons as a category, without assessing subtypes of palpons. This presumes that palpons located at different regions in the colony are derived from other palpons rather than each arising *de novo* by independent modification of gastrozooids, a hypothesis that itself could be challenged upon closer histological examination of palpon diversity.

We reconstruct palpons as present in the common ancestor of siphonophores (Fig. 2.5B, S2.14), retained in most species, but lost three times independently in the branches leading to *Bargmannia* and *Rudjakovia* sp., in calycophorans, and in *Marrus claudanielis* and Undescribed sp. L. It remains to be clarified if small buds associated with nectophores within the nectosome of *Bargmannia* species (Dunn, 2005) actually correspond to reduced palpons. The pyrostephid *Pyrostephos vanhoeffeni* (not sampled) has modified tentacle-less palpons (termed oleocysts), but the relationship between this species and *Rudjakovia* sp. is not known, so the exact patterns of loss within *Pyrostephidae* (here provisionally including *Rudjakovia* sp.) remain unclear. Within the calycophorans, one species *Stephanophyes superba* (not included in this phylogeny) has polyp-like zooids
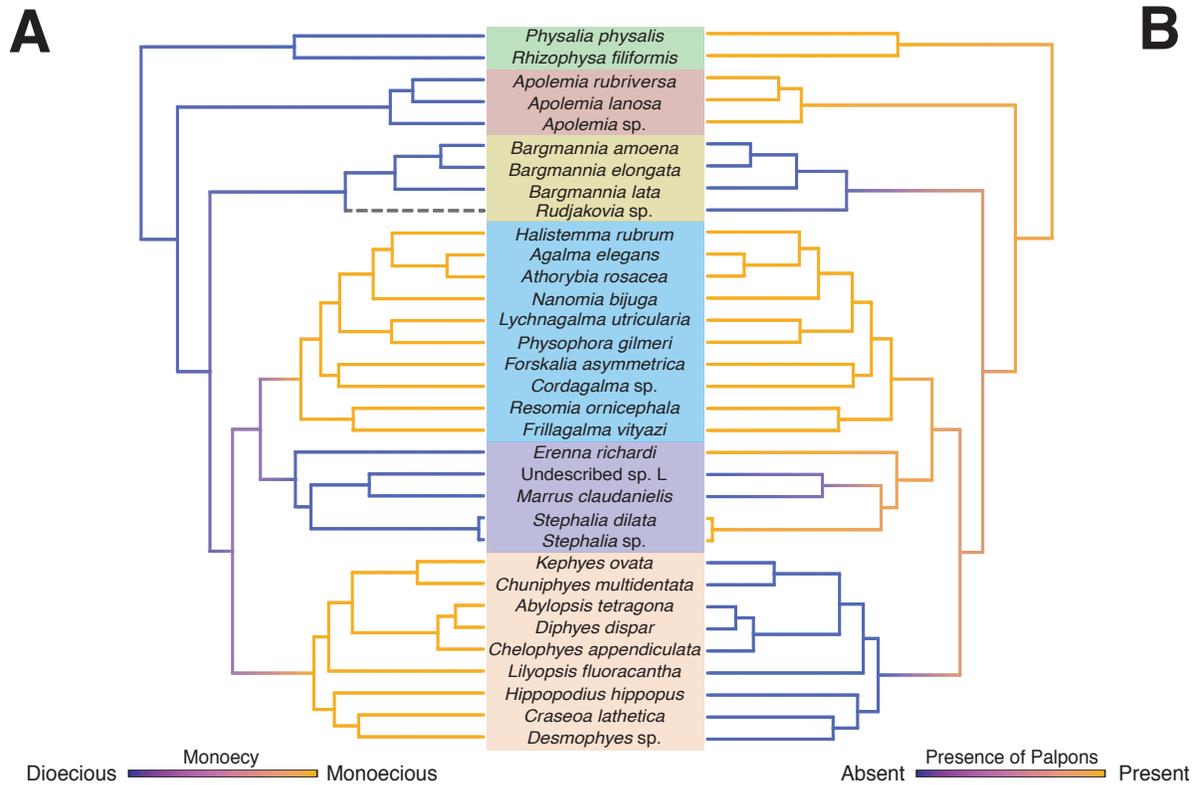
Figure 2.5: Stochastic mapping reconstruction on the ML tree of the evolutionary history of (A) sexual mode, whether a colony is monoecious or dioecious and (B) presence/absence of palpons (modified reduced gastrozooids). The color gradients show the reconstructed probability estimate of the discrete character states along the branches. Intermediate values reflect uncertainty. The grey dashed branch leading to *Rudjakovia* sp. indicates that the sexual mode of this species is unknown.

that have been described as palpons (Totton, 1965), but the exact identity of this zooid is not clear and needs further morphological examination.

2.4.4 The Gain and Loss of the Pneumatophore

The pneumatophore (Fig. 2.2A) is a gas-filled float located at the anterior end of the colony, which helps the colony maintain its orientation in the water column, and plays a role in flotation in the case of the cystonects (Church et al., 2015b; Mackie, 1974; Totton, 1965). It is not a zooid, as it is not formed by budding but by invagination at the aboral end of the planula during early development (Carré, 1969; Garstang, 1946; Leloup, 1935). Recent descriptions of the neural arrangement in the pneumatophore of *Nanomia bijuga* suggest it could also gather information on relative pressure changes (and thus depth changes), helping regulate geotaxis

(Church et al., 2015b). The ancestral siphonophore had a pneumatophore (Fig. 2.2B), since both cystonects and all "physonects" possess one (Fig. 2.4). The pneumatophore was lost in Calycophorae and never gained again in that clade. Calycophorans rely on the ionic balance of their gelatinous nectophores and bracts to retain posture and neutral buoyancy (Mackie, 1974).

2.4.5   The Gain and Loss of Tentilla

Gastrozooids (specialized feeding polyps) have a single tentacle attached to the base of the zooid that is used for prey capture (with the exception of *Physalia physalis*, which has separate zooids for feeding and prey capture, and rhodalids, where some tentacles are used to anchor to the substrate and do not participate in feeding). As in other cnidarians, stinging capsules, arranged in dense batteries of nematocysts, play a critical role in prey capture. In many siphonophore species these batteries are found in side branches of the tentacle, termed tentilla (Fig. 2.2A). Outside of Siphonophora, most hydrozoans bear simple tentacles without side branches. It is still an open question whether the common ancestor of Siphonophora had tentilla. The only siphonophores species regarded as lacking tentilla are *P. physalis*, *Apolemia* spp. (Fig. 2.1H), and *Bathyphysa conifera* (Fig. 2.1B). Since *B. conifera* is the only member of the Rhizophysidae (and of the *Bathyphysa* genus) lacking tentilla, we assume this is a case of secondary loss. When we reconstruct the evolution of this character on the current phylogeny, 70% of simulations support a common ancestor bearing tentilla, with two independent losses leading to *Physalia* and *Apolemia* (Fig. S2.3, S2.15). However, this leaves a 30% support for a simple-tentacled common ancestor followed by 2 independent gains of tentilla in the branches leading to Rhizophysidae and non-apolemiid codonophorans.

How we define absence of tentilla, especially for *Physalia physalis*, is also important. The tentacles of this species, when uncoiled, show very prominent, evenly spaced, bulging buttons which contain in the ectoderm all functional nematocytes (carrying mature nematocysts) used by the organism for prey capture (Hessinger and Ford, 1988; Totton, 1960). Siphonophore tentilla are complete diverticular branchings of the tentacle ectoderm, mesoglea, and gastrovascular canal (lined by endoderm). *Physalia*'s buttons enclose individual fluid-filled chambers connected by narrow channels to the tentacular canal, lined by endoderm (Bardi and Marques, 2007). This suggests they are not just ectodermal swellings, but probably reduced tentilla. When

we define *P. physalis* as tentilla bearing, the results for the character reconstruction lead to a more robust support for a tentilla-bearing common ancestor followed by independent losses of tentilla in the branch leading to Apolemiidae (Fig. S2.4, S2.16), and in *Bathyphysa conifera*. The application of phylogenetic methods to the evolution of tentillum morphology would be a crucial step towards understanding the evolution of these structures, and their relationship with the feeding ecology of siphonophores.

2.4.6   The Evolution of Vertical Habitat Use

Siphonophores are abundant predators in the pelagic realm, ranging from the surface (*Physalia physalis*) to bathypelagic depths (Fig. 2.4, S2.8, S2.20) (Mackie et al., 1987; Mapstone, 2014). The depth distribution of siphonophore populations is not always static, as some species are known to be vertical migrators, although this is within a relatively narrow depth range (<100m) (Pugh, 1984). Some species such as *Nanomia bijuga* exhibit synchronous diel migration patterns (Barham, 1966). Using the present phylogeny, we reconstructed the median depth changes along the phylogeny under a Brownian Motion model (Fig. S2.8 and S2.20), which had the strongest AICc support (compared to non-phylogenetic distributions, and to Ohrnstein-Uhlenbeck). This model indicates a mesopelagic most recent common ancestor, with several independent transition events to epipelagic and bathypelagic waters. There was only a single transition to benthic lifestyle on the branch of Rhodaliidae, and a single transition to a pleustonic lifestyle on the branch of *P. physalis*. There is evidence that habitat depth is conserved within some clades, with the exception of Calycophorae which have diversified across the water column (Fig. S2.8 and S2.20). Under the ML topology, depth appears to be phylogenetically conserved in Euphysonectae after the split between Clade A (shallow-living species) and Group B (deep-dwelling species) (Fig. S2.8), while under the BI-CAT topology, a mesopelagic common ancestor is predicted, with a transition to epipelagic waters in Clade A (Fig. S2.20); however several shallow-living species that likely belong in Group B were not included in this analysis. The present sampling is also not sufficient to capture significant variation in depth distributions between closely related species. Previous studies have shown that many species that are collected at the same locality are found to occupy discrete, largely non-overlapping depth distributions, including between species that are closely related (Pugh, 1974). This suggests that vertical habitat use is more labile than it appears and

may be an important mechanism in siphonophore ecology. The observed variation in depth distribution could be attributed to any of the correlated environmental variables (i.e. temperature, chlorophyll, oxygen). Temperature has been hypothesized to impose physiological limits to the dispersal of some clausophyid siphonophores (Grossmann et al., 2015). Since most of our specimens were sampled only in the Monterey Bay region, our analyses of the local oceanographic and depth distribution data cannot disentangle the effects of these different variables on the vertical distributions.

This reconstruction (Fig. S2.8 and S2.20) only included depths recorded using an ROV, thus it excludes many other independent colonizations of the epipelagic habitat. The ROV observations are reliable below 200m, and no quantitative measurements were made on SCUBA dives. Species such as *Nanomia bijuga*, *Hippopodius hippopus*, *Athorybia rosacea*, *Diphyes dispar*, and *Chelophyes appendiculata* are often encountered blue water diving less than 20m from the surface (Fig. 2.4). We also reconstructed the median depth changes along the phylogeny using median depths of 20m for all species collected by SCUBA diving or via a shallow trawl (Fig. S2.9 and S2.21), and still find support for a mesopelagic ancestor. It should be noted, however, that *H. hippopus* and *C. appendiculata* were both collected in the bay of Villefrance-sur-mer, France, where an upwelling is known to bring deeper species closer to the surface (Nival et al., 1976). Additionally, while we are confident about many of the species IDs in the VARS dataset, it is difficult to distinguish *Kephyes ovata* and *K. hiulcus* from images alone and the distribution likely includes data points from both species. *Halistemma rubrum* distributions were obtained from cruises in the Gulf of California, where the only *Halistemma* species collected by ROV is *H. rubrum*. Where we could not be certain of species identifications in the VARS dataset, we only included a few data points from specimens that were collected and identified.

2.5 Conclusions

Using phylogenomic tools we were able to resolve deep relationships within Siphonophora with strong support. We identify the clade Euphysonectae as the sister group to Calycophorae. Our results suggest that monoecy arose at least twice, based both on phylogenetic reconstruction and differences in the way monoecy is realized in different clades. We are unable to fully capture some of the complex patterns of zooid gain

and loss within Codonophora, which will require greater taxon sampling and improved morphological under-standing of many poorly known species. The improved resolution presented in this study suggests that an important next step in understanding siphonophore evolution will be targeting molecular sampling within Euphysonectae (where we sampled 13 of 62 valid described species that likely belong to the group) and Calycophorae (where we sampled 9 species in a clade of 109 valid described species) to further resolve the internal relationships within these clades.

2.6 Supplementary Information

The following figures are included as supplementary information for this chapter. Additional files can be found at https://github.com/caseywdunn/siphonophore_phylogeny_2017 and alongside the published manuscript (doi:10.1016/j.ympev.2018.06.030).
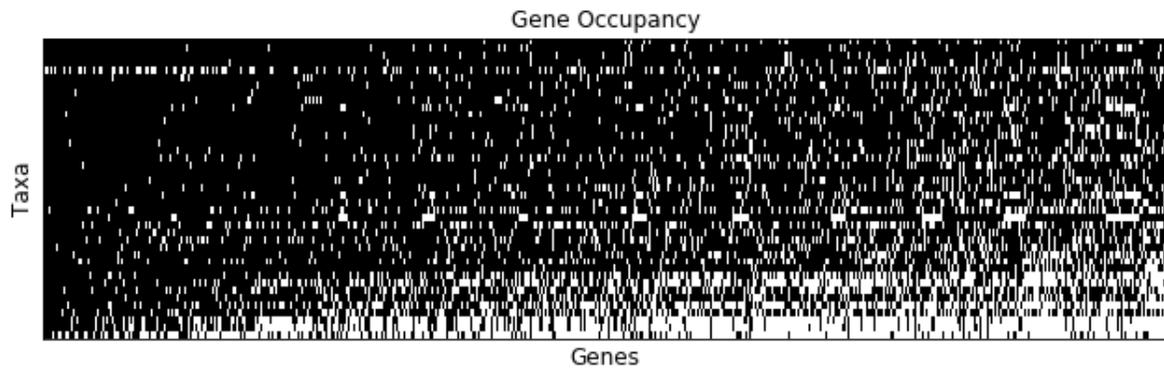
Figure S2.1: 80% gene occupancy matrix for 41 species across 1,423 genes.
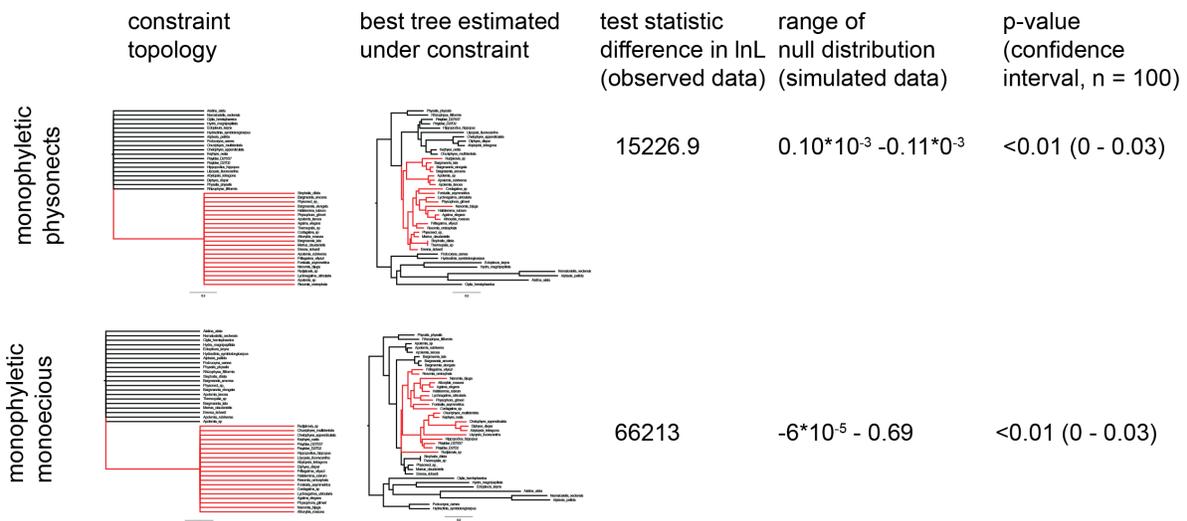
Figure S2.2: Constrained topologies specified in SOWH testing. Test statistic and p-value for each tree estimated under constraint are given.
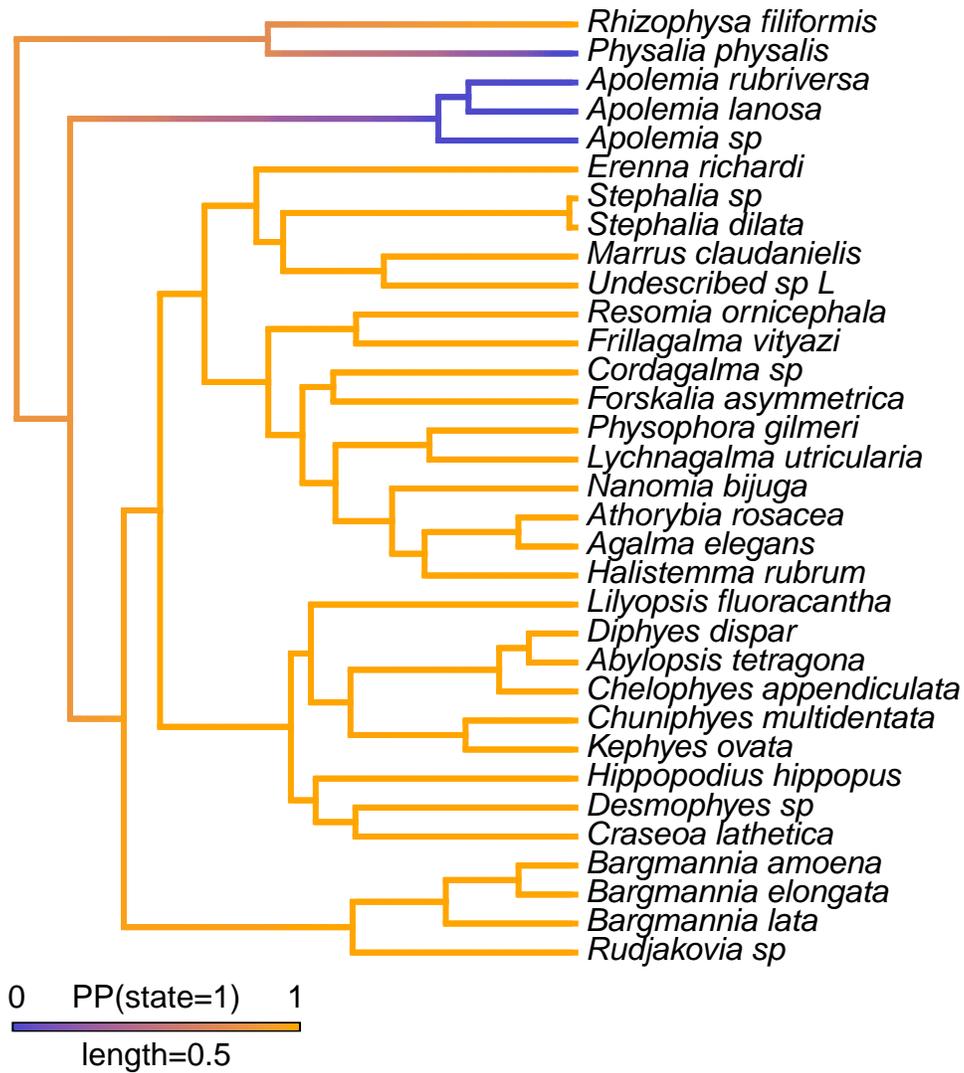
Figure S2.3: Stochastic character map of presence of tentilla with *Physalia* included as not bearing tentilla.
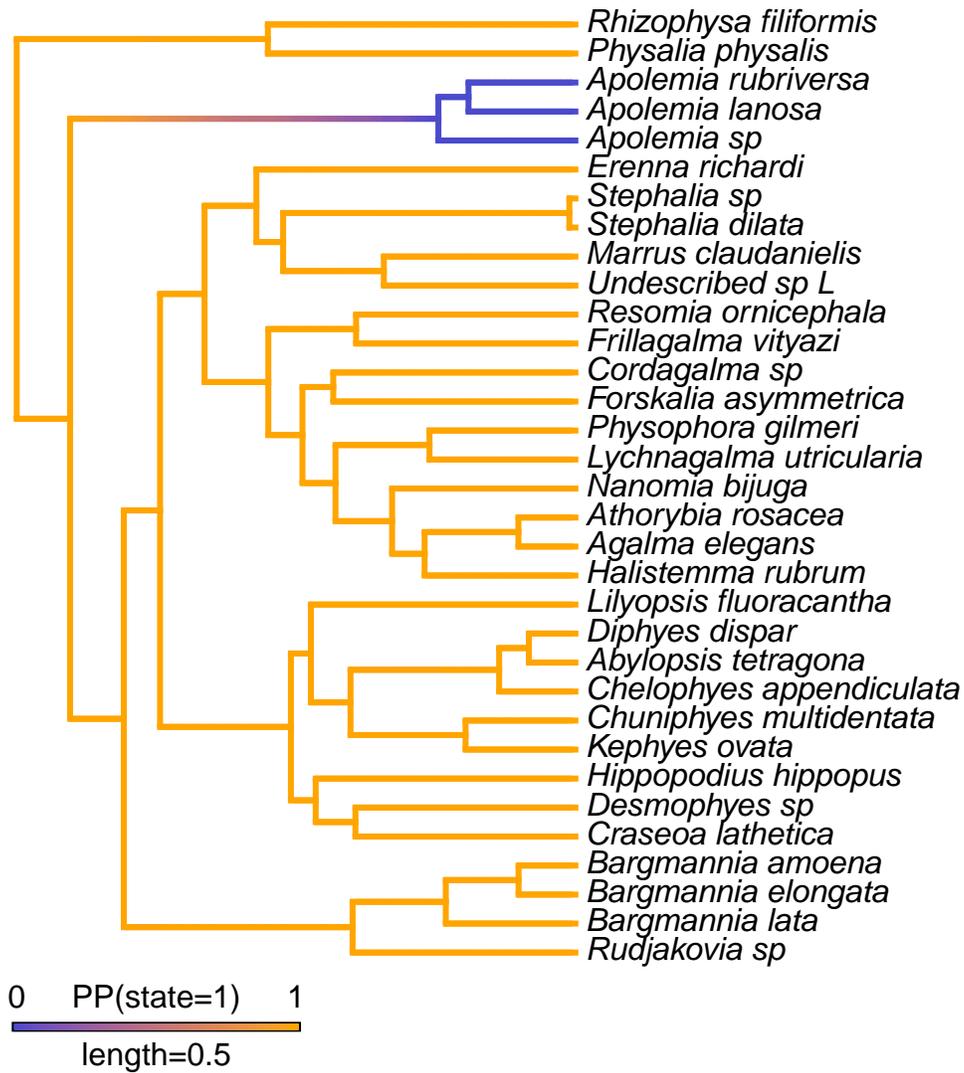
Figure S2.4: Stochastic character map of presence of tentilla with *Physalia* included as bearing tentilla.

Rhizophysa filiformis
Physalia physalis
Apolemia rubriversa
Apolemia lanosa
Apolemia sp
Erenna richardi
Stephalia sp
Stephalia dilata
Marrus claudanielis
Undescribed sp L
Resomia ornicephala
Frillagalma vityazi
Cordagalma sp
Forskalia asymmetrica
Physophora gilmeri
Lychnagalma utricularia
Nanomia bijuga
Athorybia rosacea
Agalma elegans
Halistemma rubrum
Lilyopsis fluoracantha
Diphyes dispar
Abylopsis tetragona
Chelophyes appendiculata
Chuniphyes multidentata
Kephyes ovata
Hippopodius hippopus
Desmophyes sp
Craseoa lathetica
Bargmannia amoena
Bargmannia elongata
Bargmannia lata
Rudjakovia sp

0   PP(state=1)   1

length=0.5

Figure S2.5: Stochastic character map of presence of nectosome.
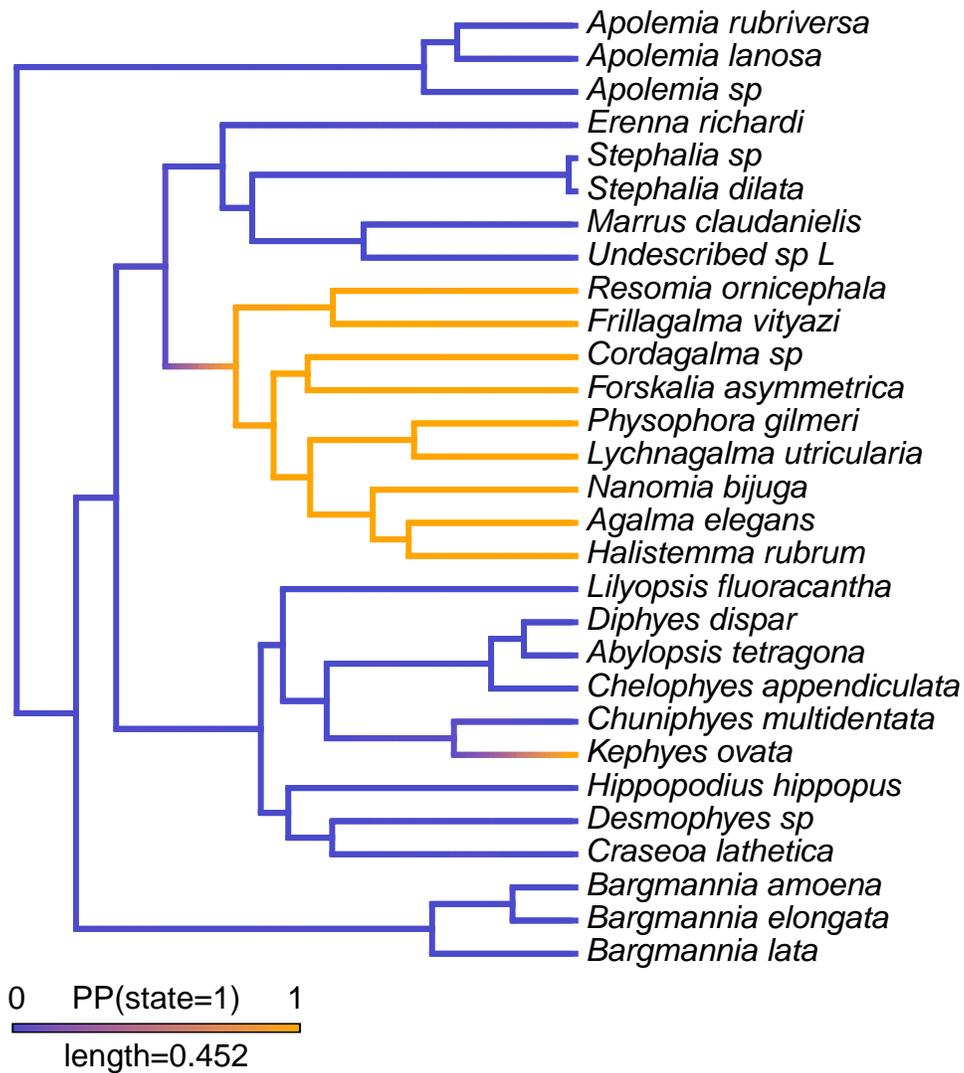
Figure S2.6: Stochastic character map of presence of a descending mantle canal in the nectophores. Cystonects and Athorybia were excluded as they do not have a nectosome.
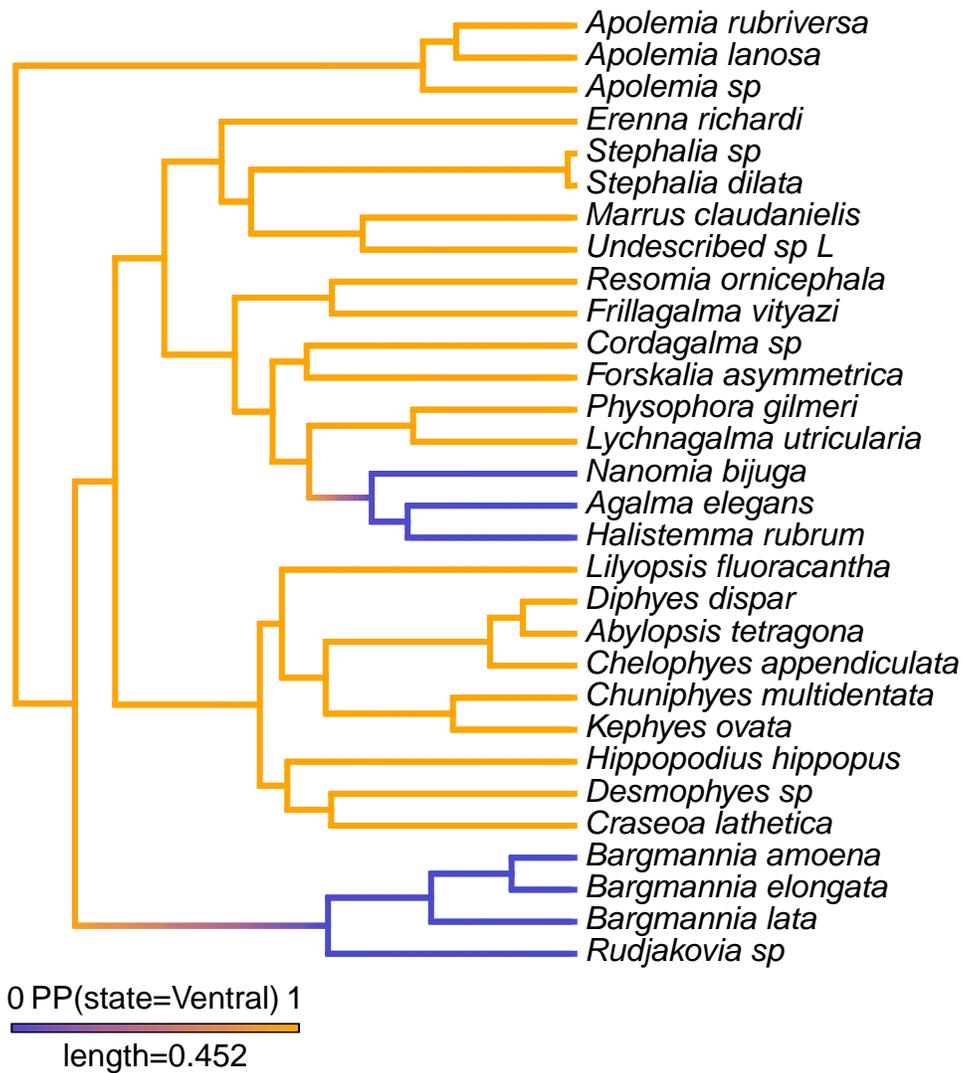
Figure S2.7: Stochastic character map for the evolution of the position of the nectosome. Cystonects were excluded as they do not have a nectosome.
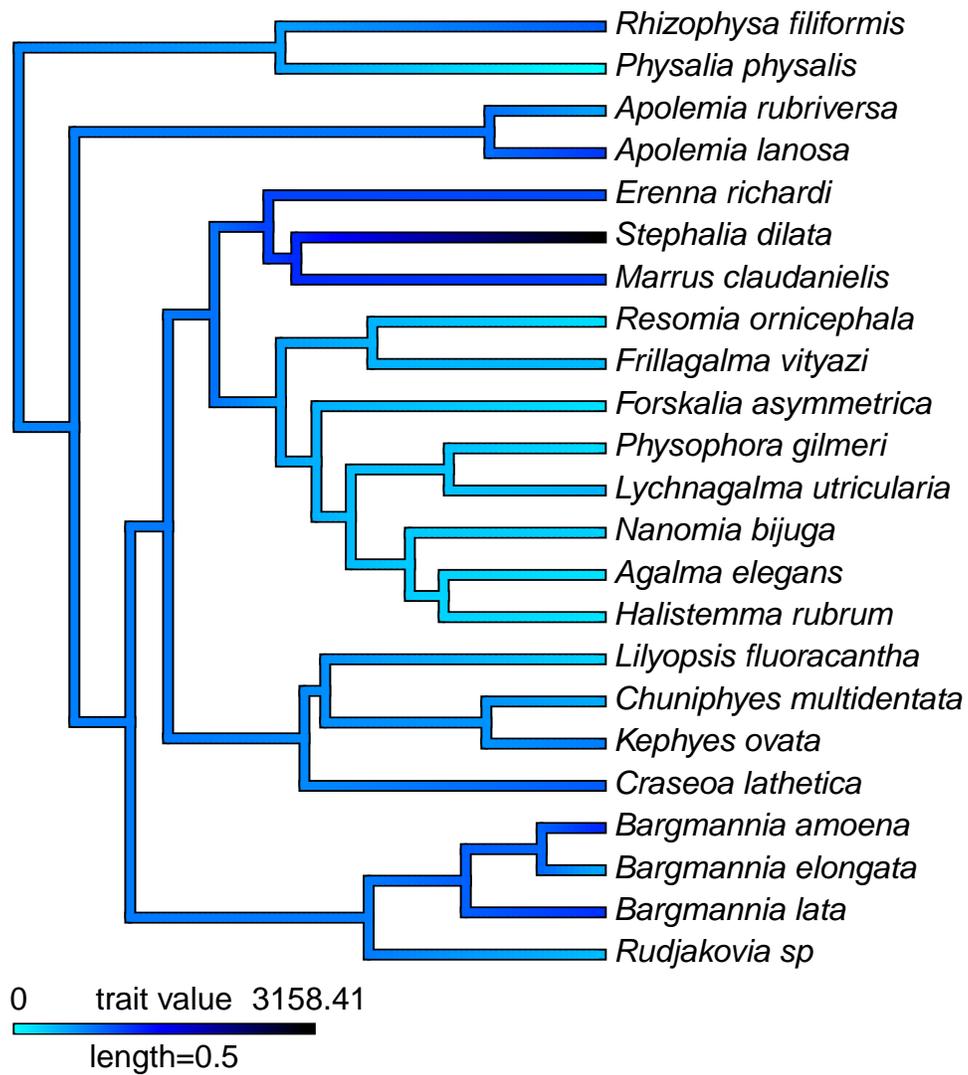
Figure S2.8: Brownian Motion character map of median depth of species observed with an MBARI ROV.
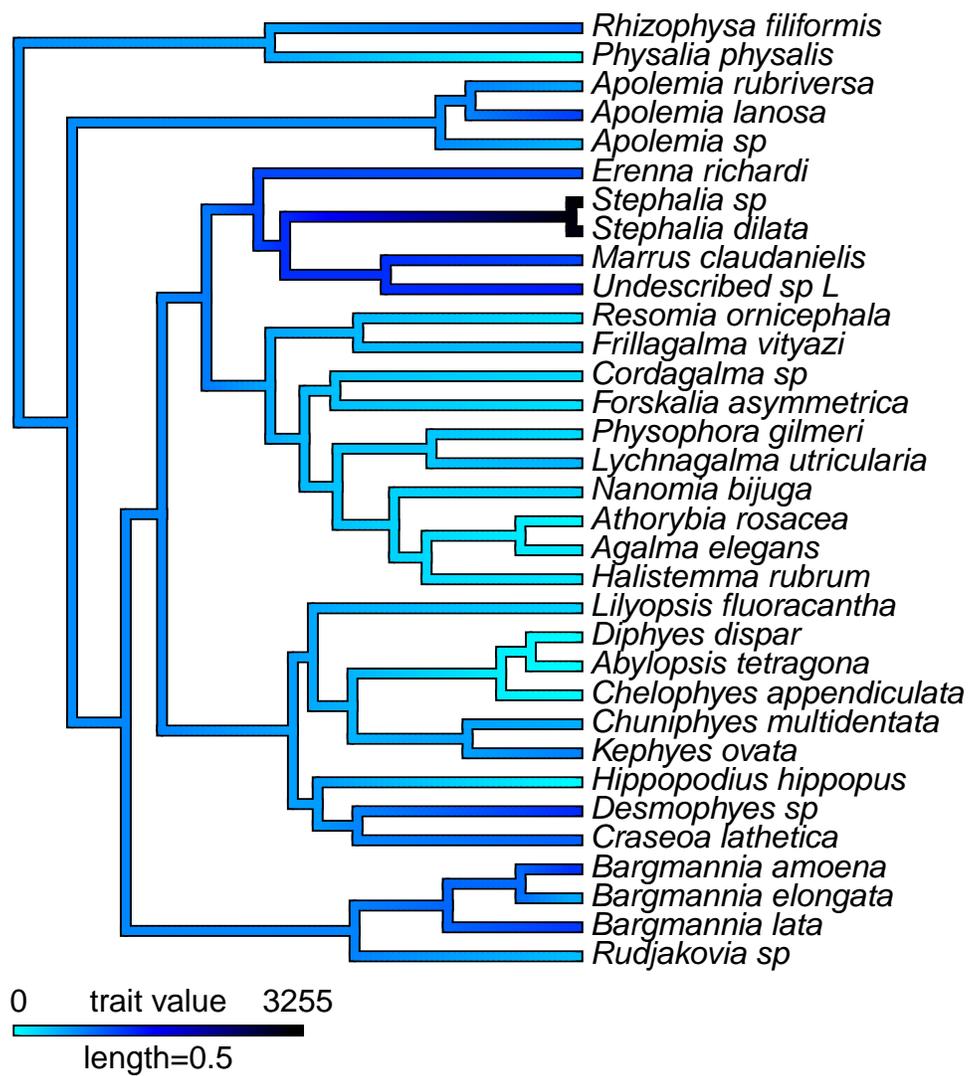
Figure S2.9: Brownian Motion character map of median depth of species including blue water diving observations.
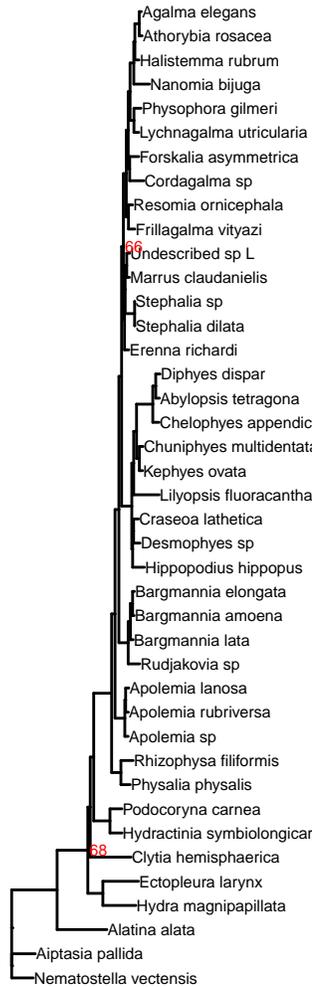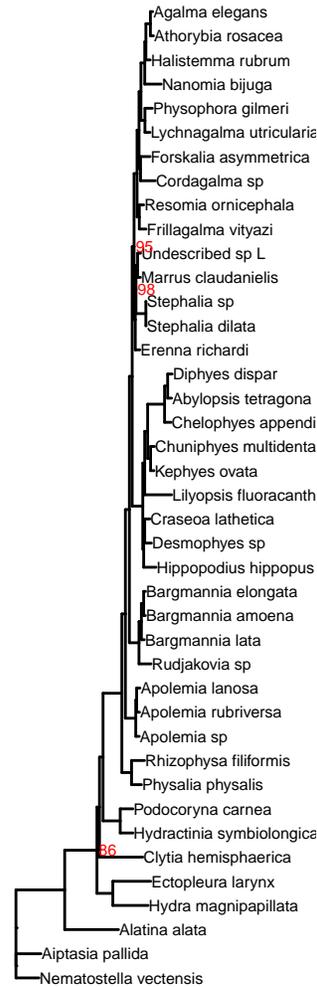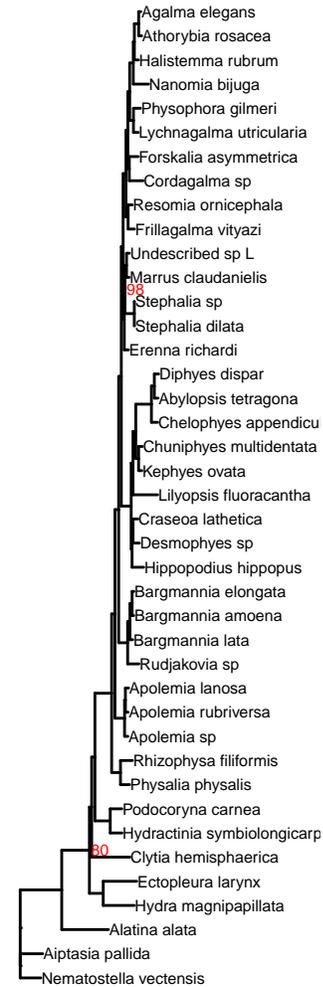
Figure S2.10: Phylogenies obtained using different models of evolution in IQTree. We used ModelFinder (Kalyaanamoorthy et al., 2017) to identify the model with the best relative model fit. We also present results for the commonly used WAG and GTR models. Unlabeled nodes have support >0.99. A. JTT + Empirically counted frequencies from alignment + FreeRate model with 7 categories (the model identified by ModelFinder as having the best fit). Log likelihood:-8113694.922; AIC:16227609.9565 . B. GTR + Optimized base frequencies by maximum-likelihood + Free rate model with 6 categories. Log likelihood: -8133157.335; AIC:16266530.7277. C. WAG + Optimized base frequencies by maximum-likelihood + Free rate model with 6 categories. Log likelihood: -8156043.772; AIC score: 16312303.6120 .
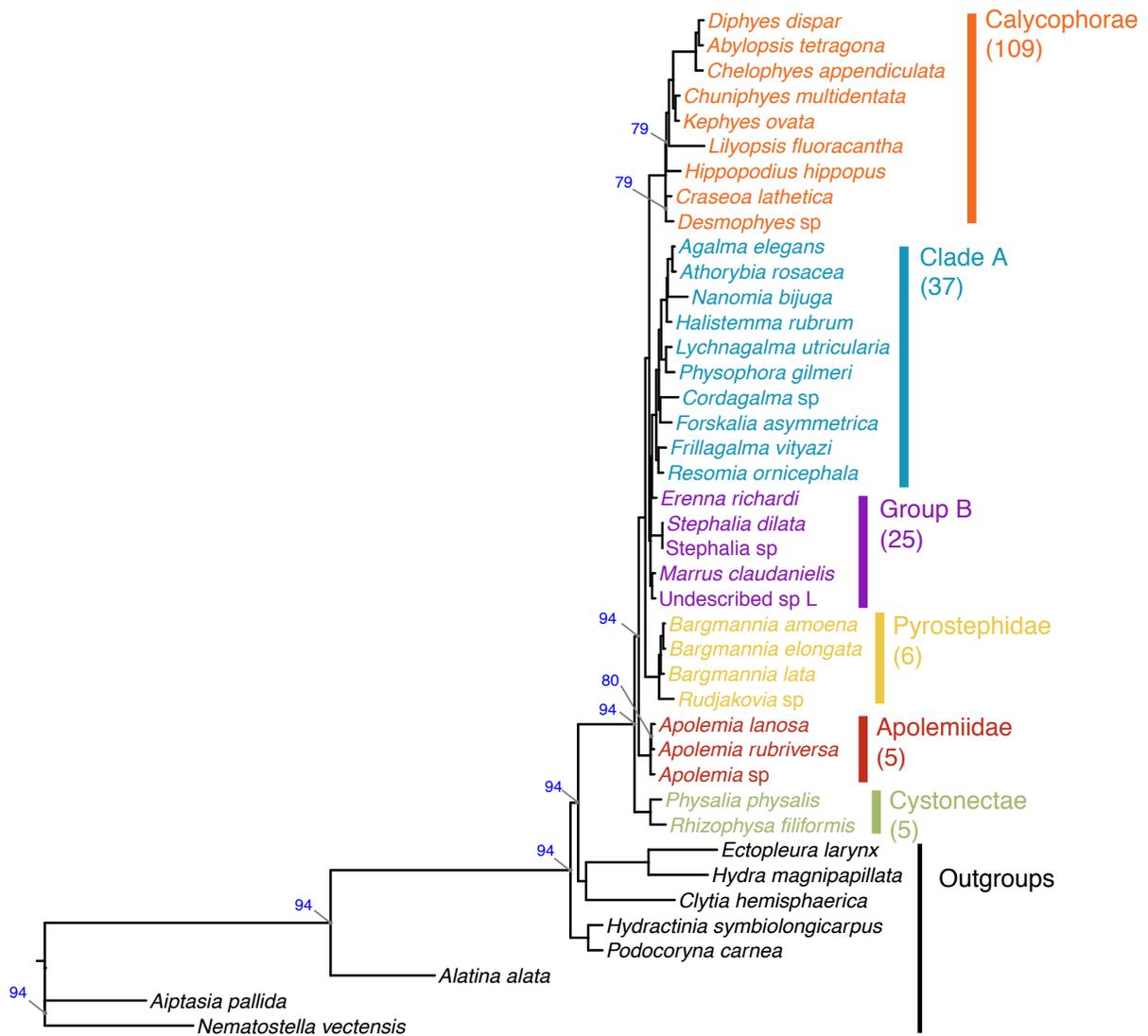
Figure S2.11: Bayesian (BI) CAT poisson phylogram with bipartition frequencies from the Bayesian posterior distribution of trees. Unlabeled nodes have support >0.99. The numbers of valid described species estimated to be in each clade based on taxonomy are shown below each clade name on the right.

Figure S2.12: Bayesian (BI) WAG and Gamma phylogram with bipartition frequencies from the Bayesian posterior distribution of trees. Unlabeled nodes have support >0.99.
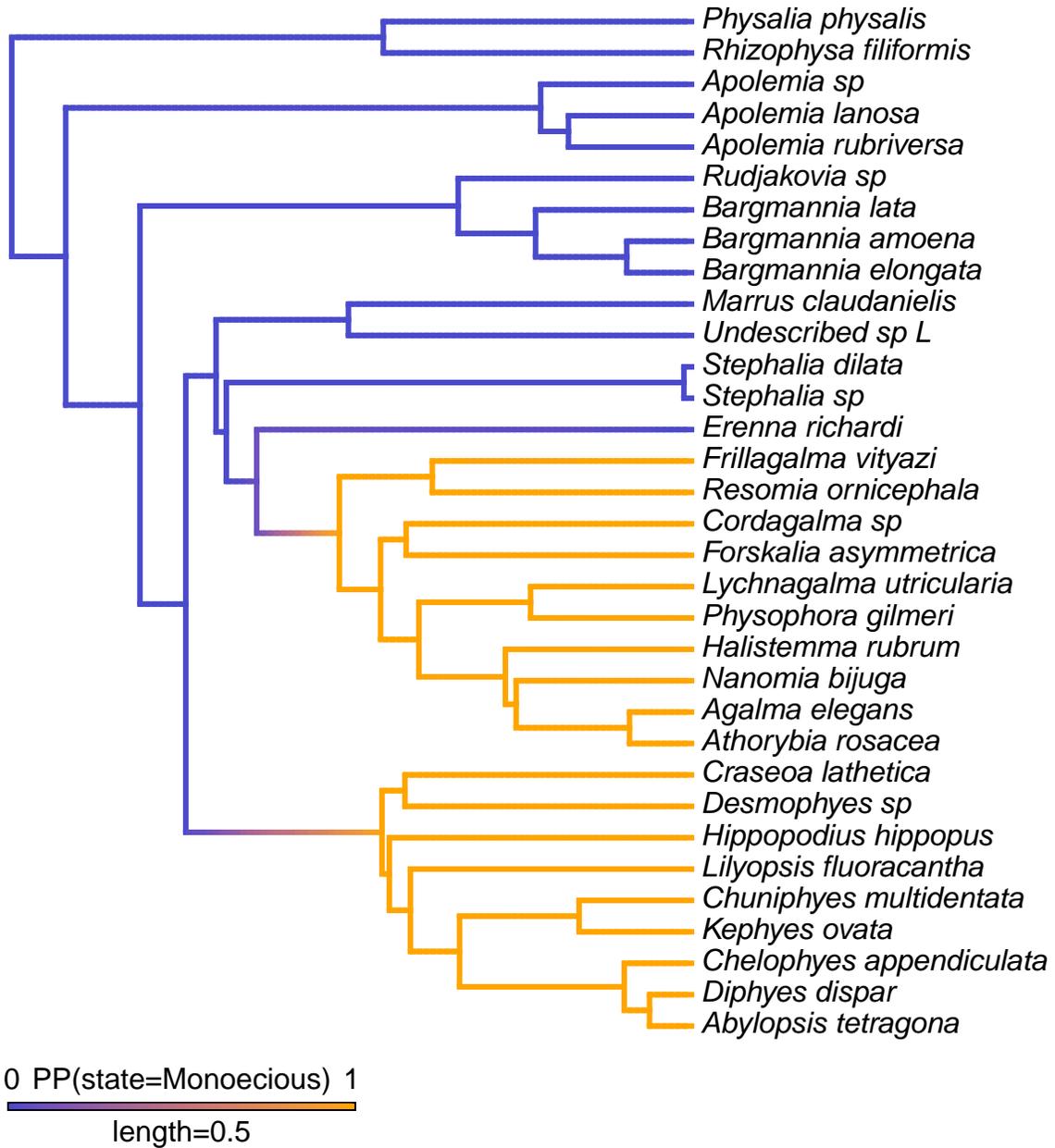
Figure S2.13: Brownian Motion character map of sexual characters on the consensus Bayesian tree topology.
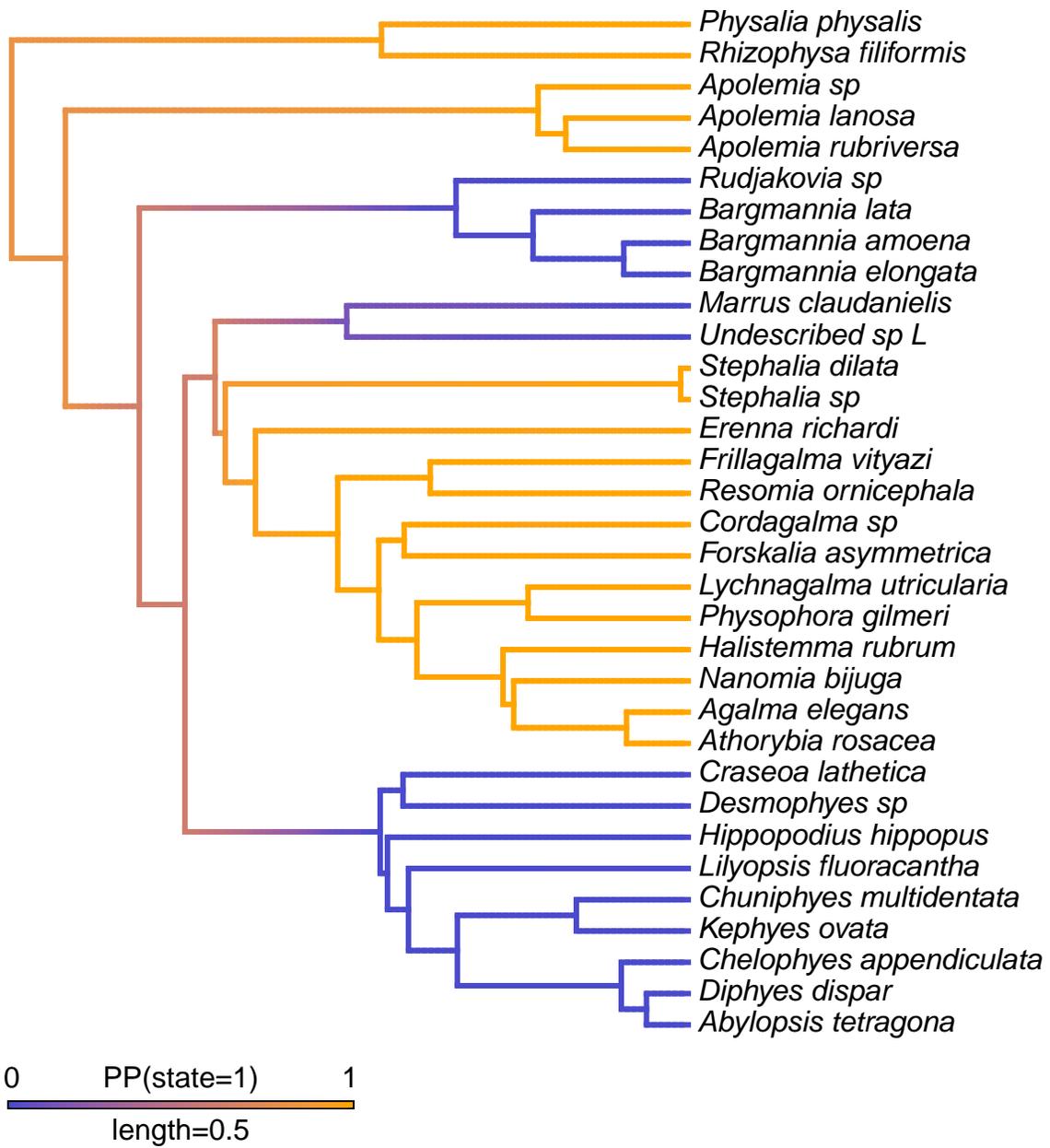
Figure S2.14: Brownian Motion character map of palpon presence/absence on the consensus Bayesian tree topology.
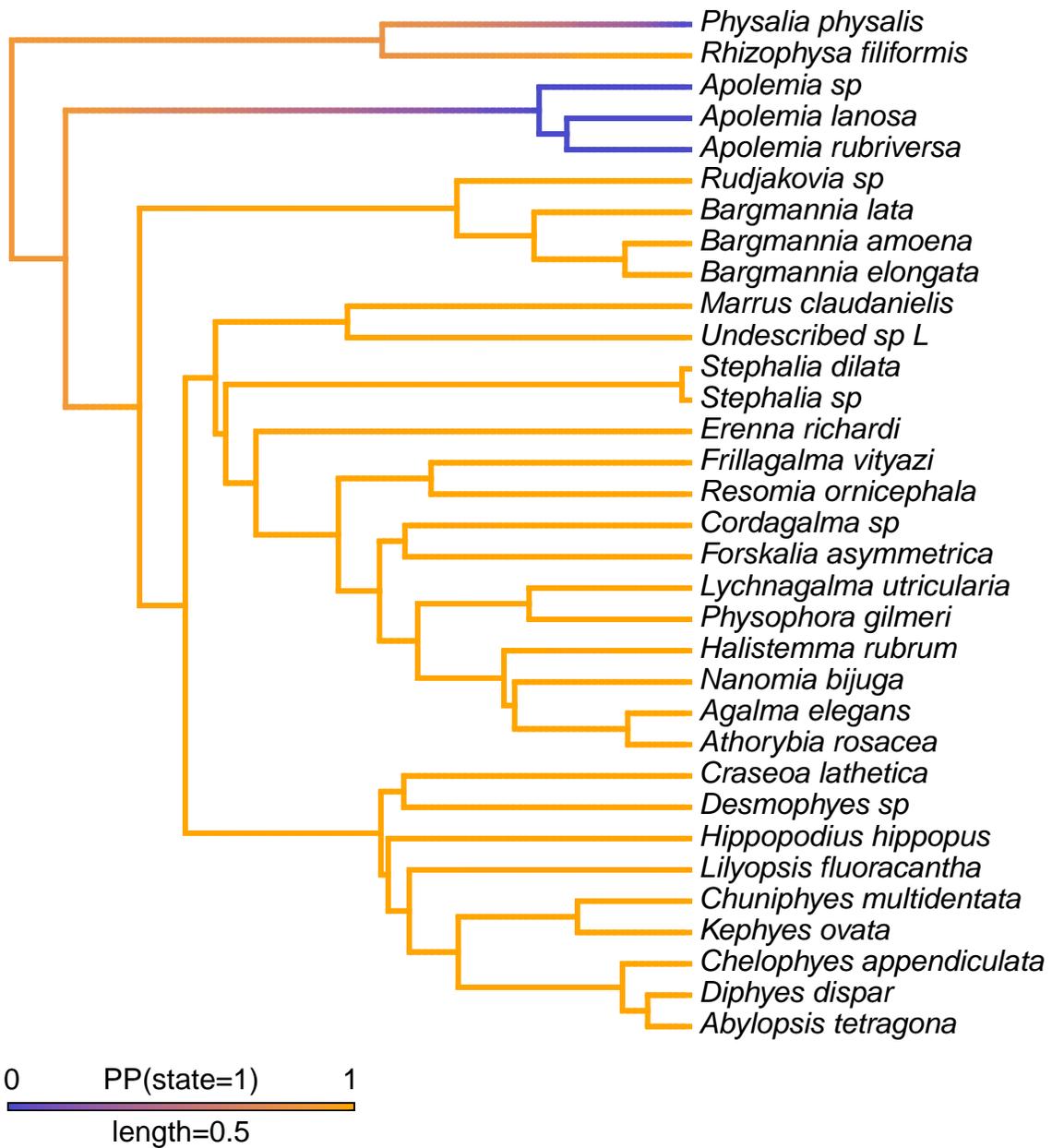
Figure S2.15: Stochastic character map of presence of tentilla with *Physalia* included as not bearing tentilla, mapped on the consensus Bayesian tree topology.
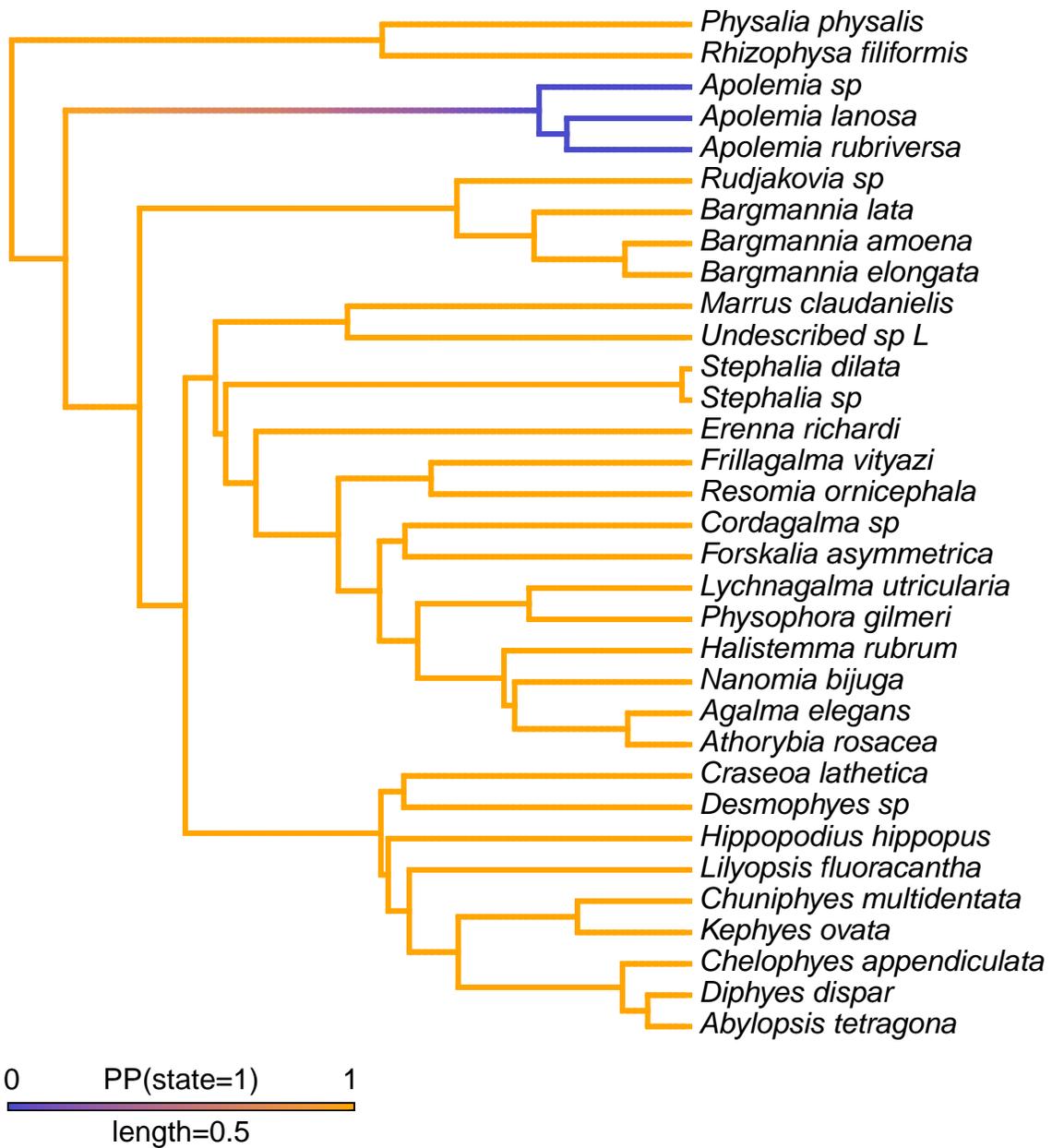
Figure S2.16: Stochastic character map of presence of tentilla with *Physalia* included as bearing tentilla, mapped on the consensus Bayesian tree topology.
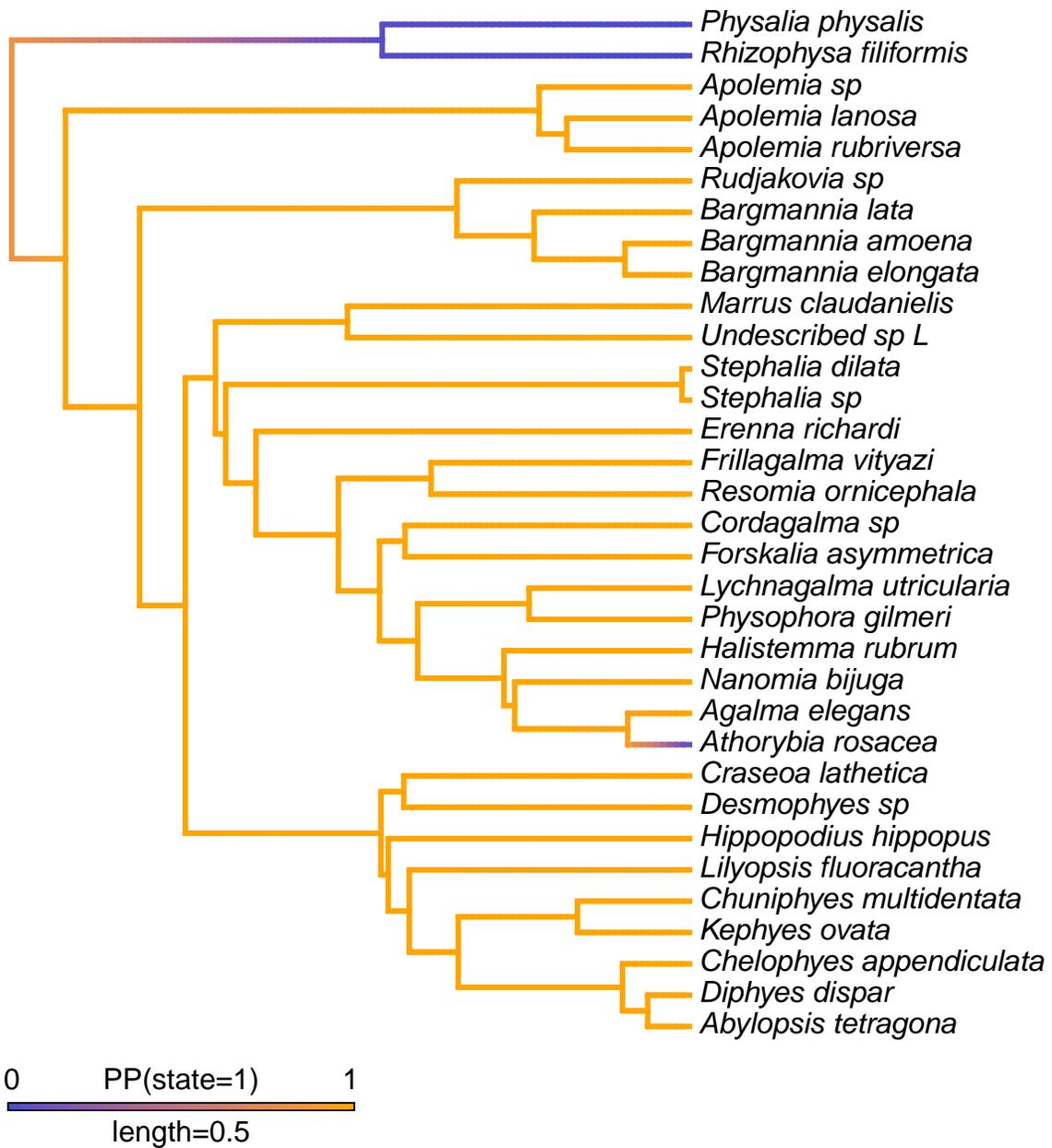
Figure S2.17: Stochastic character map of presence of nectosome, mapped on the consensus Bayesian tree topology.
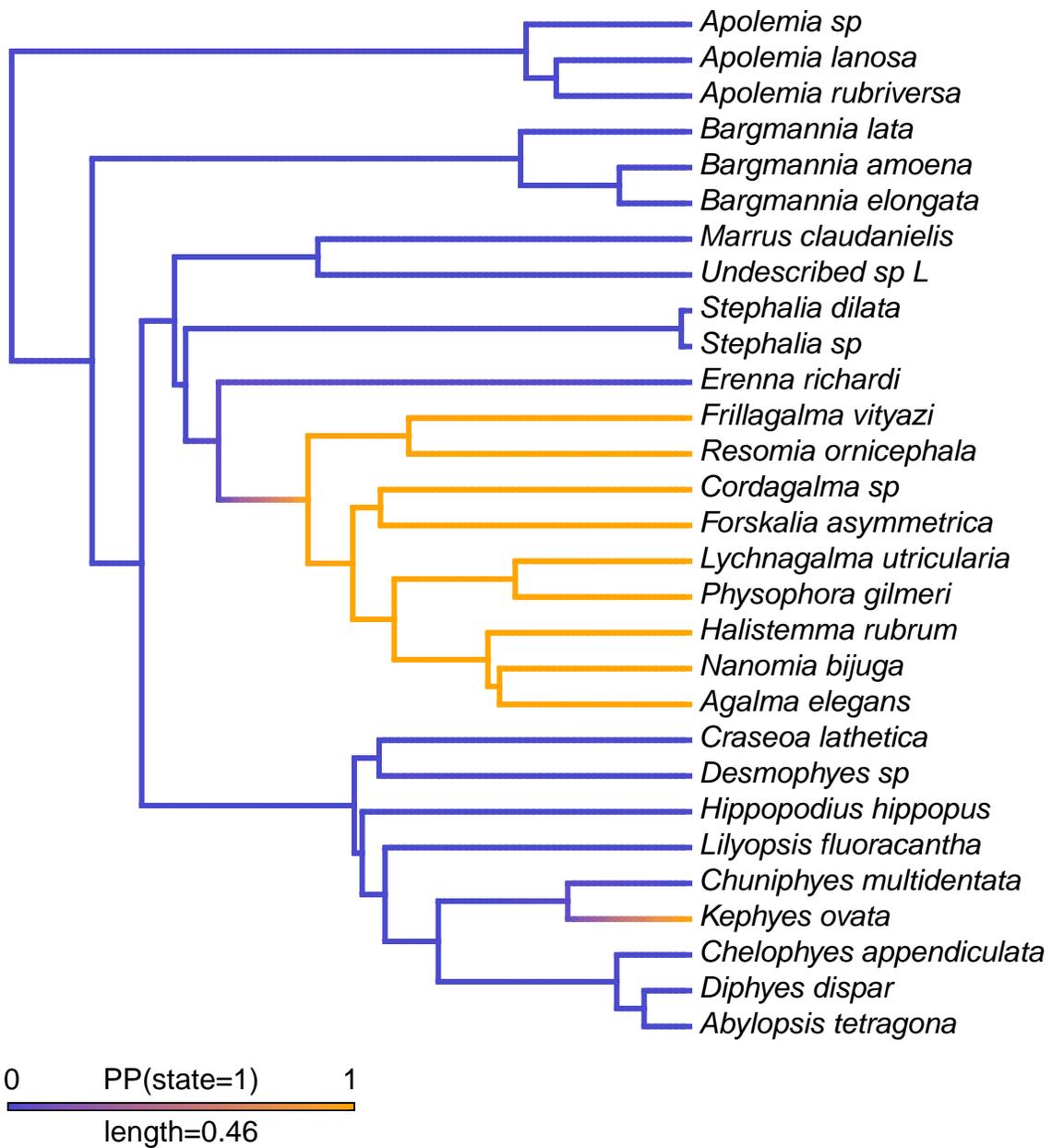
Figure S2.18: Stochastic character map of presence of a descending mantle canal in the nectophores, mapped on the consensus Bayesian tree topology.
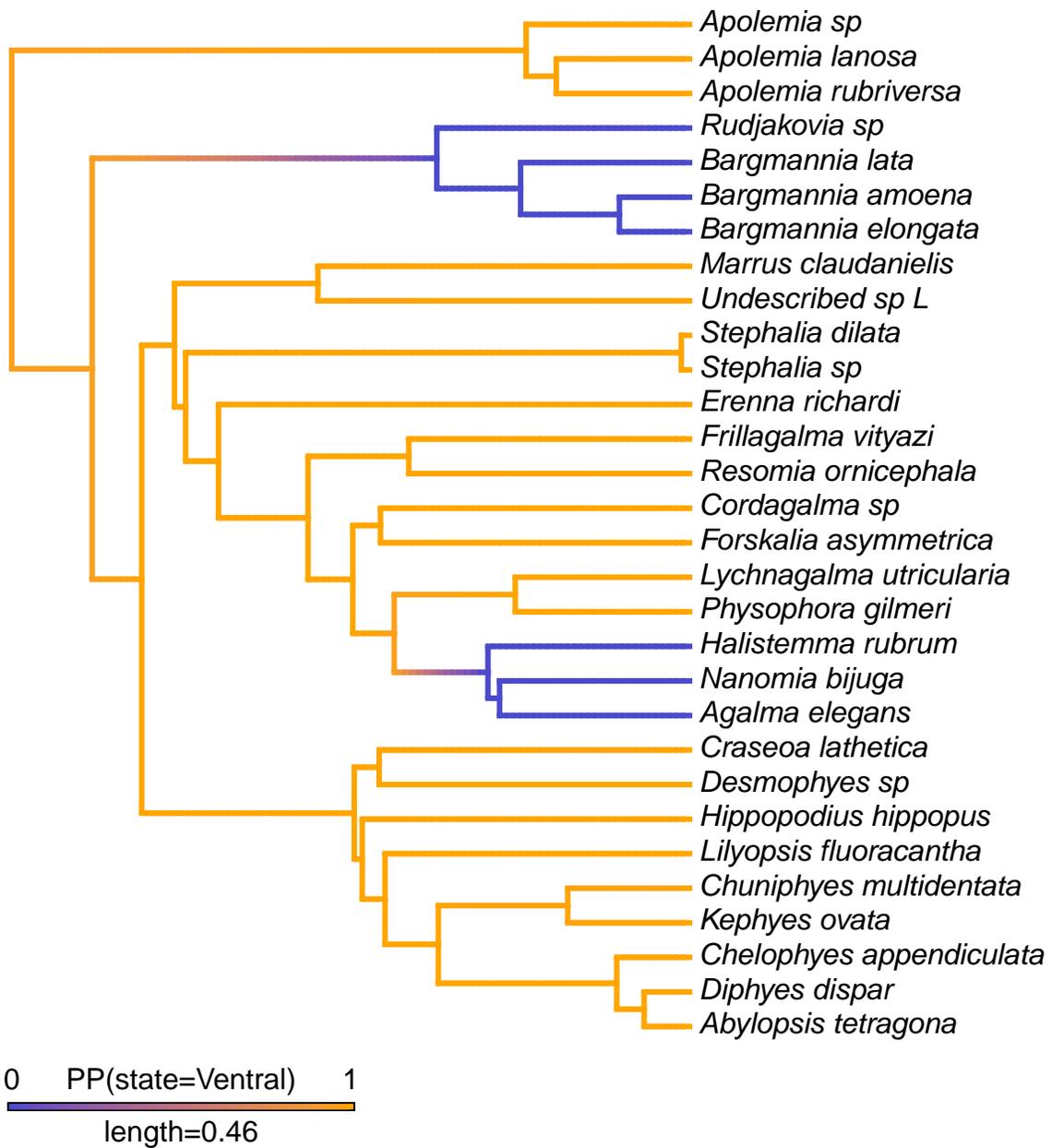
Figure S2.19: Stochastic character map for the evolution of the position of the nectosome, mapped on the consensus Bayesian tree topology.
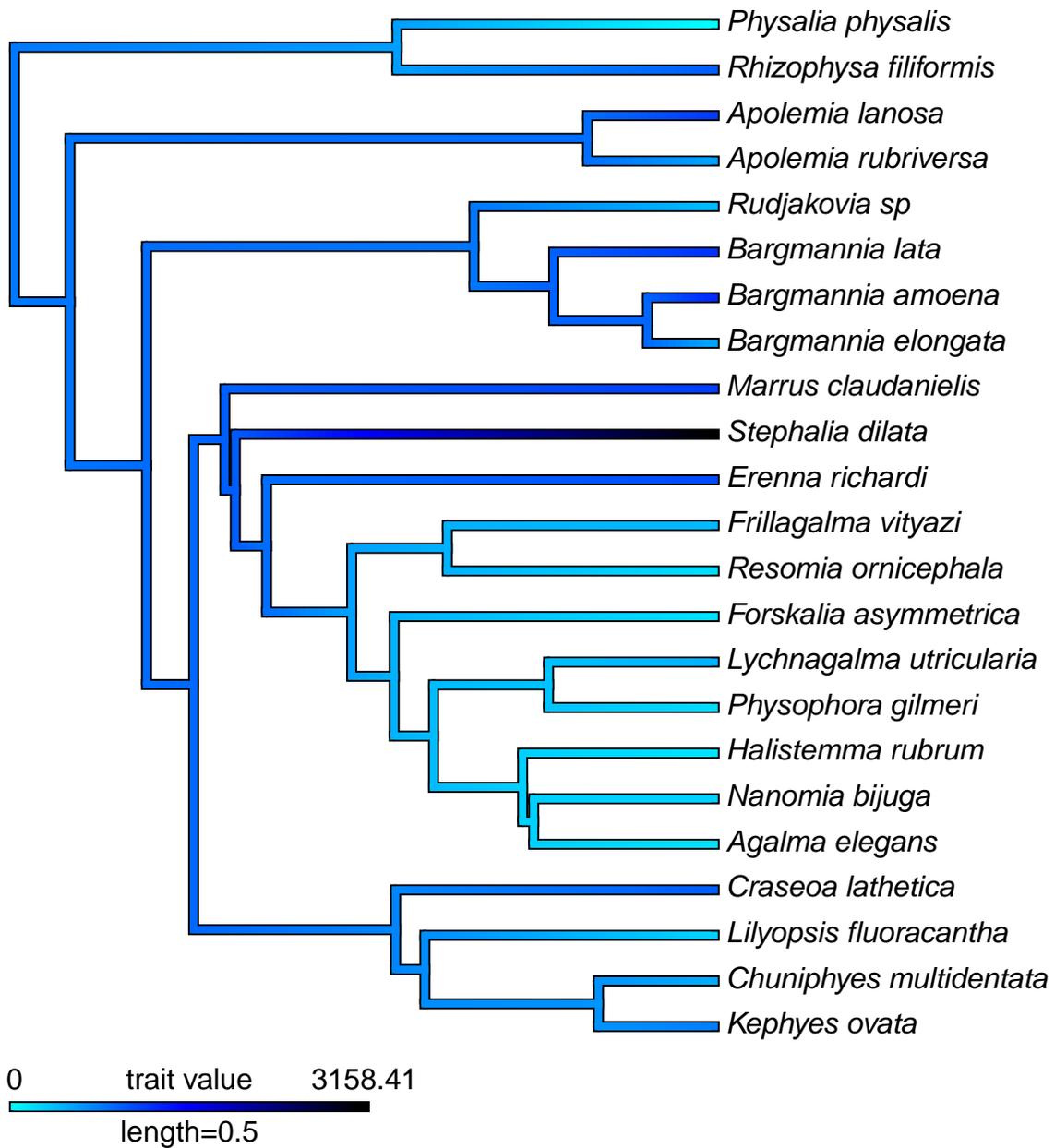
Figure S2.20: Brownian Motion character map of median depth of species observed with an MBARI ROV, mapped on the consensus Bayesian tree topology.
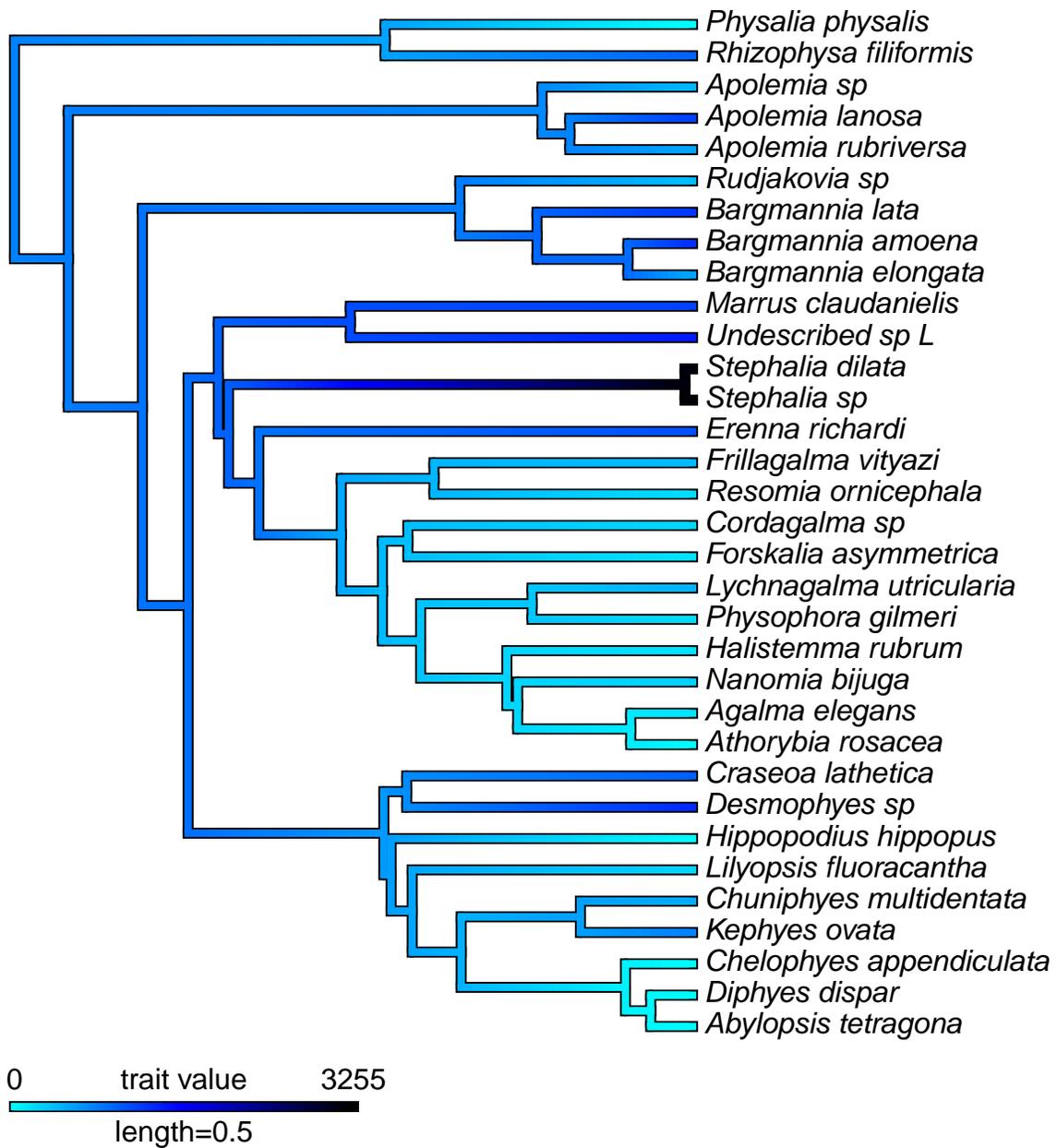
Figure S2.21: Brownian Motion character map of median depth of species including blue water diving observations, mapped on the consensus Bayesian tree topology.

Chapter 3

Molecular anatomy of siphonophores

Catriona Munro[1],‡, Stefan Siebert[2], Steven H.D. Haddock[3], Casey W. Dunn[4]

[1] Department of Ecology and Evolutionary Biology, Brown University, Providence, RI 02912, USA

[2] Department of Molecular & Cellular Biology, University of California Davis, Davis, CA 95616, USA

[3] Monterey Bay Aquarium Research Institute, Moss Landing, CA 95039, USA

[4] Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520, USA

3.1 Abstract

Siphonophores are marine hydrozoans (Cnidaria) that are highly complex colonies consisting of asexually produced bodies (zooids) that are functionally specialized for different tasks – including feeding, reproducing, swimming, and digesting. The microanatomy of siphonophore zooids and tissues has been investigated using histological methods in a few siphonophore species. Here, we use gene expression studies to build an understanding of the "molecular anatomy" of zooids in seven siphonophore species, including species whose microanatomy has not yet been described in detail. Molecular anatomical approaches enable a description of the molecular function and structure of functionally specialized zooids, and build on existing descriptions of the cellular anatomy. Using short read expression data, we describe the molecular composition of a number of zooids and one specialized tissue, the pneumatophore. In addition to finding support for previously hypothesized functions of the zooids, we suggest several candidate enzymes that may be involved in one-carbon metabolism and in the generation of carbon monoxide in the pneumatophore. We also investigate several novel zooid types in particular species, including the tentacular palpon of *Physalia physalis*, the B-palpon of *Agalma elegans*, and the two gastrozooid types in *Bargmannia elongata*. We find support for the hypothesis that the tentacular palpon is a derived gastrozooid, but do not find support that the B-palpon is distinct from the gastric palpons despite its distinct location in the colony. There are few differences in expression between the two types of gastrozooids in *B. elongata*, and it is unclear whether they perform distinct functions in the colony. Finally, we use membership in gene trees in a novel way to identify putative shared homologous genes that are specific to the function and structure of functionally specialized zooids and the pneumatophore.

3.2 Introduction

Siphonophores are a monophyletic group of marine hydrozoans within the clade Cnidaria (Zapata et al., 2015). Siphonophores are highly complex, colonial "superorganisms" consisting of asexually produced bodies (termed zooids) that are homologous to solitary free-living polyps and medusae, but that share a common gastrovascular cavity (Dunn and Wagner, 2006; Mackie, 1963, 1986; Mackie et al., 1987; Totton, 1965). The

functional specialization of siphonophore zooids has been of central interest to zoologists since the 19th century, in part because siphonophores are considered to have a greater level of functional specialization than any other colonial animal, but also because these zooids are highly interdependent (Beklemishev, 1969; Mackie, 1963).

Efforts to investigate the functional specialization of siphonophores have been limited, in part because there have been few detailed investigations of zooid structure. In the last half century, the microanatomy of siphonophore zooids and tissues has been investigated in only a handful of siphonophore species (Bardi and Marques, 2007; Carré, 1969; Church et al., 2015b; Mackie, 1960). This leaves many unknowns about how zooid structure and function differ across zooid types and species. Gene expression studies that show which genes are active in particular regions are now a natural extension of morphological studies, and the perspective they provide can be referred to as "molecular anatomy". Molecular anatomy can be especially helpful when morphological distinctions at a cellular level are either cryptic, or too poorly known to enable interpretation of structure or function, and additionally, can suggest new functional roles that microanatomy cannot predict. Descriptions of genes that are expressed in particular tissues also directly complement the descriptions of microanatomy. Recent *in situ* gene expression analyses in siphonophores have described where a small number of pre-selected genes are expressed at high spatial resolution (Church et al., 2015b; Siebert et al., 2011, 2015), but these methods are limited since they require a large number of specimens per gene and siphonophores are relatively difficult to collect. RNA-seq analyses of hand-dissected specimens (Macrander et al., 2015; Sanders et al., 2014; Sanders and Cartwright, 2015; Siebert et al., 2011), in contrast, can describe the expression of a very large number of genes at low spatial resolution. The fact that so much data is obtained from each specimen is particularly advantageous in difficult-to-collect organisms like siphonophores. An earlier RNA-seq study of gene expression in two zooid types in a single siphonophore species showed the potential of this method to better understand differences between zooids (Siebert et al., 2011). Here we expand the approach to investigate the molecular anatomy of multiple zooids across seven siphonophore species, and identify sets of genes that have expression patterns that are unique to particular zooids within species. We also identify sets of homologous genes with expression patterns that are shared in homologous zooids across species. In addition to zooids, we also describe gene expression in the pneumatophore.

69

Siphonophore colonies arise from a single embryo, which forms a primary feeding protozooid with a tentacle, and a growth zone, from which genetically identical but morphologically and functionally diverse zooids bud asexually (Carré, 1967, 1969; Carré and Carré, 1991, 1993). Unlike other colonial hydrozoans (Blackstone and Buss, 1991; Dudgeon and Buss, 1996), the development of the colony is highly consistent - functionally specialized zooids are typically found linearly along a central stem, and are precisely arranged in repeating units termed cormidia (Dunn and Wagner, 2006). The patterns of zooid arrangement within cormidia vary between species (Dunn and Wagner, 2006). Zooids are considered to be functionally specialized for particular tasks, including feeding, reproducing, defending, and swimming (see Fig. 3.1). Most siphonophore colonies consists of two major regions: the nectosome region, with a growth zone generating nectophores (swimming zooids), and the siphosome region, which is posterior to the nectosome, with another growth zone that gives rise to the rest of the zooid types, including gastrozooids (feeding zooids), palpons (circulatory/digestive zooids), bracts (defensive zooids), and gonophores (reproductive zooids). Within the two regions of growth, zooids are arranged along an anterior-posterior axis from youngest to oldest, with the youngest zooids forming as pro-buds within the growth zone, being carried posteriorly over time by the elongating stem (Dunn and Wagner, 2006). This means that different ontogenetic stages are clearly identifiable within a single colony.

Across the siphonophore phylogeny, there are multiple instances of zooid gain and loss (Dunn et al., 2005a; Munro et al., 2018). In particular, there are a number of cases where there are lineage-specific expansions of particular zooid types. For example, as seen in the nectophores of the calycophorans (Dunn et al., 2005a). The gastrozooids are another zooid that is thought to have been duplicated/modified in a lineage specific manner. One such example is in *Physalia physalis* (see chapter 1). In most siphonophores, the gastrozooid has a tentacle to capture food as well as an oral opening to ingest food. In *Physalia*, these structures and tasks are split across two types of zooids. The zooid type referred to as the gastrozooid has an oral opening and no tentacle, and another zooid, referred to as the tentacular palpon, has a tentacle but no oral opening. The tentacular palpon has been hypothesized to be a modified, mouth-less, gastrozooid that is functionally specialized for nematocyst production and prey capture. In *Bargmannia* species, two hypothesized types of gastrozooids have been identified based on size and morphology, however it is unclear if these zooids perform distinct functions within the colony (Dunn, 2005). Finally, palpons are another zooid type that are

Pneumatophore

Gas filled float - plays a role in buoyancy, floation, and geotaxis. Thought to generate carbon monoxide.

Nectophore

Medusa that is specialized for swimming. Highly muscular and propels the colony via contractions that push water jets out of the ostium

Gastrozooid

Feeding polyp - is the only zooid that ingests food, and also carries out extracellular digestion of prey. Typically has a tentacle for prey capture. Has a nematogenic region that supplies nematocysts to the tentacle.

Palpon

Derived gastrozooid. Considered to be accessory digestive zooid. Also hypothesized to play defensive and sensory functions. Typically has a reduced tentacle (palpacle).

Male gonodendron

Male gonodendra are compound structures that bear multiple **gonophores** that are medusae that contain male gametes and are the site of spermatogenesis.

Female gonodendron

Female gonodendra are compound structures that bear multiple **gonophores** that are medusae that contain female gametes and are the site of oogenesis.

Figure 3.1: Schematic of the siphonophore *Nanomia bijuga*, with highlighted zooids and pneumatophore, with explanations of known function. NGZ - nectosomal growth zone. SGZ - siphosomal growth zone. Diagram by Freya Goetz (http://commons.wikimedia.org/wiki/File:Nanomia_bijuga_whole_animal_and_growth_zones.svg)

71

very broadly defined based on morphology and behavior - they have a mouth that is used for egestion, but generally have reduced tentacles (termed palpacles) that are not involved in prey capture (Mackie and Boag, 1963; Totton, 1965). In different species, palpons may be found in different locations in the colony, and are frequently sub-categorized based on this - for example, the nectosomal palpons in *Apolemia lanosa* (Siebert et al., 2013). The exact homology of these zooids is unclear. Palpons in these locations may have arisen from other palpons or may represent a *de novo* modification of gastrozooids (see chapter 2) (Munro et al., 2018).

Historically the function of different zooid types, including those considered in this paper, have been hypothesized based on morphology, cellular anatomy, behavior, and physiology. Molecular anatomical approaches provide an additional layer, providing information on gene expression patterns that complement existing hypotheses about the function of these zooids. It also enables identification of function that cannot be identified by microanatomical descriptions. Assessing homology of these zooids is more complex, as expression patterns may be more indicative of shared function and cellular composition than evolutionary history. We address a number of open questions with these methods:

- Are there common gene expression patterns in homologous structures across species?

- What are the molecular processes that occur in these zooids and the pneumatophore, and what do these suggest about their function? For example, the exact mechanisms of intra and extra cellular digestion in gastrozooids are unknown (Mackie, 1960; Totton, 1960). Likewise, the mechanism of carbon monoxide generation in the pneumatophore remain unknown, though some substrates have been hypothesized (Wittenberg, 1960; Wittenberg et al., 1962).

- Does location in the colony impact the function of a zooid? For example, in some species palpons occur at multiple distinct locations in the colony. Though the gross morphology does not appear to differ, it has been hypothesized that these palpons may perform different functions (Dunn and Wagner, 2006).

- What are the functions of novel zooid types? And is there evidence of shared expression between these novel zooid types and the hypothesized ancestral zooid type?

Existing approaches to comparing gene expression patterns between species have focused on comparing strict 1:1 orthologs between species (Brawand et al., 2011; Breschi et al., 2016; Levin et al., 2016; Ma et al., 2018; Necsulea et al., 2014; Pankey et al., 2014; Perry et al., 2012; Sanders and Cartwright, 2015; Yang and Wang, 2013). In chapter 4, we use phylogenetic methods to explore the evolution of gene expression to better understand differences across species. Here, we take a novel approach to a narrower set of questions - Which genes show zooid-specific expression within species? Which of these zooid-specific expression patterns are conserved across species? Because the focus is on conserved features, rather than evolutionary differences, we take an explicitly non-phylogenetic, descriptive approach. First, we build gene trees of homologous genes using *de novo* transcriptomes. Then, we conduct differential gene expression analyses on all genes within species and identify a subset of genes that are significantly upregulated in particular zooids and the pneumatophore. To enable comparisons among species, we use membership in gene trees to define inclusive sets of homologous genes, and see which homologous genes have consistent expression patterns across all species (Fig. S3.1). For simplicity, throughout this chapter, we will refer to the collective of zooids and the pneumatophore as treatments, following terminology used in differential expression analyses (Love et al., 2014). While significant downregulation of genes is also interesting, for the purposes of understanding the core molecular functions and anatomy of these treatments we are focusing on genes with significantly more abundant RNA in these tissues. Using Illumina RNA sequencing methods, we identify expression patterns that are unique to particular treatments, and that may be indicative of their function, molecular processes and cellular composition. We also assess expression patterns in hypothesized novel zooid types in several species.

3.3 Methods

All scripts for the analyses are available in a git repository at https://github.com/dunnlab/siphonophoredgeproject. The most recent commit at the time of the analysis presented here was b6576c5f.
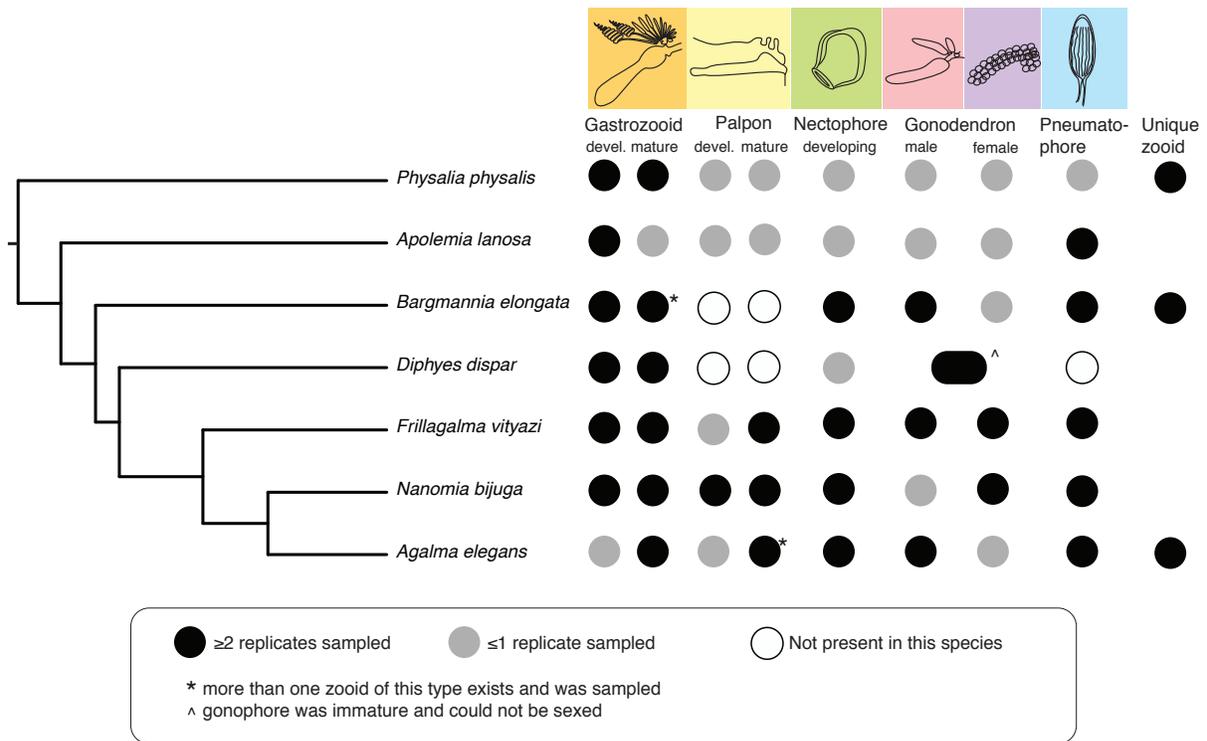
Figure 3.2: Phylogeny of the focal species sampled in this study, with details of the traits sampled for each of the species. Phylogeny modified from chapter 2, Munro et al. (2018). Black indicates that multiple replicates have been sampled, grey indicates that no or only one replicate has been sampled, and white indicates that this zooid/tissue is not present in this species. The category "unique zooid" indicates that a zooid type that is unique to this species was sampled.

3.3.1 Collecting

Specimens were collected in the north-eastern Pacific Ocean in Monterey Bay and, in the case of *Physalia physalis* the Gulf of Mexico. Specimens were collected by remotely operated vehicle (ROV) or during blue-water SCUBA dives. *Physalia* specimens were collected by hand from the beach after being freshly washed on-shore by prevailing winds. Available physical vouchers have been deposited at the Peabody Museum of Natural History (Yale University), New Haven, CT. Specimens were relaxed using 7.5% $MgCl_2$ hexahydrate in Milli-Q water at a ratio of $1/3$ $MgCl_2$ and $2/3$ seawater. Zooids were subsequently dissected from the colony and flash frozen in liquid nitrogen. Colonies were cooled to collection temperatures (e.g 4 degrees C for deep sea species) while the dissections took place. Dissections took no longer than 15-20 minutes. In the case of large colonies, the stem was cut and only partial sections of the colony were placed under the microscope at a given time. Each replicate individual represents a genetically distinct colony from the same species. Replicate specimens were of an equivalent colony size, and zooid replicates were also equivalent sizes. Larger zooid types, such as gastrozooids, were sampled as a single zooid, but smaller zooids were pooled. Pooled zooids were of a comparable maturity and sampled from the same location in a single colony.

3.3.2 Sequencing

mRNA was extracted directly from tissue using Zymo Quick RNA MicroPrep (Zymo #R1050), including a DNase step, and subsequently prepared for sequencing using the Illumina TruSeq Stranded Library Prep Kit (Illumina, #RS-122-2101). 50 base-pair single-end libraries were all sequenced on the HiSeq 2500 sequencing platform. Three sequencing runs were conducted, representing three full flow cells. To avoid potential run/lane confounding effects, where possible, libraries of multiple treatments of a single individual in a species were barcoded and pooled in a single sequencing lane, and replicate lanes of treatments from different individuals of the same species were sequenced in separate runs. Additionally, two libraries were run as technical replicates across all runs and many lanes, for a total of 20 technical replicates.

Table 3.1: Summary statistics for each of the sampled species, with details on the total number of genes in gene trees (there can be multiple genes in a gene tree), and number of unique gene trees containing genes from this species.

Species	Number of genes in gene trees	Number of unique gene trees
Diphyes dispar	4786	3261
Agalma elegans	4498	3534
Frillagalma vityazi	4171	3309
Nanomia bijuga	4166	3250
Bargmannia elongata	3945	3316
Physalia physalis	3189	2741
Apolemia lanosa	3116	2629

3.3.3 Analysis

Short read libraries were mapped to previously published transcriptomes (150bp paired end) (Munro et al., 2018) using Agalma v 2.0.0 (Dunn et al., 2013a; Guang et al., 2017), which uses a number of existing tools for transcript quantification, including RSEM (which uses Bowtie) (Langmead et al., 2009; Li and Dewey, 2011). Gene alignments were generated from the reference transcriptomes of 41 species (Munro et al., 2018) using Agalma, and subsequently PHYLDOG v.2.0 (Boussau et al., 2013) was used for simultaneous co-estimation of gene trees with the published ML species tree (Munro et al., 2018). Gene trees were filtered to exclude trees with a length threshold >2, a root depth >5, and that had more than 0.25 branches with a default length value assigned by phyldog (that are indicative of branch length=0). Using the `agalmar` package (https://github.com/caseywdunn/agalmar), we filtered the expression data to only include genes that were identified as being protein coding, and also only considered genes that are >0 in two or more treatments. Differential gene expression analyses, including normalization, were conducted in R, using the `DESeq2` package (Love et al., 2014). Libraries that were found to be outliers based on mean cook's distance were removed from the DESeq object and from downstream analyses and normalization. Testing for differential expression was conducted using the `results()` function in DESeq2. Genes were considered to be significantly differentially expressed if adjusted p-values (Bonferroni correction) were less than 0.05. Differential expression analyses were only conducted on treatments with two or more replicates.

Using pairwise differential expression between treatments within species, we were able to identify genes that

are significantly expressed and have higher expression in particular treatment types. We were also able to identify putative treatment-specific expression by identifying upregulated genes that are only found in one treatment and not significantly upregulated in others. Using the gene trees, we were also able to identify which gene-trees these genes belong to. Once gene-tree membership was established for upregulated genes, gene-tree membership was compared across all species to determine treatment-specific gene-trees containing putative homologous genes that share similar expression patterns in homologous treatments. Using gene-trees, we were also able to identify putative species-specific expression patterns - that is, expression patterns in treatments that are specific to particular species. These methods are outlined in figure S3.1. For zooid specific genes in mature gastrozooids, we excluded *Physalia physalis* and *Diphyes dispar* from the set of treatment-specific genes that are shared across species, as only one zooid (developing and mature) is sampled in these species. Similarly, for zooid specific genes in the pneumatophore, we excluded *Apolemia lanosa* from the set of treatment-specific genes that are shared across species, since pairwise comparisons were only conducted between the pneumatophore and developing gastrozooid. *Bargmannia elongata*, has two types of gastrozooid, termed white and yellow. We selected the white gastrozooid as the representative mature gastrozooid for this species, as significantly upregulated genes in the yellow gastrozooid represented an overlapping subset of the genes identified as significantly upregulated in the white gastrozooid.

GO annotations were retrieved for each of the reference translated transcriptomes (Munro et al., 2018) using the PANNZER2 web server (Törönen et al., 2018).The PANNZER2 format was modified to match the gene2GO format required for the package `topGO` (Alexa and Rahnenfuhrer, 2016). Gene set enrichment analyses were carried out within species using the R package `GOseq` (Young et al., 2010), which takes gene length into account. Over and underrepresented categories were calculated using the Wallenius approximation, and p-values were adjusted using the Benjamini and Hochberg method. Categories with an adjusted p-value below 0.05 are considered enriched. Gene set enrichment analyses were also conducted at the gene tree level, considering representative GO terms for particular gene trees. Representative GO terms were selected based on frequency of occurrence among genes in the gene tree. As gene lengths vary among species and genes in the gene tree, the `GOseq` approach could not be used, and `topGO` was used to detect GO terms that are enriched based on Fisher's exact test. This approach assumes that each gene tree has an equal probability

of having genes shared among species that are detected as differentially expressed, however results may be biased by a number of factors, including mean gene length among genes in the gene tree (Young et al., 2010).

3.4 Results and discussion

We were able to identify genes with significant differential expression that are unique to particular zooids and the pneumatophore across all species (Fig. 3.3), as well as unique to particular species (Fig. S3.10). A total of 3826 gene trees passed filtering criteria, that contain a total of 27871 homologous genes from 7 species based on sequence alignment. The number of genes per species in gene trees is shown in table 3.1). Differing numbers of treatments were sampled for each of the different species (Fig. 3.2).

3.4.1 Partitioning of expression variance among colonies, treatments, and species

Gene expression measurements can vary at a number of different scales: among sequencing lanes and runs (due to technical effects), the sampled colonies (due to genetic and environmental differences), treatments (due to tissue-specific expression), and species (due to evolutionary change). The first component of variation that we assessed was among technical replicates. The technical replicates consist of re-sequenced developing nectophores and developing gastrozooids from the same *Frillagalma vityazi* individual that were spiked in across multiple lanes and runs. Lane and run effects have been proposed as major sources of technical variability in RNA-seq data that may confound observations of biological variation (Auer and Doerge, 2010; McIntyre et al., 2011). The differences between technical replicates (Fig. S3.2) were found to be much smaller (0.39% variance of expression distance) than the differences between treatments (98.32% variance of expression distance). Differences among technical replicates of the same treatment were correlated with library size and run, not by lane.

The next variation we considered was biological variation among sampled colonies. Within species, the greatest variance was among treatments and groups of treatments, as opposed to between biological replicates
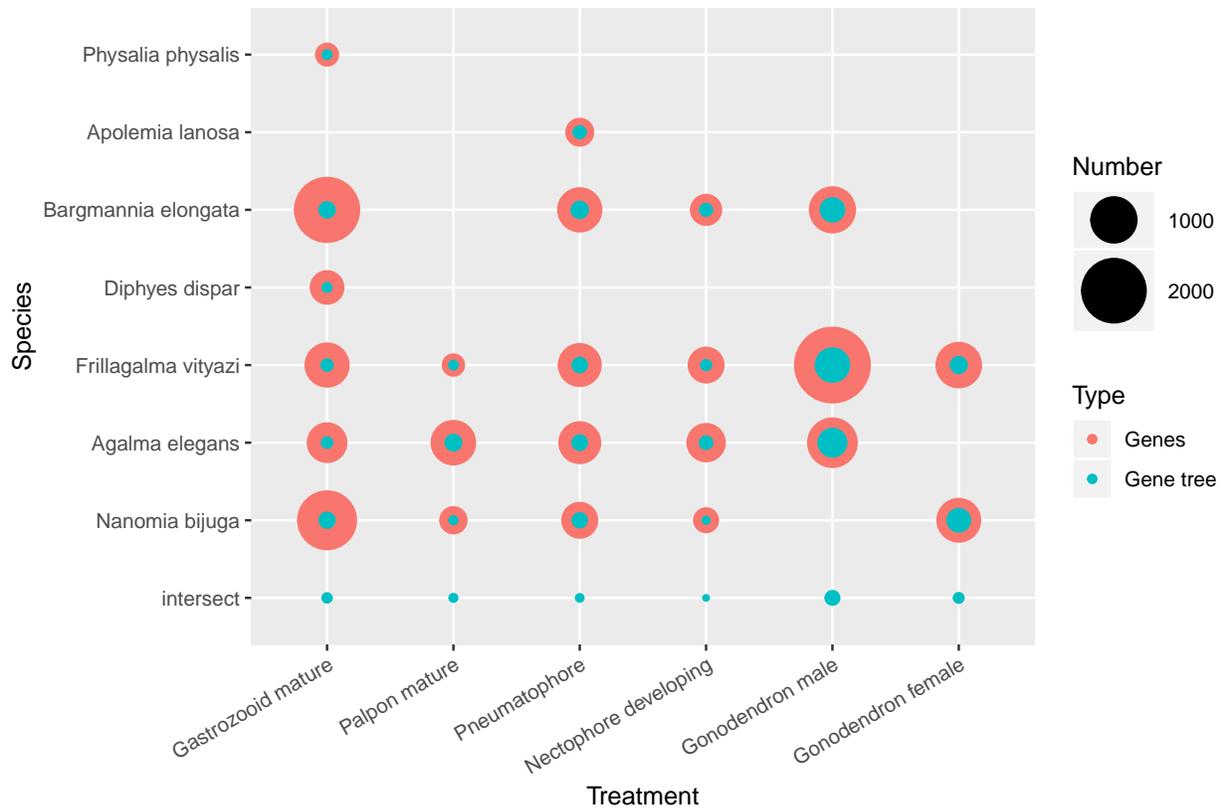
Figure 3.3: Numbers of genes (red) and gene trees containing genes (cyan) with treatment-specific expression patterns. Treatment-specific expression is where genes are upregulated in a particular treatment and are not upregulated in any other treatment within the same species. Intersect indicates the total number of gene trees with homologous genes that are shared across all species and differentially expressed exclusively in that zooid/tissue. Area of circles indicates the numbers of genes containing treatment-specific expression pattern and the number of gene trees those genes are found in. Missing values are where no treatments were sampled for this species, and does not indicate a lack of treatment-specific genes.

of the same treatment (Figs. S3.3–S3.9), with some exceptions: developing gastrozooids in *Bargmannia elongata* (Fig. S3.6); one mature gastrozooid replicate and male gonodendra in *Agalma elegans* (Fig. S3.5); and a single developing gastrozooid replicate in *Frillagalma vityazi* (Fig. S3.7). Specimens were collected in the wild at different depths and over different time periods, but despite these potentially confounding environmental factors, the major variation we observe is among treatments within species.

To identify how expression partitions among treatments, we first identify treatment specific genes. We define treatment-specific expression as genes that have more transcripts in a particular treatment of a particular species to the exclusion of all other treatments in the same species. We then identified the gene trees that contain treatment-specific genes (Fig. S3.1, steps 1 & 2). Gene trees were identified that contain treatment-specific genes and are also shared among all species (Fig. 3.3, intersect). The set of trees that can be analyzed are limited by the species with the poorest sampling and sequencing depth.

Male gonodendra, followed by female gonodendra, have the largest number of gene trees containing genes with expression patterns that are treatment-specific and shared across species (84 and 39 gene trees respectively) (Fig. 3.3, Gonodendron male and female intersect). This matches patterns identified in mammalian and avian testis, that identify more expressed protein coding, long non-coding RNA genes, and transcribed intergenic expression than any other organ (Brawand et al., 2011; Kryuchkova-Mostacci and Robinson-Rechavi, 2016; Melé et al., 2015; Necsulea et al., 2014; Soumillon et al., 2013). Organ-specific expression in mammals is highest in the testis (Kryuchkova-Mostacci and Robinson-Rechavi, 2016). The third largest set of gene trees with treatment-specific genes is found in mature gastrozooids (35 gene trees), followed by mature palpons (24 gene trees), pneumatophore (22 gene trees), and developing nectophores (11 gene trees) (Fig. 3.3, intersect).

When looking at total number of genes with significantly elevated expression we find that these patterns vary within species. As with gene trees, the male gonodendron has the largest number of treatment-specific genes in *Frillagalma vityazi* (2775 genes) (Fig. 3.3, *Frillagalma vityazi*, type = genes). However there are also many genes that have significantly higher expression in mature gastrozooids in *Nanomia bijuga* and *Bargmannia elongata* (1649 and 2035 genes respectively) that are not captured by the gene tree data set (Fig. 3.3, *Nanomia bijuga* & *Bargmmania elongata* genes vs gene trees). The smallest number of treatment-specific genes was found in mature gastrozooids in *Physalia physalis* (222 genes) and mature palpons in

Frillagalma vityazi (208 genes) (Fig. 3.3, *Physalia physalis* & *Frillagalma vityazi*, type = genes).

We identified a number of gene trees with putative species-specific expression patterns across each of the different treatments (Fig. S3.10). For the purposes of this study, we define species-specific expression as gene trees that contain significantly upregulated genes that are found in a particular species in a particular treatment, and not differentially expressed in the same treatment in other species. This differential gene expression based method, however, is biased towards species-specific expression patterns that are also particular to a treatment or treatments, and will not capture ubiquitously expressed genes that do not vary among treatments but do vary among species. Additionally, the putative species-specific gene expression also reflects genes that are simply better sampled in a particular species. Comparisons of expression among species should use phylogenetic comparative methods (see chapter 4) (Dunn et al., 2013b). In the following sections we will focus exclusively on treatment-specific gene expression patterns, while in the next chapter we consider a phylogenetic perspective.

3.4.2 Gastrozooids

Gastrozooids are polyps that are functionally specialized for feeding, and are found in all siphonophore species. Gastrozooids typically have a single tentacle attached to the base of the zooid, except in *Physalia physalis*, where gastrozooids lack a tentacle (tentacular palpons carry a tentacle - see chapter 1) (Mackie, 1960; Totton, 1960, 1965). The gastrozooid is the only zooid that ingests food (Mackie and Boag, 1963). Digestion in the gastrozooid occurs in two phases - in the first, they prey is digested extracellularly, and in the second, the particulate matter is digested intracellularly; finally, any material that cannot be digested is egested through the mouth (Mackie and Boag, 1963). The gastrozooid is typically divided into two regions: the oral hypostome and aboral basigaster (Church et al., 2015b). The basigaster has a thick ectoderm, while the hypostome, that is further divided into two regions (mid region and buccal region), consists of thickened endoderm and thin ectoderm (Church et al., 2015b; Carré, 1969). The basigaster is a site of nematogenesis, and developing nematocysts at different stages are present in the ectoderm, while the endoderm consists of absorptive cells (Church et al., 2015b). The buccal region consists of a number of folds, that is thought to enable the zooid to spread out and cover the entire prey item (Church et al., 2015b). A number of

different gland cells are present in the buccal region, including zymogen gland cells and gastric spherical cells (Carré, 1969; Church et al., 2015b). A number of different proteolytic enzymes have been identified in the gastrozooid of *Physalia physalis* (Bodansky and Rose, 1922), and the secretory cells in the buccal region are hypothesized to discharge their contents, presumably proteolytic enzymes, while feeding (Mackie, 1960).

Gene trees identified as containing treatment-specific genes in mature gastrozooids across all species were enriched for a number of GO terms (Fig. 3.3, table S3.1). These terms include regulation of proteolysis, amino sugar catabolic process, cellular response to amino acid stimulus, response to organonitrogen compound. Shared gene trees include a number of genes whose function matches the known function of the gastrozooid: Protein eva-1 homolog C (involved in heparin binding and ubiquitously expressed in human epithelial tissues (Mitsunaga et al., 2009)); members of the cysteine proteinase gene family, notably Cathepsin L (involved in digestive protein degradation (Barrett and Kirschke, 1981)); chitooligosaccharidolytic beta-N-acetylglucosaminidase (involved in the breakdown of chitin (Nagamatsu et al., 1995)); Glutathione peroxidase 7 (thought to play a role in protecting esophageal cells from reactive oxygen species generated by gastric acid and bile (Peng et al., 2011)); argininosuccinate synthase (involved in the urea synthesis in the liver, but also polyamine and creatine synthesis (Qualls et al., 2012)); aquaporin-9, a transmembrane protein involved in water, glycerol, and urea transport (Ishibashi et al., 2010); carbonic anyhydrase 6 (found in oral and gastric mucus in mammals, involved in pH maintenance and possibly taste perception (Brown et al., 1984; Kivelä et al., 1999)); and Snaclec toxins (also identified by RNAseq in sea anemone polyps (Macrander et al., 2015)). These findings in gastrozooids are an important confirmation of this method, given that gastrozooids are relatively well defined based on behavior, physiology, and histology. The expression results confirm that gastrozooids are indeed specialized for feeding, particularly extracellular digestion.

3.4.3 Palpons

Palpons are polyps that are considered to be reduced derived gastrozooids (Mackie et al., 1987; Totton, 1965). Palpons do not feed, but they have been observed to play a role in the intracellular phase of digestion, as well as in the egestion of waste materials (Mackie and Boag, 1963). Palpons are also hypothesized to play a role in defense as well as sensory functions (Totton, 1965). Palpons are thought to have been present in the common

ancestor of siphonophores, but have been lost along three branches, including to *Bargmannia* and *Rudjakovia* (a clade that includes *Bargmannia elongata*, sampled here) and calycophorans (including *Diphyes dispar*, sampled here) (Munro et al., 2018). Palpons are present in the gonodendra of *Physalia physalis*, but juvenile specimens were collected for this study, and these do not have fully developed gonodendra. Histological investigations of palpons in *Apolemia* and *Nanomia bijuga* indicate that the gastrodermal surface of the palpon is populated by funnel cells (ovoid cells with tufts of microvilli) and absorptive cells, suggesting a role in digestion and particle capture (Church et al., 2015b; Willem, 1894). In *Nanomia bijuga* there is a also reduced basigaster-like region that is the site of nematogenesis (Church et al., 2015b).

Palpon specific gene trees shared across each of the assessed species (Fig. 3.3,Table S3.2) are enriched for GO terms such as sphingolipid metabolic process, collagen biosynthetic process, regulation of hydrolase activity, regeneration, cellular response to nutrient and macromolecule metabolic process. A number of genes trees are identified as palpon specific within each of the assessed species. GO terms identified within species suggest that palpons play a role in biosythetic and metabollic processes amongst others - cellular amino acid biosynthetic process, small molecule biosynthetic process, purine ribonucleoside metabolic process, cellular metabolic compound salvage (in *Agalma elegans*); cytokine biosynthetic and metabolic process, fatty acid metabolic process, humoral immune response, metanephros morphogenesis (in *Frillagalma vityazi*); and adenosine metabolic process, immune response-regulating cell surface receptor signaling pathway involved in phagocytosis, liver morphogenesis (in *Nanomia bijuga*). Notably, a large number of gene trees were identified as being palpon specific within *Agalma elegans*, and a large number of these are also putatively species-specific (77 out of 111 gene trees) - by contrast, 0 out of 29 and 3 out of 25 gene trees are putatively species and palpon specific in *Frillagalma vityazi* and *Nanomia bijuga* respectively. Altogether, these results suggest that the palpon does indeed play a role in digestion and absorption, and possibly also in the biosynthesis of small molecules.

3.4.4 Nectophores

Nectophores are highly specialized medusae, having no gonads, tentacles, or manubrium (Totton, 1965). The sole function of the nectophore is to propel the colony (Costello et al., 2015; Mackie, 1964, 1965). Nectophores

are located in a region called the nectosome in the clade Codonophora (the group that includes *Frillagalma vityazi*, *Nanomia bijuga*, *Diphyes dispar*, *Agalma elegans*, *Bargmannia elongata* and *Apolemia lanosa*). In *Physalia physalis* and other cystonects, nectophores are located only in the gonodendra. Expression data was collected for nectophores in *Frillagalma vityazi*, *Nanomia bijuga*, *Agalma elegans* and *Bargmannia elongata*. Only developing nectophores were collected, as mature nectophores have thick mesoglea and it is a challenge to extract RNA using the same methods as the other zooids. Although only developing nectophores were collected, developing nectophores are observed to perform pumping movement and the subumbrellar muscle is thought to be active at this stage (Mackie, 1960; Steche, 1907). The subumbrellar ectoderm consists of a striated myoepithelium without a nerve plexus (Mackie, 1964, 1965). Smooth radial muscles are found at the margin, that adjust velum shape during swimming - in *Nanomia*, the radial muscles are arranged on either side of the velum, enabling the colony to swim "backwards" by directing a jet of water forwards (Mackie, 1964, 1965). The exumbrellar epithelium is non-muscular and consisting of ectodermal cells - large parts of the exumbrellar epithelium lack nerves and muscle fibers, but are nevertheless conductive in any direction (Mackie, 1965). The electrically coupled myoepithelial cells of the nectophore spread signals for synchronous contraction in a manner that is analogous to the vertebrate heart (Mackie et al., 1987). The margin of the nectophores has an inner and outer nerve ring, and a neuronal pathway along the lower side of the nectophore is hypothesized but has not been observed (Grimmelikhuijzen et al., 1986; Mackie, 1964).

Very few gene trees were identified as having expression specific to developing nectophores across *Frillagalma vityazi*, *Agalma elegans* and *Nanomia bijuga* (Fig. 3.3,Table S3.3). However enriched GO terms appear to match the known function of the nectophore, including: regulation of the force of heart contraction, ventricular cardiac muscle tissue development, mitotic actomyosin contractile ring assembly, actin-myosin filament sliding. Shared differentially expressed zooid specific gene trees include: Follistatin-related protein 1 (a glycoprotein expressed in the human heart (Shimano et al., 2011)), UDP-glucose 6-dehydrogenase (plays a number of roles, including modulating the Wnt pathway in *Drosophila* (Hacker et al., 1997) and heart valve formation in zebrafish (Walsh and Stainier, 2001)), Myosin-2 essential light chain, Tropomyosin, Neuronal Ca2+ sensor protein-1 (NCS-1) (a calcium binding protein (Boeckel and Ehrlich, 2018), regulates excitation-contraction coupling in fetal hearts in mice (Nakamura et al., 2011)), and calmodulin (a calcium

sensing protein (Chin and Means, 2000)). It is not clear why there are so few shared gene trees for developing nectophores, although it is notable that even before identifying overlap among species, fewer genes in general are significantly upregulated in this tissue (Fig. 3.3). These findings do, however, support the suggestion by Mackie (1960) that developing nectophores are indeed active before they are fully mature, and already express genes that may be involved in excitation-contraction coupling.

3.4.5 Reproductive zooids

Gonodendra are compound reproductive structures in siphonophores that have gonophores of only one sex (Totton, 1965). Gonophores are reduced medusae (Totton, 1965). In *Physalia physalis* and other cystonects, the gonodendra consist of sexual gonophores, palpons, and nectophores (Totton, 1960). In *Nanomia bijuga*, *Agalma elegans*, and *Bargmannia elongata* the female gonodendra consist only of multiple gonophores borne on a stalk (Totton, 1965). In *Frillagalma vityazi*, single male and female gonophores are borne on a short stalk (Pugh, 1998). In *Diphyes dispar*, and in other calycophorans, the gonophore attaches directly the stem (Totton, 1965). Male gonophores are packed with ectodermal sperm proginator cells, found below a thin layer of ciliated ectodermal cells (Church et al., 2015b). The central spadix consists of a gastric cavity and endodermal cells with small granules (Church et al., 2015b). Female gonophores are reduced medusae that contain either a single oocyte or multiple (Carré, 1969). The stalk of the gonodendron contains a gastrovascular cavity, and gastrodermal canals connect to each of the gonophores via the peduncle, wrapping around the oocyte (Church et al., 2015b). Some gonophores have a nerve plexus, striated muscle and nerve rings (Mackie, 1965), and in some species, such as *Diphyes dispar*, the gonophores serve a locomotory function after a eudoxid is set free (Mackie et al., 1987). In other species, the gonophores pulsate, perhaps to enable the dispersal of gametes (Mackie et al., 1987).

A large number of shared gene trees were identified as being specific to male gonodendra across all species (Fig. 3.3, Table S3.4). The GO terms enriched in this list include: cell cycle, male gamete generation, chromosome segregation, cilium organization, assembly and movement, and nuclear division. A number of these genes trees are testis and sperm specific, including T-complex-associated testis-expressed protein 1, Testicular haploid expressed gene protein-like, Sperm-associated antigen 8, Cilia- and flagella-associated

proteins; or likely relate to meiosis, e.g. homologous-pairing protein 2 homolog, Meiosis-specific nuclear structural protein 1, Dynein assembly factor 1, axonemal. Given that the function of the male gonodendron is well known as a site of spermatogenesis, these findings are not surprising but are nevertheless an important confirmation of these methods.

Fewer shared gene trees were identified as being specific to female gonodendra across all species (Fig. 3.3, Table S3.5), and the enriched GO terms appear to be less specific to oocytes or oogenesis: sensory perception, regulation of response to wounding, nervous system process, regulation of axon extension. Within each of these species there are a number of DNA/RNA binding genes, DNA replication, as well as those involved in meiotic processes and double strand break repair. Notably, in *Nanomia bijuga*, we identify a gene with blast hits to GQ-coupled Rhodopsin, suggesting possible photoreceptors within the gonophore of this species. Specialized cells have been identified in the gonad of the hydrozoan *Clytia hemispherica* that express opsin and secrete maturation inducing hormones in response to light cues (Artigas et al., 2018). It is notable that few gene trees overlap between the two sampled siphonophore species, given that there are only two species represented and both species have relatively large numbers of gene trees with genes with treatment-specific expression (Fig. 3.3). It is possible that the oocytes are at different levels of maturity within the gonophores, and this difference in maturity accounts for the lack of overlap between these two species.

The sampled *Diphyes dispar* gonophores were immature, and could not be sexed. Several genes were identified as significantly differentially expressed in *Diphyes dispar*, and within this list, gonophores share two gene trees with developing nectophores: Tropomyosin-2 and Calcium-binding protein NCS-1. Four gene trees that were identified as zooid specific to male gonodendra were identified (top blast hits: General transcription factor IIF subunit 2 , G2/mitotic-specific cyclin-B3, Polyadenylate-binding protein-interacting protein, Proline dehydrogenase), as well as four gene trees specific to female gonodendra (top blast hits: Gamete expressed 1, Fatty acid hydroperoxide lyase, Bifunctional lysine-specific demethylase and histidyl-hydroxylase MINA, CD63 antigen). As gonophores could not be sexed, we cannot be sure of the sex of the replicates within the species either. Within *Diphyes dispar* gonophores, however, we were able to identify a number of highly upregulated genes, including Green fluorescent protein and also GQ-coupled Rhodopsin genes , as in *Nanomia bijuga.* This supports the notion that there are also gonad-based opsins in siphonophores, in addition to

Clytia hemispheria – however, the exact function and location would need to be confirmed.

3.4.6   Pneumatophore

The pneumatophore is not a zooid, but is a unique siphonophore structure that is gas-filled and used as a float. In *Physalia physalis*, the pneumatophore of the mature colony is muscular and greatly enlarged, enabling the colony to float at the ocean surface and catch the wind as a sail (Totton, 1960). In other species, the pneumatophore is much smaller, and is thought to play a role in maintaining buoyancy and orientation (Mackie, 1974). The pneumatophore is thought to have been present in the common ancestor of siphonophores, but has been lost in the Calycophorae (the clade that includes *Diphyes dispar*) (Munro et al., 2018). The pneumatophore is formed as an invagination of the aboral end of the planula, and the mature pneumatophore has five distinct tissues: an external ectoderm, associated endoderm, invaginated ectoderm (forming the gas chamber, which is also surrounded by chitin), and a layer of ectodermal cells within the gas chamber that are separated from the basement membrane (Carré, 1969; Church et al., 2015b). Some species, such as *Nanomia bijuga*, have an apical pore at the top of the gas chamber from which gas can be released. The aeriform cells inside the gas chamber are thought to produce the gas that fills the gas chamber of the pneumatophore. In the planktonic species where this was measured, the composition of the float was around 90% carbon monoxide (Carré, 1969; Pickwell et al., 1964). In *Physalia physalis*, the gaseous composition of the float is 0.5-13% carbon monoxide and 15-20% oxygen - in this pleustonic species, the float is thought to be inflated with carbon monoxide initially but this is then replaced by air through diffusion (Wittenberg, 1960). The aeriform cells are packed with mitochondria, Golgi complexes and vesicles, but little smooth and rough endoplasmic reticulum (Copeland, 1968). Below the layer of densely packed mitochondria are "multivesiculate bodies" that consist of multiple membranes and spherical/oval vesicles, and are suggested to be encapsulated mitochondria, however little is known about their origin or fate (Copeland, 1968). Carbon monoxide is hypothesized to be generated in these aeriform cells by utilizing the terminal carbon of serine, in the presence of a tetrahydrofolate (Wittenberg, 1960; Wittenberg et al., 1962). The hypothesized role of serine is based on experimental data – isolated gas glands were incubated with different substrates in seawater and gas production was measured (Wittenberg, 1960); meanwhile folic acid derivatives, especially

tetrahydrofolate, are observed to have very high concentrations in the gas gland (Wittenberg et al., 1962).

A number of gene trees are identified as being specific to the pneumatophore across all species (S3.6), and these gene trees contain genes that are enriched for GO terms such as L-serine metabolic and catabolic process, tetrahydrofolate metabolic process, tetrahydrofolate interconversion, cellular amino acid metabolic process, and pigment metabolic process. Notably, some of the gene trees that are significantly upregulated in pneumatophores across all species are the mitochondrial form of serine hydroxymethyltransferase (catalyzes the reaction of L-serine and tetrahydrofolate to glycine and methylenetetrahydrofolate (Yoshida and Kikuchi, 1973)) and C-1-tetrahydrofolate synthase, cytoplasmic (involved in tetrahydrofolate interconversion (Prasannan and Appling, 2009)). In addition, D-3-phosphoglycerate dehydrogenase (cytoplasmic) is a part of the L-serine biosynthesis pathway (Fan et al., 2014; Pind et al., 2002), while Glycine cleavage system H protein, is involved in glycine and serine synthesis/catabolism in the presence of a tetrahydrofolate (Yoshida and Kikuchi, 1973). Aminomethyltransferase is part of the glycine decarboxylase complex in the mitochondrion, and catalyses the formation of methylenetetrahydrofolate, ammonia and H-protein (Fujiwara et al., 1984). Finally, Carboxypeptidase A4 cleaves hydrophobic C-terminal residues from amino acids (Tanco et al., 2010). Additionally, we identify 5-aminolevulinate synthase (mitochondrial), which converts glycine to 5-aminolevulinate, and is the first step in heme biosynthesis (Ferreira and Gong, 1995). Although we do not find other enzymes involved in heme biosynthesis or catabolism, heme catabolism is a known source of carbon monoxide in mammalian cells (Kikuchi et al., 2005).

The exact mechanism of carbon monoxide production would need to be experimentally determined. However, given the demonstrated role of L-serine in carbon monoxide production (Wittenberg, 1960), it seems likely that these enzymes may be involved either in the biosynthesis of serine or in the reactions that result in carbon monoxide production. Additionally, a number of gene trees containing tissue specific genes were identified that are involved in glycolysis and gluconeogenesis including: Fructose-bisphosphate aldolase A, Phosphoglycerate kinase (both cytoplasmic), and Tricarboxylate transport protein (mitochondrial). Carbon monoxide production is suggested to be energetically expensive, and ATP production by the many mitochondria found in the pneumatophore is likely to be important for this process (Pickwell, 1970). Copeland (1968) noted that the mitochondria in the pneumatophore are unusual and very densely packed. The mi-

tochondria have few cristae, and have a dense granular matrix within the mitochondrion (Copeland, 1968). Some of the enzymes involved in one-carbon metabolism are the mitochondrial form, and it is possible that these densely packed, unusual mitochondria are the site of carbon monoxide synthesis – however, further biochemical verification is needed.

3.4.7   Gene expression in developing zooids

Developing zooids were collected across each of the species, predominately developing gastrozooids and nectophores. In addition, we were able in some species to collect other developing zooids : developing bracts in *Frillagalma vityazi*, developing palpons in *Nanomia bijuga*, and finally developing tentacular palpons in *Physalia physalis*. Bracts are zooids that play a role in defense/buoyancy in codonophorans. And tentacular palpons are a tentacle bearing zooid found only in this species.

Only one gene tree was found to contain significantly upregulated genes across all developing gastrozooids but not mature gastrozooids across species - LIM homeobox transcription factor 1-beta. This gene tree does not contain genes that have zooid specific expression patterns. Developing gastrozooids were pooled, so multiple developmental stages were present within the sample, however, the lack of shared expression across all species may also be due to the fact that the developing gastrozooids are at different stages of development in different species and are not comparable. Very few genes were found to be significantly upregulated in developing gastrozooids vs. mature gastrozooids in some species (22 genes in *Diphyes dispar*, 14 in *Physalia physalis*, 12 in *Nanomia bijuga*, 33 when compared with white mature gastrozooids in *Bargmannia elongata*), however in others, a larger number of genes were identified as upregulated in developing gastrozooids (89 when compared with yellow mature gastrozooids in *Bargmannia elongata*, 334 in *Frillagalma vityazi*). By contrast, there were many genes identified as significantly upregulated in mature gastrozooids relative to developing gastrozooids across all species. The larger number of significantly expressed genes in *Frillagalma vityazi* may also be due to the fact that multiple technical replicates of this zooid were sequenced across different lanes and flow cells, and the expression values were collapsed for these technical replicates - as such, this zooid is more densely sampled in this species than any other. As with developing gastrozooids, only a handful of genes (31 genes) were identified as significantly upregulated in developing palpons as compared

to mature palpons in *Nanomia bijuga*. Similarly, very few genes were identified as being upregulated in developing as compared to mature tentacular palpons (17 genes) in *Physalia physalis*. This suggests that a large number of genes are expressed only in mature zooids that relate to their specific function, and many of the genes that are abundant during development remain abundant in the mature zooid.

A large number of gene trees (274) were found to contain shared genes among developing gastrozooids and developing nectophores. These gene trees were identified by looking at the intersection of significantly differentially upregulated gastrozooid genes from all species with significantly differentially expressed nectophore genes from all species, but likely include larger contributions from some species (e.g. *Bargmannia elongata*, *Frillagalma vityazi*) than others. Of this, 18 gene trees were identified as being significantly upregulated uniquely in developing gastrozooids and nectophores and not in mature zooids or tissues (see table S3.7). The GO terms that were enriched in this set include protein O-linked glycosylation, regulation of gene expression, mechanosensory behavior, maintenance of meristem identity. Notably, Protein Wnt-4 was found to be upregulated, although other members of the Wnt family, Wnt-5b and Wnt-3, were identified as being significantly differentially expressed in these developing zooids (just not uniquely). This suggests that these Wnt genes in particular may play an important role in patterning and development in siphonophores. There is a wide diversity of Wnt genes in Cnidaria, and expression patterns in other species suggest that different Wnt genes play different roles in development and play an important role in anterior-posterior patterning and endoderm specification (Kusserow et al., 2005; Lee et al., 2006; Momose et al., 2008). Within the (274) identified gene trees that are shared among developing gastrozooids and developing nectophores (but not unique to these zooids), some genes are found to be upregulated in embryonic and/or regenerating *Nematostella vectensis* polyps, including: Transcription factor SOX-14, Forkhead box protein P1, Probable C-mannosyltransferase, Heparan sulfate 2-O-sulfotransferase hst-2 , Myelin expression factor 2, Glucoside xylosyltransferase 2 (Warner et al., 2018).

Overall, 29 gene trees and 325 genes were identified as being significantly upregulated only in developing palpons within *Nanomia bijuga*. Among the genes that were found to be upregulated in developing palpons, but not uniquely expressed, 49 gene trees were identified that are shared with the set of genes upregulated in both developing gastrozooids and developing nectophores. These genes included forkhead transcription

factors, Wnt genes, and also many of the genes identified in mature gastrozooids, including digestive enzymes. In the developing bracts of *Frillagalma vityazi*, 384 gene trees and 3338 genes were identified as being significantly upregulated in developing bracts relative to other zooids and tissues. Of this, 87 gene trees are shared with developing gastrozooids and developing nectophores, and an additional 5 overlap with genes that are identified as being unique to developing nectophores and gastrozooids. Finally, 1018 genes and 125 gene trees were significantly upregulated in developing tentacular palpons relative to all other sampled *Physalia physalis* zooids. A number of genes were identified as overlapping with the set of differentially expressed nectophore and gastrozooid genes, including: RNA-binding protein Musashi homolog 2 and ELAV-like protein 4 that are associated with developing neurons and stem cells (Marlow et al., 2009). Only a single gene was found to be expressed in developing nectophores and gastrozooids, developing *Nanomia* palpons, developing *Frillagalma* bracts, and developing *Physalia* tentacular palpons: Heparan sulfate 2-O-sulfotransferase hst-2, a gene that is involved in regulation of cell migration and axon guidance. However, when *Physalia* tentacular palpons are excluded, a slightly larger set of genes are found to be shared among these developing zooids (see table S3.8).

We are not able to learn much about development within specific zooids with this coarse, pooled approach, because there is so much overlap in expression between mature and pooled developing zooids of the same type. Many of these shared genes relate to the specific function of the zooid, and also include genes that are involved in growth patterning and axis specification that are also expressed in the mature zooid. Given the highly precise, ordered formation of buds within the siphosomal growth zone (Fig 3.1), a better approach for understanding the dynamics of gene expression during development would be to sample replicates of each bud separately in ontogenetic order and compare these expression patterns across species. Homologizing differently sized buds of different zooids across species may still be difficult. However, we are able to identify expression patterns that are consistent among all developing zooids. The larger set of 274 gene trees identified as shared among developing nectophores and gastrozooids provides a more comprehensive list of genes that are likely involved in cell fate commitment and determination regardless of zooid type.
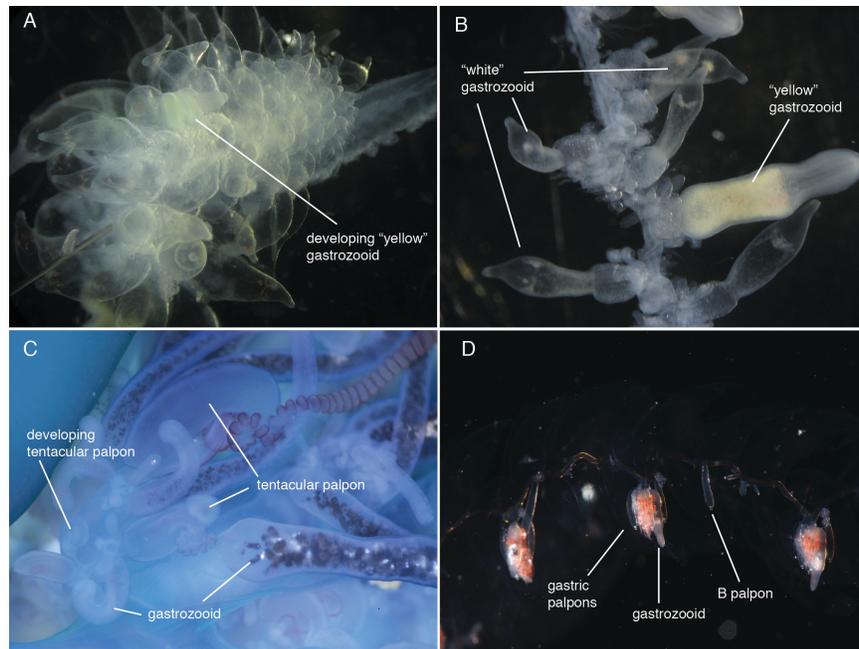
Figure 3.4: Unique siphonophore zooids that were sampled for differential gene expression. A. *Bargmannia elongata* with developing "yellow" gastrozooid surrounded by developing "white" gastrozooids. B. *Bargmannia elongata* stem with mature "white" and "yellow" gastrozooids. C. Zooids in *Physalia physalis*, including multiple developing stages of gastrozooids, tentacular palpon and tentacle; the most mature form of either zooid is not shown. D. Stem of *Agalma elegans*, with gastric palpons, B-palpon and gastrozooid shown.

3.4.8 Novel zooid types

In siphonophores, there are several instances of lineage-specific zooid diversification events. Here we discuss gene expression patterns between the novel zooid type and the hypothesized ancestral type in three species. In *Bargmannia elongata* there are two morphologically distinct gastrozooids, that we termed "white" and "yellow" gastrozooids (Fig. 3.4A and 3.4B). The "yellow" gastrozooid is larger and darker and occurs as the 7th-10th gastrozooid on the stem (Dunn, 2005). In the Portuguese man of war, *Physalia physalis*, the gastrozooid is unique compared to other gastrozooids in other species - it has a mouth, but no tentacle, and the basigaster region is greatly reduced (Totton, 1960; Mackie, 1960). Meanwhile the tentacle is associated with another zooid, the tentacular palpon (see chapter 1) (3.4C) (Bardi and Marques, 2007; Totton, 1960; Haeckel, 1888). In *P. physalis*, both the gastrozooid and the tentacular palpon are considered to be subfunctionalized from an ancestral gastrozooid type (see chapter 1). Finally, in *Agalma elegans*, there are thought to be at least two different palpon types: gastric palpons that arise at the base of the peduncle of the gastrozooid, and a palpon called the B-palpon (3.4D) (Dunn and Wagner, 2006). The distinction

between these two types of palpon is based on the location of these zooids - the gastrozooid is typically the last element of each cormidium, but based on the budding sequence, Dunn and Wagner (2006) propose that the enlarged B-palpon is the last element in *A. elegans*. Each of these cases represents a different type of novelty: in *Bargmannia elongata* the distinction between zooids was made based on size and color but not on obvious differences in function, in *P. physalis* the gastrozooids and tentacular palpons differ structurally and functionally, and finally in *A. elegans*, gastric palpons and B palpons differ only in colony location, development, and possibly size.

A large number of genes were identified as being significantly differentially expressed between mature gastrozooids and the tentacular palpons in *Physalia physalis* (978 genes in the mature tentacular palpon and 571 mature genes in the gastrozooid, representing 4.7% and 2.7% of all genes respectively). A number of genes were found to be differentially expressed in the mature tentacular palpon relative to all other tissues, of which, 1038 genes were found to have expression patterns that were not shared with the gastrozooid. 1133 genes were found to have expression patterns that were not shared with the tentacular palpon. The GO terms enriched in the tentacular palpon included a number of terms: proteolysis, glutathione catabolic and biosynthetic processes, gamma-glutamyl-peptidase activity, chondroitin sulfate proteoglycan biosynthetic process, peptide cross-linking, regulation of endopeptidase activity, cellular response to amino acid stimulus, carbohydrate metabolic process. In particular, Poly-gamma-glutamate and Chondroitin sulfate have been identified as being present in developing nematocysts in *Hydra* (Szczepanek et al., 2002; Yamada et al., 2007). Upregulated tentacular palpon genes include multiple different Gamma-glutamyl hydrolases, Gamma-glutamyltranspeptidase 1 and Glutathione-specific gamma-glutamylcyclotransferase 1, as well as Chondroitin sulfate synthase 1 and Chondroitin sulfate N-acetylgalactosaminyltransferase 1. This supports the notion of the tentacular palpon representing a zooid that specializes in nematogenesis. Nevertheless, the tentacular palpon does appear to also play a role in proteolysis. Notably, 16 gene-trees identified as containing significantly upregulated genes in the tentacular palpon overlap with gene-trees identified as containing gastrozooid specific genes in other species, including Solute carrier family 22 member 18, Protein eva-1 homolog C, Carbonic anhydrase 6, Glutathione peroxidase 1, Snaclec jerdonibitin subunit alpha, amongst others. By contrast, none of the gene-trees identified in the *Physalia physalis* gastrozooid overlap

with those identified as being gastrozooid specific across all other species. The genes upregulated in the gastrozooid nevertheless support the known digestive function of this zooid, and enriched GO terms included: receptor-mediated endocytosis, protein maturation, proteolysis, chitin metabolic process, G-protein coupled receptor signaling pathway, cytolysis, L-amino acid transport amongst others. Altogether, these findings provide support for the hypothesis that the tentacular palpon is a derived gastrozooid that is functionally specialized for nematogenesis, which is also consistent with its role as the tentacle bearing zooid.

Between the white mature gastrozooid and the yellow mature gastrozooid, very few significant expression differences were identified (15 genes were up in "white" mature gastrozooids relative to 15 genes in "yellow" gastrozooids, representing 0.038% and 0.1% of all genes respectively). No GO terms were found to be significantly enriched in either zooid relative to each other. Among genes that were found to be significantly expressed in either "white" or "yellow" gastrozooids relative to all other tissues, 958 genes were unique to "yellow" gastrozooids and not found in "white" gastrozooids, and 1184 genes were found in "white" gastrozooids and not found in "yellow" gastrozooids. These zooids both shared a large number of unique gene trees (79 gene-trees). The functional differences between these two zooids are not clear – some of the gene-trees identified in the white gastrozooid are involved in exocytosis, in cell polarity and immune response, while the yellow gastrozooid included gene-trees involved in endocytosis, immune response and Fc-gamma receptor signalling.

Finally, the number of differentially expressed genes between the B palpon and gastric palpons are incredibly small (13 and 10 genes respectively, representing 0.043% and 0.031% of all genes). Genes were identified that are upregulated in B palpons relative to all other zooids (gastric palpons were excluded). Of this, 687 genes were found to be unique to the B palpon. Most of these genes overlapped with those found specifically in the gastric palpon (628 genes), however, 59 genes were not shared with the gastric palpon. These genes that were found in the B palpon and not the gastric palpon were not found to be significantly enriched for any GO terms.

These findings raise an important question: what constitutes a novel zooid type? Different zooid types are defined based on morphological differences, functional differences, and differences in the location of the zooid in the colony. Large differences are identified between the tentacular palpon and the gastrozooid, and

differential expression patterns in the tentacular palpon suggest that this zooid may indeed be a derived gastrozooid that is functionally specialized for nematocyst production. By contrast, very few genes were found to differ between the two types of gastrozooid in *Bargmannia elongata* or between the two types of palpon in *Agalma elegans.* In both cases, significant expression differences between the two zooids represented less than 0.1% of all genes. By contrast, significantly expressed genes between the male gonodendron and the "yellow" mature gastrozooid in *B. elongata* represent 1.6% and 5.5% of all genes, while significantly expressed genes between mature gastrozooids and gastric palpons in *A. elegans* represent 1.3% and 1.1% of all genes.

These results suggest that the tentacular palpon is a clear example of a novel zooid type that is unique to *Physalia physalis.* There isn't strong evidence with these data that the B palpon and gastric palpons in *Agalma elegans* are sufficiently different to constitute a novel zooid type, and these findings suggest that location within the colony is not necessarily sufficient to designate a novel zooid type. Finally, while the difference between the two gastrozooid types in *Bargmannia elongata* was not large, different genes were identified as being differentially expressed in either type relative to all other zooids. This suggests that these two gastrozooids may indeed be distinct zooid types, although they are functionally and morphologically very similar to one another. Greater sequencing coverage, and also functional work within this species may help clarify the nature of these differences.

3.5   Conclusions

In this study, we take a more inclusive approach to the evaluation of genes across species - instead of considering only strict 1:1 orthologs, we use gene trees as the unit of comparison, enabling the identification of homologous sets of genes whose expression is unique to particular zooids or species. As with analyses with strict orthologs, by focusing on patterns that are shared across all species these results are necessarily limited by the most poorly sampled species. This limitation is not unique to this particular method. As with all comparative work, a large number of diverse species are required, but also a large number of different, well-sampled treatments are required in order to sample the full space of expression diversity. In species

with only two different treatments, treatment and species-specific differences were hard to identify, due to the fact that significant differential expression is determined in a pairwise manner.

Using gene trees as the unit of comparison has some distinct benefits over traditional approaches - it opens up opportunities to investigate expression evolution in particular treatments within a particular gene tree, as opposed to within the context of the species phylogeny. Gene trees represent a hypothesis about the evolutionary relationships between genes, and typically consist not only of speciation events, representing genes that are descended from a common ancestor within a species, but also of duplication events, genes that share a common ancestor within a genome. By identifying gene trees with expression patterns that are unique to particular zooids within species, or shared across all species, it is possible to interrogate the nature of these expression differences within the evolutionary context of the genes themselves. I will discuss this in chapter 4.

It was possible to identify expression patterns that are putatively unique to particular treatments, and that make sense in the context of their hypothesized function. There are a number of different ways to subset the list of expressed genes among treatments that are more restrictive or inclusive (e.g. considering the intersection of gene trees among treatments vs the union of gene trees among treatments). Here, we consider shared expression across a particular zooid as being the intersection of expression across all species for that zooid. As such, it is possible that some of these treatment-specific genes are not in fact specific to particular treatments, but shared among a subset of treatments in particular species in a manner that is not captured here. Nevertheless, it was possible to identify genes with significant expression patterns that indicate an abundance of RNA expression in that treatment relative to others, in a manner that is conserved across all species. Regardless of whether they are truly treatment-specific, they are likely important for the functioning and structuring of that particular treatment. For example, among the putative gastrozooid specific genes were a number of proteolytic genes and genes that are likely involved in the extracellular digestion of prey items. In addition, we were able to identify a number of enzymes likely involved in serine biosynthesis and in the pathway likely involved in the generation of carbon monoxide in the float (Wittenberg, 1960; Wittenberg et al., 1962).

The identification of developmental processes that shape particular zooids was difficult with this dataset. It

was possible to identify sets of genes that are common to all developing zooids that are likely involved in patterning and specification, however there were few expression differences among mature and developing zooids of the same type. Many expression patterns appear to be shared among developing and mature gastrozooids. Sampling only the youngest developing buds, or a time series of developing buds, may provide more information about patterning processes that are unique to zooid-specific patterning during development.

Using the descriptive knowledge gained from these expression analyses, it was also possible to test hypotheses about the function of novel zooid types within particular lineages relative to a hypothesized ancestral type, and also to test whether location in the colony has any bearing on the identity and function of particular zooid types. These findings suggest that, at least in the case of the B-palpon, there is little expression difference between the B-palpon and the gastric palpon - indicating that although the B-palpon is located in a distinct location within a cormidium, the underlying molecular functions of this zooid is consistent with that of the gastric palpon. As palpons are all considered to be derived gastrozooids, it may be difficult to disentangle cases of *de novo* modification of gastrozooids, as has been hypothesized here (Dunn and Wagner, 2006). Very few differences were found between "white" and "yellow" gastrozooids, and it is unclear what, if any, functional differences may exist between these two morphologically distinct zooids. Finally, expression findings within the tentacular palpon of *Physalia physalis* suggest that this novel zooid type is indeed functionally specialized for nematocyst production, and is likely also be a derived gastrozooid.

3.6   Supplementary information
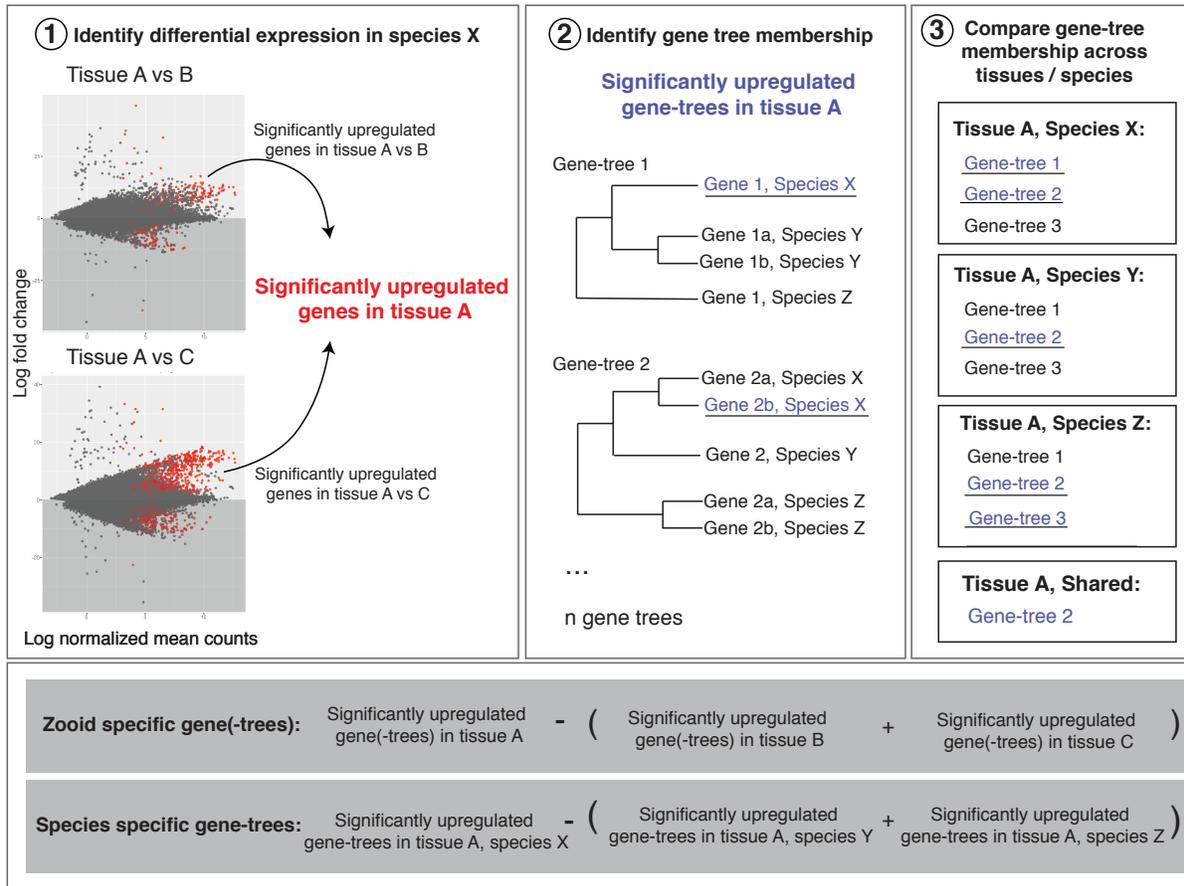
Figure S3.1: Methods used to identify upregulated genes in particular zooids within species, enabling comparisons across species. Step 1 outlines identification of significantly differentially expressed genes in one tissue relative to all other tissues. Step 2 indicates that significantly expressed genes are found within particular gene-trees. Step 3 indicates how gene tree membership can be compared across all species to identify shared expression patterns in particular tissues.
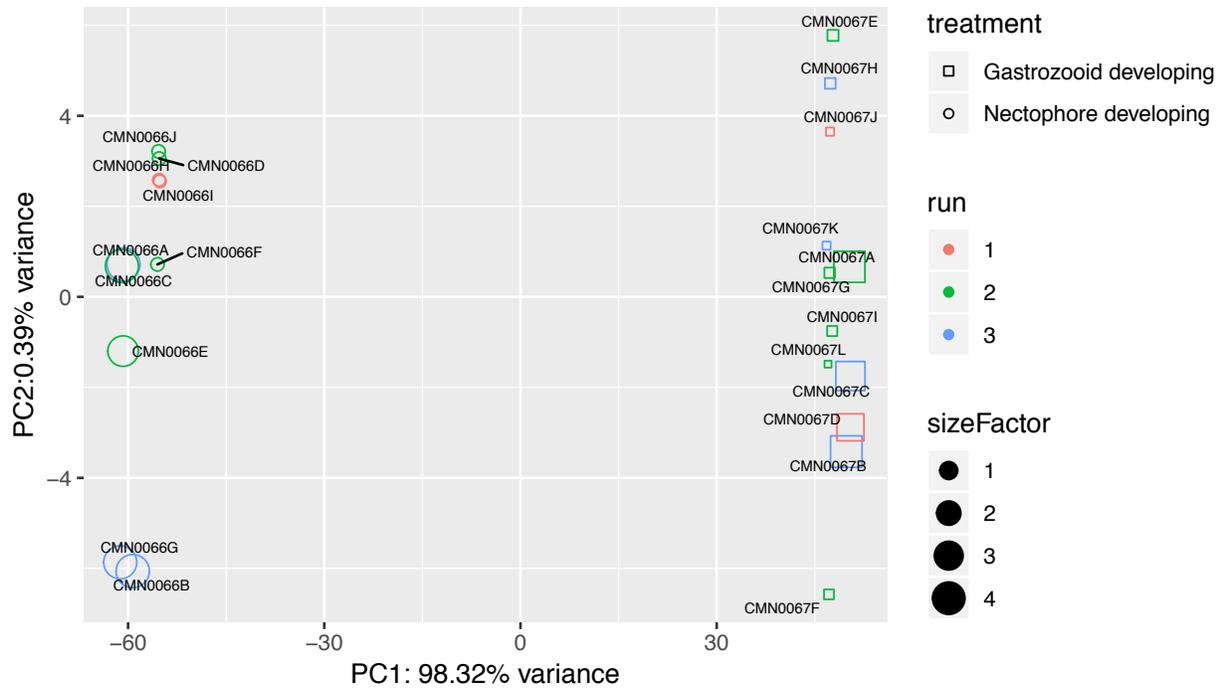
Figure S3.2: PCA of regularized log transformed expression counts of technical replicates of two different zooids from different runs and lanes. Color indicates the run number, shape indicates the zooid, and size factor indicates number of genes sequenced.
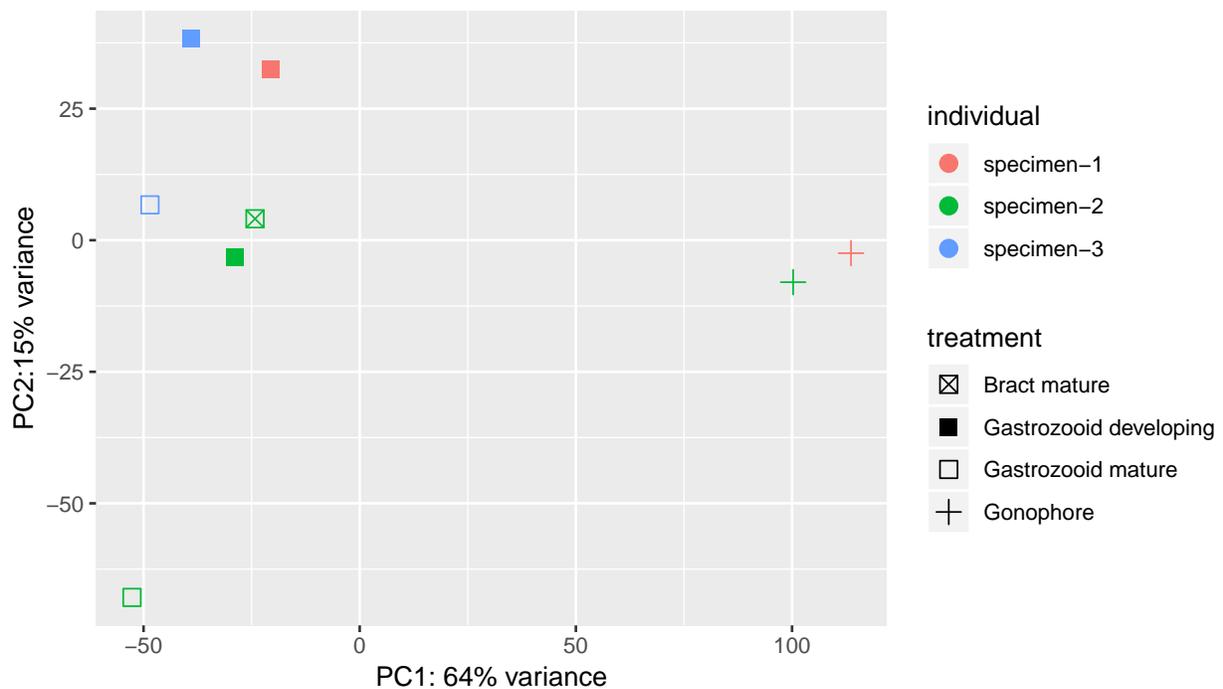
Figure S3.3: PCA of regularized log transformed expression counts of treatments in *Diphyes dispar*. Color indicates the replicate number, shape indicates the zooid.
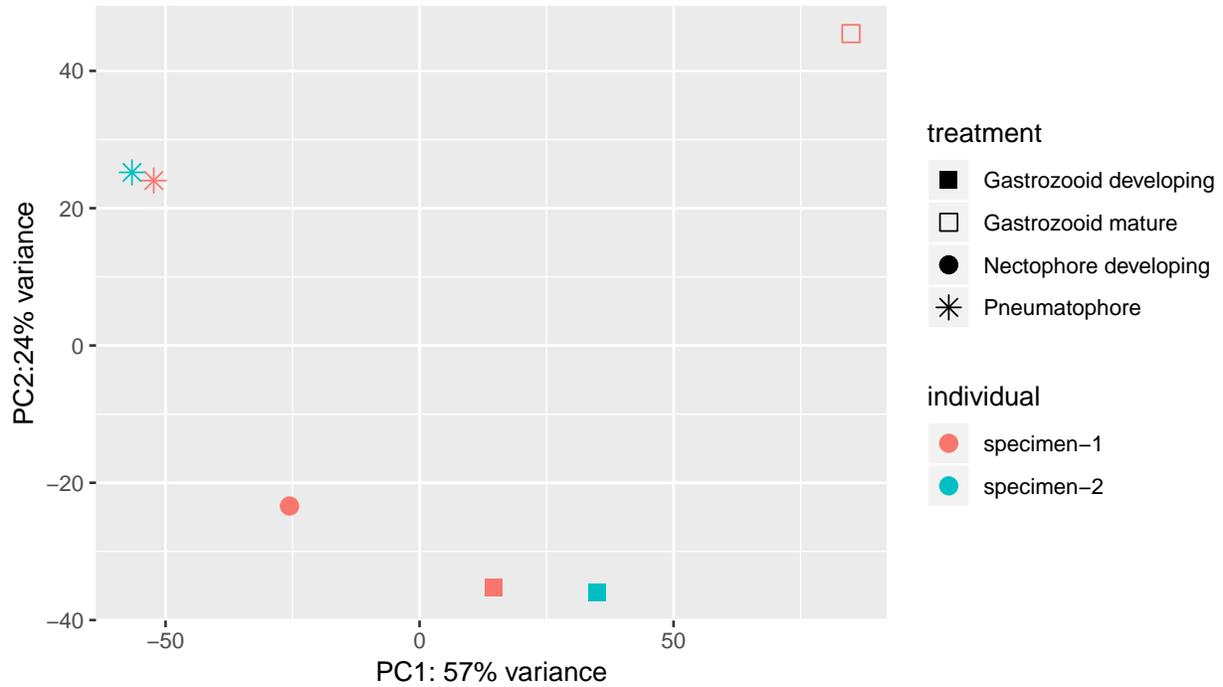
Figure S3.4: PCA of regularized log transformed expression counts of treatments in *Apolemia lanosa*. Color indicates the replicate number, shape indicates the zooid. Single replicates are included in the PCA, but not considered in differential expression analyses.
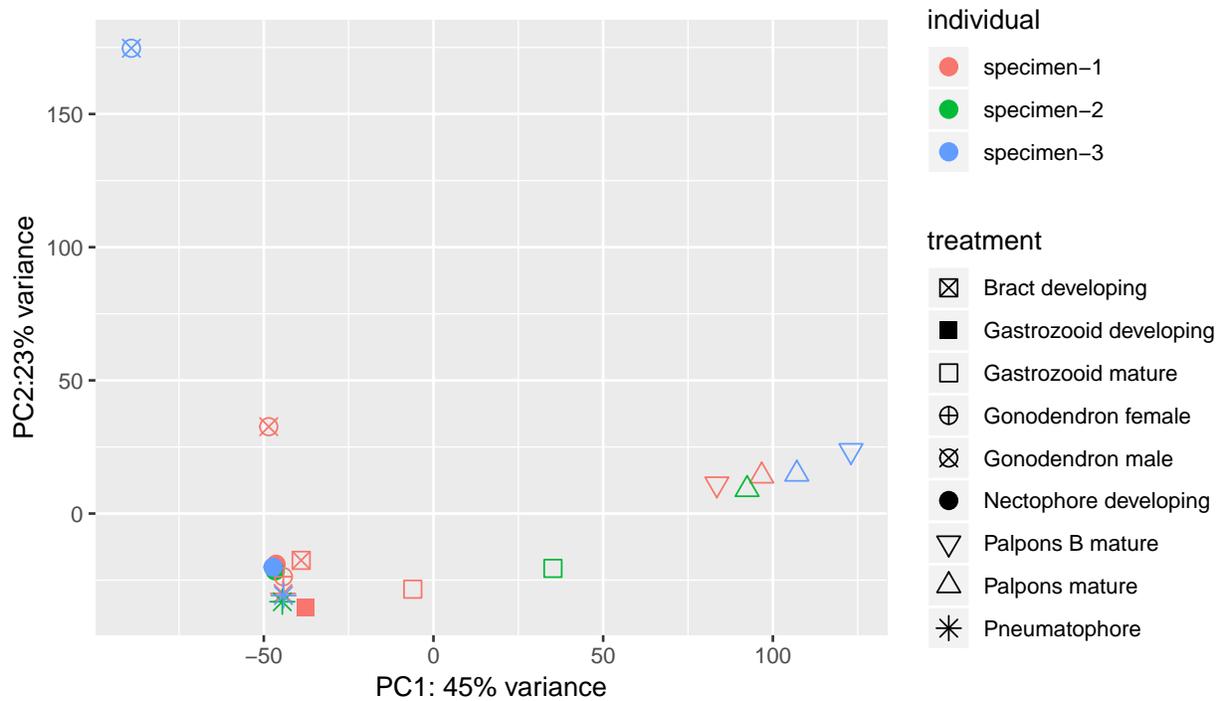
Figure S3.5: PCA of regularized log transformed expression counts of treatments in *Agalma elegans*. Color indicates the replicate number, shape indicates the zooid. Single replicates are included in the PCA, but not considered in differential expression analyses.
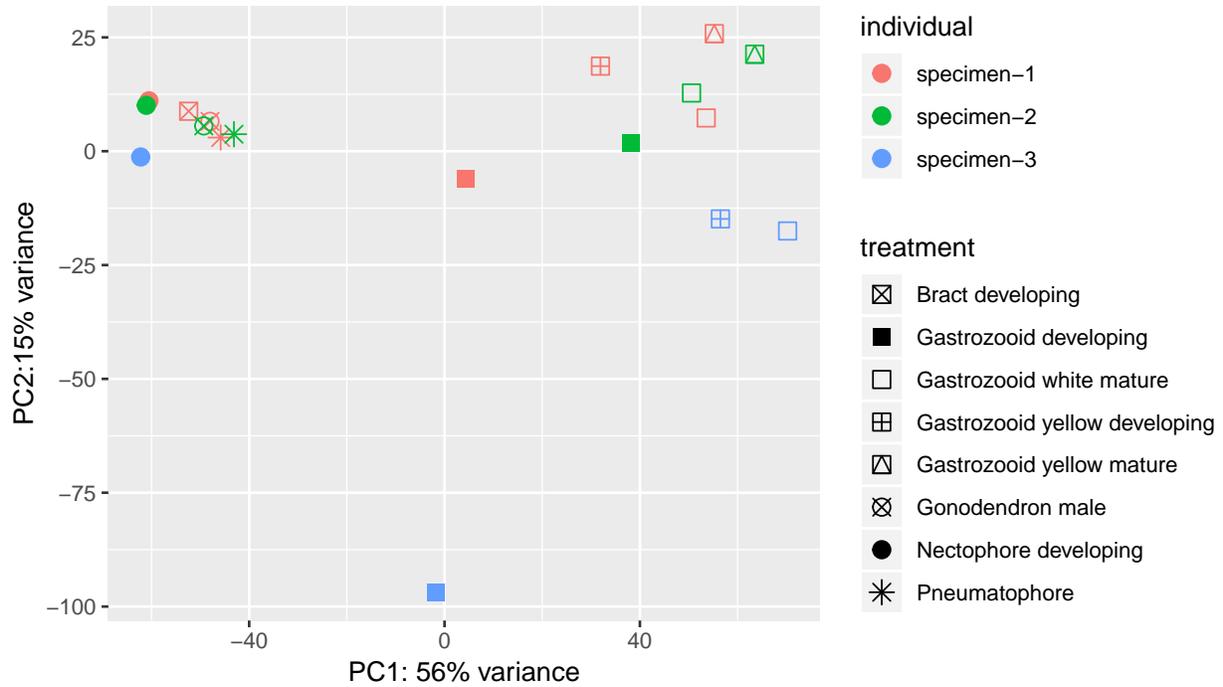
Figure S3.6: PCA of regularized log transformed expression counts of treatments in *Bargmannia elongata*. Color indicates the replicate number, shape indicates the zooid. Single replicates are included in the PCA, but not considered in differential expression analyses.
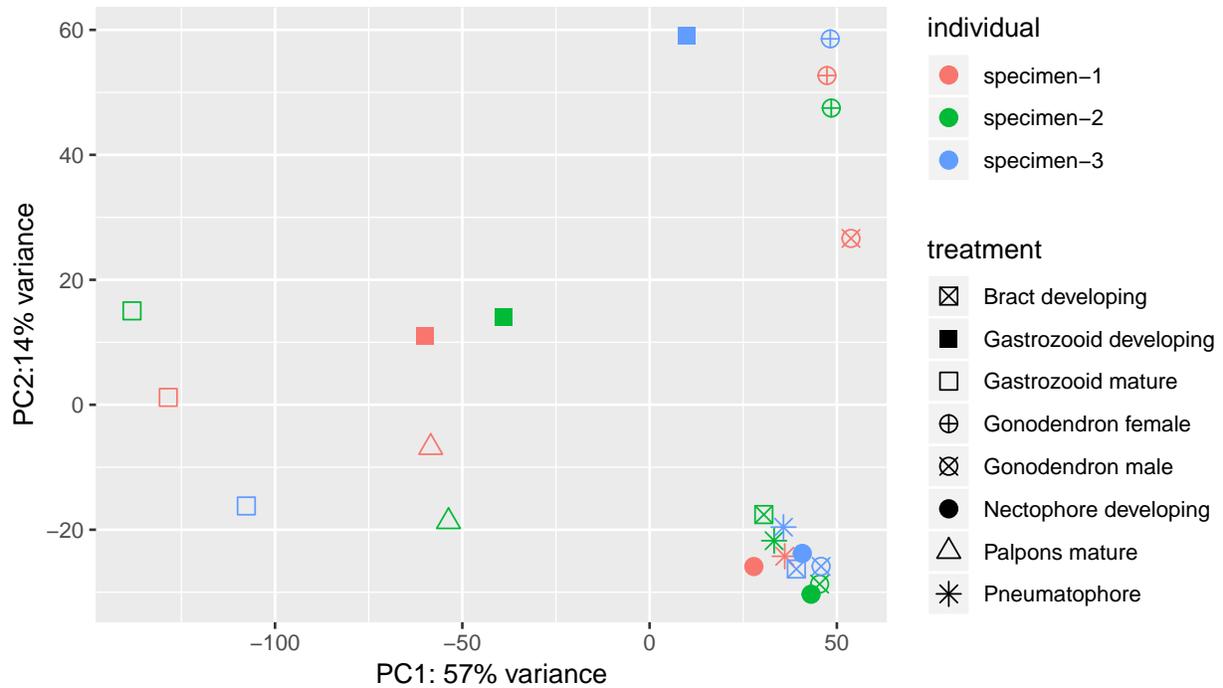
Figure S3.7: PCA of regularized log transformed expression counts of treatments in *Frillagalma vityazi*. Color indicates the replicate number, shape indicates the zooid. Single replicates are included in the PCA, but not considered in differential expression analyses.
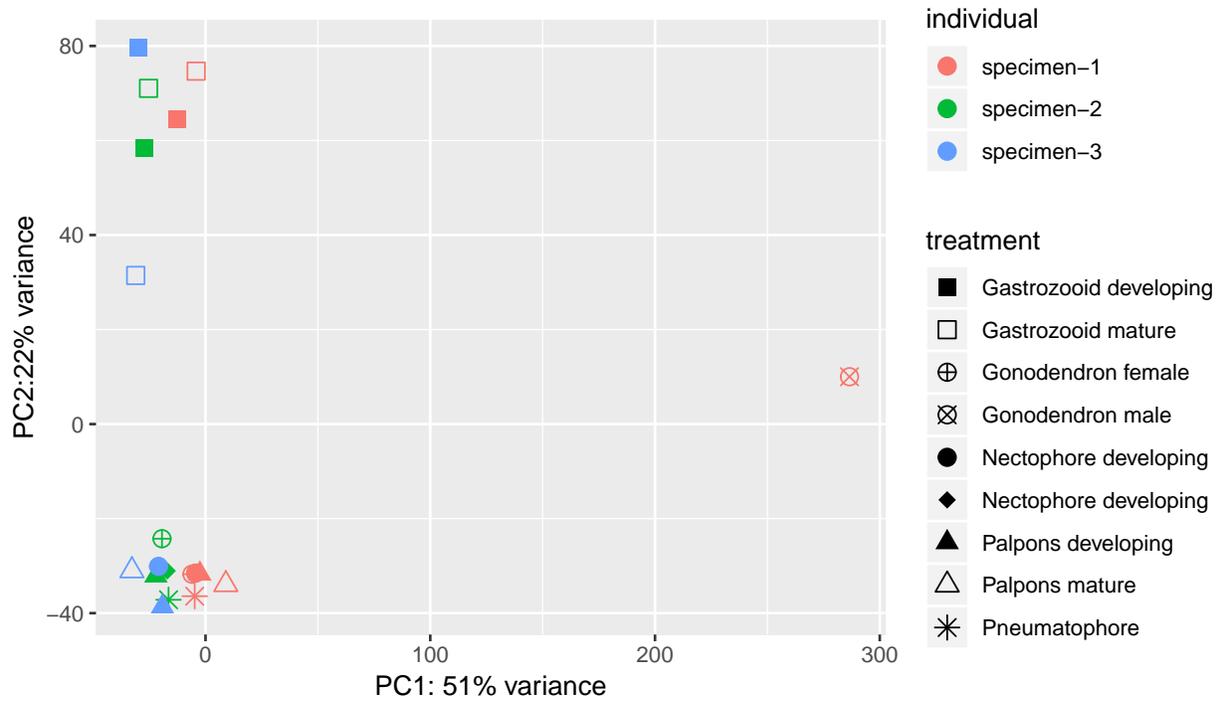
Figure S3.8: PCA of regularized log transformed expression counts of treatments in *Nanomia bijuga*. Color indicates the replicate number, shape indicates the zooid. Single replicates are included in the PCA, but not considered in differential expression analyses.
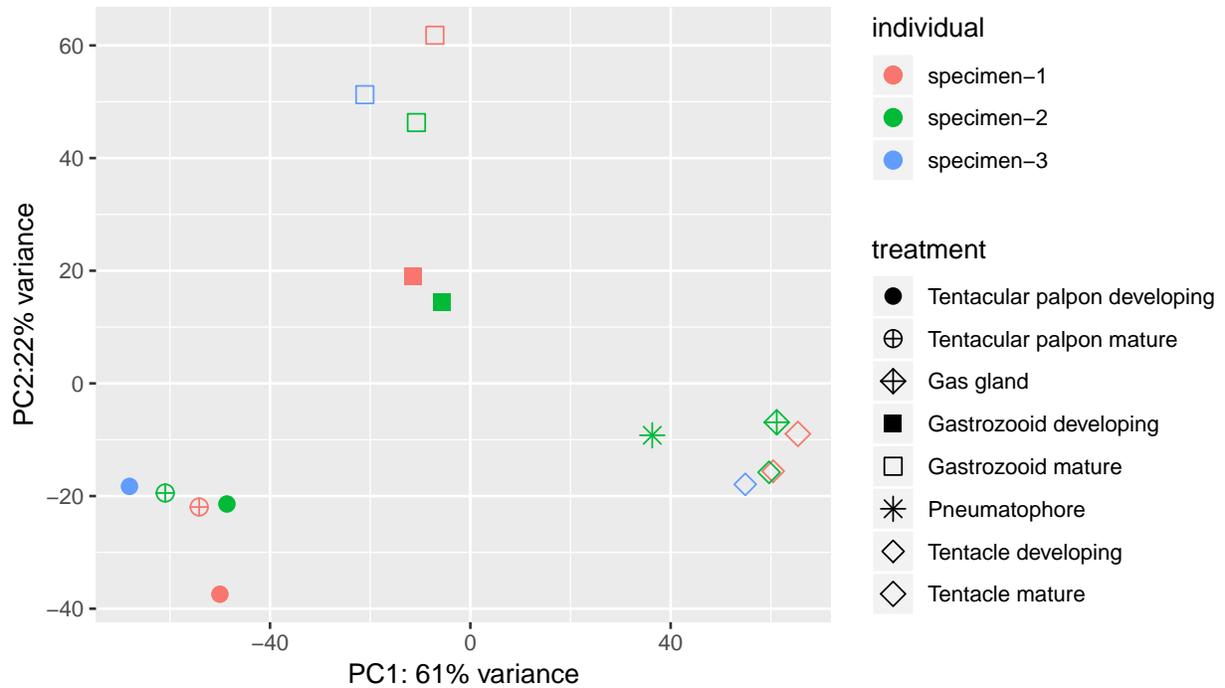
Figure S3.9: PCA of regularized log transformed expression counts of treatments in *Physalia physalis*. Color indicates the replicate number, shape indicates the zooid. Single replicates are included in the PCA, but not considered in differential expression analyses.
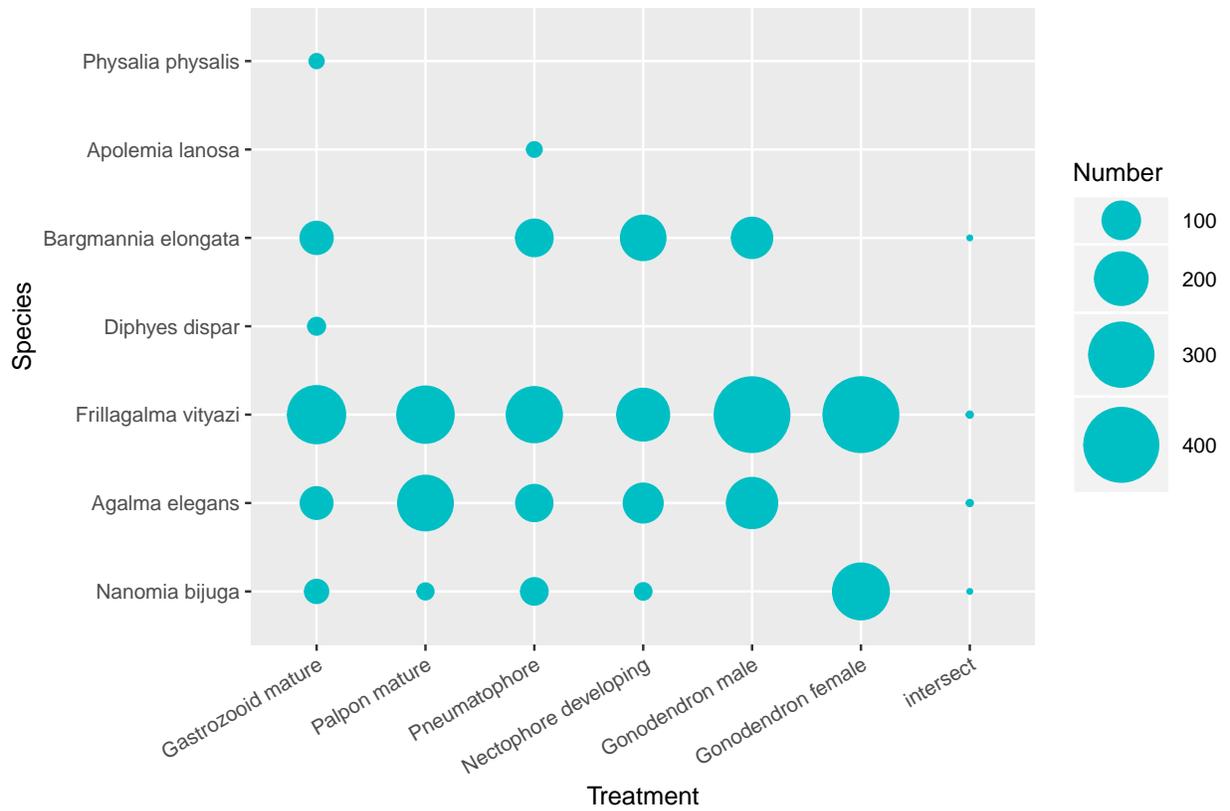
Figure S3.10: Number of gene trees containing genes that have species-specific expression patterns within a treatment. Species-specific patterns indicate that the gene tree containing significantly upregulated genes in a particular treatment is unique to that species, and gene tree membership is not found in the same treatment in other species. Intersect indicates the total number of gene trees with homologous genes that are shared across all zooids of a particular species and that are differentially expressed exclusively in that species. Area of circle indicates number of gene or gene trees. Missing values are where no treatments were sampled for this species, and does not indicate a lack of treatment-specific genes.

Table S3.1: Gene trees containing homologous genes that are significantly upregulated specifically in mature gastrozooids, across all sampled species. Gene trees are unique gene tree identifiers (first 8 alphanumeric characters). Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5.

Gene tree	Most common Blast hit	Associated GO terms
0a72d074	P58659|EVA1C_MOUSE Protein eva-1 homolog C	carbohydrate binding, proteolysis, integral component of membrane, metallocarboxypeptidase activity, zinc ion binding
0af70276	Q9JI85|NUCB2_RAT Nucleobindin-2	calcium ion binding
11bcbf3d	P49010|HEXC_BOMMO Chitooligosaccharidolytic beta-N-acetylglucosaminidase	beta-N-acetylhexosaminidase activity, carbohydrate metabolic process, plasma membrane, rhodopsin biosynthetic process, N-acetyl-beta-D-galactosaminidase activity
18376115	Q80T32|AGRD1_MOUSE Adhesion G-protein coupled receptor D1	G-protein coupled receptor activity, cell surface receptor signaling pathway, integral component of membrane, G-protein coupled receptor signaling pathway, GTPase activity
246df25c	Q9GRC0|MOS_PATPE Serine/threonine-protein kinase mos {ECO:0000250|UniProtKB:P10741}	protein kinase activity, protein phosphorylation, nucleus, integral component of membrane, ATP binding
38ba0571	Q9H4G4|GAPR1_HUMAN Golgi-associated plant pathogenesis-related protein 1	extracellular region, integral component of membrane

Table S3.1: Gene trees containing homologous genes that are significantly upregulated specifically in mature gastrozooids, across all sampled species. Gene trees are unique gene tree identifiers (first 8 alphanumeric characters). Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
39df45ec	P29025|CHI1_RHINI Chitinase 1	hydrolase activity, hydrolyzing O-glycosyl compounds, carbohydrate metabolic process, extracellular region, chitin metabolic process, fungal-type cell wall
3c9d60c1	Q5SPB6|CHAC1_DANRE Glutathione-specific gamma-glutamylcyclotransferase 1 {ECO:0000250|UniProtKB:Q9BUX1}	gamma-glutamylcyclotransferase activity, glutathione catabolic process, cytosol, transferase activity, ATP-binding cassette (ABC) transporter complex
3f9aaaea	Q10572|FOX1_CAEEL Sex determination protein fox-1	regulation of RNA splicing, RNA binding, nucleus, RNA splicing, mRNA metabolic process
4242981d	P55112|NAS4_CAEEL Zinc metalloproteinase nas-4	metalloendopeptidase activity, proteolysis, extracellular region, zinc ion binding, meprin A complex
43144cfc	F1NPQ2|MINP1_CHICK Multiple inositol polyphosphate phosphatase 1 {ECO:0000250|UniProtKB:Q9UNW1}	phosphatase activity, dephosphorylation
4917b4dc	Q920A5|RISC_MOUSE Retinoid-inducible serine carboxypeptidase	serine-type carboxypeptidase activity, proteolysis, cytosol, extracellular region, negative regulation of blood pressure

Table S3.1: Gene trees containing homologous genes that are significantly upregulated specifically in mature gastrozooids, across all sampled species. Gene trees are unique gene tree identifiers (first 8 alphanumeric characters). Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
4e262d69	Q28CK1|NR2C1_XENTR Nuclear receptor subfamily 2 group C member 1	steroid hormone receptor activity, steroid hormone mediated signaling pathway, nucleus, sequence-specific DNA binding, zinc ion binding
57d71123	P43135|COT2_MOUSE COUP transcription factor 2	steroid hormone receptor activity, steroid hormone mediated signaling pathway, nucleus, nuclear receptor activity, intracellular receptor signaling pathway
5804bf2e	O75629|CREG1_HUMAN Protein CREG1	cofactor binding
67b9ea48	Q8MJ14|GPX1_PIG Glutathione peroxidase 1	glutathione peroxidase activity, response to oxidative stress, extracellular region, cellular oxidant detoxification, sperm plasma membrane
6950be86	Q92820|GGH_HUMAN Gamma-glutamyl hydrolase	gamma-glutamyl-peptidase activity, proteolysis, extracellular space, vacuole, glutamine metabolic process
6ca4a033	Q9VAS7|INX3_DROME Innexin inx3	gap junction, ion transport, plasma membrane, integral component of membrane
768423e1	C3YWU0|FUCO_BRAFL Alpha-L-fucosidase	alpha-L-fucosidase activity, fucose metabolic process, outer acrosomal membrane, sperm plasma membrane, fucose binding

Table S3.1: Gene trees containing homologous genes that are significantly upregulated specifically in mature gastrozooids, across all sampled species. Gene trees are unique gene tree identifiers (first 8 alphanumeric characters). Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
794cb9c1	P50059|SODM2_LEPBY Superoxide dismutase [Mn] 2	superoxide dismutase activity, removal of superoxide radicals, metal ion binding, oxidation-reduction process
7bfbb823	Q865C0|CAH6_CANLF Carbonic anhydrase 6	protein tyrosine phosphatase activity, peptidyl-tyrosine dephosphorylation, integral component of membrane, carbonate dehydratase activity, extracellular space
835be6a5	O94766|B3GA3_HUMAN Galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase 3	galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase activity, Golgi membrane, protein glycosylation, metal ion binding, UDP-galactose:beta-N-acetylglucosamine beta-1,3-galactosyltransferase activity
8f1eae37	Q9BZ76|CNTP3_HUMAN Contactin-associated protein-like 3	hyaluronoglucosaminidase activity, integral component of plasma membrane, signal transduction, carbohydrate binding, transmembrane signaling receptor activity
94ea92be	P42674|BP10_PARLI Blastula protease 10	metalloendopeptidase activity, proteolysis, extracellular region, molting cycle, collagen and cuticulin-based cuticle, membrane

Table S3.1: Gene trees containing homologous genes that are significantly upregulated specifically in mature gastrozooids, across all sampled species. Gene trees are unique gene tree identifiers (first 8 alphanumeric characters). Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
951dae3e	Q54TR1|CFAD_DICDI Counting factor associated protein D	cysteine-type peptidase activity, proteolysis, lysosome, extracellular space, endopeptidase activity
96fd200e	Q96BI1|S22AI_HUMAN Solute carrier family 22 member 18	transporter activity, transmembrane transport, integral component of membrane
9fdfd8e1	Q99LJ6|GPX7_MOUSE Glutathione peroxidase 7	glutathione peroxidase activity, response to oxidative stress, nucleoplasm, cellular oxidant detoxification, mitochondrion
b5bfde57	O43315|AQP9_HUMAN Aquaporin-9	channel activity, transmembrane transport, integral component of membrane, water transport, water transmembrane transporter activity
bee0565f	Q2L6K8|CNPY4_DANRE Protein canopy 4	integral component of membrane
c484b73c	Q66I24|ASSY_DANRE Argininosuccinate synthase	argininosuccinate synthase activity, arginine biosynthetic process, myelin sheath, cytoplasm, ATP binding
c6c26a79	P24367|PPIB_CHICK Peptidyl-prolyl cis-trans isomerase B	peptidyl-prolyl cis-trans isomerase activity, protein peptidyl-prolyl isomerization, extracellular space, protein folding, U4/U6 snRNP

Table S3.1: Gene trees containing homologous genes that are significantly upregulated specifically in mature gastrozooids, across all sampled species. Gene trees are unique gene tree identifiers (first 8 alphanumeric characters). Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
dbb5d983	Q9VYN8|TENA_DROME Teneurin-a	Notch binding, Notch signaling pathway, integral component of membrane, calcium ion binding, multicellular organism development
de1b3200	D1MGU0|SLA_PROJR Snaclec jerdonibitin subunit alpha	integral component of membrane
e2a890b6	A8TX70|CO6A5_HUMAN Collagen alpha-5(VI) chain	collagen trimer, extracellular matrix structural constituent, cell adhesion, basement membrane, extracellular matrix organization
ea1145b1	Q71RP1|HPSE_RAT Heparanase	hydrolase activity, acting on glycosyl bonds, membrane

Table S3.2: Gene trees containing homologous genes that are significantly upregulated specifically in mature palpons, across all sampled species. Gene trees are unique gene tree identifiers (first 8 alphanumeric characters). Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5.

Gene tree	Most common Blast hit	Associated GO terms
0d3dee21	Q9H1J7|WNT5B_HUMAN Protein Wnt-5b	Wnt signaling pathway, signaling receptor binding, extracellular region, cell surface, cytoplasm
20c8f78e	Q96HU8|DIRA2_HUMAN GTP-binding protein Di-Ras2	GTPase activity, signal transduction, membrane, GTP binding, intracellular
29ec8b35	P55112|NAS4_CAEEL Zinc metalloproteinase nas-4	metalloendopeptidase activity, proteolysis, membrane, molting cycle, collagen and cuticulin-based cuticle, zinc ion binding
2e64a020	Q9Y5Z4|HEBP2_HUMAN Heme-binding protein 2	protein glycosylation, Golgi membrane, carbohydrate binding, transferase activity, transferring glycosyl groups, integral component of membrane
3e82172e	P56839|PEPM_MYTED Phosphoenolpyruvate phosphomutase	phosphoenolpyruvate mutase activity, organic phosphonate biosynthetic process, integral component of membrane, phosphonopyruvate hydrolase activity, lyase activity
521540b9	Q6GM78|ASGL1_XENLA Isoaspartyl peptidase/L-asparaginase	asparaginase activity, proteolysis, cytoplasm, asparagine catabolic process via L-aspartate, beta-aspartyl-peptidase activity

Table S3.2: Gene trees containing homologous genes that are significantly upregulated specifically in mature palpons, across all sampled species. Gene trees are unique gene tree identifiers (first 8 alphanumeric characters). Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
5881d87d	Q9M883|SC5D2__ARATH Putative Delta(7)-sterol-C5(6)-desaturase 2	iron ion binding, lipid biosynthetic process, integral component of membrane, oxidoreductase activity, oxidation-reduction process
5c293bef	P86009|RBP__DRONO Riboflavin-binding protein {ECO:0000303|PubMed:19416692, ECO:0000312|EMBL:BAH22358.1}	folic acid binding, cellular response to folic acid, anchored component of external side of plasma membrane, folic acid receptor activity, folic acid import across plasma membrane
7811af98	Q6P6S2|S39AB__RAT Zinc transporter ZIP11	metal ion transmembrane transporter activity, metal ion transport, integral component of membrane, transmembrane transport, ribosome
95266985	Q6DRG7|MYPT1__DANRE Protein phosphatase 1 regulatory subunit 12A	integral component of membrane
97d9144e	A2AJ76|HMCN2__MOUSE Hemicentin-2	calcium ion binding, microfibril, coronary vasculature development, aorta development, ventricular septum development

Table S3.2: Gene trees containing homologous genes that are significantly upregulated specifically in mature palpons, across all sampled species. Gene trees are unique gene tree identifiers (first 8 alphanumeric characters). Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
9d493bcf	P82968|MCPI_MELCP Four-domain proteases inhibitor	calcium ion binding, proteolysis, secretory granule, negative regulation of brain-derived neurotrophic factor receptor signaling pathway, negative regulation of collateral sprouting
9e47f304	P12256|PAC_LYSSH Penicillin acylase	hydrolase activity
a430b949	Q9JJ09|NPT2B_RAT Sodium-dependent phosphate transport protein 2B	sodium-dependent phosphate transmembrane transporter activity, sodium-dependent phosphate transport, plasma membrane, apical part of cell, brush border
a8b2c15e	Q5R5H1|METK2_PONAB S-adenosylmethionine synthase isoform type-2	methionine adenosyltransferase activity, S-adenosylmethionine biosynthetic process, cytosol, hyphal cell wall, one-carbon metabolic process
b1242116	Q5AF03|HSP31_CANAL Glyoxalase 3 {ECO:0000303|PubMed:24302734}	glutamine metabolic process, peptidase activity, spindle pole, transferase activity, microtubule associated complex
bab216d1	Q01984|HNMT_RAT Histamine N-methyltransferase	methyltransferase activity, methylation

Table S3.2: Gene trees containing homologous genes that are significantly upregulated specifically in mature palpons, across all sampled species. Gene trees are unique gene tree identifiers (first 8 alphanumeric characters). Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
d5e15e94	Q569D5|SBP1_XENTR Selenium-binding protein 1	selenium binding, fibrillar center, protein transport, methanethiol oxidase activity, protein binding
d9bf9133	Q9TWL9|COMA_CONMA Conodipine-M alpha chain	arachidonate transport, phospholipase A2 activity, extracellular region, icosanoid secretion, calcium ion binding
e24836ac	Q74FW6|TSAL_GEOSL L-threonine ammonia-lyase {ECO:0000303|PubMed:18245290}	L-threonine ammonia-lyase activity, threonine catabolic process, integral component of membrane, serine racemase activity, pyridoxal phosphate binding
e912f53d	Q8HXW6|PPT1_MACFA Palmitoyl-protein thioesterase 1	protein depalmitoylation, palmitoyl-(protein) hydrolase activity, lysosome, synaptic vesicle, membrane raft
f170ed32	Q5RCR9|CPPED_PONAB Serine/threonine-protein phosphatase CPPED1	hydrolase activity, integral component of membrane
f4b58822	Q03168|ASPP_AEDAE Lysosomal aspartic protease	aspartic-type endopeptidase activity, lysosome, proteolysis, retinal pigment epithelium development, methyltransferase activity

Table S3.2: Gene trees containing homologous genes that are significantly upregulated specifically in mature palpons, across all sampled species. Gene trees are unique gene tree identifiers (first 8 alphanumeric characters). Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
f4ddea00	Q60648|SAP3_MOUSE Ganglioside GM2 activator	ganglioside catabolic process, enzyme activator activity, positive regulation of catalytic activity

Table S3.3: Gene trees containing homologous genes that are significantly upregulated specifically in developing nectophores, across all sampled species. Gene trees are unique gene tree identifiers (first 8 alphanumeric characters). Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5.

Gene tree	Most common Blast hit	Associated GO terms
0d62beb1	Q92172|TEF_CHICK Transcription factor VBP	DNA binding transcription factor activity, regulation of transcription, DNA-templated
0ec57764	B3EWZ6|MLRP2_ACRMI MAM and LDL-receptor class A domain-containing protein 2	metalloendopeptidase activity, proteolysis, zinc ion binding
22817f82	Q8WPW2|PDX1_SUBDO Pyridoxal 5'-phosphate synthase subunit SNZERR	pyridoxal phosphate biosynthetic process, pyridoxal 5'-phosphate synthase (glutamine hydrolysing) activity, extracellular region, integral component of membrane, GMP synthase (glutamine-hydrolyzing) activity
38d40121	O54068|UDG_RHIME UDP-glucose 6-dehydrogenase {ECO:0000305}	UDP-glucose 6-dehydrogenase activity, polysaccharide biosynthetic process, cytoplasm, NAD binding, oxidation-reduction process
4dc89744	Q6IMZ0|NFIL3_RAT Nuclear factor interleukin-3-regulated protein	circadian rhythm, DNA binding transcription factor activity, nucleus, transcription by RNA polymerase II, immune response
67cef73c	Q62632|FSTL1_RAT Follistatin-related protein 1	calcium ion binding, extracellular region, response to starvation, heparin binding, maintenance of gastrointestinal epithelium

Table S3.3: Gene trees containing homologous genes that are significantly upregulated specifically in developing nectophores, across all sampled species. Gene trees are unique gene tree identifiers (first 8 alphanumeric characters). Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
6dd04572	A6PWV5|ARI3C_MOUSE AT-rich interactive domain-containing protein 3C	transcriptional activator activity, RNA polymerase II transcription regulatory region sequence-specific DNA binding, transcription by RNA polymerase II, nucleus, RNA polymerase II regulatory region sequence-specific DNA binding, positive regulation of transcription by RNA polymerase II
b0652cd2	P54357|MLC2_DROME Myosin-2 essential light chain	calcium ion binding
be9e810b	Q09711|NCS1_SCHPO Calcium-binding protein NCS-1	calcium ion binding
ca40564c	Q9U5M4|TPM2_PODCA Tropomyosin-2	calcium ion binding
fd5c7110	P27166|CALM_STYLE Calmodulin	calcium ion binding, regulation of the force of heart contraction, A band, ventricular cardiac muscle tissue morphogenesis, I band

Table S3.4: Gene trees containing homologous genes that are significantly upregulated specifically in male gonodendra, across all sampled species. Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5.

Gene tree	Most common Blast hit	Associated GO terms
0398a7b2	Q8YV57|Y2124_NOSS1 Uncharacterized WD repeat-containing protein all2124	ubiquitin-protein transferase activity, protein ubiquitination, nuclear lumen, phosphoserine residue binding, nuclear SCF ubiquitin ligase complex
0b27f565	P36124|SET3_YEAST SET domain-containing protein 3	meiotic sister chromatid cohesion, synaptonemal complex, metal ion binding, regulation of homologous chromosome segregation, meiotic recombination checkpoint
0b8f6994	Q9WU79|PROD_MOUSE Proline dehydrogenase 1, mitochondrial	proline dehydrogenase activity, proline catabolic process, mitochondrion, FAD binding, calcium ion binding
129c6664	Q8AVY1|ODF3A_XENLA Outer dense fiber protein 3	spermatid development, cytoplasmic microtubule, sensory perception of sound
13367504	Q5NVA9|EME1_PONAB Crossover junction endonuclease EME1	intracellular protein transport, membrane, intracellular
14192967	Q8NEA4|FBX36_HUMAN F-box only protein 36	nucleic acid binding
18cf30da	A7E320|UHRF1_BOVIN E3 ubiquitin-protein ligase UHRF1	nucleus, ligase activity, maintenance of DNA methylation, metal ion binding, protein autoubiquitination

Table S3.4: Gene trees containing homologous genes that are significantly upregulated specifically in male gonodendra, across all sampled species. Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
1a05fb97	Q03123|T2FB_XENLA General transcription factor IIF subunit 2	transcription factor TFIIF complex, transcription initiation from RNA polymerase II promoter, helicase activity, transcription factor activity, core RNA polymerase II binding, DNA binding
1a682459	P31335|PUR9_CHICK Bifunctional purine biosynthesis protein PURH	phosphoribosylaminoimidazolecarboxamide formyltransferase activity, purine nucleotide biosynthetic process, cytosol, IMP cyclohydrolase activity, plasma membrane
1b0be91c	A6H639|TCTE1_MOUSE T-complex-associated testis-expressed protein 1	sperm motility, sperm flagellum, protein binding, cytoskeleton, cytoplasm
1b0cd646	Q4R6T7|IQUB_MACFA IQ and ubiquitin-like domain-containing protein	acrosomal vesicle, smoothened signaling pathway, protein binding, cilium assembly, thiol-dependent ubiquitinyl hydrolase activity
1c51ca60	Q13257|MD2L1_HUMAN Mitotic spindle assembly checkpoint protein MAD2A	mitotic spindle assembly checkpoint, DNA binding

Table S3.4: Gene trees containing homologous genes that are significantly upregulated specifically in male gonodendra, across all sampled species. Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
1cfe5476	Q5XTS1|PLPL8_RABIT Calcium-independent phospholipase A2-gamma	lipid catabolic process, hydrolase activity, plastid, jasmonic acid biosynthetic process, oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen
1e0c6538	Q8VE62|PAIP1_MOUSE Polyadenylate-binding protein-interacting protein 1	RNA binding
20961f20	P51892|DNLI1_XENLA DNA ligase 1	DNA ligation involved in DNA repair, DNA ligase (ATP) activity, nucleus, mitochondrion, DNA biosynthetic process
2430403f	A7RRJ0|FEN1_NEMVE Flap endonuclease 1 {ECO:0000255|HAMAP-Rule:MF_03140}	DNA replication, removal of RNA primer, 5'-flap endonuclease activity, nucleolus, nucleoplasm, 5'-3' exonuclease activity
245c79e8	Q4KLQ5|WDR76_XENLA WD repeat-containing protein 76	cellular response to DNA damage stimulus, nucleus, DNA binding, regulation of DNA damage checkpoint
24d12038	Q071E0|KT5AA_DANRE N-lysine methyltransferase KMT5A-A {ECO:0000305}	histone-lysine N-methyltransferase activity, histone lysine methylation, nucleus, chromosome, zinc ion binding
25014d3c	Q9D552|SPT17_MOUSE Spermatogenesis-associated protein 17	calmodulin binding, cytoplasm, integral component of membrane

Table S3.4: Gene trees containing homologous genes that are significantly upregulated specifically in male gonodendra, across all sampled species. Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
254bc081	Q7ZUS1|VRK1_DANRE Serine/threonine-protein kinase VRK1	protein kinase activity, protein phosphorylation, nucleolus, Golgi stack, M band
257a3d0c	P51987|CCNB_HYDVD G2/mitotic-specific cyclin-B	nucleus, T cell homeostasis, patched binding, thymus development, histone kinase activity
2a23c574	P13439|UMPS_MOUSE Uridine 5'-monophosphate synthase	orotidine-5'-phosphate decarboxylase activity, 'de novo' pyrimidine nucleobase biosynthetic process, nucleus, orotate phosphoribosyltransferase activity, 'de novo' UMP biosynthetic process
2a652d29	Q9CXE6|XRCC3_MOUSE DNA repair protein XRCC3	DNA-dependent ATPase activity, DNA repair, DNA recombinase mediator complex, replication fork, perinuclear region of cytoplasm
2e882f31	P33610|PRI2_MOUSE DNA primase large subunit	DNA primase activity, DNA replication, synthesis of RNA primer, alpha DNA polymerase:primase complex, 4 iron, 4 sulfur cluster binding, metal ion binding

Table S3.4: Gene trees containing homologous genes that are significantly upregulated specifically in male gonodendra, across all sampled species. Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
34e601d8	Q2TBT5|RNH2A_BOVIN Ribonuclease H2 subunit A	RNA-DNA hybrid ribonuclease activity, RNA phosphodiester bond hydrolysis, endonucleolytic, ribonuclease H2 complex, nucleus, cytosol
3ab52344	A7MBP4|IFT46_DANRE Intraflagellar transport protein 46 homolog {ECO:0000250|UniProtKB:Q9DB07, ECO:0000312|ZFIN:ZDB-GENE-080102-3}	intraciliary transport, cilium, protein C-terminus binding, intraciliary transport particle B, centrosome
3d54eea1	Q6GL41|MCM4_XENTR DNA replication licensing factor mcm4	MCM complex, DNA replication initiation, DNA helicase activity, nucleus, DNA duplex unwinding
3f0b5da3	Q5ZJJ8|UBCP1_CHICK Ubiquitin-like domain-containing CTD phosphatase 1	phosphoprotein phosphatase activity, protein dephosphorylation, nucleus, metal ion binding, intracellular organelle lumen
40123c01	Q5ZMD2|ANKY2_CHICK Ankyrin repeat and MYND domain-containing protein 2	regulation of smoothened signaling pathway, metal ion binding, translation initiation factor activity, kinase activity, translational initiation
412b3825	Q8K4K3|TRIB2_MOUSE Tribbles homolog 2	protein kinase activity, protein phosphorylation, nucleus, transcription corepressor activity, ATP binding

Table S3.4: Gene trees containing homologous genes that are significantly upregulated specifically in male gonodendra, across all sampled species. Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
442c47a7	Q9H4K1|RIBC2_HUMAN RIB43A-like with coiled-coils protein 2	protein binding, nucleus, integral component of membrane, Golgi apparatus
462bc974	Q6ZMY6|WDR88_HUMAN WD repeat-containing protein 88	Prp19 complex, histone modification, ubiquitin-protein transferase activity, nuclear lumen, peptidyl-lysine modification
4c753b2d	Q9CQ66|TC1D2_MOUSE Tctex1 domain-containing protein 2	formation of cytoplasmic translation initiation complex, eukaryotic 43S preinitiation complex, translation initiation factor activity, eukaryotic 48S preinitiation complex, eukaryotic translation initiation factor 3 complex
4cf49bec	Q91ZY6|HOP2_RAT Homologous-pairing protein 2 homolog	reciprocal meiotic recombination, ligand-dependent nuclear receptor transcription coactivator activity, nucleoplasm, DBD domain binding, glucocorticoid receptor binding
50168093	A4IJ21|MNS1_XENTR Meiosis-specific nuclear structural protein 1	left/right axis specification, motile cilium, identical protein binding, positive regulation of cilium assembly, cilium organization
51e6a213	A6H8T2|CASC1_DANRE Protein CASC1	RNA binding, translational initiation, integral component of membrane, phosphorylation, kinase activity

Table S3.4: Gene trees containing homologous genes that are significantly upregulated specifically in male gonodendra, across all sampled species. Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
5427e70f	B1H283|LEXM_RAT Lymphocyte expansion molecule	arachidonate transport, phospholipase A2 activity, extracellular region, integral component of membrane, icosanoid secretion
57d7290b	Q8WXX5|DNJC9_HUMAN DnaJ homolog subfamily C member 9	heat shock protein binding, social behavior, extracellular space, nucleoplasm, unfolded protein binding
5b269bec	Q9D439|CFA53_MOUSE Cilia- and flagella-associated protein 53 {ECO:0000312|MGI:MGI:1921703}	cilium movement, microtubule motor activity, myosin complex, microtubule binding, cilium assembly
5c7656be	Q5HZL1|ERI2_XENLA ERI1 exoribonuclease 2	nucleic acid binding, DNA metabolic process, integral component of membrane, zinc ion binding, nucleic acid phosphodiester bond hydrolysis
5f9b163d	Q4V7B5|CC105_RAT Coiled-coil domain-containing protein 105	integral component of membrane
647ec194	Q5T2R2|DPS1_HUMAN Decaprenyl-diphosphate synthase subunit 1	isoprenoid biosynthetic process, transferase activity, extrinsic component of mitochondrial inner membrane, metallopeptidase activity, neuroblast development

Table S3.4: Gene trees containing homologous genes that are significantly upregulated specifically in male gonodendra, across all sampled species. Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
6514fb45	Q16831|UPP1_HUMAN Uridine phosphorylase 1	uridine phosphorylase activity, nucleotide catabolic process, cytoplasm, nucleoside metabolic process, protein serine/threonine phosphatase activity
67c16013	Q8NEP3|DAAF1_HUMAN Dynein assembly factor 1, axonemal	motile cilium assembly, cilium, dynein complex binding, extracellular ligand-gated ion channel activity, plasma membrane bounded cell projection cytoplasm
6978a2e0	O95359|TACC2_HUMAN Transforming acidic coiled-coil-containing protein 2	syntaxin binding
6fb437d1	A7SUU7|SCC4_NEMVE MAU2 chromatid cohesion factor homolog	mitotic sister chromatid cohesion
76d84696	P69341|PARN_BOVIN Poly(A)-specific ribonuclease PARN	nucleic acid binding
7d49fdf3	A1L2F3|NUSAP_DANRE Nucleolar and spindle-associated protein 1	establishment of mitotic spindle localization, spindle, microtubule, mitotic cytokinesis
83d26072	Q6NU40|CTF18_XENLA Chromosome transmission fidelity protein 18 homolog	ATP binding, positive regulation of DNA-directed DNA polymerase activity, Ctf18 RFC-like complex, nucleoplasm, cytosol

Table S3.4: Gene trees containing homologous genes that are significantly upregulated specifically in male gonodendra, across all sampled species. Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
8543de4d	Q9BWT3|PAPOG_HUMAN Poly(A) polymerase gamma	RNA polyadenylation, polynucleotide adenylyltransferase activity, nucleus, RNA 3'-end processing, RNA binding
902836f0	Q96T60|PNKP_HUMAN Bifunctional polynucleotide phosphatase/kinase	kinase activity, phosphorylation, nucleus, polynucleotide 3' dephosphorylation, polynucleotide 3'-phosphatase activity
939dc6c5	Q9D845|TEX9_MOUSE Testis-expressed sequence 9 protein	motile cilium, cilium organization, structural constituent of ribosome, protein binding, nucleic acid binding
93d8f9eb	P38024|PUR6_CHICK Multifunctional protein ADE2	phosphoribosylaminoimidazolesuccinocarboxamide synthase activity, 'de novo' IMP biosynthetic process, cytoplasm, extracellular exosome, identical protein binding
96b9505a	P10242|MYB_HUMAN Transcriptional activator Myb	DNA binding, regulation of transcription, DNA-templated, nucleus, regulatory region nucleic acid binding, chromatin
9c6a9096	Q5HZP1|RNH2B_XENLA Ribonuclease H2 subunit B	nucleus
9fafa5b1	A4QNE6|WDR92_XENTR WD repeat-containing protein 92	ubiquitin binding, apoptotic process, integral component of membrane, calcium ion binding

Table S3.4: Gene trees containing homologous genes that are significantly upregulated specifically in male gonodendra, across all sampled species. Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
9fbb9c95	Q9BYN7|ZN341__HUMAN Zinc finger protein 341	nucleic acid binding, nucleus, regulation of development, heterochronic, regulation of transcription, DNA-templated, RNA polymerase II transcription factor activity, sequence-specific DNA binding
a26e70b2	Q5PQR6|THEGL__RAT Testicular haploid expressed gene protein-like	NA
a7c3f37a	O54956|DPOE2__MOUSE DNA polymerase epsilon subunit 2	epsilon DNA polymerase complex, DNA-directed DNA polymerase activity, DNA-dependent DNA replication, DNA biosynthetic process, DNA binding
b23bd90b	Q9NPB8|GPCP1__HUMAN Glycerophosphocholine phosphodiesterase GPCPD1	phosphoric diester hydrolase activity, lipid metabolic process, cell, integral component of membrane, starch binding
b8947d4b	Q86Y56|DAAF5__HUMAN Dynein assembly factor 5, axonemal {ECO:0000303|PubMed:25232951, ECO:0000312|HGNC:HGNC:26013}	outer dynein arm assembly, dynein intermediate chain binding, cytoplasm, inner dynein arm assembly, cilium movement
b9c6c7d9	O93257|XRCC6__CHICK X-ray repair cross-complementing protein 5	Ku70:Ku80 complex, telomeric DNA binding, double-strand break repair via nonhomologous end joining, damaged DNA binding, telomere maintenance

126

Table S3.4: Gene trees containing homologous genes that are significantly upregulated specifically in male gonodendra, across all sampled species. Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
bead6f90	Q3V0Q6|SPAG8_MOUSE Sperm-associated antigen 8 {ECO:0000312|MGI:MGI:3056295}	microtubule binding, positive regulation of transcription by RNA polymerase II, nucleus, cytoplasm
c529c787	Q4V7T8|ROP1L_XENLA Ropporin-1-like protein	sperm capacitation, motile cilium, kinase activity, flagellated sperm motility, sperm cytoplasmic droplet
c6aea7fc	Q9UL16|CFA45_HUMAN Cilia- and flagella-associated protein 45 {ECO:0000312|HGNC:HGNC:17229}	nucleoplasm, Rab GTPase binding, intracellular protein transport, motile cilium, 3'-5' exonuclease activity
c94ca1ad	A6QM04|ZWILC_BOVIN Protein zwilch homolog	RZZ complex, mitotic cell cycle checkpoint
c9a77110	Q5F4A1|G2E3_CHICK G2/M phase-specific E3 ubiquitin-protein ligase	metal ion binding, ligase activity
cd1157f4	Q6DRC3|CU059_DANRE UPF0769 protein C21orf59 homolog	cell projection morphogenesis, cilium, phosphatidylinositol phosphate 5-phosphatase activity, cytoskeleton, cytosol
cea5838f	Q5M939|HAT1_RAT Histone acetyltransferase type B catalytic subunit	chromatin silencing at telomere, histone acetyltransferase activity, chromosome, telomeric region, histone binding, histone acetylation
d1942cfd	Q9DAQ4|CB081_MOUSE Uncharacterized protein C2orf81 homolog	NA

Table S3.4: Gene trees containing homologous genes that are significantly upregulated specifically in male gonodendra, across all sampled species. Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
d1fcb0b4	Q5SF07|IF2B2__MOUSE Insulin-like growth factor 2 mRNA-binding protein 2	RNA binding
d2f3a65d	Q9CR92|CCD96__MOUSE Coiled-coil domain-containing protein 96	microtubule organizing center, NAD-dependent histone deacetylase activity (H3-K14 specific), histone H3 deacetylation, transmembrane signaling receptor activity, motile cilium
d3007a51	Q9W719|HPRT__CHICK Hypoxanthine-guanine phosphoribosyltransferase	hypoxanthine phosphoribosyltransferase activity, IMP salvage, cytoplasm, guanine phosphoribosyltransferase activity, purine ribonucleoside salvage
d416d596	Q95YJ5|TXND3__CIOIN Thioredoxin domain-containing protein 3 homolog	nucleoside diphosphate kinase activity, UTP biosynthetic process, cell, GTP biosynthetic process, CTP biosynthetic process
d6d619b8	Q567I9|CB5D1__DANRE Cytochrome b5 domain-containing protein 1	motile cilium, metal ion binding, chromatin organization, centrosome cycle, histone demethylase activity (H3-K27 specific)
d94cf1b0	Q86UC2|RSPH3__HUMAN Radial spoke head protein 3 homolog	cilium, protein binding, regulation of transcription by RNA polymerase II, protein transport, RNA polymerase II regulatory region sequence-specific DNA binding

Table S3.4: Gene trees containing homologous genes that are significantly upregulated specifically in male gonodendra, across all sampled species. Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
e3a93af8	Q9BZI7|REN3B_HUMAN Regulator of nonsense transcripts 3B	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay, nucleic acid binding, integral component of membrane, ATP binding, nucleus
e7138cb5	P39963|CCNB3_CHICK G2/mitotic-specific cyclin-B3	nucleus, regulation of G2/M transition of mitotic cell cycle, protein kinase binding, regulation of cyclin-dependent protein serine/threonine kinase activity, cyclin-dependent protein serine/threonine kinase regulator activity
e7370119	Q9QXL7|NDK7_RAT Nucleoside diphosphate kinase 7	nucleoside diphosphate kinase activity, UTP biosynthetic process, cytoplasm, GTP biosynthetic process, CTP biosynthetic process
ef453843	Q99741|CDC6_HUMAN Cell division control protein 6 homolog	DNA replication initiation, nucleus, kinase binding, ATP binding, cell division
f853d2dd	Q8TC29|ENKUR_HUMAN Enkurin	sperm principal piece, calmodulin binding, nucleoside metabolic process, SH3 domain binding, acrosomal vesicle
f9501b29	Q8R3P7|CLUA1_MOUSE Clusterin-associated protein 1	intraciliary transport particle B, cilium assembly, protein binding, DNA-directed 5'-3' RNA polymerase activity, centrosome

Table S3.4: Gene trees containing homologous genes that are significantly upregulated specifically in male gonodendra, across all sampled species. Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
fb2506d2	Q8C5T8|CC113_MOUSE Coiled-coil domain-containing protein 113	centriolar satellite, cilium assembly, transferase activity, transferring hexosyl groups, nucleoplasm, cytosol
ff704c86	Q96M32|KAD7_HUMAN Adenylate kinase 7	cytidylate kinase activity, nucleotide phosphorylation, cytosol, adenylate kinase activity, ATP binding

Table S3.5: Gene trees containing homologous genes that are significantly upregulated specifically in female gonodendra, across all sampled species. Gene trees are unique gene tree identifiers (first 8 alphanumeric characters). Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5.

Gene tree	Most common Blast hit	Associated GO terms
03c017ae	Q9H8H3|MET7A_HUMAN Methyltransferase-like protein 7A	methyltransferase activity, methylation, lipid droplet, menaquinone biosynthetic process, integral component of membrane
05ac1b29	Q08E12|PSF3_BOVIN DNA replication complex GINS protein PSF3	DNA replication, nucleus, chromatin binding, protein binding, intracellular organelle lumen
0b8f0865	Q8IYB1|M21D2_HUMAN Protein MB21D2	transmembrane transporter activity, transmembrane transport, integral component of membrane
0b932bc3	Q9SZJ2|GRDP2_ARATH Glycine-rich domain-containing protein 2 {ECO:0000303|PubMed:25653657}	integral component of membrane
1c43e027	Q8IUF8|MINA_HUMAN Bifunctional lysine-specific demethylase and histidyl-hydroxylase MINA	transcriptional repressor activity, RNA polymerase II transcription factor binding, negative regulation of transcription by RNA polymerase II, nucleolus, transcription factor complex, identical protein binding
21ac1304	Q1LWC2|T106B_DANRE Transmembrane protein 106B	integral component of membrane, zinc ion binding

Table S3.5: Gene trees containing homologous genes that are significantly upregulated specifically in female gonodendra, across all sampled species. Gene trees are unique gene tree identifiers (first 8 alphanumeric characters). Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
2817db6e	K4CF70|HPL_SOLLC Fatty acid hydroperoxide lyase, chloroplastic {ECO:0000303|PubMed:10859201}	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, oxidation-reduction process, monooxygenase activity, iron ion binding, heme binding
2a42d0a8	Q6NRS2|PQLC1_XENLA PQ-loop repeat-containing protein 1	integral component of membrane
2c8034de	D3ZAT9|FAXC_RAT Failed axon connections homolog	calcium-dependent phospholipid binding, cellular protein modification process, integral component of membrane, transferase activity, calcium ion binding
3fc69fd3	P97478|COQ7_MOUSE 5-demethoxyubiquinone hydroxylase, mitochondrial {ECO:0000255|HAMAP-Rule:MF_03194}	ubiquinone biosynthetic process, extrinsic component of mitochondrial inner membrane, oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, NAD(P)H as one donor, and incorporation of one atom of oxygen, 2-octoprenyl-3-methyl-6-methoxy-1,4-benzoquinone hydroxylase activity, metal ion binding

Table S3.5: Gene trees containing homologous genes that are significantly upregulated specifically in female gonodendra, across all sampled species. Gene trees are unique gene tree identifiers (first 8 alphanumeric characters). Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
43fc4673	A6MEY4|PA2B_BUNFA Basic phospholipase A2 BFPA	arachidonic acid secretion, phospholipase A2 activity, extracellular region, other organism presynaptic membrane, phospholipase A2 activity consuming 1,2-dioleoylphosphatidylethanolamine)
4566e958	Q20191|NAS13_CAEEL Zinc metalloproteinase nas-13	metalloendopeptidase activity, proteolysis, extracellular region, integral component of membrane, molting cycle, collagen and cuticulin-based cuticle
49643a6d	Q92541|RTF1_HUMAN RNA polymerase-associated protein RTF1 homolog	Cdc73/Paf1 complex, transcription elongation from RNA polymerase II promoter, DNA binding, histone modification
5766c0ff	P27607|PGH2_CHICK Prostaglandin G/H synthase 2	prostaglandin-endoperoxide synthase activity, cyclooxygenase pathway, neuron projection, peroxidase activity, endoplasmic reticulum
5d1cc94b	Q496A3|SPAS1_HUMAN Spermatogenesis-associated serine-rich protein 1	NA

Table S3.5: Gene trees containing homologous genes that are significantly upregulated specifically in female gonodendra, across all sampled species. Gene trees are unique gene tree identifiers (first 8 alphanumeric characters). Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
61fabd63	Q8WV22|NSE1_HUMAN Non-structural maintenance of chromosomes element 1 homolog	SUMO ligase complex, DNA repair, condensed chromosome, chromosomal part, intracellular signal transduction
6a447d6a	Q766D5|B4GN4_MOUSE N-acetyl-beta-glucosaminyl-glycoprotein 4-beta-N-acetylgalactosaminyltransferase 1	acetylgalactosaminyltransferase activity, Golgi cisterna membrane, catalytic activity, acting on a glycoprotein, CCR4-NOT complex, integral component of membrane
6b7eec9e	A7RJI7|CCZ1_NEMVE Vacuolar fusion protein CCZ1 homolog	vesicle, vesicle-mediated transport, guanyl-nucleotide exchange factor activity, lysosomal membrane, calcium ion binding
6c8be0b1	P08911|ACM5_RAT Muscarinic acetylcholine receptor M5	G-protein coupled receptor activity, G-protein coupled receptor signaling pathway, integral component of membrane
7188f9b2	Q8R087|B4GT7_MOUSE Beta-1,4-galactosyltransferase 7	transferase activity, transferring glycosyl groups, carbohydrate metabolic process, Golgi apparatus, proteoglycan metabolic process, negative regulation of fibroblast proliferation
7b09595c	Q681K7|GEX1_ARATH Protein GAMETE EXPRESSED 1	integral component of membrane

Table S3.5: Gene trees containing homologous genes that are significantly upregulated specifically in female gonodendra, across all sampled species. Gene trees are unique gene tree identifiers (first 8 alphanumeric characters). Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
81ef7a76	Q8N9W6|BOLL_HUMAN Protein boule-like	germ cell development, mRNA binding
85654715	Q9BZ19|ANR60_HUMAN Ankyrin repeat domain-containing protein 60	signal transduction
86fe85d3	Q5RDQ3|AMERL_PONAB AMMECR1-like protein	nucleus, protein binding, integral component of membrane
a314c672	Q5SP85|CC85A_MOUSE Coiled-coil domain-containing protein 85A	translation release factor activity, translational termination
a58ddc99	B3EWZ5|MLRP1_ACRMI MAM and LDL-receptor class A domain-containing protein 1	serine-type endopeptidase activity, proteolysis, extracellular region, chorion, integral component of membrane
ad67bb2e	Q66IF1|REEP6_DANRE Receptor expression-enhancing protein 6	integral component of membrane
b313d2fe	Q8VD72|TTC8_MOUSE Tetratricopeptide repeat protein 8	non-motile cilium assembly, BBSome, RNA polymerase II repressing transcription factor binding, carbohydrate binding, photoreceptor connecting cilium

Table S3.5: Gene trees containing homologous genes that are significantly upregulated specifically in female gonodendra, across all sampled species. Gene trees are unique gene tree identifiers (first 8 alphanumeric characters). Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
b41444ab	O54862|MBTP2_CRIGR Membrane-bound transcription factor site-2 protease	metalloendopeptidase activity, proteolysis, integral component of membrane, positive regulation of transcription from RNA polymerase II promoter in response to endoplasmic reticulum stress, cytoplasm
b542e974	Q91Z46|DUS7_MOUSE Dual specificity protein phosphatase 7	protein tyrosine/serine/threonine phosphatase activity, inactivation of MAPK activity, cytoplasm, MAP kinase phosphatase activity, protein tyrosine phosphatase activity
c4d4b4e1	P87139|YDM9_SCHPO Uncharacterized RING finger protein C57A7.09	ligase activity, vesicle docking involved in exocytosis, intracellular, zinc ion binding, integral component of membrane
d15151b1	Q2IBC1|CAV1_RHIFE Caveolin-1	caveola assembly, caveola, Golgi membrane
dfbd2ffc	Q80V23|ZNF32_MOUSE Zinc finger protein 32	nucleic acid binding, nucleus, regulation of transcription, DNA-templated, zinc ion binding, aminoacylase activity

Table S3.5: Gene trees containing homologous genes that are significantly upregulated specifically in female gonodendra, across all sampled species. Gene trees are unique gene tree identifiers (first 8 alphanumeric characters). Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
e03d8f40	A7RWC9|ITPA_NEMVE Inosine triphosphate pyrophosphatase {ECO:0000255|HAMAP-Rule:MF_03148}	NADH pyrophosphatase activity, nucleoside triphosphate catabolic process, cytoplasm, nucleoside-triphosphate diphosphatase activity, deoxyribonucleoside triphosphate metabolic process
e9284b34	G0LXV8|LATA_LATHA Alpha-latrotoxin-Lh1a	protein kinase activity, protein phosphorylation, outer membrane-bounded periplasmic space, integral component of membrane, ATP binding
e970ec46	Q62132|PTPRR_MOUSE Receptor-type tyrosine-protein phosphatase R	protein tyrosine phosphatase activity, peptidyl-tyrosine dephosphorylation, cytoskeleton, integral component of membrane, cytoskeletal protein binding
ea739905	Q9QXV3|ING1_MOUSE Inhibitor of growth protein 1	RNA binding, integral component of membrane
fa75a48c	Q7T3C7|RT4I1_DANRE Reticulon-4-interacting protein 1 homolog, mitochondrial	mitochondrion, zinc ion binding, oxidation-reduction process, oxidoreductase activity, mushroom body development

Table S3.5: Gene trees containing homologous genes that are significantly upregulated specifically in female gonodendra, across all sampled species. Gene trees are unique gene tree identifiers (first 8 alphanumeric characters). Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
fbbf7d98	P08962|CD63_HUMAN CD63 antigen	transcription factor TFIID complex, translation initiation factor activity, translational initiation, metallopeptidase activity, zinc ion binding

Table S3.6: Gene trees containing homologous genes that are significantly upregulated specifically in pneumatophores, across all sampled species. Gene trees are unique gene tree identifiers (first 8 alphanumeric characters). Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5.

Gene tree	Most common Blast hit	Associated GO terms
0b3262b6	Q16658|FSCN1_HUMAN Fascin	protein binding, bridging, actin filament organization, cytoskeleton, actin filament binding, cytoplasm
127d770e	P29176|FOSX_MSVFR Transforming protein v-Fos/v-Fox	regulation of transcription by RNA polymerase II, DNA binding transcription factor activity, transcription factor complex, nucleus, DNA binding
18143833	P62749|HPCL1_RAT Hippocalcin-like protein 1	calcium ion binding, dendrite, negative regulation of calcium ion import across plasma membrane, regulation of 1-phosphatidylinositol 4-kinase activity, axon
290a917d	P34897|GLYM_HUMAN Serine hydroxymethyltransferase, mitochondrial	glycine hydroxymethyltransferase activity, glycine biosynthetic process from serine, cytoplasm, membrane-bounded organelle, pyridoxal phosphate binding
3a867a06	Q9D110|MTHFS_MOUSE 5-formyltetrahydrofolate cyclo-ligase	5-formyltetrahydrofolate cyclo-ligase activity, ATP binding, metal ion binding, transferase activity

Table S3.6: Gene trees containing homologous genes that are significantly upregulated specifically in pneumatophores, across all sampled species. Gene trees are unique gene tree identifiers (first 8 alphanumeric characters). Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
419216f8	Q805B4|TISDB_XENLA mRNA decay activator protein ZFP36L2-B {ECO:0000305}	metal ion binding, 3'-UTR-mediated mRNA destabilization, P-body, mRNA 3'-UTR binding, cytosol
54f43b3d	P05065|ALDOA_RAT Fructose-bisphosphate aldolase A	fructose-bisphosphate aldolase activity, glycolytic process, nuclear heterochromatin, protein binding, cytoplasm
678499ff	P51903|PGK_CHICK Phosphoglycerate kinase	phosphoglycerate kinase activity, glycolytic process, cytoplasm, ATP binding, intracellular membrane-bounded organelle
6ea146c3	Q8CGB3|UACA_MOUSE Uveal autoantigen with coiled-coil domains and ankyrin repeats	fatty-acyl-CoA binding, nucleoside metabolic process, H4/H2A histone acetyltransferase complex, protein kinase activity, protein phosphorylation
7600cdff	A5A6P1|SERA_PANTR D-3-phosphoglycerate dehydrogenase	phosphoglycerate dehydrogenase activity, L-serine biosynthetic process, myelin sheath, NAD binding, microbody
7fcca8c2	P28764|SODM_LISMO Superoxide dismutase [Mn]	superoxide dismutase activity, removal of superoxide radicals, metal ion binding, oxidation-reduction process

Table S3.6: Gene trees containing homologous genes that are significantly upregulated specifically in pneumatophores, across all sampled species. Gene trees are unique gene tree identifiers (first 8 alphanumeric characters). Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
88cfb0ee	Q9UI42|CBPA4_HUMAN Carboxypeptidase A4	metallocarboxypeptidase activity, proteolysis, extracellular space, external side of cell wall, integrin complex
957fa139	Q9H4G4|GAPR1_HUMAN Golgi-associated plant pathogenesis-related protein 1	extracellular region, sterol transport, sterol binding, fungal-type cell wall, Golgi cisterna membrane
9a89f142	P48728|GCST_HUMAN Aminomethyltransferase, mitochondrial {ECO:0000305}	aminomethyltransferase activity, glycine catabolic process, mitochondrion, transaminase activity, endoplasmic reticulum signal peptide binding
9a8d317e	P13797|PLST_HUMAN Plastin-3	actin binding, actin filament bundle, actin filament network formation, actin filament bundle assembly, actin filament
9c7c8fa3	P11586|C1TC_HUMAN C-1-tetrahydrofolate synthase, cytoplasmic	formate-tetrahydrofolate ligase activity, oxidation-reduction process, cytoplasm, methylenetetrahydrofolate dehydrogenase (NADP+) activity, 10-formyltetrahydrofolate biosynthetic process

Table S3.6: Gene trees containing homologous genes that are significantly upregulated specifically in pneumatophores, across all sampled species. Gene trees are unique gene tree identifiers (first 8 alphanumeric characters). Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
9cd4df05	P79110|TXTP_BOVIN Tricarboxylate transport protein, mitochondrial	mitochondrial citrate transmembrane transport, citrate transmembrane transporter activity, integral component of membrane, mitochondrion, organelle membrane
9d78a5b4	P29321|EPHA8_RAT Ephrin type-A receptor 8	protein kinase activity, protein phosphorylation, endoplasmic reticulum membrane, peptidyl-tyrosine modification, ATP binding
ac0c0e6a	D9IQ16|GXN_ACRMI Galaxin {ECO:0000312|EMBL:ADI50283.1}	integral component of membrane
bd053c4f	Q9WY55|GCSH_THEMA Glycine cleavage system H protein {ECO:0000255|HAMAP-Rule:MF_00272}	glycine cleavage complex, glycine decarboxylation via glycine cleavage system, CDP-diacylglycerol-glycerol-3-phosphate 3-phosphatidyltransferase activity, oxidoreductase activity, mitochondrion
cb1c0dac	Q9WUK5|INHBC_RAT Inhibin beta C chain	growth factor activity, regulation of signaling receptor activity, extracellular region

Table S3.6: Gene trees containing homologous genes that are significantly upregulated specifically in pneumatophores, across all sampled species. Gene trees are unique gene tree identifiers (first 8 alphanumeric characters). Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
eab34a97	P43090|HEM0_OPSTA 5-aminolevulinate synthase, erythroid-specific, mitochondrial	5-aminolevulinate synthase activity, mitochondrial matrix, protoporphyrinogen IX biosynthetic process, pyridoxal phosphate binding, nucleoplasm

Table S3.7: Gene trees containing homologous genes that are significantly upregulated specifically in developing nectophores and developing gastrozooids, across all sampled species. Gene trees are unique gene tree identifiers (first 8 alphanumeric characters). Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5.

Gene tree	Most common Blast hit	Associated GO terms
18d8e5db	Q9CY21|WBS22_MOUSE Probable 18S rRNA (guanine-N(7))-methyltransferase	rRNA (guanine-N7)-methylation, rRNA (guanine) methyltransferase activity, nucleolus, protein heterodimerization activity, aspartate-semialdehyde dehydrogenase activity
25e08ffd	Q9Y2G5|OFUT2_HUMAN GDP-fucose protein O-fucosyltransferase 2	fucose metabolic process, transferase activity, transferring glycosyl groups, Golgi apparatus, protein O-linked fucosylation, catalytic activity, acting on a protein
39ec6ccf	Q9UH92|MLX_HUMAN Max-like protein X	transcription factor binding, regulation of transcription, DNA-templated, nuclear membrane, protein dimerization activity, nucleoplasm
44f712fb	Q8C854|MYEF2_MOUSE Myelin expression factor 2	RNA binding
48f3d05a	Q8N983|RM43_HUMAN 39S ribosomal protein L43, mitochondrial	ribosome, structural constituent of ribosome, mitochondrial matrix, mitochondrial protein complex
4d5e7ee8	Q5ZI78|TFIP8_CHICK Tumor necrosis factor alpha-induced protein 8	regulation of apoptotic process

Table S3.7: Gene trees containing homologous genes that are significantly upregulated specifically in developing nectophores and developing gastrozooids, across all sampled species. Gene trees are unique gene tree identifiers (first 8 alphanumeric characters). Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
54862130	Q4R4S8|MED20_MACFA Mediator of RNA polymerase II transcription subunit 20	mediator complex, RNA polymerase II transcription cofactor activity, regulation of transcription by RNA polymerase II, transcription, DNA-templated
5bd308f9	Q8W257|PFI_PTIFI Polyenoic fatty acid isomerase	catalase activity, cellular oxidant detoxification, integral component of membrane, isomerase activity, phosphopantetheine binding
5ec90d33	Q5RBZ2|MEP50_PONAB Methylosome protein 50	protein kinase activity, protein phosphorylation, nuclear SCF ubiquitin ligase complex, ADP binding, kinesin complex
6ef0eac9	Q810K9|GXLT2_MOUSE Glucoside xylosyltransferase 2	transferase activity, transferring glycosyl groups, protein O-linked glycosylation, integral component of membrane, protein phosphatase 4 complex, negative regulation of Notch signaling pathway

Table S3.7: Gene trees containing homologous genes that are significantly upregulated specifically in developing nectophores and developing gastrozooids, across all sampled species. Gene trees are unique gene tree identifiers (first 8 alphanumeric characters). Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
922eeea1	O17645|HST2_CAEEL Heparan sulfate 2-O-sulfotransferase hst-2	sulfotransferase activity, protein peptidyl-prolyl isomerization, integral component of membrane, FK506 binding, heat shock protein binding
9b4c6f8c	Q2PZI1|D19L1_HUMAN Probable C-mannosyltransferase DPY19L1	protein C-linked glycosylation via tryptophan, mannosyltransferase activity, nuclear inner membrane, mannosylation, spermatid development
bbee3e8c	Q5W1J5|FOXP1_XENLA Forkhead box protein P1	sequence-specific DNA binding, nucleus, regulation of transcription, DNA-templated, DNA binding transcription factor activity, protein homodimerization activity
cdefcda2	O35972|RM23_MOUSE 39S ribosomal protein L23, mitochondrial	structural constituent of ribosome, ribosome, translation, RNA binding, mitochondrial gene expression
f4d33655	Q04892|SOX14_MOUSE Transcription factor SOX-14	nucleus, DNA binding, regulation of transcription, DNA-templated, DNA binding transcription factor activity, integral component of membrane

Table S3.7: Gene trees containing homologous genes that are significantly upregulated specifically in developing nectophores and developing gastrozooids, across all sampled species. Gene trees are unique gene tree identifiers (first 8 alphanumeric characters). Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
f866d8f6	O75844|FACE1_HUMAN CAAX prenyl protease 1 homolog	CAAX-box protein processing, metalloendopeptidase activity, integral component of membrane, intrinsic component of endoplasmic reticulum membrane, prenylated protein catabolic process
fac0c29e	P49337|WNT4_CHICK Protein Wnt-4	Wnt signaling pathway, signaling receptor binding, extracellular region, cytoplasm, multicellular organism development
fc6946ce	P30352|SRSF2_CHICK Serine/arginine-rich splicing factor 2	RNA binding, regulation of RNA splicing, nuclear speck, chromatin, phospholipase A2 activity

Table S3.8: Gene trees containing homologous genes that are significantly upregulated in developing *Nanomia bijuga* palpons, developing *Frillagalma vityazi* bracts, and developing gastrozooids and nectophores across all sampled species. Gene trees are unique gene tree identifiers (first 8 alphanumeric characters). Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5.

Gene tree	Most common Blast hit	Associated GO terms
0af70276	Q9JI85|NUCB2_RAT Nucleobindin-2	calcium ion binding
11bcbf3d	P49010|HEXC_BOMMO Chitooligosaccharidolytic beta-N-acetylglucosaminidase	beta-N-acetylhexosaminidase activity, carbohydrate metabolic process, plasma membrane, rhodopsin biosynthetic process, N-acetyl-beta-D-galactosaminidase activity
38d40121	O54068|UDG_RHIME UDP-glucose 6-dehydrogenase {ECO:0000305}	UDP-glucose 6-dehydrogenase activity, polysaccharide biosynthetic process, cytoplasm, NAD binding, oxidation-reduction process
3e82172e	P56839|PEPM_MYTED Phosphoenolpyruvate phosphomutase	phosphoenolpyruvate mutase activity, organic phosphonate biosynthetic process, integral component of membrane, phosphonopyruvate hydrolase activity, lyase activity
3f9aaaea	Q10572|FOX1_CAEEL Sex determination protein fox-1	regulation of RNA splicing, RNA binding, nucleus, RNA splicing, mRNA metabolic process
413f8455	A3KCL7|OXDD_PIG D-aspartate oxidase	D-amino-acid oxidase activity, D-amino acid metabolic process, peroxisome, FAD binding, integral component of membrane

Table S3.8: Gene trees containing homologous genes that are significantly upregulated in developing *Nanomia bijuga* palpons, developing *Frillagalma vityazi* bracts, and developing gastrozooids and nectophores across all sampled species. Gene trees are unique gene tree identifiers (first 8 alphanumeric characters). Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
4242981d	P55112|NAS4_CAEEL Zinc metalloproteinase nas-4	metalloendopeptidase activity, proteolysis, extracellular region, zinc ion binding, meprin A complex
4566e958	Q20191|NAS13_CAEEL Zinc metalloproteinase nas-13	metalloendopeptidase activity, proteolysis, extracellular region, integral component of membrane, molting cycle, collagen and cuticulin-based cuticle
57048fad	P21251|CALM_STIJA Calmodulin	calcium ion binding, cytoskeleton, phosphorylation, extracellular exosome, oxidation-reduction process
5766c0ff	P27607|PGH2_CHICK Prostaglandin G/H synthase 2	prostaglandin-endoperoxide synthase activity, cyclooxygenase pathway, neuron projection, peroxidase activity, endoplasmic reticulum
58654818	P56101|DNJC5_TETCF DnaJ homolog subfamily C member 5 {ECO:0000250|UniProtKB:Q9H3Z4}	synaptic vesicle, negative regulation of neuron apoptotic process, unfolded protein binding, heat shock protein binding, zinc ion binding

149

Table S3.8: Gene trees containing homologous genes that are significantly upregulated in developing *Nanomia bijuga* palpons, developing *Frillagalma vityazi* bracts, and developing gastrozooids and nectophores across all sampled species. Gene trees are unique gene tree identifiers (first 8 alphanumeric characters). Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
7811af98	Q6P6S2|S39AB_RAT Zinc transporter ZIP11	metal ion transmembrane transporter activity, metal ion transport, integral component of membrane, transmembrane transport, ribosome
9e47f304	P12256|PAC_LYSSH Penicillin acylase	hydrolase activity
a430b949	Q9JJ09|NPT2B_RAT Sodium-dependent phosphate transport protein 2B	sodium-dependent phosphate transmembrane transporter activity, sodium-dependent phosphate transport, plasma membrane, apical part of cell, brush border
a58ddc99	B3EWZ5|MLRP1_ACRMI MAM and LDL-receptor class A domain-containing protein 1	serine-type endopeptidase activity, proteolysis, extracellular region, chorion, integral component of membrane
b1242116	Q5AF03|HSP31_CANAL Glyoxalase 3 {ECO:0000303|PubMed:24302734}	glutamine metabolic process, peptidase activity, spindle pole, transferase activity, microtubule associated complex
f27a4758	A6GYX5|DAPF_FLAPJ Diaminopimelate epimerase {ECO:0000255|HAMAP-Rule:MF_00197}	diaminopimelate epimerase activity, diaminopimelate metabolic process, cytoplasm, lysine biosynthetic process

Table S3.8: Gene trees containing homologous genes that are significantly upregulated in developing *Nanomia bijuga* palpons, developing *Frillagalma vityazi* bracts, and developing gastrozooids and nectophores across all sampled species. Gene trees are unique gene tree identifiers (first 8 alphanumeric characters). Blast hit is the most frequent blast hit for the gene tree. GO terms are limited to the first 5. *(continued)*

Gene tree	Most common Blast hit	Associated GO terms
f4de86f6	O70273|EHF_MOUSE ETS homologous factor	epithelial cell proliferation, sequence-specific DNA binding, nucleus, epithelial cell differentiation, DNA binding transcription factor activity
fbbf7d98	P08962|CD63_HUMAN CD63 antigen	transcription factor TFIID complex, translation initiation factor activity, translational initiation, metallopeptidase activity, zinc ion binding

Chapter 4

Evolution of gene expression across species and tissues in Siphonophora

Catriona Munro[1]‡, Felipe Zapata[2], Mark Howison[3], Steven H.D. Haddock[4], Casey W. Dunn[5]

[1] Department of Ecology and Evolutionary Biology, Brown University, Providence, RI 02912, USA

[2] Department of Ecology and Evolutionary Biology, University of California Los Angeles, Los Angeles, CA 90095, USA

[3] Watson Institute for International and Public Affairs, Brown University, Providence, RI 02912, USA

[4] Monterey Bay Aquarium Research Institute, Moss Landing, CA 95039, USA

[5] Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520, USA

4.1 Abstract

Differences in gene expression are thought to be an important component of phenotypic diversity within and among species, and one central question in comparative functional genomics is how gene expression variance partitions among different tissues and organs across species. In this study, we apply phylogenetic comparative methods to investigate evolutionary changes in gene expression of siphonophore tissues across species. Expression values are mapped to gene phylogenies, considering the complex evolutionary histories of homologous genes and enabling a comparison of genes that arose via a speciation or duplication event. We also introduce methods to isolate branches in gene trees that correspond to a branch in the species tree, enabling an independent comparison of the same branch across gene trees. The global distribution of change across branches in gene trees that correspond to a branch in a species tree can be explained by the interaction between shared changes on species-tree branches and treatments (zooids and the pneumatophore). Changes across a given branch within a gene tree are highly correlated among treatments, with male and female gonodendra showing the greatest divergence in evolutionary change across a branch. There are more lineage and treatment specific changes in expression in mature palpons along branches leading to *Agalma elegans* and *Nanomia bijuga.* Finally, we use these data to test hypotheses about the ortholog conjecture. We fail to reject the hypothesis that there is no difference in the evolution of expression following speciation and duplication events in all but one treatment. Palpons are the exception, and we observe greater evolutionary change in expression across branches following duplication events as compared to following speciation events in this zooid type. These findings suggest that palpons have greater functional diversity than was previously thought.

4.2 Introduction

Gene expression is an important component of phenotypic diversity within and among units of biological organization, including cells, tissues, organs, and species (Brawand et al., 2011; Gilad et al., 2005; King and Wilson, 1975; Rifkin et al., 2003). There are a number of ways in which gene expression can change, including in the relative magnitude of expression, or temporal or spatial shifts in expression (Carroll, 2005;

Romero et al., 2012; Wray et al., 2003). With high throughput RNA-seq data, we are able to quantify the relative expression level of a large number of genes within and among species and homologous organs. One category of questions in comparative genomics is whether, in closely related species, most genes have low expression variance among tissues and high expression variance among species, or high expression variance among tissues and low expression variance among species. Based primarily on qualitative distance measures and principal component analysis (PCA) of expression across 1:1 orthologs, most studies suggest that gene expression is more conserved among homologous organs of the same type across different species than among different organs within the same species; suggesting that molecular, cellular and developmental pathways are highly conserved among species (Brawand et al., 2011; Breschi et al., 2016; Clarke et al., 2017; Gilad and Mizrahi-Man, 2015; Khaitovich et al., 2004; Merkin et al., 2012). While others found the opposite result (Lin et al., 2014; Pankey et al., 2014; Tschopp et al., 2014; Yang and Wang, 2013), and suggest the majority of genes vary little among organs within a species but have diverged significantly between species (although see Gilad and Mizrahi-Man, 2015; Gu, 2015; Sudmant et al., 2015). However, these global patterns of gene expression fail to account for gene-specific patterns, which may vary in organ and species specificity, and clustering patterns may be driven by the behavior of a small subset of genes (Breschi et al., 2016). Genes that show high variance across species and low variance across organs are frequently ubiquitously expressed genes (Breschi et al., 2016). These questions seek to understand differences in expression levels in orthologous genes among organs and species, but do not explicitly test this in a phylogenetic context. In this study, we use phylogenetic comparative methods (PCMs) to compare gene expression patterns between different siphonophore zooids, and one specialized tissue, and map expression values directly onto gene trees, in order to investigate evolutionary changes gene expression of different tissues across branches in gene trees.

Only a handful of studies have applied PCMs to compare gene expression values across species (Clarke et al., 2017; Chang and Duda, 2014; Chen et al., 2017; Eng et al., 2009; Gu, 2004; Oakley et al., 2005; Rohlfs and Nielsen, 2015; Stern and Crandall, 2018; Whitney et al., 2011). PCMs were developed to address challenges in comparing trait data across species, particularly the non-independence of traits due to the evolutionary history of the species (Dunn et al., 2013b; Felsenstein, 1985, 2008; FitzJohn, 2012; Grafen, 1989; Revell et al., 2012; Pagel, 1999; Uyeda and Harmon, 2014). Multiple pairwise comparisons are frequently used in

comparative genomic work, however re-analysis of the same data using PCMs has been shown to support very different conclusions; underscoring the importance of considering species relationships when making statements about evolutionary processes (Dunn et al., 2018).

Analyses of gene expression that do take phylogeny into account typically focus on strict orthologs and map expression values onto the species tree (Chen et al., 2017; Oakley:2005ky; Rohlfs and Nielsen, 2015; Stern and Crandall, 2018), while other analyses use the expression values themselves to build neighbor-joining trees (Brawand et al., 2011; Stern and Crandall, 2018). However, in addition to understanding species relationships, understanding the homology of genes is critical for understanding the evolution of gene expression. Gene phylogenies are often used to study species relationships and the history of gene duplication and loss, but they are more useful than that – they can also be used to study the evolution of gene traits such as expression. Gene phylogenies are hypotheses about the evolutionary relationship of gene sequences to one another, and gene topology reflects a number of evolutionary events, including speciation events, gene duplication and loss, and molecular evolution. The terms 'ortholog' and 'paralog' were proposed as a means to distinguish the relationships between genes, and are defined by their evolutionary history – orthologous genes are obtained by speciation events, where the common ancestor of orthologous genes is a single gene in the common ancestor of the species being considered, while paralogous genes are the product of duplication events, where the common ancestor of paralogous genes is a single gene in a genome (Fitch, 1970, 2000). However, gene phylogenies often show a complicated history of gene relationships, and the term 'paralog' may apply to a duplication event that occurred within a single species, or a duplication event that occurred in an ancestral genome and is shared among its descendants (Sonnhammer and Koonin, 2002). Sonnhammer and Koonin (2002) suggested the terms 'in-paralogs', 'out-paralogs' and 'co-orthologs' to address these different histories, but these terms also have limitations (see Dunn and Munro, 2016).

Many analyses of comparative expression avoid these complicated evolutionary histories by focusing exclusively on strict 1:1 orthologs, that is, gene families that show no evidence of duplication events. The common ancestor of these genes is assumed to be the gene found in the genome of the common ancestor of the species that are considered in the study. There are a number of reasons for a focus on strict 1:1 orthologs (Dunn and Munro, 2016), some of these reasons are practical – it greatly simplifies analyses when complex gene

histories are not considered; but it also reflects implicit or explicit assumptions about orthologs and paralogs, for example, that strict orthologs are not impacted by neofunctionalization or subfunctionalization events that are hypothesized to be more frequent following duplication (Force et al., 1999; Hughes, 1994; Lynch and Force, 2000; Ohno, 1970), or they have more conserved function relative to paralogs (this hypothesis is termed the ortholog conjecture) (Gabaldon and Koonin, 2013; Nehrt et al., 2011). A number of studies have tested the predictions of the ortholog conjecture, with some finding weak support for conserved function in orthologs, and others finding no support (Altenhoff et al., 2012; Chen and Zhang, 2012; Dunn et al., 2018; Kryuchkova-Mostacci and Robinson-Rechavi, 2016; Nehrt et al., 2011; Yanai et al., 2004b). A recent comparative phylogenetic analysis of gene expression patterns in mammalian organs suggests that phylogenetic distance is a better predictor of gene expression similarity than whether the evolutionary history of the gene reflects a duplication or speciation event (Dunn et al., 2018). In addition, strict orthology is also not necessarily an indicator of simpler evolutionary histories, and may in fact represent a higher duplicate loss rate (Dunn and Munro, 2016; De Smet et al., 2013). By focusing on strict orthologs, there is a risk of introducing ascertainment biases by only focusing on single copy genes, and in addition, we throw away important expression data by avoiding complex gene histories.

In this study, we map gene expression of different siphonophore zooids, and one specialized tissue (the pneumatophore, a gas filled float), onto gene phylogenies. To simplify terms, we will refer to the collective of zooids (asexually produced bodies) and the pneumatophore (a specialized tissue) as treatments, following terminology that is often used in differential expression analyses (Love et al., 2014). Our approach differs significantly from previous approaches – we apply phylogenetic comparative methods and suggest an approach to assessing gene expression patterns that takes the complex evolutionary history of genes into account, as well as the evolutionary history of species. Instead of discussing gene relationships in terms of 'orthologs' and 'paralogs', particularly 1:1 orthologs, we will refer to speciation and duplication events, selecting branches based on their evolutionary history and relationship to the species tree. We investigate the evolution of gene expression among zooids across species, and also test hypotheses about the evolution of gene expression across siphonophore species.

Siphonophores are hydrozoans (Cnidaria), with unique morphology and development in comparison to other

hydrozoans. Like many other hydrozoans, they are colonial and have asexually produced zooids (bodies) that are are attached to a stem or stolon and have a shared gastrovascular cavity. Most siphonophores are pelagic (with the exception of a clade of secondarily benthic siphonophores, and a single pleustonic species), and unlike other hydrozoans, the whole colony is detached from the seafloor and moves through the use of propulsive zooids or contractions of the stem. Other hydrozoans have functional specialization, but siphonophores have more functionally specialized zooids than any other animal, with zooids that are specialized for tasks such as swimming (nectophore), feeding (gastrozooid), reproducing (gonodendra, male and female), and defending/excreting (palpons) (Mackie et al., 1987; Totton, 1965). The high level of functional specialization and interdependence of zooids within a siphonophore colony led Mackie (1963) to call siphonophores 'super-organisms', and suggested that multicellular zooids, that are homologous to solitary free living individuals, are analogous to organs in other metazoan clades.

4.3   Material and methods

All scripts for the analyses are available in a git repository at https://github.com/dunnlab/siphonophore_ compexpression. The most recent commit at the time of the analysis presented here was b9daae9d.

4.3.1   Analysis

Differential gene expression libraries (50bp short read) were obtained from 7 siphonophore species (*Nanomia bijuga*, *Agalma elegans*, *Frillagalma vityazi*, *Diphyes dispar*, *Bargmannia elongata*, *Apolemia lanosa*, *Physalia physalis*), and from 5 zooids and one specialized tissue (collectively, we will refer to these zooids/tissue as treatments) for which there are libraries for at least two biological replicates in two species (pneumatophore, developing gastrozooid, mature gastrozooid, developing nectophore, mature palpon, male gonodendron, female gonodendron). Collection, extraction, and sequencing methods are outlined in chapter 3.

The differential gene expression (DGE) libraries were mapped to previously published reference transcriptomes (see chapter 2) (Munro et al., 2018) using Agalma v 2.0.0 (Dunn et al., 2013a; Guang et al., 2017). The expression pipeline of Agalma uses a number of existing tools (Langmead et al., 2009; Li and Dewey, 2011).
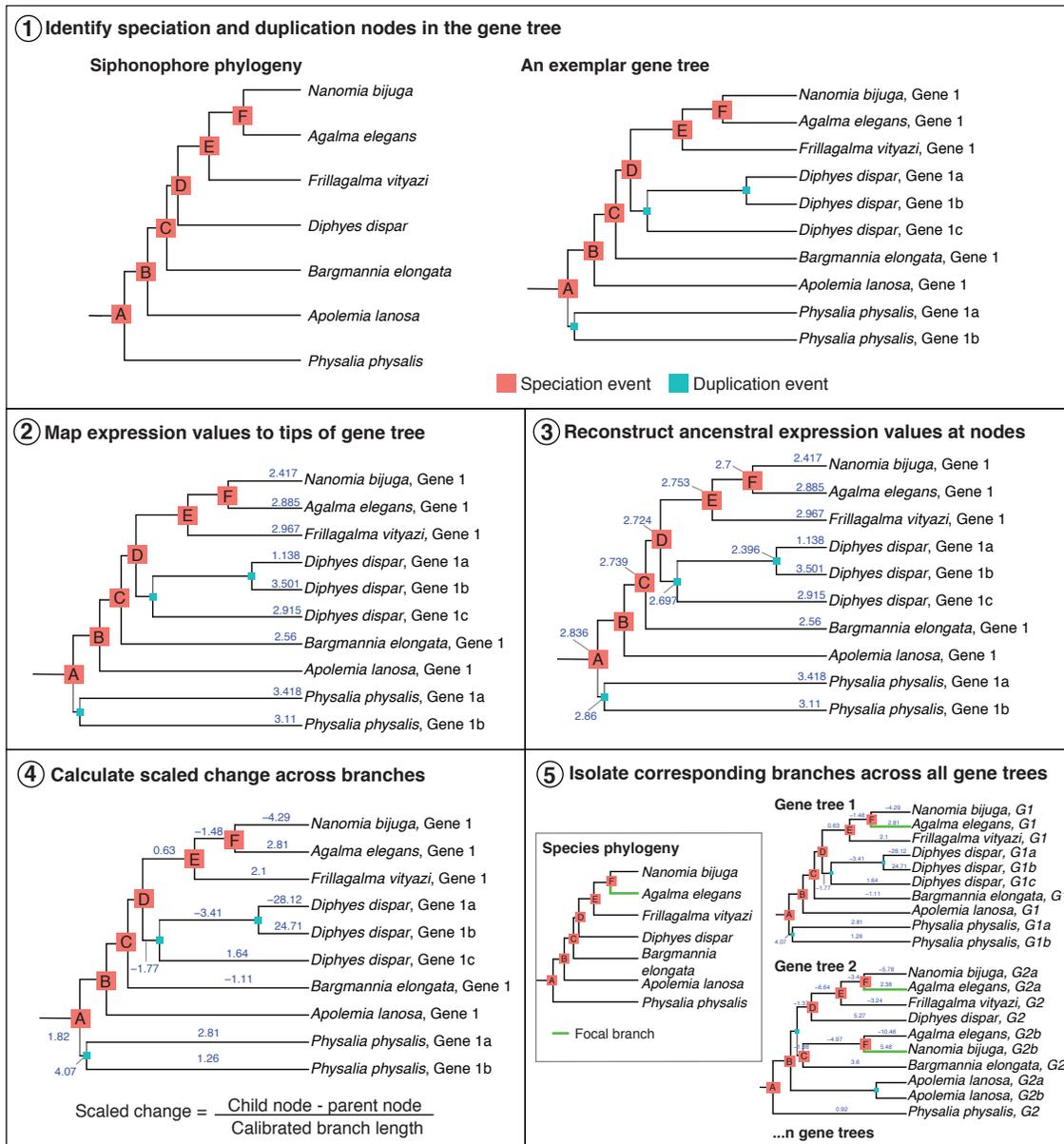
Figure 4.1: Methods used to identify changes in expression across branches in gene trees. In step 1, we have labelled each of the nodes in the siphonophore phylogeny, and identified equivalent speciation nodes across every gene tree (an exemplar is shown here). Step 2, we map expression values to the tips (TPM10K). Step 3, reconstruct ancestral trait expression values at all nodes where expression data is available. Step 4, we calculate scaled change in expression across branches and divide by the branch length. Branch length is calibrated to the species tree branch lengths. Step 5, we identify branches in gene trees that correspond to equivalent branches in the species tree. There may be more than one branch in a gene tree that corresponds to the same branch in the species tree.

Gene alignments were generated from the reference transcriptomes of 41 species (Munro et al., 2018) using Agalma, and subsequently PHYLDOG v.2.0 (Boussau et al., 2013) was used for simultaneous co-estimation of gene trees with the published ML species tree (Munro et al., 2018) enabling annotation of speciation and duplication events at the nodes of each gene tree (Fig. 4.1, step 1). Phylogenetic analyses were also conducted in R using `geiger`, `ape`, `phytools`, `Rphylopars`, and `hutan` (Paradis et al., 2004; Church et al., 2015a; Goolsby et al., 2017; Harmon et al., 2007; Revell, 2012). Phylogenetic trees were visualized in R using `ggtree` and `treeio` (Yu et al., 2017). Linear models were constructed using lm(), and wilcox tests were carried out using the function wilcox.test() in base R. See Supplementary Information for R package version numbers.

Gene trees were filtered to exclude trees with a length threshold >2, a root depth >5, and that had more than 0.25 branches with a default length value assigned by phyldog (that are indicative of branch length=0). Tips without expression values for any of the treatments were pruned out of the tree. Gene trees with fewer than three expression values at the tips were discarded, retaining only trees with three or more values. Additionally, only trees with one or more speciation events were retained, as speciation events are used for time calibrations. The gene trees were then time calibrated to the species tree using chronos() in the `ape` package, so that the branch lengths were scaled to the same equivalent length across all gene trees (Paradis et al., 2004). Some gene trees could not be calibrated against the node constraints from the species tree and were discarded.

Using the `agalmar` package (https://github.com/caseywdunn/agalmar), we filtered out genes that were flagged as being rRNA, and selected only protein coding genes. We also only considered genes that were greater than 0 in at least two libraries. After filtering, expression values were normalized using a method we are calling transcripts per million 10K (TPM10K). For gene i of a given species, TPM is typically calculated as (Li et al., 2009a):

$$TPM_i = \frac{10^6 \times \theta_i}{\ell_i \times \sum_{i=1}^{n} \frac{\theta_j}{\ell_j}}$$

Where θ_i is the number of the mapped reads to gene i, ℓ_i is the effective length of the gene, and n is the number of genes in the reference. The intent of this measure is to make libraries comparable within a single species. The sum of TPM values within a library is 10^6, and the mean is $\frac{10^6}{n}$. One implication of this is that TPM values are not directly comparable across species, since in practice n differs across species. If this were not accounted for, then it could appear, for example, that genes all have lower expression in a species with a more complete reference transcriptome and higher n. To account for differences in means among species, we use a new measure, TPM10K, that accounts for differences in n:

$$TPM10K_i = \frac{TPM_i \times n}{10^4}$$

Where the sum of TPM10K values within a library is $10^2 \times n$ and the mean is 10^2. By multiplying by n we are able to account for different sequencing depths among species, and ensure a common mean. As n is large, we divide by an arbitrary number (in this case 10^4) in order to reduce the magnitude of the expression value.

We then took the mean TPM10K value for each gene across replicates of the same treatment within a species and applied a log transformation. Using gene trees with expression values for each gene within a species at the tips, maximum likelihood ancestral trait values were generated at the nodes using the anc.recon() function in `Rphylopars` assuming a Brownian model of evolution (Fig. 4.1, step 2 & 3) (Goolsby et al., 2017). As not all zooids are present in all of the species, the trees were pruned down to the subset of tips with expression values for ancestral trait reconstructions. Node values were then added back to the unpruned tree with all of the reconstructed expression values. Change in expression was measured across a branch by taking the difference between a parent node and a child node, and then this difference is scaled by branch length (Fig. 4.1, step 4).

4.4 Results and Discussion

4.4.1 Overview of expression change

The 6767 gene trees from our previous phylogenetic analyses of Siphonophora were the starting point for the trees considered here (chapter 2) (Munro et al., 2018). The gene trees containing tips for 41 species were pruned down to include only the focal 7 species for which expression is considered here. Gene trees were then filtered to only consider those with expression data for at least three tips, and those that could be successfully time calibrated to the species tree, resulting in 3826 gene trees. There are 27871 tips with expression data across these gene trees. The internal nodes on these gene trees consist of 15949 speciation events and 8096 duplication events.

We first considered the evolutionary change in expression along branches in the gene tree that correspond to branches in the species tree. These are identified as gene tree branches that have parent and child nodes that are both speciation events and that correspond to speciation events and branches in the species tree (Fig. 4.1, step 5). This method enables the selection of specific branches within gene trees that are equivalent to branches within the species tree, and are thus comparable with one another across all gene trees. Unlike a strict 1:1 ortholog approach, this approach considers equivalent branches that are descended from speciation events, but that have more complex evolutionary histories. For example, due to deeper gene duplication events, gene trees often contain multiple branches that correspond to the same branch in the species tree (Fig. 4.1, step 5). Our method allows us to consider all of these branches. Strict ortholog methods would discard some or all data impacted in this way by duplication events. Each speciation node in a gene tree is assigned an identifier that corresponds to the equivalent node in the species phylogeny. Each branch in the species tree is given a unique letter, as shown in figure 4.2, and the corresponding branches in gene trees are given the same letter. A simulated (null) data set of random expression values was also obtained by using a Brownian motion model with empirically derived mean and standard-deviation values for each gene tree.

The distribution of expression changes along branches in gene trees that correspond to equivalent branches in the species tree are shown in figure 4.2. The variance structure of the empirical data matches that of the simulated data (Fig. S4.1), suggesting that the size of standard deviation of change around the mean that

162
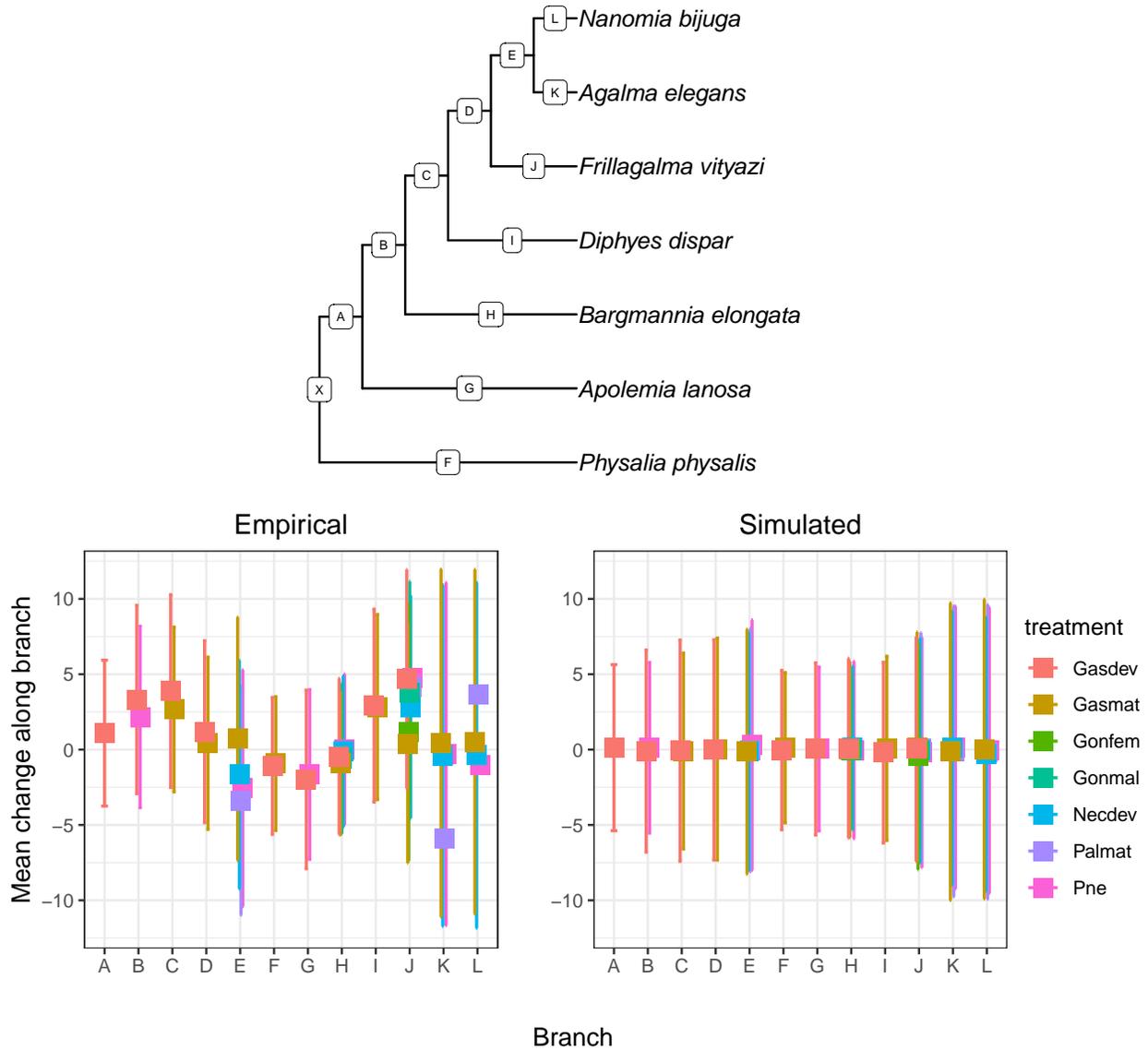
Figure 4.2: Mean changes along branches across all gene trees that correspond to branches in the species tree, error bar is one standard deviation. Top panel: species phylogeny with branch IDs given as numbers. Lower panel: Distribution of changes along a branch in a gene tree, showing mean change in the empirical dataset and in the BM simulation. Branch refers to the branch ID in the species tree. Treatment type is coded in colour, Gasdev = Developing Gastrozooid, Gasmat= Mature gastrozooid, Gonfem= Female gonodendron, Gonmal= Male gonodendron, Necdev= Developing nectophore, Palmat = Mature palpon, Pne= pneumatophore.

is observed in figure 4.2 (lengths of the bars) is a function of the underlying structure of the tree and the ancestral trait reconstruction methods that were used. High variance is especially pronounced in branches in the gene tree that correspond to K and L that have a shared parent node and lead to sister taxa *Nanomia bijuga* and *Agalma elegans* (Fig. S4.1). These two taxa are the only sister tip values that were sampled in this study. The mean values of change along branches do, however, differ from the null expectation (Fig. 4.2). Using linear models we find that expression change along branches in gene trees that correspond to branches in the species tree can be explained by significant differences among treatments, branches, and the interaction of the two ($p < 2.318445e-31$, $p < 6.164e-182$, $p=0$, respectively, two-tailed anova). The null models, based on simulated values on the gene trees under BM ($p=0.7298$, $p=0.9716$, $p=0.7232$), and also based on random reshuffling of changes on branches ($p=0.7543$, $p=0.2172$, $p=0.7081$), show mean changes of zero across all tissues and branches.

Among these global patterns of change across branches in gene trees that correspond to specific branches in species trees, differences in mean change vary among treatments and branches. In developing gastrozooids, mean change across branches B, C, J and I are much higher than in the simulated data suggesting that a subset of branches within gene trees show large lineage-specific increases in expression in this zooid at this stage of development relative to others within the same gene tree (Fig. 4.2). Small but positive shifts in the mean are also seen along branches A and D. The nature of these calculations means that an equal and opposite trend can be seen along sister branches. A negative shift in mean change across branches in gene trees are seen along branches F and G. For mature gastrozooids, a large positive shift in mean change is seen along branches C and I, and there is a negative shift in mean change across branch F. Small deviations from zero are seen across all other branches. Mature gastrozooids and developing gastrozooids show similar patterns in the distribution and mean change across the same branch, except across branch J. Notably, *Frillagalma vitazi* is the only species sampled here that has more than hypothesized one gastrozooid type (see chapter 3) (Dunn, 2005). Pneumatophores show a large positive shift in mean change across branches B and J, and show a negative shift in mean change across G, E and L. Moderate changes are seen in the mean change across branches in developing nectophores: with increases across J and an equal and opposite decrease across J. The most dramatic shifts in mean change can be seen in mature palpons – a large number

of positive changes in expression are observed along branch L and J, while a large number of negative changes are observed along branches K and E. Taken together, these global patterns of change across branches in gene trees (that correspond to species tree branches) suggest that a subset of gene tree branches may be driving large lineage specific changes in expression in particular treatments.

Due to unequal sampling, female gonodendra were collected for *Agalma elegans* and *Frillagalma vityazi*, and male gonodendra were collected for *Nanomia bijuga*, *Frillagalma vityazi* and *Bargmannia elongata*. Due to the structure of the sampling across branches, the changes in gonodendron expression across branches L and K cannot be observed. However, these values can be observed along branches J and H, for male gonodendra, and just branch J for female gonodendra. These results indicate large positive shifts relative to 0 in mean change across branch J for male gonodendra, and a slight negative change across branch H. A small positive shift in mean change is observed across branch J for female gonodendra.

4.4.2 Expression change across specific branches within gene trees

We were then interested in whether changes across particular gene tree branches are treatment specific, or shared across treatments. We investigated the co-variance structure of changes across particular branches (that correspond to species tree branches) within gene trees among different treatments (Figs. 4.3, S4.2). For change across the same branch within a gene tree that corresponds to a particular branch in the species tree, there is a positive correlation between pairwise comparisons of treatments, however the strength of these correlations vary among treatments (Figs. 4.3, S4.2. For branch J, for example, which is the branch with the most sampled treatments, there are differences in the strength of pairwise correlations of changes across the same branch in the same gene tree across different treatments (Fig. 4.3). Notably, male and female gonodendra have the weakest correlation with one another (r=0.5). This matches patterns observed in amniotes, where mean gene expression divergence relative to all other organs is greatest in testis, potentially due to sexual-selection, as well as being a site of widespread genomic transcription in spermatocytes and spermatids, potentially as a result of chromatin remodeling (Brawand et al., 2011; Kryuchkova-Mostacci and Robinson-Rechavi, 2016; Necsulea et al., 2014; Melé et al., 2015; Soumillon et al., 2013). For branches that correspond to branch J in the species tree, male gonodendra and developing gastrozooids are also
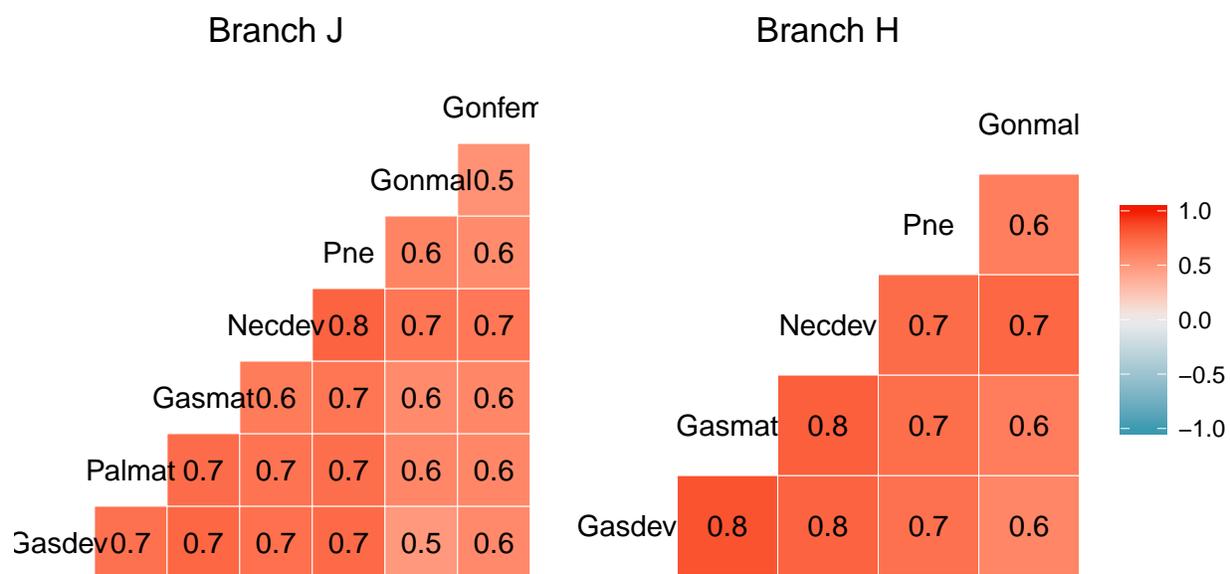
Figure 4.3: Covariance of changes across the same branch across all gene trees, where the branch corresponds to branch J or branch H in the species tree. Gasdev - developing gastrozooid. Gasmat - mature gastrozooid. Palmat - mature palpon. Necdev - developing nectophore. Gonmal - male gonodendron. Gonfem - female gonodendra.

weakly correlated with regards to change across the same branch in a gene tree. Both male and female gonodendra are more correlated with developing nectophores than they are with any other treatment. This is notable, because gonophores (reproductive zooids within gonodendra) are hypothesized to be derived medusae, while nectophores are highly specialized medusae (Totton, 1965). Similar patterns are also seen among treatments in branch H, although female gonodendra were not sampled here, and male gonodendra are more correlated with developing gastrozooids across this branch. Unfortunately, due to the limited sampling of male and female gonodendra, our power to investigate general sex-specific patterns of expression across multiple branches is limited.

In sum, changes across a given branch within a gene tree are highly correlated among treatments, with male and female gonodendra showing the greatest divergence in evolutionary change across branches J and H. Most evolutionary changes across a branch in a gene tree are highly consistent among treatments. However, as we know from the global patterns, there are a subset of branches (that correspond to specific branches in the species tree) that show large changes in a treatment specific manner. Next, we consider instances of treatment *and* branch specific change.

By filtering on large positive (> 1) changes across branches in particular treatments, while also selecting for neutral negative changes across the same branch for other treatments, we are able to identify genes within gene trees that show lineage specific positive shifts in one treatment that is coincident with lineage specific neutral or negative shifts in all other treatments (Fig. 4.4 Top). Likewise, we are also able to identify lineage specific negative shifts ($< \text{-}1$) in one treatment that are coincident with lineage specific neutral or positive shifts in all other treatments (Fig. 4.4 Bottom). The number of observed changes were scaled by the number of sampled treatments for that branch, as branches with greater sampling (J) would appear to have fewer lineage and treatment specific changes.

Across gene trees, a large number of branches have positive changes in mature palpons along the branch that corresponds to branch L in the species tree. These changes were found in gene trees that are enriched for GO terms for organ induction/lung morphogenesis, double strand break repair and mRNA-containing ribonucleoprotein complex export from nucleus. In total, 208 branches in 207 gene trees are found to have this pattern. The enrichment for organ induction/lung morphogenesis is likely due to identified changes
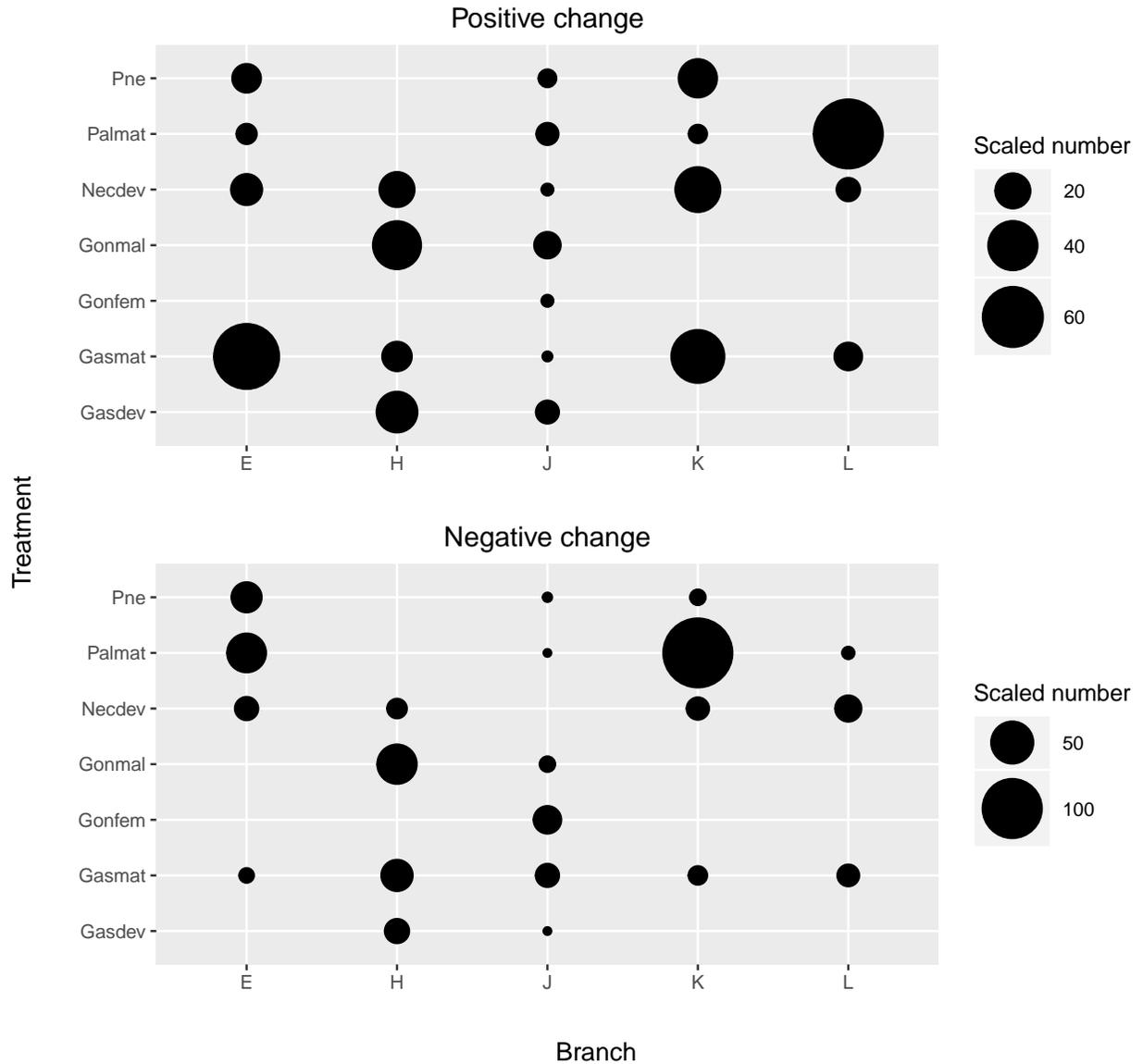
Figure 4.4: Number of changes across a branch that are specific to a particular treatment, where number indicates the number of branches in gene trees (that correspond to a specific species tree branch) with a particular change, scaled by the number of sampled treatments per branch. Top: Change is positive in treatment where it is negative in all others. Bottom: Change is negative in treatment where it is neutral or positive in all others.

across a branch within gene trees of Wnt 3, Frizzled-8, Forkhead box (Fox) P1, which are involved in lung patterning/morphogenesis in Bilateria, but are also implicated in anterior/posterior patterning in Cnidaria (Magie et al., 2005). While it might be expected, given the structure of the phylogeny, that many branch K branches with negative change specific to mature palpons would be found in the same gene trees as those identified as having positive change across branch L branches – only 68 out of 340 gene trees are common between those that are specific and positive across L and those that are specific and negative across K. The genes that have decreased expression across branch K in palpons relative to all other treatments were enriched for GO terms such as circulatory system development and mechanoreceptor differentiation. The reasons for the large number of lineage and palpon specific changes across branches K and L is not clear. *Agalma elegans* has palpons in multiple locations along the stem, near the gastrozooid (gastric palpons), associated with the female gonodendra (female-associated palpon) and the B-palpon, while *Nanomia bijuga* has one "type" of palpon, and male and female gonodendra appear at the base of this palpon (Dunn and Wagner, 2006). Additionally, a large number of gene tree branches that correspond to branch E have positive changes across these branches that are specific to mature gastrozooids. This indicates an increase in expression specifically in gastrozooids across the branch that leads to the node that represents the common ancestor of *Nanomia bijuga* and *Agalma elegans*. These branches were found in gene trees that are enriched for a number of GO terms, including macromolecule modification and chromatin organization. It is also not clear what the reasons are for the large number of lineage specific changes in gastrozooids leading to the Agalmatidae.

It is important to note that a change in expression patterns that is unique to one treatment does not necessarily mean that expression is becoming more or less treatment-specific. Treatment-specific expression is typically a statement about differences in the levels of expression among treatments at the tips or nodes of the gene trees. For example, treatment-specific expression is where expression of a gene is high in one treatment and low in all other treatments. Change across a branch looks at differences in expression values of only one treatment between parent and child nodes (some node values may be terminal nodes, i.e. tips), and each of the treatments are considered independently of one another (Fig. 4.1). Large positive or negative changes in expression across a branch represent lineage specific changes within a particular treatment. Here, we do not consider differences in levels of expression *among* species, and as such, do not discuss treatment-

specific expression, but instead consider treatment-specific patterns of expression change – that is, changes in expression across a specific branch or branches in gene trees that are unique to a particular treatment .

4.4.3 Testing the ortholog conjecture

The ortholog conjecture posits that orthologs have more conserved function than paralogs (Nehrt et al., 2011). Function is a broad term, but is defined here as change in gene expression across a branch of a gene tree. Our null hypothesis is that there is no difference in the evolution of expression between speciation events and duplication events, and that therefore the changes along branches resulting from speciation events and duplication events are drawn from a similar distribution. By contrast, the alternative hypothesis is that there is a higher rate of absolute change after duplication events as compared with speciation events. In this case, the distribution of absolute change in expression across the branch would be expected to be higher in branches resulting from duplication events. These predictions are the same as those outlined by Dunn et al. (2018), but differ slightly in their implementation – Dunn et al. (2018) were considering changes in phylogenetic independent contrasts of the summary statistic tau. Here, we are considering change in expression, within particular treatments, across branches that descend from either speciation or duplication events. We discarded the top and bottom deciles of branch lengths in both the empirical and simulated data set, to ensure that these findings are not skewed by very short or very long branch lengths. For expression within developing gastrozooids, mature gastrozooids, pneumatophore, male gonodendra, and female gonodendra, we do not find increased evolutionary change in expression following duplication events compared to speciation events in both the empirical and simulated (null) data sets (wilcoxon test failed to reject the null, p value = 1), as well as in the developing nectophore (empirical p=0.013, simulated p=1). However, for expression values in the mature palpons ($p < 4.1763e-20$), we did find support for the ortholog conjecture; and failed to reject the null hypothesis in the simulated data (p=0.074). This finding was not sensitive to the exclusion of branches originating from deep duplication events that preceded the evolution of the siphonophores (mature palpons, $p<2.930203e-17$). The difference was seen most clearly in expression changes among mature palpons on branches leading to *Nanomia bijuga* (p = 0.00083). These patterns still hold when branches originating from deep duplication events are excluded.

The basis of the ortholog conjecture is the idea that gene duplication may be responsible for generating evolutionary novelty, and therefore the function of orthologs is more conserved than paralogs (Nehrt et al., 2011). A number of hypotheses have been put forward to explain the maintenance of gene duplicates after duplication events, including neofunctionalization (evolution of novel function in one duplicate), subfunctionalization (division of functions among duplicates), and conservation (conservation of function in both duplicates) (Ohno, 1970; Lynch and Force, 2000; Force et al., 1999; Hahn, 2009). The duplication-degeneration-complementation (DDC) model is the most favored model, suggesting maintenence of duplicates is driven mostly by degenerative mutations, with the first mutational event leading to the subfunctionalization of one of the copies, and the second mutation leading either to subfunctionalization, neofunctionalization, or nonfunctionalization of the other copy (Lynch and Force, 2000; Force et al., 1999). Support for the DDC model is mixed, and it is suggested that subfunctionalization and neofunctionalization of expression may be highly context dependent (Hahn, 2009; Dunn et al., 2018; Huminiecki and Wolfe, 2004). Our findings here support this suggestion of context dependence, as we find support for the ortholog conjecture only in one of the treatments.

Finally, it is important to note that these comparisons considered global patterns of expression change across branches. It would be interesting to investigate patterns of expression change across branches following speciation and duplication events within gene trees, and to identify particular branches within gene trees that show greater evolutionary change following duplication events. Unfortunately, there is not enough power to test these hypotheses at the level of the gene tree, as there are a limited number of tips and speciation/duplication branches within the gene trees. Investigations within gene trees would, however, be possible with much greater species sampling.

4.4.4   Comparison to other methods

With the expansion of functional genomic tools, including RNA-seq and single cell sequencing methods, we are able to look not only at how genomic variation gives rise to phenotypic diversity in a single species or organism, but also at how functional genomic variation shapes phenotypic diversity across a number (>2) of closely and distantly related species to understand broader evolutionary patterns and processes (Brawand

et al., 2011; Barbosa-Morais et al., 2012; Breschi et al., 2016; Clarke et al., 2017; Macrander et al., 2016; Merkin et al., 2012; Necsulea et al., 2014; Perry et al., 2012; Levin et al., 2016; Sudmant et al., 2015; Ma et al., 2018; Yang and Wang, 2013; Zhang et al., 2014). The methods developed here are broadly applicable to a range of functional genomic data, in addition to existing RNA-seq data sets in other species.

This approach presents several solutions to past limitations. We apply ancestral trait reconstruction methods to reconstruct expression values at nodes in gene trees and calculate changes across branches. This has several advantages – we are able to take species relationships into account, and we are also able to overcome sampling issues at the tips, as expression values of different treatments can be reconstructed at deep internal nodes, even where there may be inconsistent sampling at the tips. We also propose a novel solution to identifying comparable genes across species. Where past approaches identify strict orthologs *before* conducting analyses, we use information from all genes within a gene tree regardless of their evolutionary history of duplication or speciation. Subsequently, we are then able to use information at the nodes to select branches in gene trees that correspond to particular branches in the species tree. This enables the selection of branches that are descended from speciation events, but that do not necessarily belong to gene trees where only single copy genes are found. This enables us to consider vastly more genes than we would be able to with a strict ortholog approach. Additionally, we are also able to compare expression following duplication and speciation events. Finally, by using gene trees, we are able to consider expression evolution within the context of the gene phylogeny, which may differ significantly from the species phylogeny. This enables us to consider the evolution of gene expression in context, considering not only the evolutionary relationships between genes, but also between species.

There are, however, limitations to this approach as applied here. Here we consider changes among species within a particular treatment, and do not directly investigate differences in the magnitude of expression among treatments across species, or treatment-specificity. Other authors have used summary statistics to capture treatment-specific expression, such as tau (Yanai et al., 2004a), which ranges from 0 (there is no specificity in expression) to 1 (expression is highly specific to a tissue), however these statistics provide no directionality (i.e. specific to *which* tissue) (Dunn et al., 2018; Kryuchkova-Mostacci and Robinson-Rechavi, 2016). For example, a gene may have a tau of 1 and be highly specific to nectophores in species A, and

also have tau of 1 and be highly specific to gastrozooids in species B, which makes it a highly problematic summary statistic for interpreting patterns and mechanisms of expression evolution (Dunn et al., 2018). Understanding the evolution of tissue specificity along branches in gene trees will require the development of novel summary statistics that take specificity and directionality into account. Once these summary statistics are developed, it would be possible to use the data and methods developed here to calculate these summary statistics at the nodes and the tips, and identify specific shifts in treatment specificity in particular lineages within gene trees.

Secondly, as with all methods that rely on mapping to reference transcriptomes rather than genomes, this approach is limited by the quality of the reference transcriptomes. Not all reference transcriptomes were sequenced to equal depth among species, and this has important effects on the presence or absence of genes from particular species within the gene tree. This not only has an effect on the representation of expression values, but also impacts the power to investigate patterns of expression among branches within a gene tree. However, with genome sequencing becoming cheaper and more readily available, the widespread availability of reference genomes will help alleviate many of these issues. Reference genomes will also improve gene models, enabling the distinction of different alleles of the same gene from paralogous genes, this in turn will improve the quality of the gene trees.

4.5   Conclusions

Palpons emerge as the only treatment to support the ortholog conjecture, and we observe greater evolutionary change in expression in these zooids across branches following duplication events as compared to following speciation events. These expression results may provide support for subfunctionalization or neofunctionalization of duplicated genes within this zooid. Among branches that follow speciation events (and correspond to branches in the species tree), palpons also have the largest number of branches with lineage and treatment specific changes in branches that correspond to branch L (positive) and branch K (negative). Branch L is the branch leading to *Nanomia bijuga*, while branch K leads to *Agalma elegans* – the nature of the differences between palpons in these two species is not clear. Differential gene expression could not

identify significant differences between the B-palpon and the gastric palpons (chapter 3), but it appears that there are a number of expression differences between gastric palpons in *Agalma elegans* and palpons in *Nanomia bijuga*. Notably, many of genes identified as showing lineage and treatment specific change across branch L are those involved in patterning and morphogenesis.

Palpons have historically been defined not by their unique features, but by what they lack (a mouth, a well developed basigaster, and other conspicuous morphological features) relative to gastrozooids (Totton, 1965). They often dissociate from specimens and do not preserve well upon fixation, which means they are often not described as thoroughly as other zooids. Palpons are thought to be present at the common ancestor of siphonophores, however they have been lost multiple times across Siphonophora (chapter 2) (Munro et al., 2018). It is notable that unlike many other zooid types such as gastrozooids or gonophores that play a more conserved role within the colony, palpons have been observed to perform a wide range of functions in different siphonophore species. While many other zooid types such as gastrozooids or gonophores are thought to play relatively well-defined and conserved role within the colony, various functions have been ascribed to palpons in different siphonophore species. These include digestion, absorption, and egestion in *Nanomia bijuga* and *Forskalia* (Mackie and Boag, 1963), and also defense in *Physophora hydrostatica* (Totton, 1965). This latter case is particularly striking – the palpons of *Physophora hydrostatica* vigorously attack objects placed in their vicinity (Totton, 1965).

The lack of knowledge about morphological traits specific to palpons and the sense that they don't have a consistent narrow function can give the sense that they are not as interesting as other zooids. The expression results presented here suggest that palpons should be viewed in a very different light – they may have greater functional diversity than previously thought and are more fundamental to siphonophore biology than has been appreciated. The large changes in palpon expression that we find along particular branches may be associated with lineage-specific adaptations. The fact that they are only zooid in which we find support for the ortholog conjecture suggests that neo-functionalization or sub-functionalization may play a bigger role in these zooids than others, consistent with more evolutionary change in their function across the siphonophore phylogeny.

The phylogenetic gene-tree based methods we introduced here allowed us to consider vastly more genes than

in strict ortholog approaches. Although there are several limitations, including a dependence on high quality reference transcriptomes, these results are highly encouraging. We now have a powerful framework which can be applied to investigate patterns of gene expression among species and potentially other summary statistics of expression. Here, we specifically investigate changes within particular treatments, but another equally important question is investigating patterns of treatment specificity of expression among species. These technical limitations will be eased by sequencing advances, and with greater access to high quality genomes for multiple species, our abilities to apply these methods and compare expression across species will be even stronger. These methods will become more relevant as higher quality references become available. Additionally, with better gene models and more complete coverage, our power to investigate patterns of change across branches within specific gene trees will also be greater. Finally, an obvious next step will be to apply these methods to address the potential drivers of these changes, by modelling specific evolutionary scenarios on gene trees.

4.6 Supplementary Information
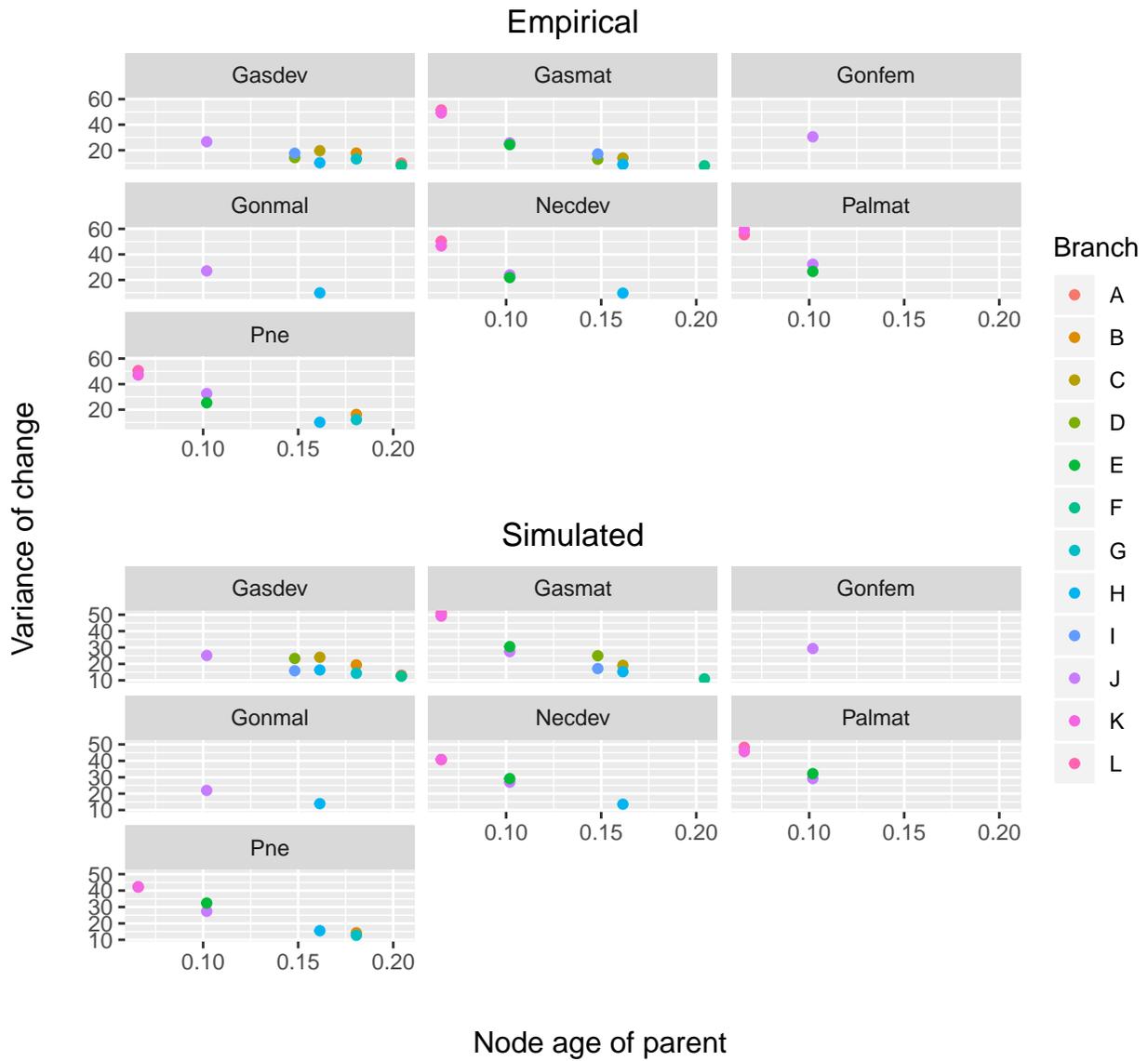
Figure S4.1: Variance of change across a branch plotted against the age of the parent node. Branch ID is coded in colour, values are separated by treatment. Gasdev = Developing Gastrozooid, Gasmat= Mature gastrozooid, Gonfem= Female gonodendron, Gonmal= Male gonodendron, Necdev= Developing nectophore, Palmat = Mature palpon, Pne= pneumatophore.
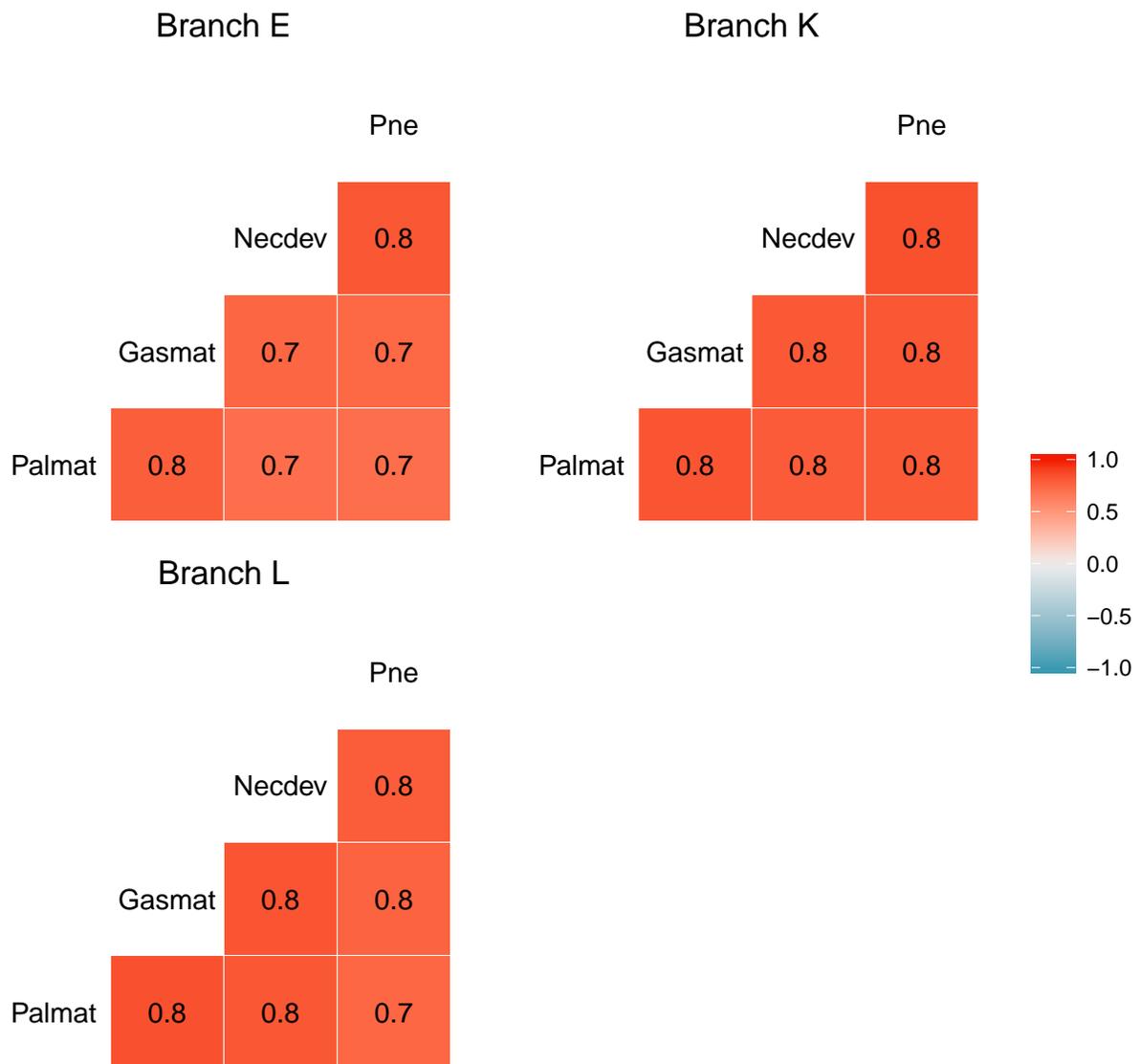
Figure S4.2: Covariance of changes across the same branch across all gene trees, where the branch corresponds to branch E, K or L in the species tree. Gasdev - developing gastrozooid. Gasmat - mature gastrozooid. Palmat - mature palpon. Necdev - developing nectophore.

Bibliography

Agassiz, A. and Mayer, A. G. (1902). Reports on the scientific results of the expedition to the Tropical Pacific, 1899-1900. III. The Medusae. *Memoirs of the Museum of Comparative Zoology at Harvard University*, 26:139–176.

Alexa, A. and Rahnenfuhrer, J. (2016). *topGO: Enrichment Analysis for Gene Ontology*. R package version 2.30.1.

Altenhoff, A. M., Studer, R. A., Robinson-Rechavi, M., and Dessimoz, C. (2012). Resolving the ortholog conjecture: orthologs tend to be weakly, but significantly, more similar in function than paralogs. *PLoS Computational Biology*, 8(5):e1002514.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3):403–410.

Araya, J. F., Aliaga, J. A., and Araya, M. E. (2016). On the distribution of *Physalia physalis*(Hydrozoa: Physaliidae) in Chile. *Marine Biodiversity*, 46(3):731–735.

Artigas, G. Q., Lapébie, P., Leclère, L., Takeda, N., Deguchi, R., Jékely, G., Momose, T., and Houliston, E. (2018). A gonad-expressed opsin mediates light-induced spawning in the jellyfish *Clytia*. *Elife*, 7:e29555.

Auer, P. L. and Doerge, R. (2010). Statistical design and analysis of RNA-Seq data. *Genetics*, 185(2):405–416.

Babcock, H. L. (1938). The Sea-Turtles of the Bermuda Islands, with a Survey of the present state of the Turtle Fishing Industry. *Proceedings of the Zoological Society of London*, 107(4):595–602.

Barbosa-Morais, N. L., Irimia, M., Pan, Q., Xiong, H. Y., Gueroussov, S., Lee, L. J., Slobodeniuc, V., Kutter, C., Watt, S., Colak, R., et al. (2012). The evolutionary landscape of alternative splicing in vertebrate species. *Science*, 338(6114):1587–1593.

Bardi, J. and Marques, A. C. (2007). Taxonomic redescription of the Portuguese man-of-war, *Physalia physalis* (Cnidaria, Hydrozoa, Siphonophorae, Cystonectae) from Brazil. *Iheringia, Série Zoologia*, 97:425 – 433.

Barham, E. G. (1966). Deep scattering layer migration and composition: observations from a diving saucer. *Science*, 151(3716):1399–1403.

Barrett, A. J. and Kirschke, H. (1981). [41] Cathepsin B, cathepsin H, and cathepsin L. In *Methods in Enzymology*, volume 80, pages 535–561. Elsevier.

Bassot, J.-M., Bilbaut, A., Mackie, G. O., Passano, L. M., and De Ceccatty, M. P. (1978). Bioluminescence and other responses spread by epithelial conduction in the siphonophore *Hippopodius*. *The Biological Bulletin*, 155(3):473–498.

Beklemishev, W. N. (1969). *Principles of Comparative Anatomy of Invertebrates. Volume I. Promorphology.* Oliver & Boyd, Edinburgh.

Bidigare, R. R. and Biggs, D. C. (1980). The role of sulfate exclusion in buoyancy maintenance by siphonophores and other oceanic gelatinous zooplankton. *Comparative Biochemistry and Physiology Part A: Physiology*, 66(3):467–471.

Bieri, R. (1966). Feeding preferences and rates of the snail, *Ianthina prolongata*, the barnacle, *Lepas anserifera*, the nudibranchs, *Glaucus atlanticus* and *Fiona pinnata*, and the food web in the marine neuston. *Publications of the Seto Marine Biological Laboratory*, 14:161–170.

Biggs, D. C. and Harbison, G. R. (1976). The siphonophore *Bathyphysa sibogae* Lens and Van Riemsdijk, 1908, in the Sargasso Sea, with notes on its natural history. *Bulletin of Marine Science*, 26(1):14–18.

Bingham, F. O. and Albertson, H. D. (1974). Observations on beach strandings of the *Physalia* (Portuguese-man-of-war) community. *Veliger*, 17(2):220–224.

Blackstone, N. W. and Buss, L. W. (1991). Shape variation in hydractiniid hydroids. *The Biological Bulletin*, 180(3):394–405.

Bodansky, M. and Rose, W. (1922). Comparative studies of digestion. I The digestive enzymes of coelenterates. *American Journal of Physiology*, pages 473–481.

Boeckel, G. R. and Ehrlich, B. E. (2018). NCS-1 is a regulator of calcium signaling in health and disease. *Biochimica et Biophysica Acta - Molecular Cell Research*.

Boussau, B., Szöllősi, G. J., Duret, L., Gouy, M., Tannier, E., and Daubin, V. (2013). Genome-scale coestimation of species and gene trees. *Genome Research*, 23(2):323–330.

Brawand, D., Soumillon, M., Necsulea, A., Julien, P., Csárdi, G., Harrigan, P., Weier, M., Liechti, A., Aximu-Petri, A., Kircher, M., Albert, F. W., Zeller, U., Khaitovich, P., Grützner, F., Bergmann, S., Nielsen, R., Pääbo, S., and Kaessmann, H. (2011). The evolution of gene expression levels in mammalian organs. *Nature*, 478(7369):343–348.

Breschi, A., Djebali, S., Gillis, J., Pervouchine, D. D., Dobin, A., Davis, C. A., Gingeras, T. R., and Guigó, R. (2016). Gene-specific patterns of expression variation across organs and species. *Genome Biology*, 17(1):151.

Brown, D., Garcia-Segura, L., and Orci, L. (1984). Carbonic anhydrase is associated with taste buds in rat tongue. *Brain Research*, 324(2):346–348.

Carré, C. (1967). Le developpement larvaire d'*Abylopsis tetragona*. *Cahiers de Biologie Marine*, 8:185–193.

Carré, C. (1979). Sur le genre *Sulculeolaria* Blainville, 1834 (Siphonophora, Calycophorae, Diphyidae). *Annales de l'Institut océanographique (Paris)*, 55:27–48.

Carré, C. and Carré, D. (1991). A complete life cycle of the calycophoran siphonophore *Muggiaea kochi* (Will) in the laboratory, under different temperature conditions: ecological implications. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences*, 334(1269):27–32.

Carré, C. and Carré, D. (1993). Ordre des siphonophores. In Grassé, P.-P., editor, *Traité de Zoologie. Anatomie, Systématique, Biologie*, pages 523–596. Paris:Masson.

Carré, D. (1969). Etude histologique du developpement de *Nanomia bijuga* (Chiaje, 1841), Siphonophore Physonecte, Agalmidae. *Cahiers de Biologie Marine*, 10:325–341.

Carré, D. and Carré, C. (2000). Origin of germ cells, sex determination, and sex inversion in medusae of the genus Clytia (Hydrozoa, Leptomedusae): The influence of temperature. *Journal of Experimental Zoology*, 287(3):233–242.

Carroll, S. B. (2005). Evolution at two levels: on genes and form. *PLoS Biology*, 3(7):e245.

Cartwright, P., Bowsher, J., and Buss, L. W. (1999). Expression of a Hox gene, Cnox-2, and the division of labor in a colonial hydroid. *Proceedings of the National Academy of Sciences of the United States of America*, 96(5):2183–2186.

Cartwright, P., Evans, N. M., Dunn, C. W., Marques, A., Miglietta, M. P., Schuchert, P., and Collins, A. G. (2008). Phylogenetics of Hydroidolina (Hydrozoa: Cnidaria). *Journal of the Marine Biological Association of the United Kingdom*, 88(08):1663.

Cartwright, P. and Nawrocki, A. M. (2010). Character evolution in Hydrozoa (phylum Cnidaria). *Integrative and Comparative Biology*, 50:456–472.

Chang, D. and Duda, T. F. (2014). Application of community phylogenetic approaches to understand gene expression: differential exploration of venom gene space in predatory marine gastropods. *BMC Evolutionary Biology*, 14(1):123.

Chapman, J. A., Kirkness, E. F., Simakov, O., Hampson, S. E., Mitros, T., Weinmaier, T., Rattei, T., Balasubramanian, P. G., Borman, J., Busam, D., Gee, L., Kibler, D. F., Law, L., Lindgens, D., Martinez, D. E., Peng, J., Wigge, P. A., Bertulat, B., Guder, C., Nakamura, Y., Ozbek, S., Watanabe, H., Khalturin, K., Hemmrich, G., Franke, A., Augustin, R., Fraune, S., Hayakawa, E., Hayakawa, S., Hirose, M., Hwang, J. S., Ikeo, K., Nishimiya-Fujisawa, C., Ogura, A., Takahashi, T., Steinmetz, P. R. H., Zhang, X., Aufschnaiter, R., Marie-Eder, K., Anne-Gorny, K., Salvenmoser, W., Heimberg, A. M., Wheeler, B. M., Peterson, K. J., Böttger, A., Tischler, P., Wolf, A., Gojobori, T., Remington, K. A., Strausberg, R. L., Venter, C. J., Technau, U., Hobmayer, B., Bosch, T. C. G., Holstein, T. W., Fujisawa, T., Bode,

H. R., David, C. N., Rokhsar, D. S., and Steele, R. E. (2010). The dynamic genome of Hydra. *Nature*, 464(7288):592.

Chen, J., Swofford, R., Johnson, J., Cummings, B. B., Rogel, N., Lindblad-Toh, K., Haerty, W., di Palma, F., and Regev, A. (2017). A quantitative model for characterizing the evolutionary history of mammalian gene expression. *bioRxiv*.

Chen, X. and Zhang, J. (2012). The Ortholog Conjecture Is Untestable by the Current Gene Ontology but Is Supported by RNA Sequencing Data. *PLoS Computational Biology*, 8(11):e1002784.

Chin, D. and Means, A. R. (2000). Calmodulin: a prototypical calcium sensor. *Trends in Cell Biology*, 10(8):322–328.

Choy, C. A., Haddock, S. H. D., and Robison, B. H. (2017). Deep pelagic food web structure as revealed by in situ feeding observations. *Proceedings of the Royal Society B: Biological Sciences*, 284(1868).

Chun, C. (1887). Zur Morphologie der Siphonophoren. 2. Ueber die postembryonale Entwickelung von *Physalia*. *Zoologischer Anzeiger*, 10:557–577.

Chun, C. (1891). Die Canarischen Siphonophoren I. *Stephanophyes superba* und die Familie der Stephanophyiden. *Senckenberg Gesellschaft für Naturforschung*, 16:553–627.

Church, S. H., Ryan, J. F., and Dunn, C. W. (2015a). Automation and Evaluation of the SOWH Test with SOWHAT. *Systematic Biology*, 64(6):1048–1058.

Church, S. H., Siebert, S., Bhattacharyya, P., and Dunn, C. W. (2015b). The histology of *Nanomia bijuga* (hydrozoa: Siphonophora). *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 324(5):435–449.

Clark, A. M. (1970). The Portuguese Man-of-War. *Marine Observer*, 40:63–65.

Clark, F. E. and Lane, C. E. (1961). Composition of float gases of *Physalia physalis*. *Proceedings of the Society for Experimental Biology and Medicine*, 107(3):673–674.

Clarke, T. H., Garb, J. E., Haney, R. A., Chaw, R. C., Hayashi, C. Y., and Ayoub, N. A. (2017). Evolutionary shifts in gene expression decoupled from gene duplication across functionally distinct spider silk glands. *Scientific Reports*, 7(1):8393.

Copeland, D. E. (1968). Fine Structures Of The Carbon Monoxide Secreting Tissue In The Float Of Portuguese Man-of-war *Physalia physalis* L. *The Biological Bulletin*, 135(3):486–500.

Costello, J. H., Colin, S. P., Gemmell, B. J., Dabiri, J. O., and Sutherland, K. R. (2015). Multi-jet propulsion organized by clonal development in a colonial siphonophore. *Nature Communications*, 6:8158.

De Smet, R., Adams, K. L., Vandepoele, K., Van Montagu, M. C., Maere, S., and Van de Peer, Y. (2013). Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. *Proceedings of the National Academy of Sciences of the United States of America*, 110(8):2898–2903.

Dudgeon, S. R. and Buss, L. W. (1996). Growing with the flow: on the maintenance and malleability of colony form in the hydroid Hydractinia. *The American Naturalist*, 147(5):667–691.

Dunn, C., Pugh, P., and Haddock, S. (2005a). Molecular Phylogenetics of the Siphonophora (Cnidaria), with Implications for the Evolution of Functional Specialization. *Systematic Biology*, 54(6):916–935.

Dunn, C. W. (2005). Complex colony-level organization of the deep-sea siphonophore *Bargmannia elongata* (Cnidaria, Hydrozoa) is directionally asymmetric and arises by the subdivision of pro-buds. *Developmental Dynamics*, 234(4):835–845.

Dunn, C. W., Howison, M., and Zapata, F. (2013a). Agalma: an automated phylogenomics workflow. *BMC Bioinformatics*, 14(1):330.

Dunn, C. W., Luo, X., and Wu, Z. (2013b). Phylogenetic analysis of gene expression. *Integrative and Comparative Biology*, 53(5):847–856.

Dunn, C. W. and Munro, C. (2016). Comparative genomics and the diversity of life. *Zoologica Scripta*, 45(S1):5–13.

Dunn, C. W., Pugh, P. R., and Haddock, S. H. D. (2005b). Molecular Phylogenetics of the Siphonophora (Cnidaria), with Implications for the Evolution of Functional Specialization. *Systematic Biology*, 54(6):916–935.

Dunn, C. W. and Wagner, G. P. (2006). The evolution of colony-level development in the Siphonophora (Cnidaria: Hydrozoa). *Development Genes and Evolution*, 216(12):743–754.

Dunn, C. W., Zapata, F., Munro, C., Siebert, S., and Hejnol, A. (2018). Pairwise comparisons across species are problematic when analyzing functional genomic data. *Proceedings of the National Academy of Sciences of the United States of America*, pages E409–E417.

Eng, K. H., Bravo, H. C., and Keleş, S. (2009). A phylogenetic mixture model for the evolution of gene expression. *Molecular Biology and Evolution*, 26(10):2363–2372.

Enright, A. J., Van Dongen, S., and Ouzounis, C. A. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Research*, 30(7):1575–1584.

Fan, J., Teng, X., Liu, L., Mattaini, K. R., Looper, R. E., Vander Heiden, M. G., and Rabinowitz, J. D. (2014). Human phosphoglycerate dehydrogenase produces the oncometabolite D-2-hydroxyglutarate. *ACS Chemical Biology*, 10(2):510–516.

Felsenstein, J. (1985). Phylogenies and the Comparative Method. *The American Naturalist*, 125(1):1–15.

Felsenstein, J. (2008). Comparative Methods with Sampling Error and Within-Species Variation: Contrasts Revisited and Revised. *The American Naturalist*, 171(6):713–725.

Ferreira, G. C. and Gong, J. (1995). 5-Aminolevulinate synthase and the first step of heme biosynthesis. *Journal of Bioenergetics and Biomembranes*, 27(2):151–159.

Fidler, A. L., Vanacore, R. M., Chetyrkin, S. V., Pedchenko, V. K., Bhave, G., Yin, V. P., Stothers, C. L., Rose, K. L., McDonald, W. H., Clark, T. A., Borza, D., Steele, R., Ivy, M., Aspirnauts, Hudson, J., and Hudson, B. (2014). A unique covalent bond in basement membrane is a primordial innovation for tissue evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 111(1):331–336.

Fitch, W. M. (1970). Distinguishing Homologous from Analogous Proteins. *Systematic Biology*, 19(2):99–113.

Fitch, W. M. (2000). Homology: a personal view on some of the problems. *Trends in Genetics*, 16(5):227 – 231.

FitzJohn, R. G. (2012). Diversitree: comparative phylogenetic analyses of diversification in R. *Methods in Ecology and Evolution*, 3(6):1084–1092.

Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*, 151(4):1531–1545.

Fujiwara, K., Okamura-Ikeda, K., and Motokawa, Y. (1984). Mechanism of the glycine cleavage reaction. Further characterization of the intermediate attached to H-protein and of the reaction catalyzed by T-protein. *Journal of Biological Chemistry*, 259(17):10664–10668.

Gabaldon, T. and Koonin, E. V. (2013). Functional and evolutionary implications of gene orthology. *Nature Reviews Genetics*, 14(5):360–366.

Garstang, W. (1946). The morphology and relations of the Siphonophora. *Quarterly Journal of Microscopical Science*, 87(2):103–193.

Gilad, Y. and Mizrahi-Man, O. (2015). A reanalysis of mouse ENCODE comparative gene expression data. *F1000Research*, 4.

Gilad, Y., Rifkin, S. A., Bertone, P., Gerstein, M., and White, K. P. (2005). Multi-species microarrays reveal the effect of sequence divergence on gene expression profiles. *Genome Research*, 15(5):674–680.

Goolsby, E. W., Bruggeman, J., and Ané, C. (2017). Rphylopars: fast multivariate phylogenetic comparative methods for missing data and within-species variation. *Methods in Ecology and Evolution*, 8(1):22–27.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., and Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7):644–652.

Grafen, A. (1989). The phylogenetic regression. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 326(1233):119–157.

Grimmelikhuijzen, C. J. P., Spencer, A. N., and Carré, D. (1986). Organization of the nervous system of physonectid siphonophores. *Cell and Tissue Research*, 246:463–479.

Grossmann, M. M., Nishikawa, J., and Lindsay, D. J. (2015). Diversity and community structure of pelagic cnidarians in the celebes and sulu seas, southeast asian tropical marginal seas. *Deep-Sea Research Part I: Oceanographic Research Papers*, 100:54–63.

Gu, X. (2004). Statistical framework for phylogenomic analysis of gene family expression profiles. *Genetics*, 167(1):531–542.

Gu, X. (2015). Understanding tissue expression evolution: from expression phylogeny to phylogenetic network. *Briefings in Bioinformatics*, 17(2):249–254.

Guang, A., Howison, M., Zapata, F., Lawrence, C. E., and Dunn, C. (2017). Revising transcriptome assemblies with phylogenetic information in Agalma 1.0. *bioRxiv*.

Hacker, U., Lin, X., and Perrimon, N. (1997). The Drosophila sugarless gene modulates Wingless signaling and encodes an enzyme involved in polysaccharide biosynthesis. *Development*, 124(18):3565–3573.

Haddock, S. H. D., Dunn, C. W., and Pugh, P. R. (2005). A re-examination of siphonophore terminology and morphology, applied to the description of two new prayine species with remarkable bio-optical properties. *Journal of the Marine Biological Association of the United Kingdom*, 85(3):695–707.

Haeckel, E. (1888). Report on the Siphonophorae collected by HMS Challenger during the years 1873-1876. *Report of the Scientific Results of the voyage of H.M.S. Challenger. Zoology*, 28:1–380.

Hahn, M. W. (2009). Distinguishing among evolutionary models for the maintenance of gene duplicates. *Journal of Heredity*, 100(5):605–617.

Harmon, L. J., Weir, J. T., Brock, C. D., Glor, R. E., and Challenger, W. (2007). GEIGER: investigating evolutionary radiations. *Bioinformatics*, 24(1):129–131.

Herring, P. J. (1971). Biliprotein coloration of *Physalia physalis*. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry*, 39(4):739–746.

Hessinger, D. A. and Ford, H. M. (1988). Ultrastructure of the small cnidocyte of the Portuguese man-of-war (*Physalia physalis*) tentacle. In Hessinger, D. A. and Lenhoff, H. M., editors, *The Biology of Nematocysts*. Academic Press, San Diego, USA.

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution*, 35(2):518–522.

Huelsenbeck, J. P., Nielsen, R., and Bollback, J. P. (2003). Stochastic mapping of morphological characters. *Systematic Biology*, 52(2):131–158.

Hughes, A. L. (1994). The evolution of functionally novel proteins after gene duplication. *Proceedings of the Royal Society B: Biological Sciences*, 256(1346):119–124.

Huminiecki, L. and Wolfe, K. H. (2004). Divergence of spatial gene expression profiles following species-specific gene duplications in human and mouse. *Genome research*, 14(10a):1870–1879.

Huxley, T. H. (1859). *The Oceanic Hydrozoa: a description of the Calycophoridae and Physophoridae observed during the voyage of HMS Rattlesnake 1846-1850*. The Ray Society, London.

Iosilevskii, G. and Weihs, D. (2009). Hydrodynamics of sailing of the Portuguese man-of-war *Physalia physalis*. *Journal of the Royal Society Interface*, 6(36):613–626.

Ishibashi, K., Kondo, S., Hara, S., and Morishita, Y. (2010). The evolutionary aspects of aquaporin family. *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology*, 300(3):R566–R576.

Jacobs, W. (1937). Beobachtungen Über das Schweben der Siphonophoren. *Zeitschrift für vergleichende Physiologie*, 24(4):583–601.

Jenkins, R. L. (1983). Observations on the commensal relationship of *Nomeus gronovii* with *Physalia physalis*. *Copeia*, 1983(1):250–252.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., von Haeseler, A., and Jermiin, L. S. (2017). Modelfinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14(6):587.

Kato, K. (1933). Is *Nomeus* a harmless Inquilinus of *Physalia*? *Proceedings of the Imperial Academy*, 9(9):537–538.

Katoh, K. and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, 30(4):772–780.

Kayal, E., Bentlage, B., Cartwright, P., Yanagihara, A. A., Lindsay, D. J., Hopcroft, R. R., and Collins, A. G. (2015). Phylogenetic analysis of higher-level relationships within hydroidolina (cnidaria: Hydrozoa) using mitochondrial genome data and insight into their mitochondrial transcription. *PeerJ*, 3:e1403.

Kayal, E., Bentlage, B., Pankey, M. S., Ohdera, A. H., Medina, M., Plachetzki, D. C., Collins, A. G., and Ryan, J. F. (2018). Phylogenomics provides a robust topology of the major cnidarian lineages and insights on the origins of key organismal traits. *BMC Evolutionary Biology*, 18(1):68.

Kayal, E., Roure, B., Philippe, H., Collins, A. G., and Lavrov, D. V. (2013). Cnidarian phylogenetic relationships as revealed by mitogenomics. *BMC Evolutionary Biology*, 13(1):5.

Kerr, A. M., Baird, A. H., and Hughes, T. P. (2011). Correlated evolution of sex and reproductive mode in corals (anthozoa: Scleractinia). *Proceedings of the Royal Society B: Biological Sciences*, 278(1702):75–81.

Khaitovich, P., Weiss, G., Lachmann, M., Hellmann, I., Enard, W., Muetzel, B., Wirkner, U., Ansorge, W., and Pääbo, S. (2004). A neutral model of transcriptome evolution. *PLoS Biology*, 2(5):e132.

Kikuchi, G., Yoshida, T., and Noguchi, M. (2005). Heme oxygenase and heme degradation. *Biochemical and Biophysical Research Communications*, 338(1):558–567.

King, M.-C. and Wilson, A. C. (1975). Evolution at two levels in humans and chimpanzees. *Science*, 188(4184):107–116.

Kivelä, J., Parkkila, S., Parkkila, A.-K., Leinonen, J., and Rajaniemi, H. (1999). Salivary carbonic anhydrase isoenzyme VI. *Journal of Physiology*, 520(2):315–320.

Kryuchkova-Mostacci, N. and Robinson-Rechavi, M. (2016). Tissue-specificity of gene expression diverges slowly between orthologs, and rapidly between paralogs. *PLoS Computational Biology*, 12(12):e1005274.

Kusserow, A., Pang, K., Sturm, C., Hrouda, M., Lentfer, J., Schmidt, H. A., Technau, U., von Haeseler, A., Hobmayer, B., Martindale, M. Q., et al. (2005). Unexpected complexity of the Wnt gene family in a sea anemone. *Nature*, 433(7022):156.

Lamarck, J. B. (1801). *Système des animaux sans vertèbres*. Deterville, Paris.

Lane, C. E. (1960a). The Portuguese man-of-war. *Scientific American*, 202(3):158–171.

Lane, C. E. (1960b). The toxin of *Physalia* nematocysts. *Annals of the New York Academy of Sciences*, 90(3):742–750.

Lane, C. E. and Dodge, E. (1958). The toxicity of *Physalia* nematocysts. *The Biological Bulletin*, 115(2):219–226.

Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4):357–359.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10(3):R25.

Larimer, J. L. and Ashby, E. A. (1962). Float gases, gas secretion and tissue respiration in the Portuguese man-of-war, *Physalia*. *Journal of Cellular and Comparative Physiology*, 60(1):41–47.

Lartillot, N., Lepage, T., and Blanquart, S. (2009). PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics*, 25(17):2286–2288.

Lee, P. N., Pang, K., Matus, D. Q., and Martindale, M. Q. (2006). A WNT of things to come: evolution of Wnt signaling and polarity in cnidarians. In *Seminars in Cell & Developmental Biology*, volume 17, pages 157–167. Elsevier.

Lehnert, E. M., Burriesci, M. S., and Pringle, J. R. (2012). Developing the anemone *Aiptasia* as a tractable

model for cnidarian-dinoflagellate symbiosis: the transcriptome of aposymbiotic *A. pallida*. *BMC Genomics*, 13(1):271.

Leloup, E. (1935). Les siphonophores de la rade de Villefranche-sur-Mer (Alpes Maritimes, France). *Mémoires du Musée royal d'histoire naturelle de Belgique*, 11:1–12.

Lenhoff, H. M. and Schneiderman, H. A. (1959). The chemical control of feeding in the Portuguese man-of-war, *Physalia physalis* L. and its bearing on the evolution of the Cnidaria. *The Biological Bulletin*, 116(3):452–460.

Levin, M. (2005). Left–right asymmetry in embryonic development: a comprehensive review. *Mechanisms of Development*, 122(1):3–25.

Levin, M., Anavy, L., Cole, A. G., Winter, E., Mostov, N., Khair, S., Senderovich, N., Kovalev, E., Silver, D. H., Feder, M., Fernandez-Valverde, S. L., Nakanishi, N., Simmons, D., Simakov, O., Larsson, T., Liu, S.-Y., Jerafi-Vider, A., Yaniv, K., Ryan, J. F., Martindale, M. Q., Rink, J. C., Arendt, D., Degnan, S. M., Degnan, B. M., Hashimshony, T., and Yanai, I. (2016). The mid-developmental transition and the evolution of animal body plans. *Nature*, 531(7596):637–641.

Li, B. and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, 12(1):323.

Li, B., Ruotti, V., Stewart, R. M., Thomson, J. A., and Dewey, C. N. (2009a). RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics*, 26(4):493–500.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009b). The sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16):2078–2079.

Lin, S., Lin, Y., Nery, J. R., Urich, M. A., Breschi, A., Davis, C. A., Dobin, A., Zaleski, C., Beer, M. A., Chapman, W. C., et al. (2014). Comparison of the transcriptional landscapes between human and mouse tissues. *Proceedings of the National Academy of Sciences of the United States of America*, 111(48):17224–17229.

Lindsay, D. (2005). Planktonic communities below 2000 m depth. *Bulletin of the Plankton Society of Japan (Japan)*.

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12):550.

Lynch, M. and Force, A. (2000). The probability of duplicate gene preservation by subfunctionalization. *Genetics*, 154(1):459–473.

Ma, S., Avanesov, A. S., Porter, E., Lee, B. C., Mariotti, M., Zemskaya, N., Guigo, R., Moskalev, A. A., and Gladyshev, V. N. (2018). Comparative transcriptomics across 14 Drosophila species reveals signatures of longevity. *Aging Cell*, page e12740.

Mackie, G. O. (1960). Studies on *Physalia physalis* (L.). Part 2. Behavior and histology. *Discovery Reports*, 30:371–407.

Mackie, G. O. (1963). Siphonophores, Bud Colonies, and Superorganism. In Dougherty, E., editor, *The Lower Metazoa*, pages 329–337. University of California Press, Berkeley.

Mackie, G. O. (1964). Analysis of locomotion of a siphonophore colony. *Proceedings of the Royal Society B: Biological Sciences*, 159:366–391.

Mackie, G. O. (1965). Conduction in the nerve-free epithelia of siphonophores. *American Zoologist*, 5:366–391.

Mackie, G. O. (1974). Locomotion, flotation, and dispersal. In Muscatine, L. and Lenhoff, H. M., editors, *Coelenterate Biology: Reviews and New Perspectives*, pages 313–357. Academic Press, New York, USA.

Mackie, G. O. (1986). From aggregates to integrates: physiological aspects of modularity in colonial animals. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences*, 313(1159):175–196.

Mackie, G. O. and Boag, D. A. (1963). Fishing, feeding and digestion in siphonophores. *Pubblicazioni della Stazione Zoologica di Napoli*, 33:178–96.

Mackie, G. O., Pugh, P. R., and Purcell, J. E. (1987). Siphonophore biology. *Advances in Marine Biology*, 24:97–262.

Macrander, J., Broe, M., and Daly, M. (2016). Tissue-Specific Venom Composition and Differential Gene Expression in Sea Anemones. *Genome Biology and Evolution*, 8(8):2358–2375.

Macrander, J., Brugler, M. R., and Daly, M. (2015). A RNA-seq approach to identify putative toxins from acrorhagi in aggressive and non-aggressive *Anthopleura elegantissima* polyps. *BMC Genomics*, 16(1):221.

Magie, C. R., Pang, K., and Martindale, M. Q. (2005). Genomic inventory and expression of Sox and Fox genes in the cnidarian Nematostella vectensis. *Development Genes and Evolution*, 215(12):618–630.

Mapstone, G. M. (2014). Global diversity and review of Siphonophorae (Cnidaria: Hydrozoa). *PLoS One*, 9(2):e87737.

Marlow, H. Q., Srivastava, M., Matus, D. Q., Rokhsar, D., and Martindale, M. Q. (2009). Anatomy and development of the nervous system of Nematostella vectensis, an anthozoan cnidarian. *Developmental Neurobiology*, 69(4):235–254.

McIntyre, L. M., Lopiano, K. K., Morse, A. M., Amin, V., Oberg, A. L., Young, L. J., and Nuzhdin, S. V. (2011). RNA-seq: technical variability and sampling. *BMC Genomics*, 12(1):293.

Melé, M., Ferreira, P. G., Reverter, F., DeLuca, D. S., Monlong, J., Sammeth, M., Young, T. R., Goldmann, J. M., Pervouchine, D. D., Sullivan, T. J., et al. (2015). The human transcriptome across tissues and individuals. *Science*, 348(6235):660–665.

Merkin, J., Russell, C., Chen, P., and Burge, C. B. (2012). Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science*, 338(6114):1593–1599.

Mitsunaga, K., Harada-Itadani, J., Shikanai, T., Tateno, H., Ikehara, Y., Hirabayashi, J., Narimatsu, H., and Angata, T. (2009). Human C21orf63 is a heparin-binding protein. *Journal of Biochemistry*, 146(3):369–373.

Momose, T., Derelle, R., and Houliston, E. (2008). A maternally localised Wnt ligand required for axial patterning in the cnidarian Clytia hemisphaerica. *Development*, 135(12):2105–2113.

Munro, C., Siebert, S., Zapata, F., Howison, M., Damian-Serrano, A., Church, S. H., Goetz, F. E., Pugh, P. R., Haddock, S. H. D., and Dunn, C. W. (2018). Improved phylogenetic resolution within Siphonophora (Cnidaria) with implications for trait evolution. *Molecular Phylogenetics and Evolution*, 127:823–833.

Nagamatsu, Y., Yanagisawa, I., Kimoto, M., Okamoto, E., and Koga, D. (1995). Purification of a chi-tooligosaccharidolytic β-N-acetylglucosaminidase from Bombyx mori larvae during metamorphosis and the nucleotide sequence of its cDNA. *Bioscience, Biotechnology, and Biochemistry*, 59(2):219–225.

Nakamura, T. Y., Jeromin, A., Mikoshiba, K., and Wakabayashi, S. (2011). Neuronal calcium sensor-1 promotes immature heart function and hypertrophy by enhancing Ca2+ signals. *Circulation Research*, pages CIRCRESAHA–111.

Necsulea, A., Soumillon, M., Warnefors, M., Liechti, A., Daish, T., Zeller, U., Baker, J. C., Grützner, F., and Kaessmann, H. (2014). The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature*, 505(7485):635.

Nehrt, N. L., Clark, W. T., Radivojac, P., and Hahn, M. W. (2011). Testing the Ortholog Conjecture with Comparative Functional Genomic Data from Mammals. *PLoS Computational Biology*, 7(6):e1002073.

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*, 32(1):268–274.

Nival, P., Gostan, J., Malara, G., and Chara, R. (1976). Evolution du plancton dans la baie de Villefranche-sur-Mer ala fin du printemps (mai et juin 1971), 2. Biomasse de phytoplancton et production primarie. *Vie Milieu*, 26:47–76.

Oakley, T. H., Gu, Z., Abouheif, E., Patel, N. H., and Li, W.-H. (2005). Comparative methods for the analysis of gene-expression evolution: an example using yeast functional genomic data. *Molecular Biology and Evolution*, 22(1):40–50.

Ohno, S. (1970). *Evolution by Gene Duplication*. Springer-Verlag, Berlin.

Okada, Y. K. (1932). Développement post-embryonnaire de la Physalie Pacifique. *Memoirs of the College of Science, Kyoto Imperial University, Series B*, 8(1):1–27.

Okada, Y. K. (1935). Les jeunes Physalies. *Memoirs of the College of Science, Kyoto Imperial University, Series B*, 10(5):407–410.

Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*, 401(6756):877–884.

Pagès, F., Gonzàlez, H. E., Ramòn, M., Sobarzo, M., and Gili, J.-M. (2001). Gelatinous zooplankton assemblages associated with water masses in the Humboldt Current System, and potential predatory impact by *Bassia bassensis* (Siphonophora: Calycophorae). *Marine Ecology Progress Series*, 210:13–24.

Pankey, M. S., Minin, V. N., Imholte, G. C., Suchard, M. A., and Oakley, T. H. (2014). Predictable transcriptome evolution in the convergent and complex bioluminescent organs of squid. *Proceedings of the National Academy of Sciences of the United States of America*, 111(44):E4736–E4742.

Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20:289–290.

Peng, D., Belkhiri, A., Hu, T., Chaturvedi, R., Asim, M., Wilson, K. T., Zaika, A., and El-Rifai, W. (2011). Glutathione peroxidase 7 protects against oxidative DNA damage in oesophageal cells. *Gut*, pages gutjnl–2011.

Perry, G. H., Melsted, P., Marioni, J. C., Wang, Y., Bainer, R., Pickrell, J. K., Michelini, K., Zehr, S., Yoder, A. D., Stephens, M., et al. (2012). Comparative RNA sequencing reveals substantial genetic variation in endangered primates. *Genome Research*, 22(4):602–610.

Philippe, H., Derelle, R., Lopez, P., Pick, K., Borchiellini, C., Boury-Esnault, N., Vacelet, J., Renard, E., Houliston, E., Quéinnec, E., Da Silva, C., Winker, P., Le Guyader, H., Leys, S., Jackson, D., Schreiber, F., Erpenbeck, D., Morgenstern, B., Wörheide, G., and Manuel, M. (2009). Phylogenomics revives traditional views on deep animal relationships. *Current Biology*, 19(8):706–712.

Pickwell, G. V. (1970). The physiology of carbon monoxide production by deep-sea coelenterates: causes and consequences. *Annals of the New York Academy of Sciences*, 174(1):102–115.

Pickwell, G. V., Barham, E. G., and Wilton, J. W. (1964). Carbon monoxide production by a bathypelagic siphonophore. *Science*, 144(3620):860–862.

Pind, S., Slominski, E., Mauthe, J., Pearlman, K., Swoboda, K. J., Wilkins, J. A., Sauder, P., and Natowicz, M. R. (2002). V490M, a common mutation in 3-phosphoglycerate dehydrogenase deficiency, causes enzyme deficiency by decreasing the yield of mature enzyme. *Journal of Biological Chemistry*, 277(9):7136–7143.

Pontin, D. R. and Cruickshank, R. H. (2012). Molecular phylogenetics of the genus *Physalia* (Cnidaria: Siphonophora) in New Zealand coastal waters reveals cryptic diversity. *Hydrobiologia*, 686(1):91–105.

Prasannan, P. and Appling, D. R. (2009). Human mitochondrial C1-tetrahydrofolate synthase: submitochondrial localization of the full-length enzyme and characterization of a short isoform. *Archives of biochemistry and biophysics*, 481(1):86–93.

Prieto, L., Macías, D., Peliz, A., and Ruiz, J. (2015). Portuguese Man-of-War (*Physalia physalis*) in the Mediterranean: A permanent invasion or a casual appearance? *Scientific Reports*, 5:11545.

Pugh, P. R. (1974). The vertical distribution of the siphonophores collected during the SOND cruise, 1965. *Journal of the Marine Biological Association of the United Kingdom*, 54(1):25–90.

Pugh, P. R. (1983). Benthic Siphonophores: A Review of the Family Rhodaliidae (Siphonophora, Physonectae). *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences*, 301:165–300.

Pugh, P. R. (1984). The diel migrations and distributions within a mesopelagic community in the north east Atlantic. 7. Siphonophores. *Progress in Oceanography*, 13(3-4):461–489.

Pugh, P. R. (1998). A re-description of *Frillagalma vityazi* Daniel 1966 (Siphonophorae, Agalmatidae). *Scientia Marina*, 62(3):233–245.

Pugh, P. R. (2003). A revision of the family Forskaliidae (Siphonophora, Physonectae). *Journal of Natural History*, 37(11):1281–1327.

Pugh, P. R. (2006). The taxonomic status of the genus Moseria (Siphonophora, Physonectae). *Zootaxa*, 1343:1–42.

Pugh, P. R. (2016). A synopsis of the Family Cordagalmatidae fam. nov.(Cnidaria, Siphonophora, Physonectae). *Zootaxa*, 4095(1):1–64.

Pugh, P. R., Pagès, F., and Boorman, B. (1997). Vertical distribution and abundance of pelagic cnidarians in the eastern Weddell Sea, Antarctica. *Journal of the Marine Biological Association of the United Kingdom*, 77(2):341–360.

Purcell, J. E. (1981a). Dietary composition and diel feeding patterns of epipelagic siphonophores. *Marine Biology*, 65(1):83–90.

Purcell, J. E. (1981b). Feeding ecology of *Rhizophysa eysenhardti*, a siphonophore predator of fish larvae. *Limnology and Oceanography*, 26(3):424–432.

Purcell, J. E. (1984). Predation on fish larvae by *Physalia physalis*, the Portuguese man of war. *Marine Ecology Progress Series*, 19(1):189–191.

Putnam, N. H., Srivastava, M., Hellsten, U., Dirks, B., Chapman, J., Salamov, A., Terry, A., Shapiro, H., Lindquist, E., Kapitonov, V. V., Jurka, J., Genikhovich, G., Grigoriev, I. V., Lucas, S. M., Steele, R. E., Finnerty, J. R., Technau, U., Martindale, M. Q., and Rokhsar, D. S. (2007). Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science*, 317(5834):86–94.

Qualls, J. E., Subramanian, C., Rafi, W., Smith, A. M., Balouzian, L., DeFreitas, A. A., Shirey, K. A., Reutterer, B., Kernbauer, E., Stockinger, S., et al. (2012). Sustained generation of nitric oxide and control of mycobacterial infection requires argininosuccinate synthase 1. *Cell Host & Microbe*, 12(3):313–323.

Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3(2):217–223.

Revell, L. J., Mahler, D. L., and Neto, P. P. (2012). A new phylogenetic method for identifying exceptional phenotypic diversification. *Evolution*, 66(1):135–146.

Rifkin, S. A., Kim, J., and White, K. P. (2003). Evolution of gene expression in the Drosophila melanogaster subgroup. *Nature Genetics*, 33(2):138.

Rohlfs, R. V. and Nielsen, R. (2015). Phylogenetic ANOVA: The Expression Variance and Evolution Model for Quantitative Trait Evolution. *Systematic biology*, 64(5):695–708.

Romero, I. G., Ruvinsky, I., and Gilad, Y. (2012). Comparative studies of gene expression and the evolution of gene regulation. *Nature Reviews Genetics*, 13(7):505.

Sanders, S. M. and Cartwright, P. (2015). Interspecific differential expression analysis of RNA-Seq data yields insight into life cycle variation in hydractiniid hydrozoans. *Genome Biology and Evolution*, 7(8):2417–2431.

Sanders, S. M., Shcheglovitova, M., and Cartwright, P. (2014). Differential gene expression between functionally specialized polyps of the colonial hydrozoan *Hydractinia symbiolongicarpus* (Phylum Cnidaria). *BMC Genomics*, 15(1):406.

Schlining, B. M. and Stout, N. J. (2006). MBARI's video annotation and reference system. In *OCEANS 2006*, pages 1–5. IEEE.

Schuchert, P. (2003). Hydroids (Cnidaria, Hydrozoa) of the Danish expedition to the Kei Islands. *Steenstrupia*, 27(2):137–256.

Schuchert, P. (2018). World Hydrozoa Database. Siphonophorae. Accessed through: World Register of Marine Species.

Shimano, M., Ouchi, N., Nakamura, K., van Wijk, B., Ohashi, K., Asaumi, Y., Higuchi, A., Pimentel, D. R., Sam, F., Murohara, T., et al. (2011). Cardiac myocyte follistatin-like 1 functions to attenuate hypertrophy following pressure overload. *Proceedings of the National Academy of Sciences of the United States of America*, page 201108559.

Siebert, S., Goetz, F. E., Church, S. H., Bhattacharyya, P., Zapata, F., Haddock, S. H. D., and Dunn, C. W. (2015). Stem cells in *Nanomia bijuga* (Siphonophora), a colonial animal with localized growth zones. *EvoDevo*, 6(1):22.

Siebert, S., Pugh, P. R., Haddock, S. H. D., and Dunn, C. W. (2013). Re-evaluation of characters in Apolemiidae (Siphonophora), with description of two new species from Monterey Bay, California. *Zootaxa*, 3702(3):201–232.

Siebert, S., Robinson, M. D., Tintori, S. C., Goetz, F., Helm, R. R., Smith, S. A., Shaner, N., Haddock, S. H. D., and Dunn, C. W. (2011). Differential Gene Expression in the Siphonophore *Nanomia bijuga* (Cnidaria) Assessed with Multiple Next-Generation Sequencing Workflows. *PLoS ONE*, 6(7):e22953.

Sonnhammer, E. L. L. and Koonin, E. V. (2002). Orthology, paralogy and proposed classification for paralog subtypes. *Trends in Genetics*, 18(12):619 – 620.

Soumillon, M., Necsulea, A., Weier, M., Brawand, D., Zhang, X., Gu, H., Barthes, P., Kokkinaki, M., Nef, S., Gnirke, A., et al. (2013). Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. *Cell Reports*, 3(6):2179–2190.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, 22(21):2688–2690.

Steche, O. (1907). Die Genitalanlagen der Rhizophysalien. *Zeitschrift fur Wissenschaftliche Zooiogie*, 86:134–171.

Steche, O. (1910). Die knospungsgesetz und der Bau der Anhangsgruppen von Physalia. In *Festschrift zum sechzigsten geburtstag Richard Hertwigs*, pages 356–372. Verlag Von Gustave Fischer, Jena.

Stern, D. B. and Crandall, K. A. (2018). The evolution of gene expression underlying vision loss in cave animals. *Molecular Biology and Evolution*, page msy106.

Sudmant, P. H., Alexis, M. S., and Burge, C. B. (2015). Meta-analysis of RNA-seq expression data across species, tissues and studies. *Genome Biology*, 16(1):287.

Sukumaran, J. and Holder, M. T. (2010). DendroPy: a Python library for phylogenetic computing. *Bioinformatics*, 26(12):1569–1571.

Swofford, D. L., Olsen, G. J., Waddell, P. J., and Hillis, D. M. (1996). Phylogenetic inference. In Hillis D M, Moritz C, M. B. K., editor, *Molecular systematics. 2nd ed.* Sunderland (MA): Sinauer Associates.

Szczepanek, S., Cikala, M., and David, C. N. (2002). Poly-γ-glutamate synthesis during formation of nematocyst capsules in Hydra. *Journal of Cell Science*, 115(4):745–751.

Talavera, G. and Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology*, 56(4):564–577.

Tanco, S., Zhang, X., Morano, C., Avilés, F. X., Lorenzo, J., and Fricker, L. D. (2010). Characterization of the substrate specificity of human carboxypeptidase A4 and implications for a role in extracellular peptide processing. *Journal of Biological Chemistry*, 285(24):18385–18396.

Thompson, T. E. and Bennett, I. (1969). *Physalia* nematocysts: utilized by mollusks for defense. *Science*, 166(3912):1532–1533.

Törönen, P., Medlar, A., and Holm, L. (2018). PANNZER2: a rapid functional annotation web server. *Nucleic Acids Research*, gky350.

Totton, A. K. (1960). Studies on *Physalia physalis* (L.). Part 1. Natural history and morphology. *Discovery Reports*, 30:301–368.

Totton, A. K. (1965). *A synopsis of the Siphonophora*. British Museum (Natural History).

Totton, A. K. and Mackie, G. O. (1956). Dimorphism in the Portuguese man-of-war. *Nature*, 177(4502):290.

Tschopp, P., Sherratt, E., Sanger, T. J., Groner, A. C., Aspiras, A. C., Hu, J. K., Pourquié, O., Gros, J., and Tabin, C. J. (2014). A relative shift in cloacal location repositions external genitalia in amniote evolution. *Nature*, 516(7531):391.

Uyeda, J. C. and Harmon, L. J. (2014). A Novel Bayesian Method for Inferring and Interpreting the Dynamics of Adaptive Landscapes from Phylogenetic Comparative Data. *Systematic Biology*, 63(6):902–918.

Valdés, Á. and Campillo, O. A. (2004). Systematics of pelagic aeolid nudibranchs of the family Glaucidae (Mollusca, Gastropoda). *Bulletin of Marine Science*, 75(3):381–389.

Walsh, E. C. and Stainier, D. Y. (2001). UDP-glucose dehydrogenase required for cardiac valve formation in zebrafish. *Science*, 293(5535):1670–1673.

Warner, J. F., Guerlais, V., Amiel, A. R., Johnston, H., Nedoncelle, K., and Röttinger, E. (2018). NvERTx:

a gene expression database to compare embryogenesis and regeneration in the sea anemone Nematostella vectensis. *Development*, 145(10):dev162867.

Whitney, K. D., Boussau, B., Baack, E. J., and Garland, T. (2011). Drift and Genome Complexity Revisited. *PLoS Genetics*, 7(6):e1002092–5.

Willem, V. (1894). La structure du palpon chez *Apolemia uvaria* Esch., et les phénomènes de l'absorption dans ces organes. *Bulletin de l'Académie Royale de Belgique Classe des Sciences*, 27:354–363.

Williams, R. and Conway, D. V. P. (1981). Vertical distribution and seasonal abundance of *Aglantha digitale* (OF Müller)(Coelenterata: Trachymedusae) and other planktonic coelenterates in the northeast Atlantic Ocean. *Journal of Plankton Research*, 3(4):633–643.

Wilson, D. P. (1947). The Portuguese Man-of-War, *Physalia physalis*L., in British and adjacent seas. *Journal of the Marine Biological Association of the United Kingdom*, 27(1):139–172.

Wittenberg, J. B. (1960). The source of carbon monoxide in the float of the Portuguese man-of-war, *Physalia physalis* L. *Journal of Experimental Biology*, 37(4):698–705.

Wittenberg, J. B., Noronha, J. M., and Silverman, M. (1962). Folic acid derivatives in the gas gland of *Physalia physalis* L. *Biochemical Journal*, 85(1):9.

Woodcock, A. H. (1944). A theory of surface water motion deduced from the wind-induced motion of the *Physalia. Journal of Marine Research*, 5(3):196–206.

Woodcock, A. H. (1956). Dimorphism in the Portuguese man-of-war. *Nature*, 178(4527):253.

Wray, G. A., Hahn, M. W., Abouheif, E., Balhoff, J. P., Pizer, M., Rockman, M. V., and Romano, L. A. (2003). The evolution of transcriptional regulation in eukaryotes. *Molecular Biology and Evolution*, 20(9):1377–1419.

Yamada, S., Morimoto, H., Fujisawa, T., and Sugahara, K. (2007). Glycosaminoglycans in Hydra magnipapillata (Hydrozoa, Cnidaria): demonstration of chondroitin in the developing nematocyst, the sting organelle, and structural characterization of glycosaminoglycans. *Glycobiology*, 17(8):886–894.

Yanagihara, A. A., Kuroiwa, J. M. Y., Oliver, L. M., and Kunkel, D. D. (2002). The ultrastructure of nematocysts from the fishing tentacle of the Hawaiian bluebottle, *Physalia utriculus* (Cnidaria, Hydrozoa, Siphonophora). *Hydrobiologia*, 489(1-3):139–150.

Yanai, I., Benjamin, H., Shmoish, M., Chalifa-Caspi, V., Shklar, M., Ophir, R., Bar-Even, A., Horn-Saban, S., Safran, M., Domany, E., et al. (2004a). Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics*, 21(5):650–659.

Yanai, I., Graur, D., and Ophir, R. (2004b). Incongruent expression profiles between human and mouse orthologous genes suggest widespread neutral evolution of transcription control. *Omics : a journal of integrative biology*, 8(1):15–24.

Yang, R. and Wang, X. (2013). Organ evolution in angiosperms driven by correlated divergences of gene sequences and expression patterns. *The Plant Cell*, 25(1):71–82.

Yoshida, T. and Kikuchi, G. (1973). Major pathways of serine and glycine catabolism in various organs of the rat and cock. *The Journal of Biochemistry*, 73(5):1013–1022.

Young, M. D., Wakefield, M. J., Smyth, G. K., and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology*, 11(2):R14.

Yu, G., Smith, D. K., Zhu, H., Guan, Y., and Lam, T. T.-Y. (2017). ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, 8(1):28–36.

Zaitsev, Y. (1997). Neuston of seas and oceans. In Liss, P. S. and Duce, R. A., editors, *The sea surface and global change*, pages 371—382. Cambridge University Press, Cambridge.

Zapata, F., Goetz, F. E., Smith, S. A., Howison, M., Siebert, S., Church, S. H., Sanders, S. M., Ames, C. L., McFadden, C. S., France, S. C., Daly, M., Collins, A. G., Haddock, S. H. D., Dunn, C. W., and Cartwright, P. (2015). Phylogenomic analyses support traditional relationships within Cnidaria. *PLoS One*, 10(10):e0139068.

Zhang, G., Li, C., Li, Q., Li, B., Larkin, D. M., Lee, C., Storz, J. F., Antunes, A., Greenwold, M. J., Meredith, R. W., et al. (2014). Comparative genomics reveals insights into avian genome evolution and adaptation. *Science*, 346(6215):1311–1320.