

# Evolution of gene expression across species and specialized zooids in Siphonophora

Catriona Munro<sup>1,2</sup>, Stefan Siebert<sup>3</sup>, Felipe Zapata<sup>4</sup>, Mark Howison<sup>5</sup>, Steven H.D. Haddock<sup>6</sup>, Casey W. Dunn<sup>7</sup>

<sup>1</sup> Department of Ecology and Evolutionary Biology, Brown University, Providence, RI 02912, USA

<sup>2</sup> Current address: Collège de France, PSL Research University, CNRS, Inserm, Center for Interdisciplinary Research in Biology, 75005 Paris, France

<sup>3</sup> XXXXXX Insert Stefan address XXXXXXXXXXXX

<sup>4</sup> Department of Ecology and Evolutionary Biology, University of California Los Angeles, Los Angeles, CA 90095, USA

<sup>5</sup> Watson Institute for International and Public Affairs, Brown University, Providence, RI 02912, USA

<sup>6</sup> Monterey Bay Aquarium Research Institute, Moss Landing, CA 95039, USA

<sup>7</sup> Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520, USA

## Abstract

## Introduction

Gene expression is an important component of phenotypic diversity within and among units of biological organization, including cells, tissues, organs, and species [1–4]. 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 [5–7]. 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, homologous organs, and developmental timepoints. 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 [???,??,1,8–12]. Other studies found the opposite result [13–16], however re-analyses contest these findings [10,17,18]. These global patterns of gene expression, however, 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 [8]. In this study, we take a gene-centric approach using 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 [9,19–26]. 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 [27–34]. 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 [35].

Analyses of gene expression that do take phylogeny into account typically focus on strict orthologs and map expression values onto the species tree [20,24,25], while other analyses use the expression values themselves to build neighbor-joining trees [1,25]. 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.

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 [36–40]. 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 [37,41].

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 [42–45]. This leaves many unknowns about how zooid structure and function differ across zooid types and species. Recent *in situ* gene expression analyses in siphonophores have described where a small number of pre-selected genes are expressed at high spatial resolution [44,46,47], 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 [46,48–50], 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 [46].

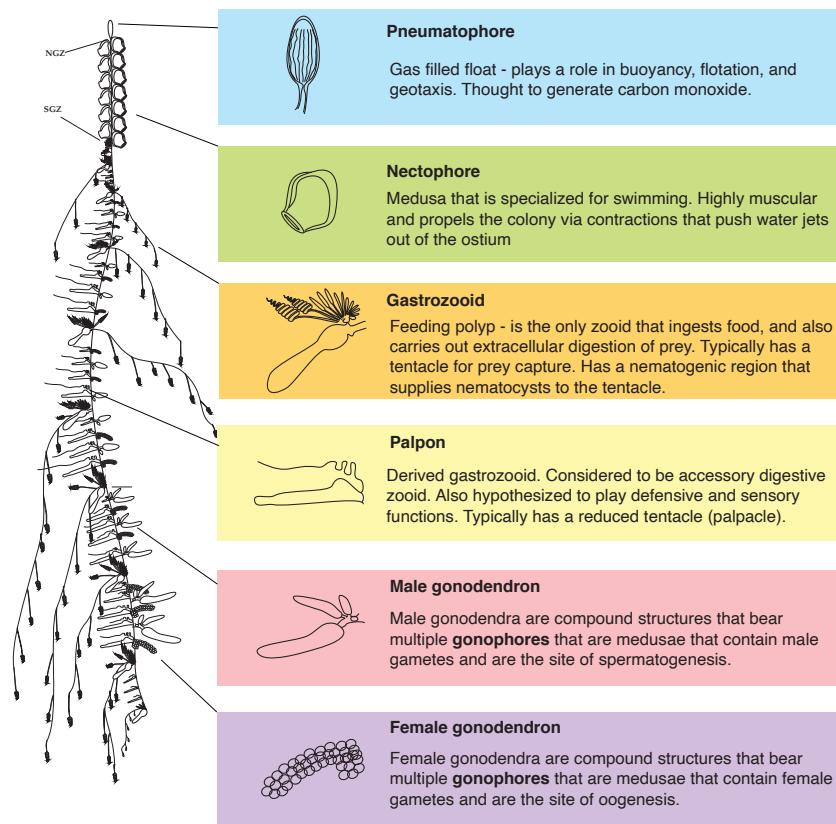


Figure 1: Schematic of the siphonophore *Nanomia bijuga*, with highlighted zooids and pneumatophore, with explanations of known function. NGZ - nectosomal growth zone. SGZ - siphosomal growth zone. Diagram by Freya Goetz ([http://commons.wikimedia.org/wiki/File:Nanomia\\_bijuga\\_whole\\_animal\\_and\\_growth\\_zones.svg](http://commons.wikimedia.org/wiki/File:Nanomia_bijuga_whole_animal_and_growth_zones.svg))

In this study we use RNA-seq data to conduct differential expression analyses among siphonophore zooids, and one specialized tissue (the pneumatophore, a gas filled float), within siphonophore species, and also map

RNA-seq data directly onto gene phylogenies in order to compare gene expression signatures within zooid types across species. We also investigate differential gene expression patterns within species, in particular focusing on three species-specific zooid types that represent lineage specific zooid diversification events.

## Results

### Differential expression within species

We sequenced mRNA from microdissected zooids from seven different siphonophore species (at least two replicate colonies) and mapped these short-read libraries to previously published transcriptomes [51]. We collected RNA-seq data from, where possible, 5 different zooids and one specialized tissue, the pneumatophore, as well as unique zooids specific to *Agalma elegans* (B palpons), *Physalia physalis* (tentacular palpon), and *Bargmannia elongata* (yellow and white gastrozooids), and where possible developing and mature zooids (Table 1). Due to species availability, we were not able to sample more than one replicate for some zooids - single replicates were excluded from downstream differential gene expression analyses (see Fig. S1).

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 [52,53]. The differences between technical replicates (Fig. S2) were found to be much smaller (0.39% variance of expression distance) than the differences between zooids (98.32% variance of expression distance). Differences among technical replicates of the same zooid were correlated with library size and run, not by lane.

The next component of variation we considered was biological variation among sampled colonies. Within species, the greatest variance was among zooids/tissues, as opposed to between biological replicates of the same zooid/tissue (Figs. S3–S9), with some exceptions, such as a developing palpon replicate in *Nanomia bijuga* (Fig. S8), and a developing gastrozooid and male gonodendra in *Frillagalma vityazi* (Fig. S7). 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 zooids/specialized tissues within species.

We identified genes that are significantly upregulated in particular zooids, and in the pneumatophore, for each of the sampled species. We also identified genes that are significantly upregulated in only one zooid and not in any other within a given species (Fig. 2, S10). The list of genes that were identified as significantly differentially expressed only in a given species and zooid, and the associated transcriptome sequence ids, top blast hits, GO terms, and gene phylogeny ids, are available as supplementary data (supplementary file 1). The full list of genes upregulated in particular zooids, based on pairwise differential expression analyses, are available in supplementary file 2.

For each species and zooid, we found GO term enrichment for biological processes consistent with the functional specialization of the zooid. In gastrozooids that are solely responsible for feeding and digestion, for example, we found these zooids to be enriched for genes involved in proteolysis, chitin, glutathione and peptide catabolism, as well as metabolism of carbohydrates. Likewise for male gonodendra, we found GO term enrichment for biological processes such as sperm flagellum, mitotic cell cycle process, DNA replication. For female gonodendra, there was GO term enrichment for a number of biological processes including mitotic cell cycle process, DNA replication (in *Agalma elegans*), as well as a number of signalling pathways and developmental processes (in *Frillagalma vityazi*), likely reflecting maternal provisioning to the egg [REF].

Within the four best sampled species , we also identified 92 differentially expressed putative transcription factors (24 out of 71 in *Bargmannia elongata*, 43 out of 75 identified in *Frillagalma vityazi*, 50 out of 79 in *Agalma elegans* and 34 out of 72 identified in *Nanomia bijuga*). Many of these transcription factors are significantly differentially expressed in several zooids regardless of developmental stage (both mature and developing zooids), and a subset were found to be significantly differentially expressed only in particular zooids and developmental stages (S11).

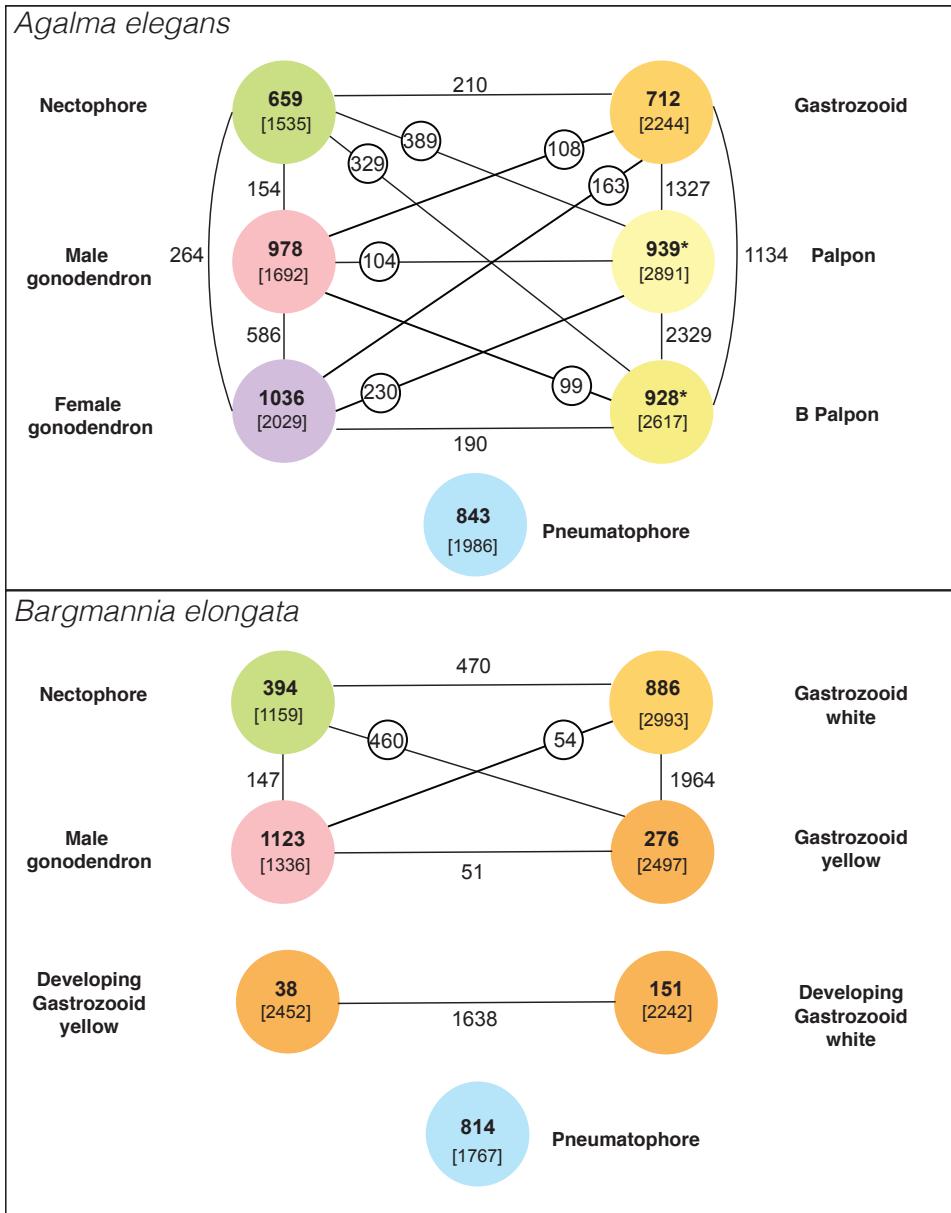


Figure 2: Differentially expressed genes in siphonophore zooids, with two species *Agalma elegans* and *Bargmannia elongata* shown here. Within the circles, the number of differentially expressed genes found uniquely within a particular zoid (bold) and the total number of differentially expressed genes identified in pairwise comparisons between zooids (square brackets). Connections between zooids show the number of highly differentially expressed genes shared among those zooids. \*Due to the very large number of shared genes among Palpons and B Palpons, this number includes genes that are shared between the Palpon and B Palpon.

Table 1: Number of genes per species (library size), number of sampled zooids (at least two replicates, including developing and mature stages), 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	Number of Zoids Sampled	Number of Genes in Gene Trees	Total Number of Gene Trees
<i>Diphyes dispar</i>	51981	3	5833	3381
<i>Frillagalma vityazi</i>	49145	8	5186	3394
<i>Nanomia bijuga</i>	36249	7	5053	3274
<i>Agalma elegans</i>	30234	7	4962	3488
<i>Bargmannia elongata</i>	38177	7	4550	3401
<i>Physalia physalis</i>	19965	5	3886	2789
<i>Apolemia lanosa</i>	18541	2	3531	2628

### Unique species-specific zooids

In siphonophores, there are several instances of lineage-specific zooid diversification events. We investigated 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. S12A and S12B). The “yellow” gastrozooid is larger and darker and occurs as the 7th-10th gastrozooid on the stem [54]. 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 [45,55]. Meanwhile the tentacle is associated with another zooid, the tentacular palpon (Fig. S12C) [42,55–57]. In *P. physalis*, both the gastrozooid and the tentacular palpon are considered to be subfunctionalized from an ancestral gastrozooid type [56]. 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 (Fig. S12D) [36]. 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 *et al* propose that the enlarged B-palpon is the last element in *A. elegans* [36]. 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* (976 genes in the mature tentacular palpon and 606 genes in the mature gastrozooid). A number of genes were found to be differentially expressed in the mature tentacular palpon relative to all other tissues, of which, 670 genes were found to have expression patterns that were not shared with the gastrozooid. In the gastrozooid 849 genes were found to have expression patterns that were not shared with the tentacular palpon.

Between the white mature gastrozooid and the yellow mature gastrozooid in *Bargmannia elongata*, few significant expression differences were identified (8 genes were up in “white” mature gastrozooids relative to 36 genes in “yellow” gastrozooids). Among genes that were found to be significantly differentially expressed in either “white” or “yellow” gastrozooids relative to all other tissues, 276 genes were unique to “yellow” gastrozooids and not found in “white” gastrozooids, and 886 genes were found in “white” gastrozooids and not found in “yellow” gastrozooids (Fig. 2, bottom panel).

Finally, in *Agalma elegans* very few differentially expressed genes were identified between the B palpon and gastric palpons (1 and 2 genes respectively. All three genes have no significant blast hit and did not map to any gene trees). Genes were identified that are upregulated in B palpons relative to all other zooids (gastric

palpons were excluded). Of this, 928 genes were found to be differentially expressed in the B palpon. Most of these genes overlapped with those found to be differentially expressed in the gastric palpon relative to all other zooids (746 genes) (Fig. 2, top panel).

### Evolution of gene expression

In order to compare expression patterns for each zooid across siphonophore species, we used transcriptome and genome data from 41 cnidarian species to generate a total of 7070 gene trees, of which 3831 gene trees passed filtering criteria. We used Orthofinder [58] to reconcile the species tree and the gene trees and identify nodes that represent speciation and duplication events (Fig. 3, step 1). Pruned and unpruned calibrated gene tree files are available as supplementary data. The number of genes represented in these gene trees is shown in table 1. The internal nodes on these gene trees consist of 20088 speciation events and 9082 duplication events. Expression values, as transcripts per million 10K (TPM10K, see methods), were mapped to the tips of the gene trees (Fig. 3, step 2), and maximum likelihood trait values were reconstructed at the nodes (Fig. 3, step 3). Changes in expression across branches were calculated based on the difference between parent and child nodes, scaled by the branch length (Fig. 3, step 4) [35]. We focused on expression in a subset of zooids and tissues that are common across the sampled species: gastrozooids (developing and mature), nectophores (developing), palpons (mature), male and female gonodendra, and the pneumatophore. An exemplar gene tree, for Wnt family genes, with expression values from two zooids mapped at the nodes and tips is shown in figure S13.

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. 3, 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. 3, 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, and the corresponding branches in gene trees are given the same letter. A simulated 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. Under this model, expression variance accumulates among lineages on the gene tree as a function of time, and is used as a model of evolution under drift, as well as some forms of natural selection [59–61].

The distribution of expression changes along branches in gene trees that correspond to equivalent branches in the species tree are shown in figure 4. The variance structure of the empirical data matches that of the simulated data (Fig. S14), suggesting that the size of standard deviation of change around the mean that is observed in figure 4 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. S14). 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 Brownian model (Fig. 4). 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 zooids/the pneumatophore, branches, and the interaction of the two ( $p < 2.2e-16$ ,  $p = 2.319e-13$ ,  $p < 2.2e-16$ , respectively, two-tailed anova). The Brownian model, based on simulated values on the gene trees under BM ( $p=0.4438$ ,  $p=0.9829$ ,  $p=0.4694$ ), and also based on random reshuffling of changes on branches ( $p=0.31218$ ,  $p=0.06787$ ,  $p=0.0.54698$ ), show that expression change along branches in gene trees that correspond to branches in the species tree cannot be explained by significant differences among zooids/the pneumatophore, branches, and the interaction of the two.

Global patterns of expression change across all branches and tissues/zoooids provides perspective on potential

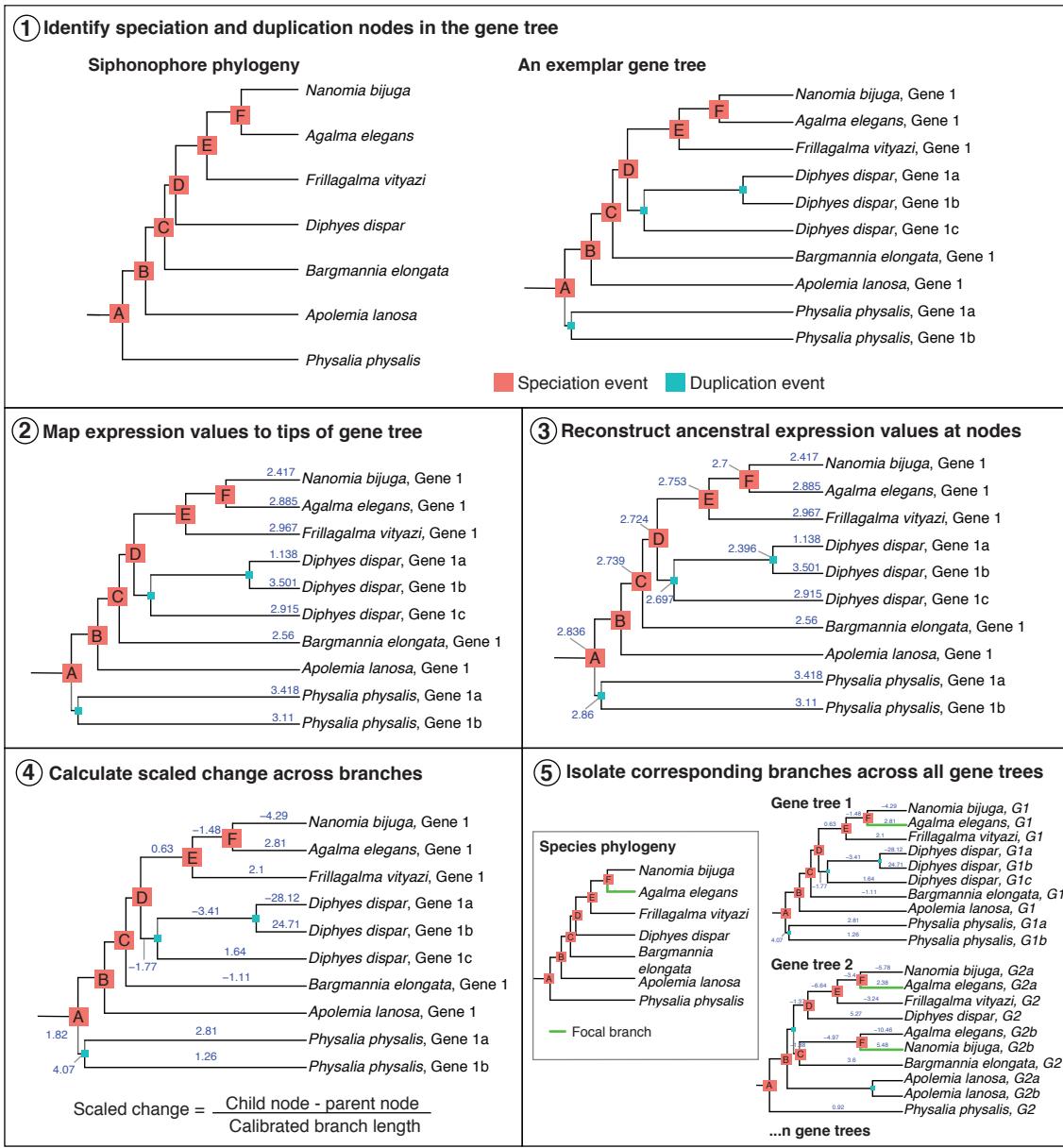


Figure 3: Methods used to identify changes in expression across branches in gene trees. In step 1, we have labelled each of the nodes in the siphonophore phylogeny, and identified equivalent speciation nodes across every gene tree (an exemplar is shown here). Step 2, we map expression values to the tips (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.

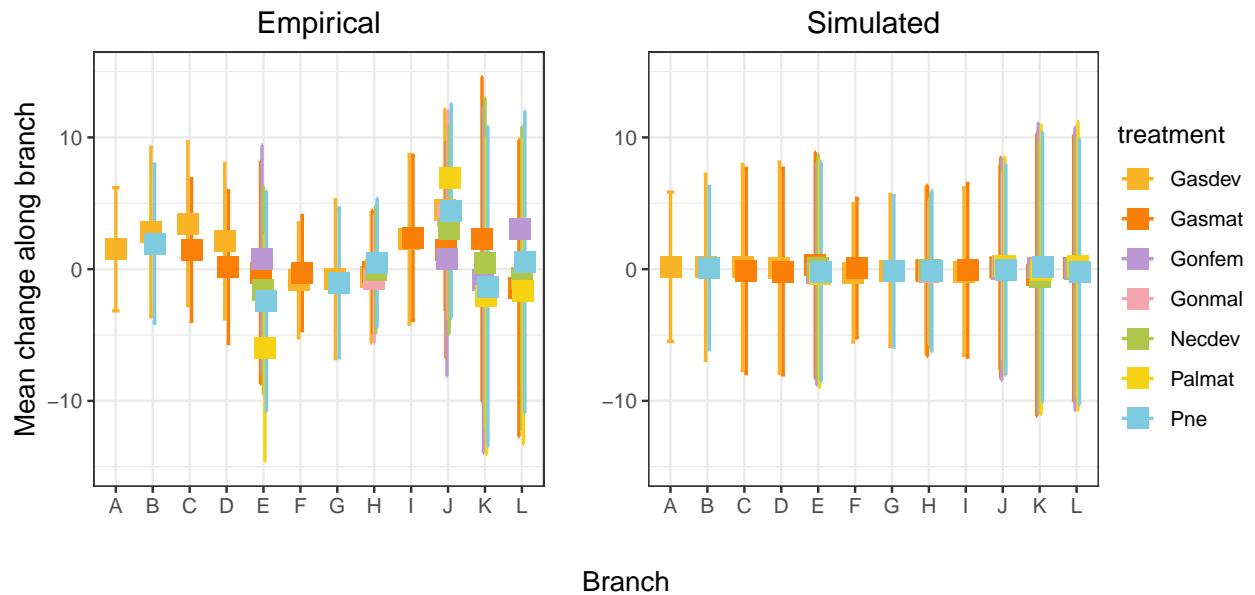
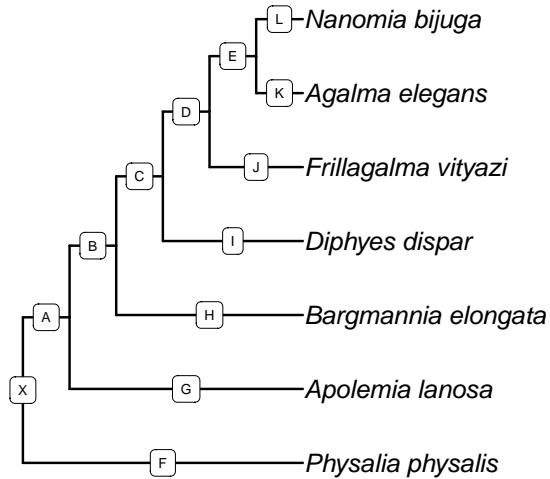


Figure 4: Mean changes along branches across all gene trees that correspond to branches in the species tree, error bar is one standard deviation. Top panel: species phylogeny with branch IDs given as numbers. Lower panel: Distribution of changes along a branch in a gene tree, showing mean change in the 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.

lineage-specific shifts in expression in each zooid/tissue. These results point to a subset of branches in gene trees for particular zooids with large increases or decreases in expression across the branch. These patterns were particularly strong for palpons, where on average changes in expression were highly negative across branches leading to the clade containing *Nanomia bijuga* and *Agalma elegans*, while on average changes in expression were highly positive across branches leading to *Frillagalma vityazi*. We identified 576 gene tree branches in 557 gene trees corresponding to branch J (branch leading to *Frillagalma vityazi*, fig 4) with large ( $>10$ ) positive changes in expression across the branch. Among these, we identified several gene families of interest that show large lineage specific increases in expression in Palpons, including genes involved in patterning and development such as Wnt receptor Frizzled-4, Wnt ligand Wnt2, and homeobox gene Otx-2. For the pneumatophore, we identified 436 gene tree branches corresponding to branch J in 428 gene trees with large ( $>10$ ) positive changes in expression across the branch, including homeobox proteins SAX-1 and Hox-B8 - the latter gene showed low or no TPM10K expression in *Nanomia bijuga*, *Agalma elegans*, and *Apolemia lanosa* pneumatophores and high expression in *Frillagalma vityazi* pneumatophores. The full list of branches and changes in expression across the branch and tissue/zoid are available in supplementary file 3.

Next, we were interested in branch level patterns of expression, focusing on whether change in expression across the same branch in a gene tree is correlated between different zooids. We investigated the co-variance structure of changes across particular branches (that correspond to species tree branches) within gene trees among different zooids/the pneumatophore (Figs. 5). 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 zooids/the pneumatophore. In particular, changes across all branches are highly correlated between developing and mature gastrozooids. Changes in expression across the same branch were also highly correlated between gastrozooids and palpons, especially so across branches corresponding to branch K (4, top panel) in the species tree (branch leading to *Agalma elegans*). Meanwhile, changes across all branches are much less correlated between any zooid/the pneumatophore and both the male or female gonodendra. These results suggest that with the exception of male and female gonodendra, in general changes in expression across gene tree branches are highly correlated among zooids - a large positive change in expression in one zooid across a gene tree branch will also typically see a large positive change in expression in another zooid across the same branch.

We were then able to identify instances of branches in gene trees where change in expression across the branch is negatively or positively correlated among different zooids. One example of possible lineage and zooid specific changes in expression across a branch is found in a putative hydrozoan specific actinoporin-like gene family, with closest blast similarity to Auger and Cone snail tereporin/coroporan toxins. In this toxin-like family, we were able to identify one clade (clade Y, Fig. 6, Fig. S15) with three branches corresponding to branches K, L and I in the species tree. The branches that correspond to I and K, leading to a *Diphyes dispar* and *Agalma elegans* gene respectively, both have large increases in mature gastrozooid expression across the branch, while the branch that corresponds to branch L (leading to *Nanomia bijuga*) has a large decrease in expression across the branch in mature gastrozooids and an independent large increase in expression for the same branch in mature palpons. For all other zooids/tissues the expression at the tips is close to 0, with only small changes in expression across the branch. In *Nanomia bijuga* the gene 976500 in Clade Y is significantly differentially expressed in mature palpons as compared to mature gastrozooids (log<sub>2</sub> Fold Change 6.83224, Bonferroni adjusted p-value 0.000567528). These results point towards a lineage and zooid specific change in expression across branch L leading to an actinoporin-like gene homolog in *Nanomia bijuga*.

This gene family shows a complex history of duplication events, with several deep duplication events prior to the divergence of siphonophores, and multiple species-specific duplication events. The exact function of these toxin-like genes is not known or characterized, however they appear to be found predominantly in gastrozooids and/or palpons. In the five duplicated *Physalia physalis* genes, expression is highest in mature gastrozooids relative to developing and mature tentacular palpons (the zooid that bears the tentacle), as well as the developing tentacle. The full phylogeny of all available hydrozoan genes is available in supplementary figure S15. The three *Clytia hemisphaerica* homologs are highly expressed specifically in the polyp head, and whole male and female medusa (blast search on Marine Invertebrate Models Database <http://marimba-test.obs-vlfr.fr/>, XLOC\_029306, XLOC\_013838, XLOC\_029344).

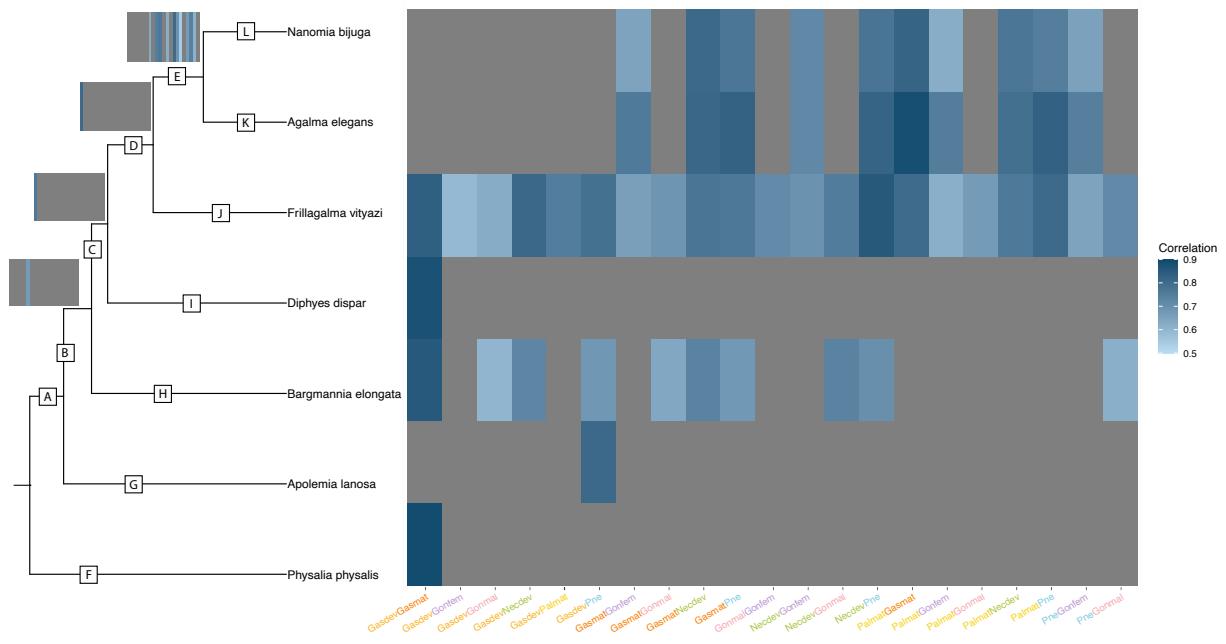


Figure 5: Siphonophore phylogeny and pairwise correlation of changes across a branch among different pairs of zooids/the pneumatophore. Correlation of changes from node to tip are shown in the plot to the right of the phylogeny. Changes between two nodes are shown adjacent to the child node. Color indicates correlation between 0.9 (red) and 0.5 (green). Treatment type is coded in colour along the x-axis, Gasdev = Developing Gastrozooid, Gasmat= Mature gastrozooid, Gonfem= Female gonodendron, Gonmal= Male gonodendron, Necdev= Developing nectophore, Palmat = Mature palpon, Pne= pneumatophore.

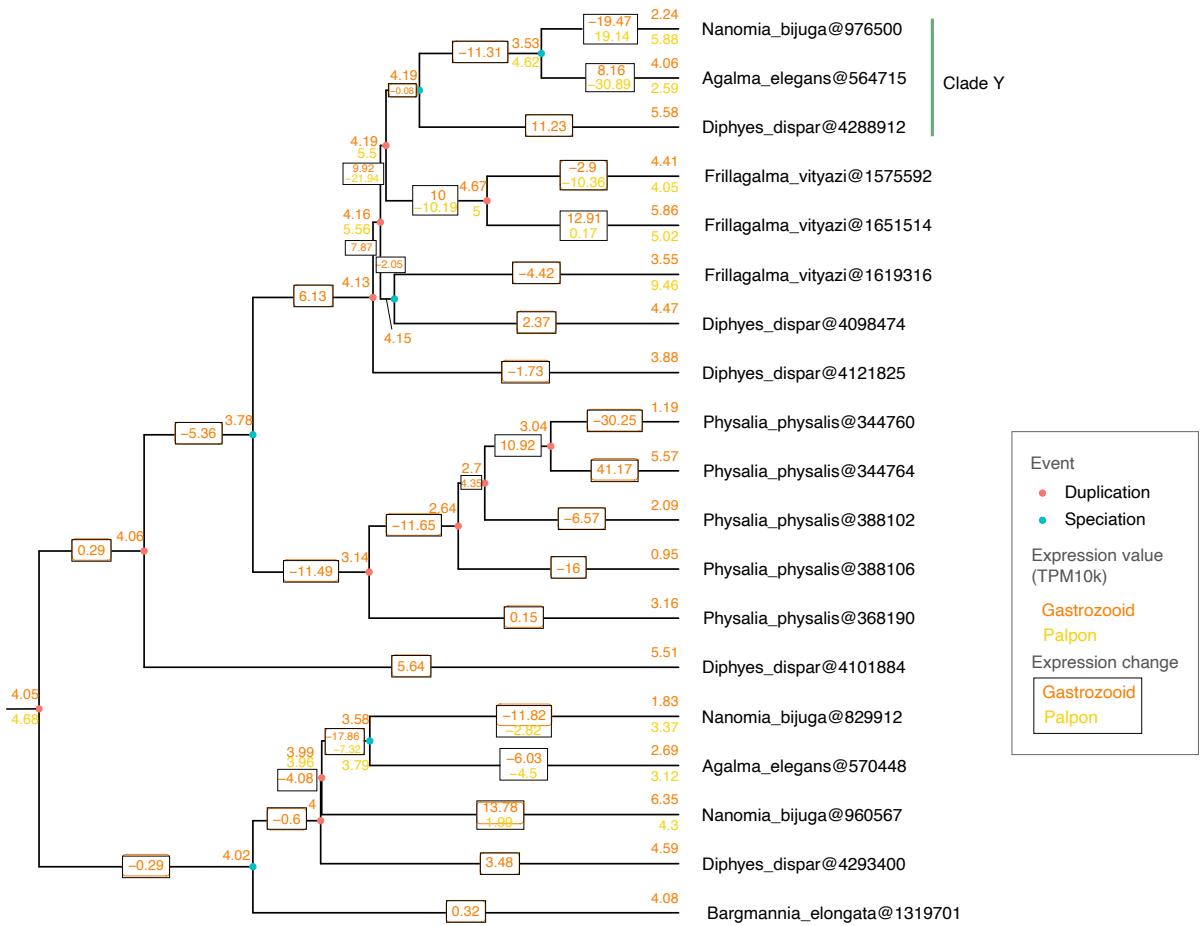


Figure 6: Pruned maximum likelihood gene tree of a putative actinoporin-like gene family found in siphonophore species with expression values (TPM10K) at the tips and nodes. Mature gastrozooid expression in orange (above node), palpon expression in yellow (below node). Changes in expression across the branch are found in black boxes along the branch, orange = mature gastrozooid, yellow = mature palpon. Red circles are duplication nodes and blue circles are speciation nodes.

## Discussion

Colonial animals, consisting of genetically identical modules that are physically and physiologically connected, can be found across the metazoan tree of life [62]. Functional specialization has evolved multiple times within colonial groups, and siphonophores in particular are suggested to have the greatest diversity of functionally specialized zooids [41,62]. Functionally specialized siphonophore zooids have been likened to animal organs, with each functionally specialized zooid possessing distinct cell-types and performing distinct roles within the colony [37]. While this analogy works well for understanding the function of zooids within the colony as a whole, the analogy falls short in terms of explaining the evolutionary origin and development of these biological units: functionally specialized zooids are evolutionarily homologous to free living organisms, they are multicellular, possess distinct cell types, and show regional subfunctionalization [40,44,45]. While the developmental mechanisms generating zooids are very different (they arise asexually from a growth zone), the evolutionary processes acting on zooids may be similar to those acting on other modular biological units such as cell type, tissue, and organ [62]. Understanding the cellular and molecular processes underlying the patterning and molecular function of functionally specialized cnidarian zooids remain an open biological question, and in recent years there has been a focus in particular on differential gene expression patterns found in different functionally specialized zooids [46,49,50]. In this study, we found a large number of genes that are differentially expressed within different siphonophore zooids, reflecting the distinct anatomy and function of these zooids. We were also able to identify potential transcription factors that are significantly differentially expressed within particular zooids, with potential homologs found in several species, which may be interesting candidates for future study in siphonophores or related cnidarian colonial groups.

Unique zooid types and zooid delimitation

Evolution of gene expression

## 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 [1,8,9,12,16,18,63–69]. 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. 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.

## Methods

All scripts for the analyses are available in a git repository at [https://github.com/dunnlab/Siphonophore\\_Expression](https://github.com/dunnlab/Siphonophore_Expression). The most recent commit at the time of the analysis presented here was XXXX.

### 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<sub>2</sub> hexahydrate in Milli-Q water at a ratio of 1/3 MgCl<sub>2</sub> 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.

### 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 zooids/tissue of a single individual in a species were barcoded and pooled in a single sequencing lane, and replicate lanes of zooids/tissue 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. Sequences, along with sampling metadata, are publically available on NCBI under BioProject ID PRJNA540747.

### Analysis

#### *Sequencing & Differential gene expression*

Short read libraries were mapped to previously published transcriptomes (150bp paired end) [51] using Agalma v 2.0.0 [70,71], which uses a number of existing tools for transcript quantification, including RSEM (which uses Bowtie) [72,73]. 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. Differential gene expression analyses, including normalization, were conducted in R, using the `DESeq2` package [74]. 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 zooids/tissue with two or more replicates.

GO annotations were retrieved for each of the reference translated transcriptomes [51] using the PANNZER2 web server [75]. The PANNZER2 format was modified to match the gene2GO format required for the package `topGO` [76]. Gene set enrichment analyses were carried out within species using the R package `Goseq` [77],

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 the 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 [77].

#### ***Phylogenetic analysis of gene expression***

Gene trees were generated using the transcriptomes and genomes from 41 species [51] using Agalma v 2.0.0 [70,71]. Following the `treefinder` step of Agalma v 2.0.0, amino acid data were exported and supplied to Orthofinder [58] for simultaneous co-estimation of gene trees with the published maximum likelihood species tree [51]. Within Orthofinder, the selected multiple sequence alignment method was MAFFT [78] and maximum likelihood tree inference method was IQ-tree with the LG+F+R4 substitution model [79].

Phylogenetic analyses were conducted in R using `geiger`, `ape`, `phytools`, `Rphylopars`, and `hutan` [80–84]. Phylogenetic trees were visualized in R using `ggtree` and `treeio` [85]. 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 (that are indicative of branch length=0). Gene trees nodes were annotated as speciation events or duplication events, based on assignment by Orthofinder [58]. Speciation nodes were subsequently assigned a node ID equivalent to the species tree node, using species tip names from the gene tree to determine the most common recent ancestor in the species tree, using the `phytools` package. Due to the use of the species-overlap method by Orthofinder, some clades of single copy genes were assigned as speciation events, although the topology is inconsistent with the species tree. To avoid time calibration issues, due to descendant nodes being assigned the same species node ID, descendant speciation nodes with the same species node ID were marked as null, and gene trees with nodes greater than 0.3 null nodes per internal node were excluded – indicating widespread topological differences with the species tree. Tips without expression values 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 [80]. Some gene trees could not be calibrated against the node constraints from the species tree and were discarded.

For comparative expression analyses, 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 [86]:

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

Where  $\theta_i$  is the number of the mapped reads to gene  $i$ ,  $\ell_i$  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 zooid/tissue 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 `Rphyloparts` assuming a Brownian model of evolution (Fig. 3, step 2 & 3) [82]. 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. 3, step 4).

## Acknowledgements

This work was supported by the National Science Foundation (DEB-1256695 and the Waterman Award). CM was also supported in part by a RI-EPSCoR Fellowship, NSF EPS-1004057. Analyses were conducted with computational resources and services at the Center for Computation and Visualization at Brown University, supported in part by the NSF EPSCoR EPS-1004057 and the State of Rhode Island. Analyses were also conducted with computational resources and services XXXXXXXX Yale CCV

## Supplementary data

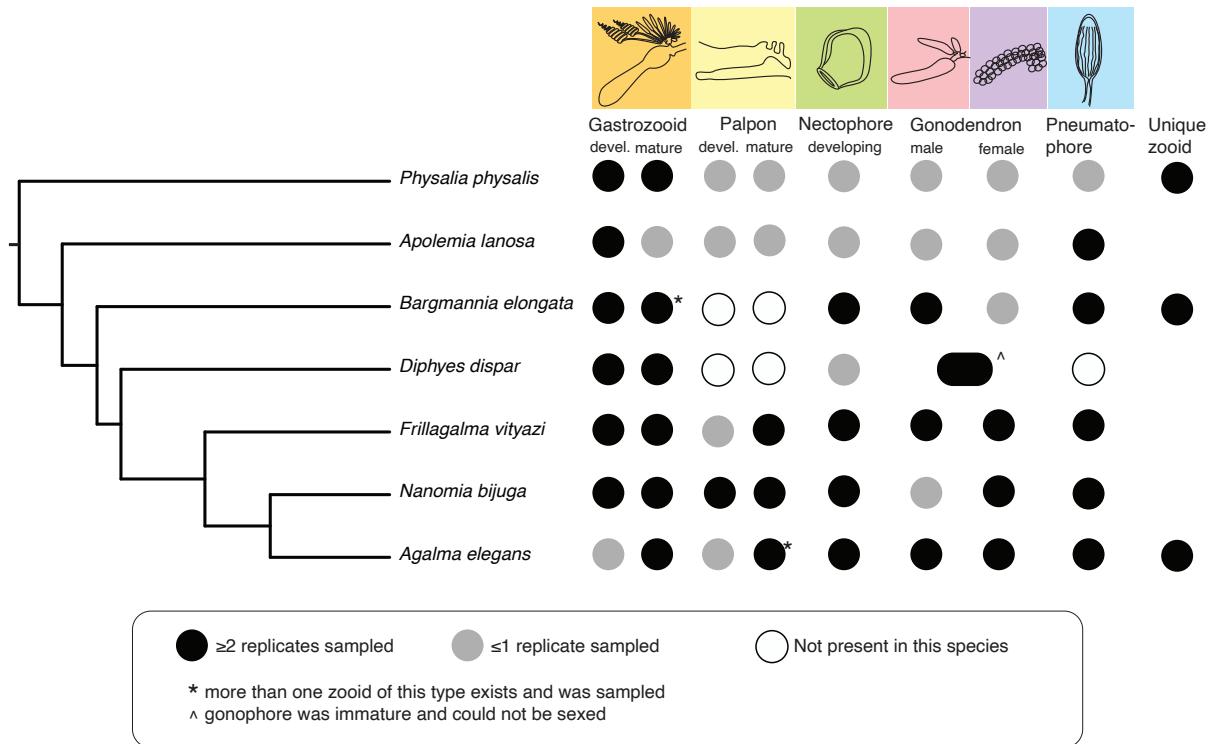


Figure S1: Phylogeny of the focal species sampled in this study, with details of the traits sampled for each of the species. Phylogeny modified from Munro et al. 2018, [51]. Black indicates that multiple replicates have been sampled, grey indicates that no or only one replicate has been sampled, and 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.

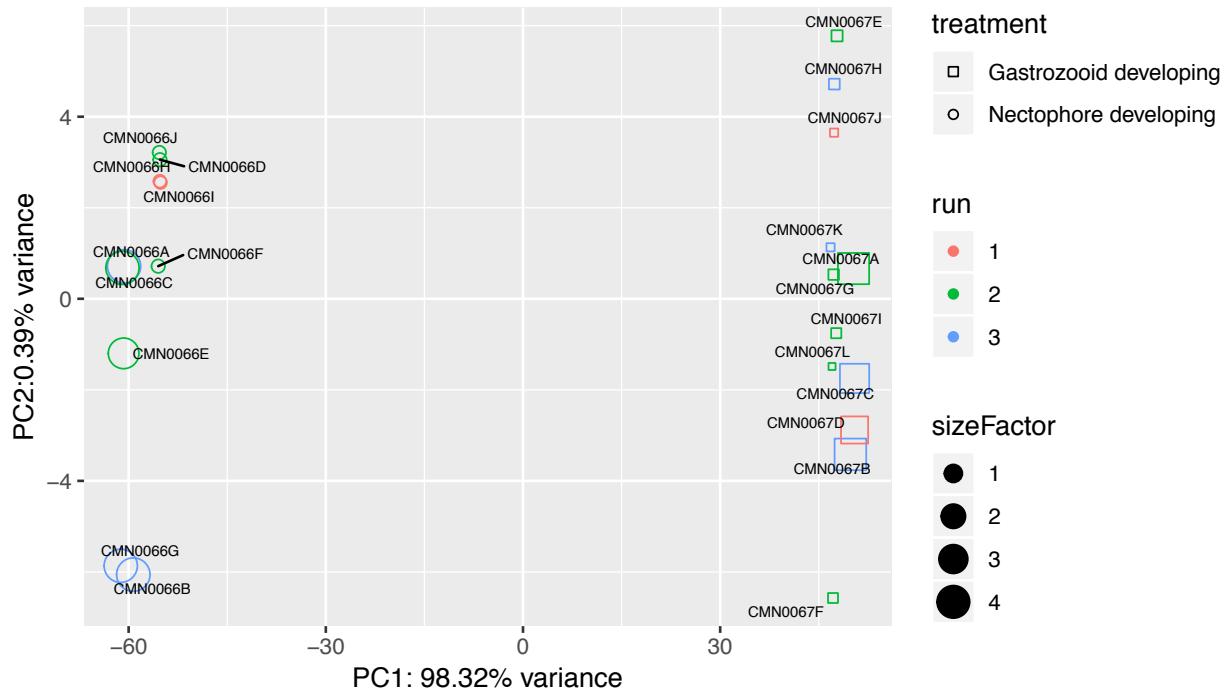


Figure S2: PCA of regularized log transformed expression counts of technical replicates of two different zooids in *Frillagalma vityazi* from different runs and lanes. Color indicates the run number, shape indicates the zooid, and size factor indicates number of genes sequenced.

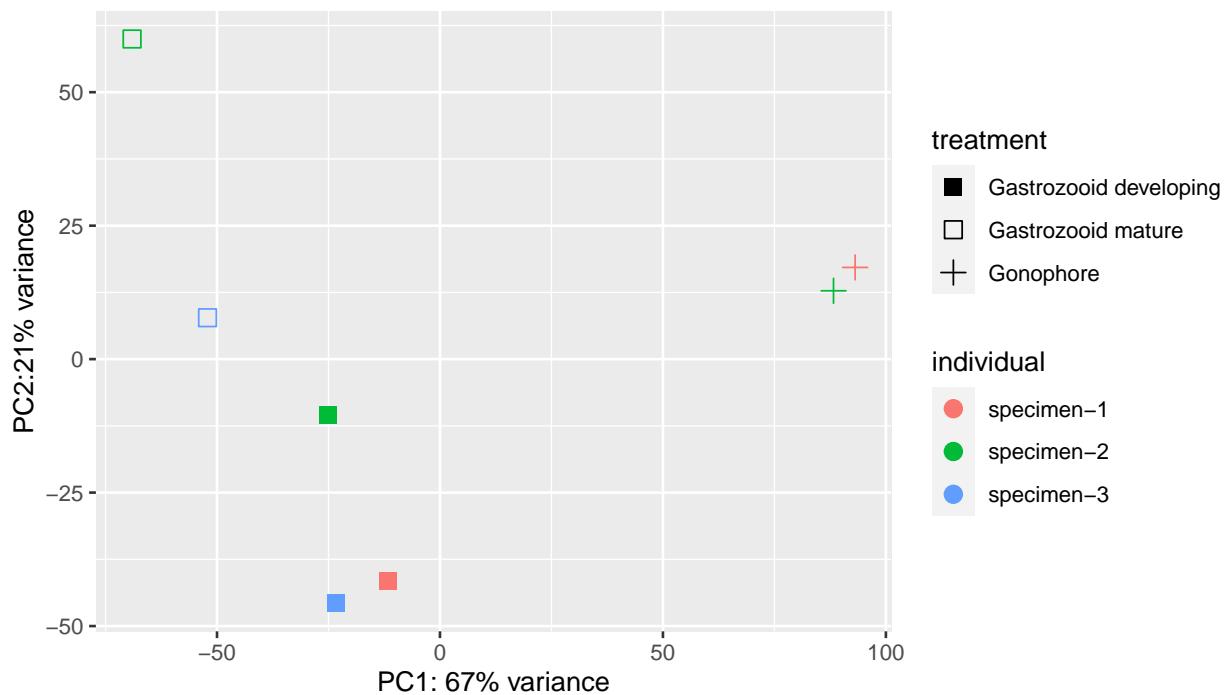


Figure S3: PCA of regularized log transformed expression counts of zooids/tissues in *Diphyes dispar*. Color indicates the replicate number, shape indicates the zooid.

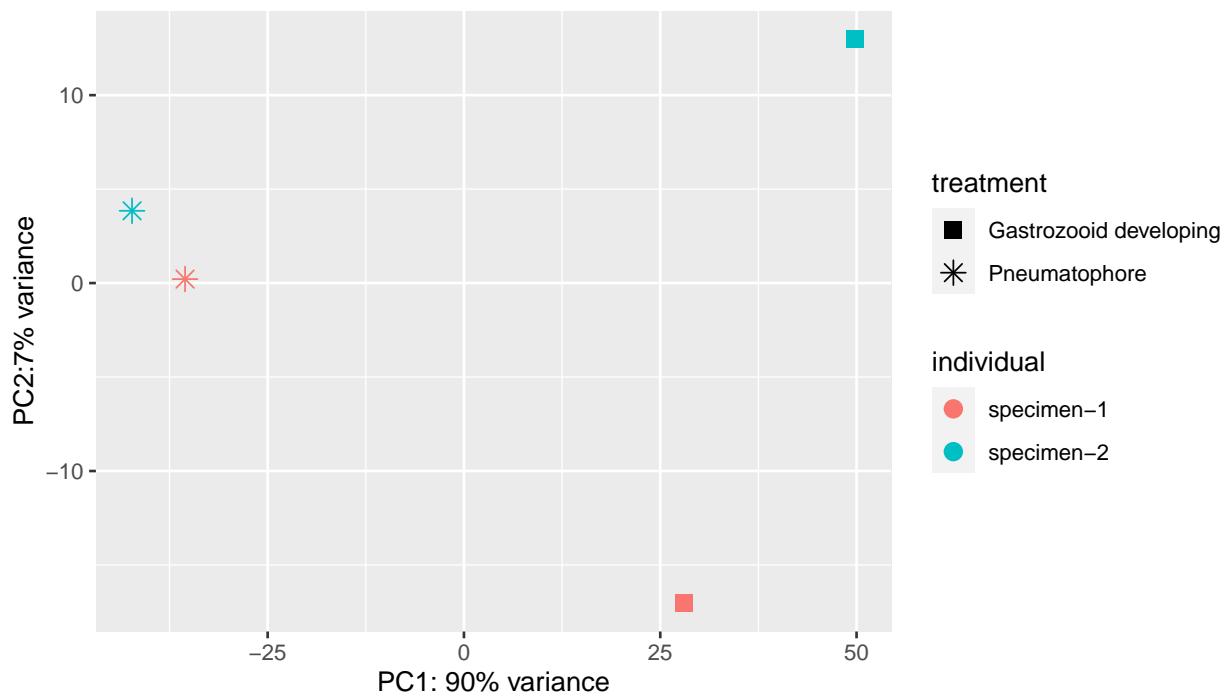


Figure S4: PCA of regularized log transformed expression counts of zooids/tissues in *Apolemia lanosa*. Color indicates the replicate number, shape indicates the zooid.

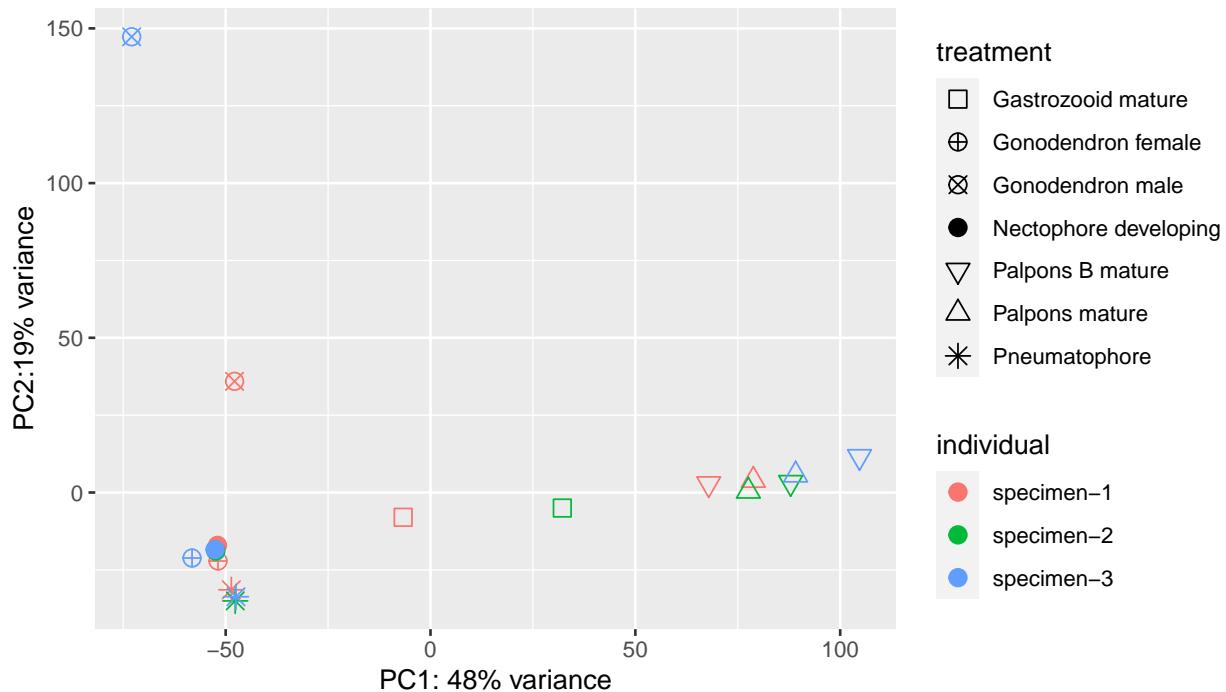


Figure S5: PCA of regularized log transformed expression counts of zooids/tissues in *Agalma elegans*. Color indicates the replicate number, shape indicates the zooid.

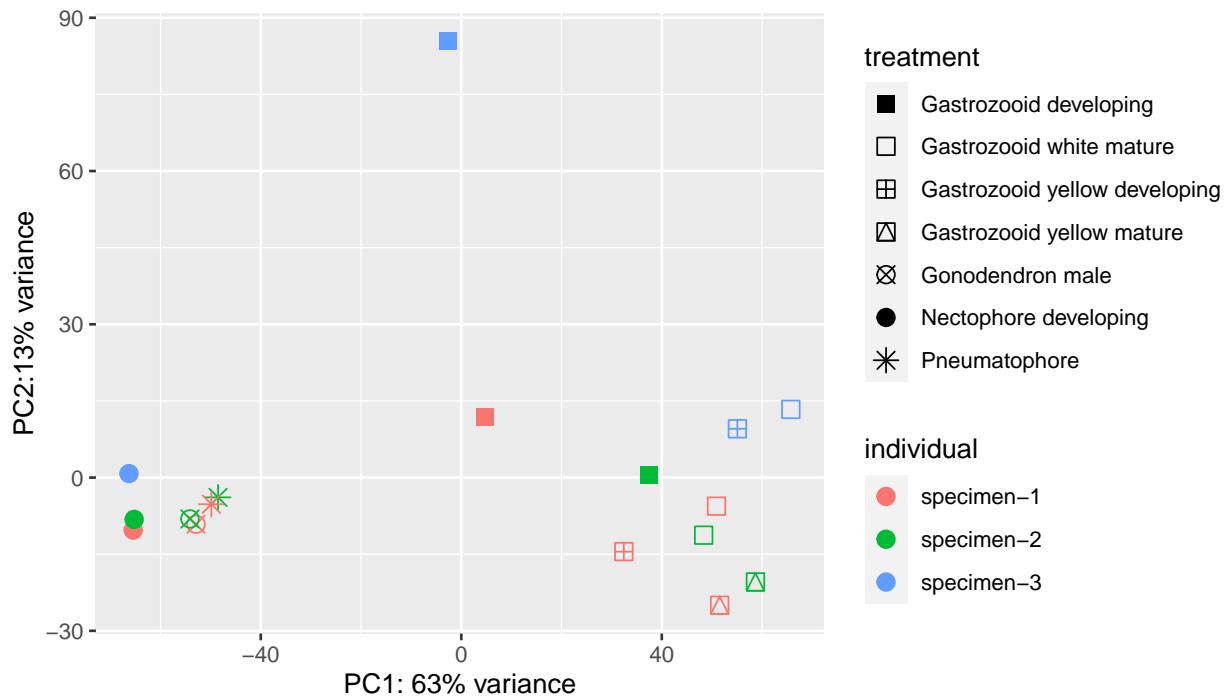


Figure S6: PCA of regularized log transformed expression counts of zooids/tissues in *Bargmannia elongata*. Color indicates the replicate number, shape indicates the zooid.

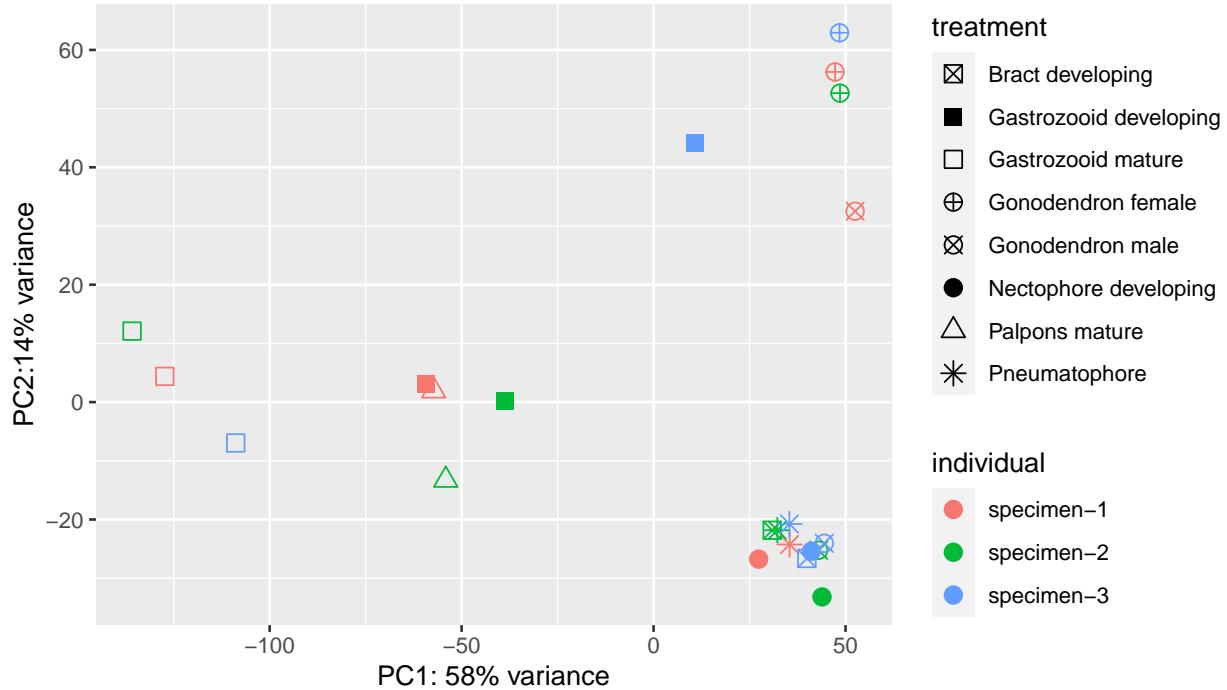


Figure S7: PCA of regularized log transformed expression counts of zooids/tissues in *Frillagalma vityazi*. Color indicates the replicate number, shape indicates the zooid.

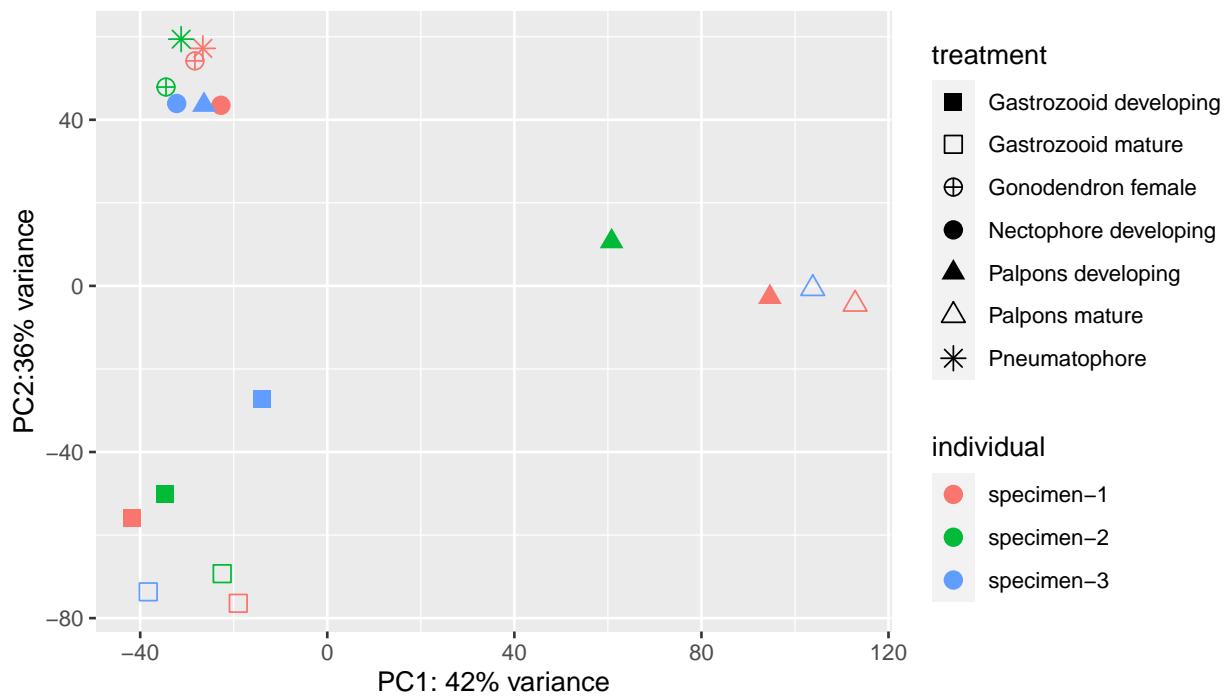


Figure S8: PCA of regularized log transformed expression counts of zooids/tissues in *Nanomia bijuga*. Color indicates the replicate number, shape indicates the zooid.

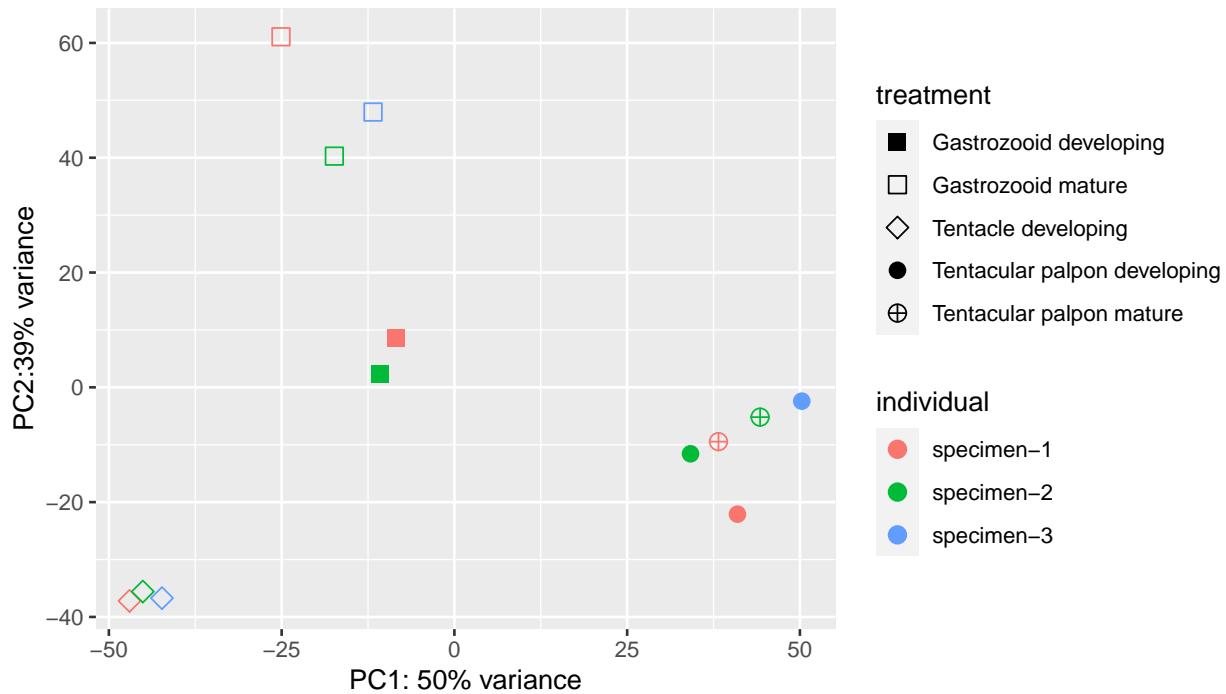


Figure S9: PCA of regularized log transformed expression counts of zooids/tissues in *Physalia physalis*. Color indicates the replicate number, shape indicates the zooid.

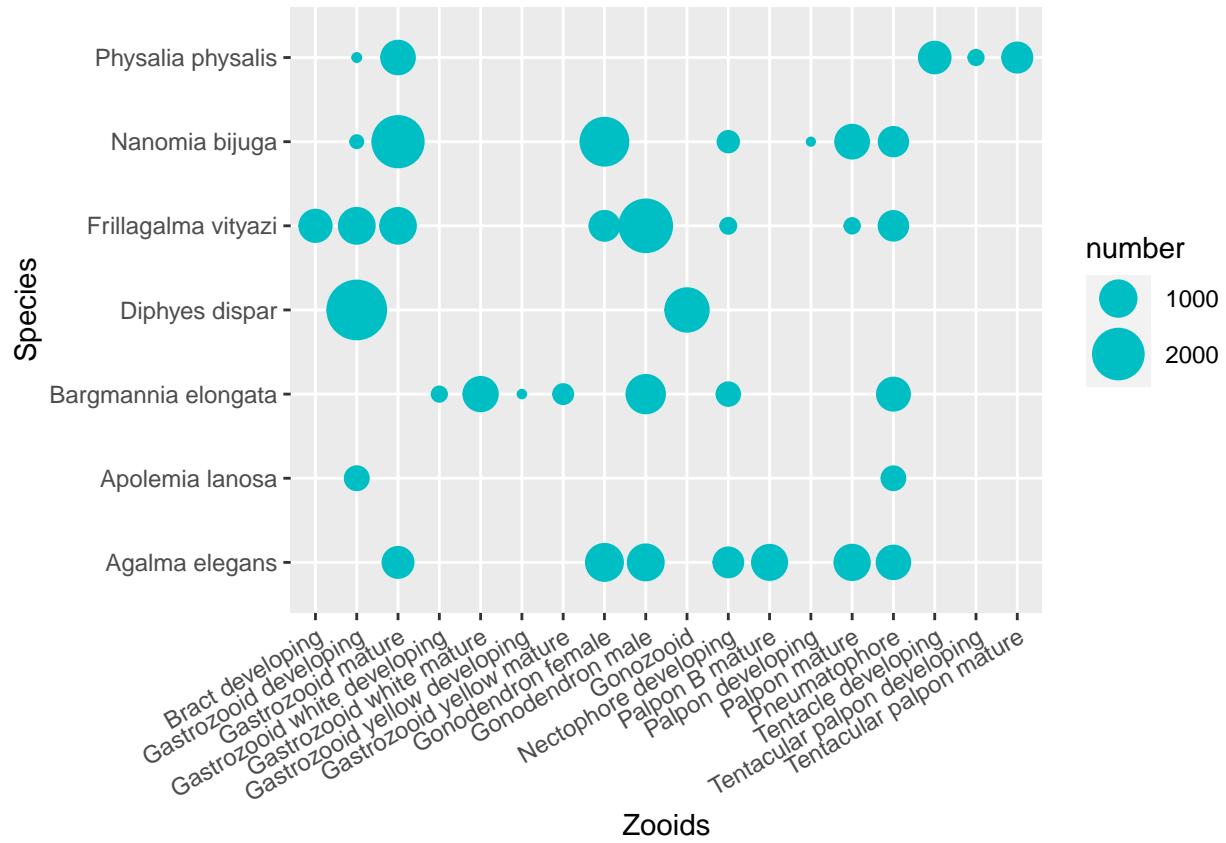


Figure S10: Number of genes identified as uniquely differentially expressed in zooids within different siphonophore species. Unique in this case means that the genes are significantly upregulated in these zooids and are not found to be upregulated in any other zooid within the same species. Note that larger numbers of genes are identified in species where fewer zooids were sampled, this is likely due to sampling differences, as opposed to biological differences among species.

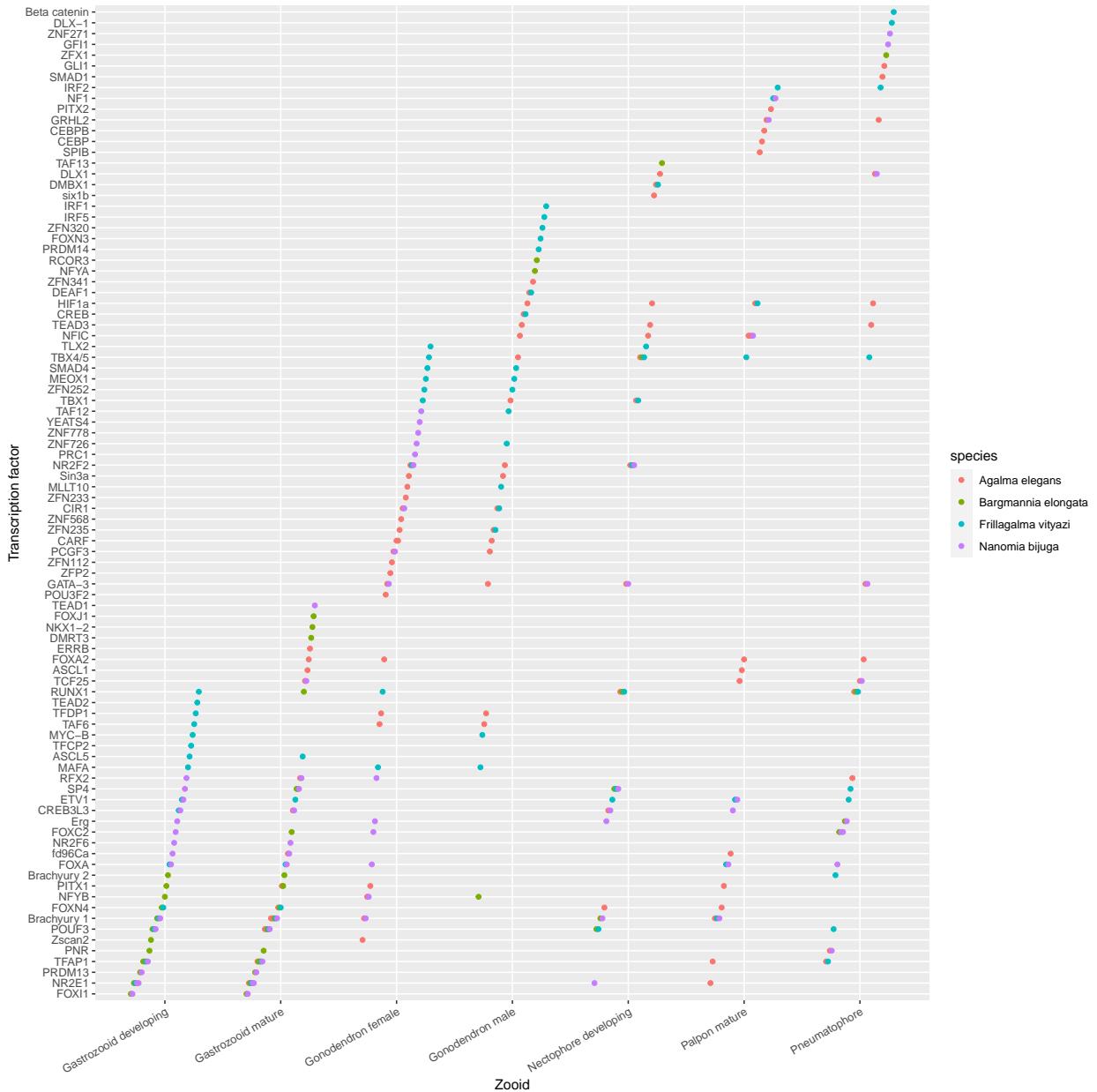


Figure S11: Putative transcription factors found to be significantly upregulated in zooids and the pneumatophore across four different siphonophore species. Transcription factor identity based on blast hit, GO DNA-binding transcription factor activity followed by manual curation. For *Bargmannia elongata* values for mature and developing gastrozooids and values for ‘white’ gastrozooids.

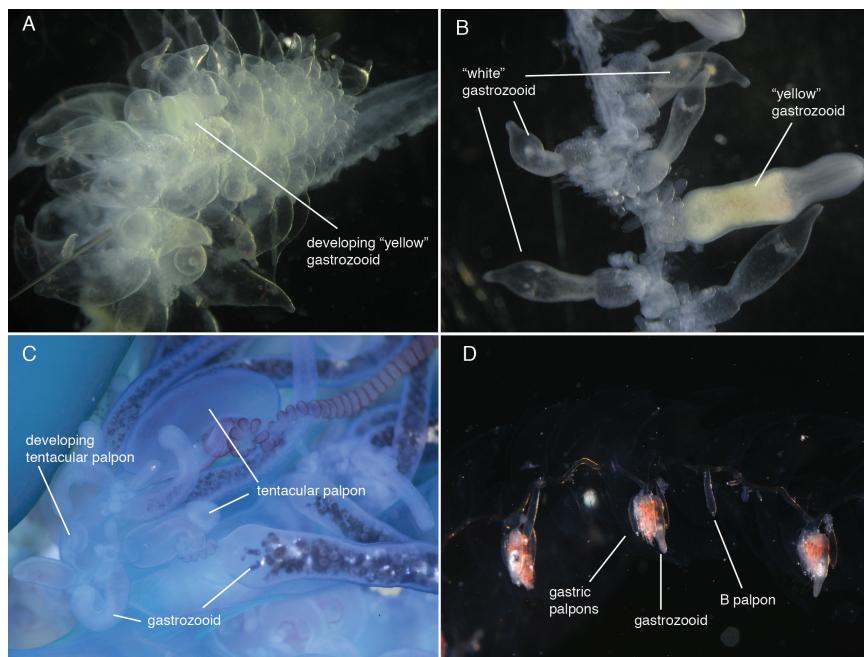


Figure S12: Unique siphonophore zooids that were sampled for differential gene expression. A. *Bargmannia elongata* with developing “yellow” gastrozooid surrounded by developing “white” gastrozooids. B. *Bargmannia elongata* stem with mature “white” and “yellow” gastrozooids. C. Zoids in *Physalia physalis*, including multiple developing stages of gastrozooids, tentacular palpon and tentacle; the most mature form of either zoid is not shown. D. Stem of *Agalma elegans*, with gastric palpons, B-palpon and gastrozooid shown.

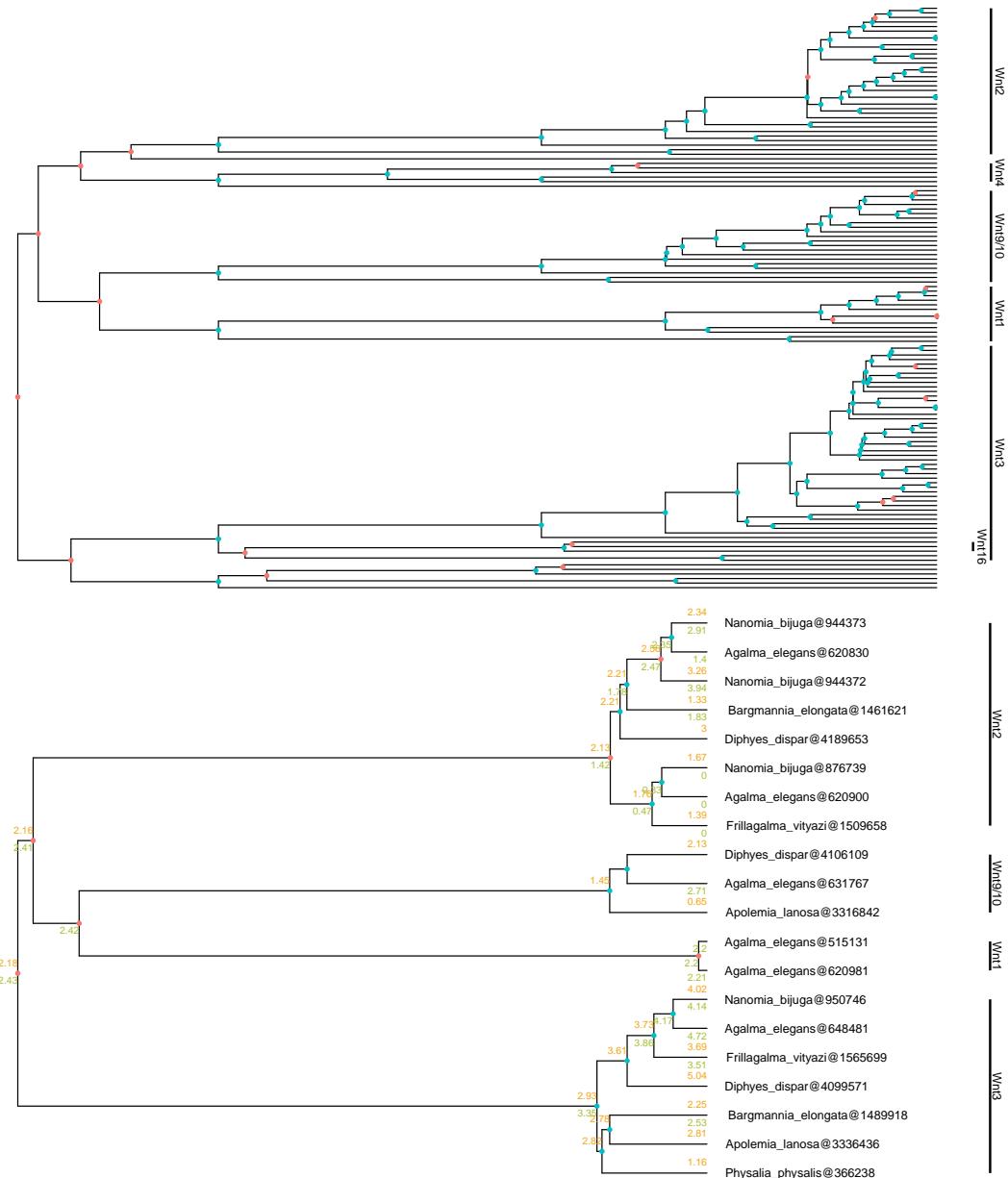


Figure S13: Top panel: Maximum likelihood gene tree of the Wnt family in Cnidaria, Wnt clades that include siphonophore sequences are highlighted. Bottom panel: Pruned gene tree of wnt genes in siphonophore species with expression values (TPM10K) at the tips and nodes. Developing gastrozooid expression in orange (above node), developing nectophore expression in green (below node). Red circles are duplication nodes and blue circles are speciation nodes.

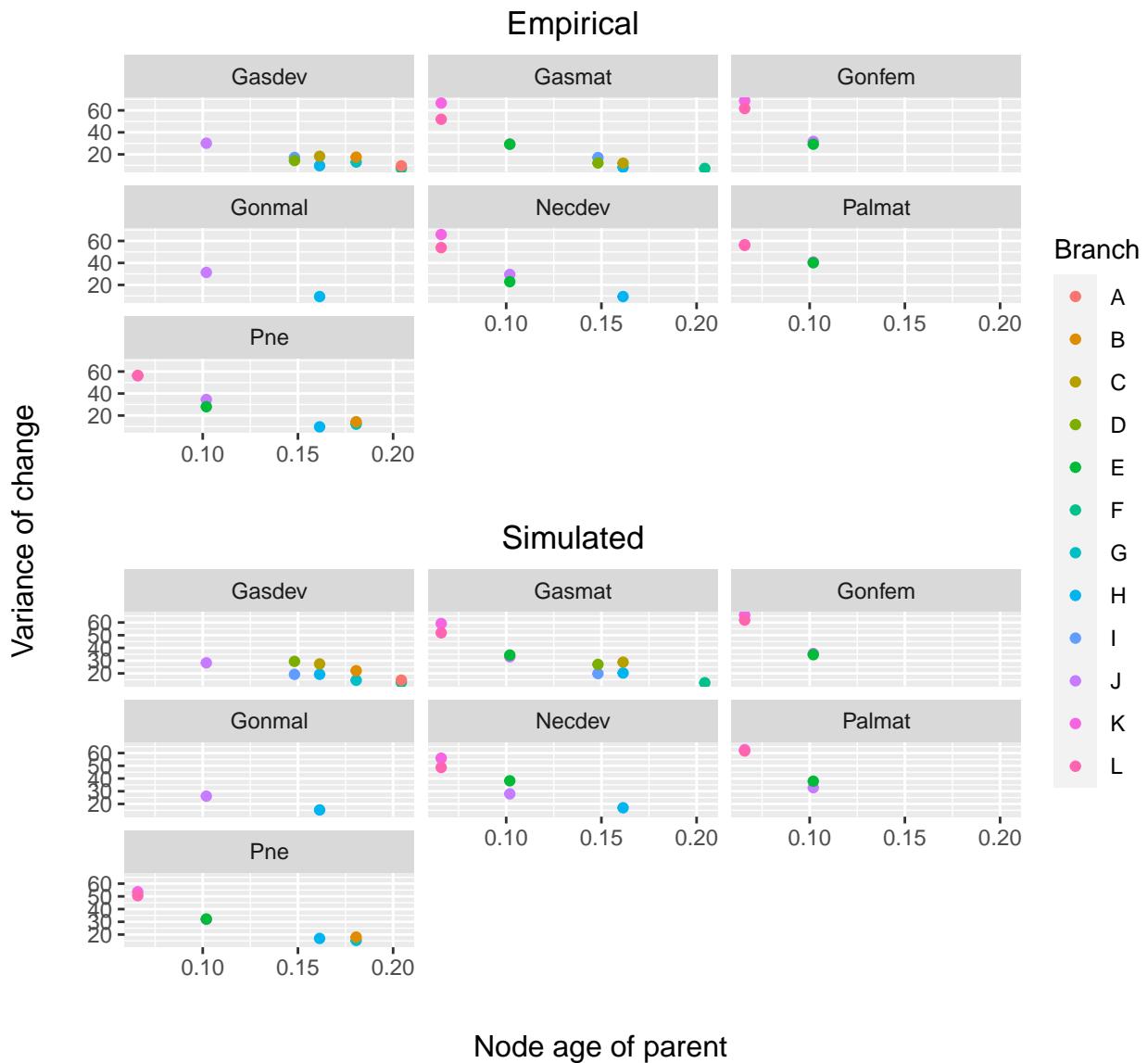


Figure S14: Variance of change across a branch plotted against the age of the parent node. Branch ID is coded in colour, values are separated by treatment. Gasdev = Developing Gastrozooid, Gasmat= Mature gastrozooid, Gonfem= Female gonodendron, Gonmal= Male gonodendron, Necdev= Developing nectophore, Palmat = Mature palpon, Pne= pneumatophore.

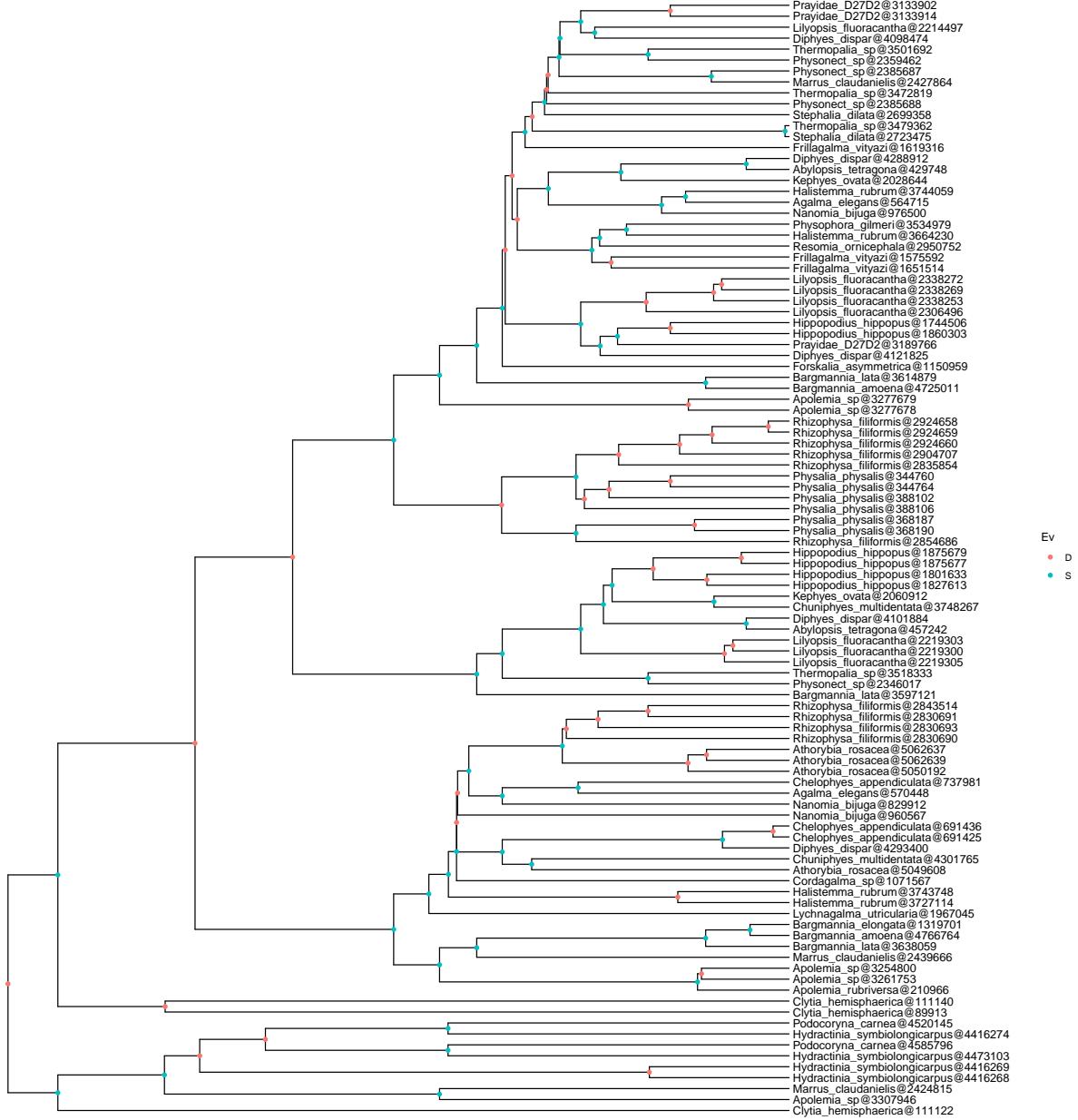


Figure S15: Maximum likelihood gene tree of the putative actinoporin-like toxin with all available species included. Red circles are duplication nodes and blue circles are speciation nodes.

## Software versions

This manuscript was computed on Sun Jul 19 18:50:34 2020 with the following R package versions.

```
R version 4.0.0 (2020-04-24)
Platform: x86_64-apple-darwin17.0 (64-bit)
Running under: macOS Catalina 10.15.4

Matrix products: default
BLAS:    /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRblas.dylib
LAPACK:  /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRlapack.dylib

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

attached base packages:
[1] grid      parallel  stats4    stats     graphics  grDevices utils
[8] datasets  methods   base

other attached packages:
[1] gtools_3.8.2           ggrepel_0.8.2
[3] kableExtra_1.1.0       corrplot_0.84
[5] ggpubr_0.3.0          lme4_1.1-23
[7] FSA_0.8.30            emmeans_1.4.6
[9] gridExtra_2.3          geiger_2.0.6.4
[11] GGally_1.5.0          PhylogeneticEM_1.4.0
[13] Matrix_1.2-18         magrittr_1.5
[15] digest_0.6.25        forcats_0.5.0
[17] stringr_1.4.0          dplyr_0.8.5
[19] purrr_0.3.4            readr_1.3.1
[21] tidyverse_1.3.0        tibble_3.0.1
[23] tidyverse_1.3.0        RColorBrewer_1.1-2
[25] picante_1.8.1          nlme_3.1-147
[27] vegan_2.5-6            lattice_0.20-41
[29] permute_0.9-5          ape_5.3
[31] fields_10.3            maps_3.3.0
[33] spam_2.5-1             dotCall164_1.0-0
[35] knitr_1.28              bookdown_0.19
[37] ggplot2_3.3.0           agalmar_0.0.0.9000
[39] hutan_0.5.1             vsn_3.56.0
[41] goseq_1.40.0            geneLenDataBase_1.24.0
[43] BiasedUrn_1.07          topGO_2.40.0
[45] SparseM_1.78            GO.db_3.11.1
[47] AnnotationDbi_1.50.0    graph_1.66.0
[49] DESeq2_1.28.1           SummarizedExperiment_1.18.1
[51] DelayedArray_0.14.0      matrixStats_0.56.0
[53] Biobase_2.48.0           GenomicRanges_1.40.0
[55] GenomeInfoDb_1.24.0      IRanges_2.22.1
[57] S4Vectors_0.26.0          BiocGenerics_0.34.0
[59] impute_1.62.0            treeio_1.12.0
[61] ggtree_2.2.1

loaded via a namespace (and not attached):
[1] tidyselect_1.1.0          RSQLite_2.2.0          combinat_0.0-8
```

```

[4] BiocParallel_1.22.0      animation_2.6          munsell_0.5.0
[7] codetools_0.2-16        preprocessCore_1.50.0 statmod_1.4.34
[10] withr_2.2.0            colorspace_1.4-1     rstudioapi_0.11
[13] ggsignif_0.6.0         labeling_0.3          GenomeInfoDbData_1.2.3
[16] mnormt_1.5-7          farver_2.0.3          bit64_0.9-7
[19] coda_0.19-3            vctrs_0.3.0          generics_0.0.2
[22] clusterGeneration_1.3.4 xfun_0.13           BiocFileCache_1.12.0
[25] R6_2.4.1                locfit_1.5-9.4       bitops_1.0-6
[28] reshape_0.8.8          assertthat_0.2.1    scales_1.1.1
[31] gtable_0.3.0            affy_1.66.0          phangorn_2.5.5
[34] rlang_0.4.6             genefilter_1.70.0   scatterplot3d_0.3-41
[37] splines_4.0.0           rtracklayer_1.48.0 rstatix_0.5.0
[40] lazyeval_0.2.2          broom_0.5.6          abind_1.4-5
[43] BiocManager_1.30.10     yaml_2.2.1           modelr_0.1.7
[46] GenomicFeatures_1.40.0  backports_1.1.7     tools_4.0.0
[49] affyio_1.58.0           ellipsis_0.3.0      Rcpp_1.0.4.6
[52] plyr_1.8.6              progress_1.2.2       zlibbioc_1.34.0
[55] RCurl_1.98-1.2          prettyunits_1.1.1   openssl_1.4.1
[58] deSolve_1.28             haven_2.2.0          cluster_2.1.0
[61] fs_1.4.1                data.table_1.12.8  openxlsx_4.1.5
[64] reprex_0.3.0             mvtnorm_1.1-0       hms_0.5.3
[67] patchwork_1.0.0          evaluate_0.14       xtable_1.8-4
[70] XML_3.99-0.3            rio_0.5.16          readxl_1.3.1
[73] compiler_4.0.0           biomaRt_2.44.0      crayon_1.3.4
[76] minqa_1.2.4             htmltools_0.4.0     mgcv_1.8-31
[79] geneplotter_1.66.0       aplot_0.0.4          expm_0.999-4
[82] lubridate_1.7.8          DBI_1.1.0           dbplyr_1.4.3
[85] subplex_1.6              MASS_7.3-51.6       rappdirs_0.3.1
[88] boot_1.3-25              car_3.0-7           cli_2.0.2
[91] quadprog_1.5-8           igraph_1.2.5         pkgconfig_2.0.3
[94] rvcheck_0.1.8            GenomicAlignments_1.24.0 numDeriv_2016.8-1.1
[97] foreign_0.8-79           xml2_1.3.2          foreach_1.5.0
[100] annotate_1.66.0          webshot_0.5.2       XVector_0.28.0
[103] estimability_1.3         rvest_0.3.5          phytools_0.7-20
[106] Biostrings_2.56.0        fastmatch_1.1-0     rmarkdown_2.1
[109] cellranger_1.1.0         tidytree_0.3.3      curl_4.3
[112] Rsamtools_2.4.0          nloptr_1.2.2.1      lifecycle_0.2.0
[115] jsonlite_1.6.1           carData_3.0-3       viridisLite_0.3.0
[118] askpass_1.1               limma_3.44.1         fansi_0.4.1
[121] pillar_1.4.4              plotrix_3.7-8       httr_1.4.1
[124] survival_3.1-12          glue_1.4.1          zip_2.0.4
[127] iterators_1.0.12         bit_1.1-15.2        stringi_1.4.6
[130] blob_1.2.1               memoise_1.1.0

```

## References

1. Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, Harrigan P, Weier M, Liechti A, Aximu-Petri A, Kircher M, Albert FW, Zeller U, Khaitovich P, Grützner F, Bergmann S, Nielsen R, Pääbo S, Kaessmann H: **The evolution of gene expression levels in mammalian organs.** *Nature* 2011, **478**:343–34810.1038/nature10532.
2. Gilad Y, Rifkin SA, Bertone P, Gerstein M, White KP: **Multi-species microarrays reveal the effect of sequence divergence on gene expression profiles.** *Genome Research* 2005, **15**:674–680.
3. King M-C, Wilson AC: **Evolution at two levels in humans and chimpanzees.** *Science* 1975, **188**:107–116.
4. Rifkin SA, Kim J, White KP: **Evolution of gene expression in the *Drosophila melanogaster* subgroup.** *Nature Genetics* 2003, **33**:138.
5. Carroll SB: **Evolution at two levels: on genes and form.** *PLoS Biology* 2005, **3**:e245.
6. Romero IG, Ruvinsky I, Gilad Y: **Comparative studies of gene expression and the evolution of gene regulation.** *Nature Reviews Genetics* 2012, **13**:505.
7. Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, Romano LA: **The evolution of transcriptional regulation in eukaryotes.** *Molecular Biology and Evolution* 2003, **20**:1377–1419.
8. Breschi A, Djebali S, Gillis J, Pervouchine DD, Dobin A, Davis CA, Gingeras TR, Guigó R: **Gene-specific patterns of expression variation across organs and species.** *Genome Biology* 2016, **17**:151.
9. Clarke TH, Garb JE, Haney RA, Chaw RC, Hayashi CY, Ayoub NA: **Evolutionary shifts in gene expression decoupled from gene duplication across functionally distinct spider silk glands.** *Scientific Reports* 2017, **7**:8393.
10. Gilad Y, Mizrahi-Man O: **A reanalysis of mouse ENCODE comparative gene expression data.** *F1000Research* 2015, **4**10.12688/f1000research.6536.1.
11. Khaitovich P, Weiss G, Lachmann M, Hellmann I, Enard W, Muetzel B, Wirkner U, Ansorge W, Pääbo S: **A neutral model of transcriptome evolution.** *PLoS Biology* 2004, **2**:e132.
12. Merkin J, Russell C, Chen P, Burge CB: **Evolutionary dynamics of gene and isoform regulation in Mammalian tissues.** *Science* 2012, **338**:1593–1599.
13. Lin S, Lin Y, Nery JR, Urich MA, Breschi A, Davis CA, Dobin A, Zaleski C, Beer MA, Chapman WC, others: **Comparison of the transcriptional landscapes between human and mouse tissues.** *Proceedings of the National Academy of Sciences of the United States of America* 2014, **111**:17224–17229.
14. Pankey MS, Minin VN, Imholte GC, Suchard MA, Oakley TH: **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* 2014, **111**:E4736–E4742.
15. Tschopp P, Sherratt E, Sanger TJ, Groner AC, Aspiras AC, Hu JK, Pourquié O, Gros J, Tabin CJ: **A relative shift in cloacal location repositions external genitalia in amniote evolution.** *Nature* 2014, **516**:391.
16. Yang R, Wang X: **Organ evolution in angiosperms driven by correlated divergences of gene sequences and expression patterns.** *The Plant Cell* 2013, **25**:71–82.
17. Gu X: **Understanding tissue expression evolution: from expression phylogeny to phylogenetic network.** *Briefings in Bioinformatics* 2015, **17**:249–254.
18. Sudmant PH, Alexis MS, Burge CB: **Meta-analysis of RNA-seq expression data across species, tissues and studies.** *Genome Biology* 2015, **16**:287.
19. Chang D, Duda TF: **Application of community phylogenetic approaches to understand gene expression: differential exploration of venom gene space in predatory marine gastropods.** *BMC Evolutionary Biology* 2014, **14**:12310.1186/1471-2148-14-123.

20. Chen J, Swofford R, Johnson J, Cummings BB, Rogel N, Lindblad-Toh K, Haerty W, Palma F di, Regev A: **A quantitative model for characterizing the evolutionary history of mammalian gene expression.** *bioRxiv* 2017, 10.1101/229096 Available: <https://www.biorxiv.org/content/early/2017/12/04/229096>.
21. Eng KH, Bravo HC, Keleş S: **A phylogenetic mixture model for the evolution of gene expression.** *Molecular Biology and Evolution* 2009, **26**:2363–237210.1093/molbev/msp149.
22. Gu X: **Statistical framework for phylogenomic analysis of gene family expression profiles.** *Genetics* 2004, **167**:531–542.
23. Oakley TH, Gu Z, Abouheif E, Patel NH, Li W-H: **Comparative methods for the analysis of gene-expression evolution: an example using yeast functional genomic data.** *Molecular Biology and Evolution* 2005, **22**:40–5010.1093/molbev/msh257.
24. Rohlf RV, Nielsen R: **Phylogenetic ANOVA: The Expression Variance and Evolution Model for Quantitative Trait Evolution.** *Systematic biology* 2015, **64**:695–70810.1093/sysbio/syv042.
25. Stern DB, Crandall KA: **The evolution of gene expression underlying vision loss in cave animals.** *Molecular Biology and Evolution* 2018,:msy106.
26. Whitney KD, Boussau B, Baack EJ, Garland T: **Drift and Genome Complexity Revisited.** *PLoS Genetics* 2011, **7**:e1002092–510.1371/journal.pgen.1002092.
27. Dunn CW, Luo X, Wu Z: **Phylogenetic analysis of gene expression.** *Integrative and Comparative Biology* 2013, **53**:847–856.
28. Felsenstein J: **Phylogenies and the Comparative Method.** *The American Naturalist* 1985, **125**:1–15.
29. Felsenstein J: **Comparative Methods with Sampling Error and Within-Species Variation: Contrasts Revisited and Revised.** *The American Naturalist* 2008, **171**:713–72510.1086/587525.
30. FitzJohn RG: **Diversitree: comparative phylogenetic analyses of diversification in R.** *Methods in Ecology and Evolution* 2012, **3**:1084–109210.1111/j.2041-210X.2012.00234.x.
31. Grafen A: **The phylogenetic regression.** *Philosophical Transactions of the Royal Society B: Biological Sciences* 1989, **326**:119–157.
32. Revell LJ, Mahler DL, Neto PP: **A new phylogenetic method for identifying exceptional phenotypic diversification.** *Evolution* 2012, **66**:135–14610.1111/j.1558-5646.2011.01435.x.
33. Pagel M: **Inferring the historical patterns of biological evolution.** *Nature* 1999, **401**:877–88410.1038/44766.
34. Uyeda JC, Harmon LJ: **A Novel Bayesian Method for Inferring and Interpreting the Dynamics of Adaptive Landscapes from Phylogenetic Comparative Data.** *Systematic Biology* 2014, **63**:902–91810.1093/sysbio/syu057.
35. Dunn CW, Zapata F, Munro C, Siebert S, Hejnol A: **Pairwise comparisons across species are problematic when analyzing functional genomic data.** *Proceedings of the National Academy of Sciences of the United States of America* 2018,:E409–E41710.1073/pnas.1707515115.
36. Dunn CW, Wagner GP: **The evolution of colony-level development in the Siphonophora (Cnidaria: Hydrozoa).** *Development Genes and Evolution* 2006, **216**:743–75410.1007/s00427-006-0101-8.
37. Mackie GO: **Siphonophores, Bud Colonies, and Superorganism.** In *The lower metazoa*. Edited by Dougherty E Berkeley: University of California Press; 1963:329–337.
38. Mackie GO: **From aggregates to integrates: physiological aspects of modularity in colonial animals.** *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences* 1986, **313**:175–196.

39. Mackie GO, Pugh PR, Purcell JE: **Siphonophore biology**. *Advances in Marine Biology* 1987, **24**:97–26210.1016/S0065-2881(08)60074-7.
40. Totton AK: *A synopsis of the Siphonophora*. British Museum (Natural History); 1965.
41. Beklemishev WN: *Principles of Comparative Anatomy of Invertebrates. Volume I. Promorphology*. Edinburgh: Oliver & Boyd; 1969.
42. Bardi J, Marques AC: **Taxonomic redescription of the Portuguese man-of-war, *Physalia physalis* (Cnidaria, Hydrozoa, Siphonophorae, Cystonectae) from Brazil**. *Iheringia, Série Zoologia* 2007, **97**:425–43310.1590/S0073-47212007000400011Available: [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0073-47212007000400011&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0073-47212007000400011&nrm=iso).
43. Carré D: **Etude histologique du développement de *Nanomia bijuga* (Chiaje, 1841), Siphonophore Physonecte, Agalmidae**. *Cahiers de Biologie Marine* 1969, **10**:325–341.
44. Church SH, Siebert S, Bhattacharyya P, Dunn CW: **The histology of *nanomia bijuga* (hydrozoa: Siphonophora)**. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* 2015, **324**:435–449.
45. Mackie GO: **Studies on *Physalia physalis* (L.). Part 2. Behavior and histology**. *Discovery Reports* 1960, **30**:371–407.
46. Siebert S, Robinson MD, Tintori SC, Goetz F, Helm RR, Smith SA, Shaner N, Haddock SHD, Dunn CW: **Differential Gene Expression in the Siphonophore *Nanomia bijuga* (Cnidaria) Assessed with Multiple Next-Generation Sequencing Workflows**. *PLoS ONE* 2011, **6**:e2295310.1371/journal.pone.0022953.g007.
47. Siebert S, Goetz FE, Church SH, Bhattacharyya P, Zapata F, Haddock SHD, Dunn CW: **Stem cells in *Nanomia bijuga* (Siphonophora), a colonial animal with localized growth zones**. *EvoDevo* 2015, **6**:22.
48. Macrander J, Brugler MR, Daly M: **A RNA-seq approach to identify putative toxins from acrorhagi in aggressive and non-aggressive *Anthopleura elegantissima* polyps**. *BMC Genomics* 2015, **16**:221.
49. Sanders SM, Shcheglovitova M, Cartwright P: **Differential gene expression between functionally specialized polyps of the colonial hydrozoan *Hydractinia symbiolongicarpus* (Phylum Cnidaria)**. *BMC Genomics* 2014, **15**:40610.1093/gbe/evv153.
50. Sanders SM, Cartwright P: **Interspecific differential expression analysis of RNA-Seq data yields insight into life cycle variation in hydractiniid hydrozoans**. *Genome Biology and Evolution* 2015, **7**:2417–243110.1186/1471-2164-15-406.
51. Munro C, Siebert S, Zapata F, Howison M, Damian-Serrano A, Church SH, Goetz FE, Pugh PR, Haddock SHD, Dunn CW: **Improved phylogenetic resolution within Siphonophora (Cnidaria) with implications for trait evolution**. *Molecular Phylogenetics and Evolution* 2018, **127**:823–83310.1016/j.ympev.2018.06.030.
52. Auer PL, Doerge R: **Statistical design and analysis of RNA-Seq data**. *Genetics* 2010, **185**:405–41610.1534/genetics.110.114983.
53. McIntyre LM, Lopiano KK, Morse AM, Amin V, Oberg AL, Young LJ, Nuzhdin SV: **RNA-seq: technical variability and sampling**. *BMC Genomics* 2011, **12**:293.
54. Dunn CW: **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* 2005, **234**:835–845.
55. Totton AK: **Studies on *Physalia physalis* (L.). Part 1. Natural history and morphology**. *Discovery Reports* 1960, **30**:301–368.

56. Munro C, Vue Z, Behringer RR, Dunn CW: **Morphology and development of the portuguese man of war, *physalia physalis***. *Scientific reports* 2019, **9**:1–12.
57. Haeckel E: **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* 1888, **28**:1–380.
58. Emms DM, Kelly S: **OrthoFinder: Phylogenetic orthology inference for comparative genomics**. *Genome biology* 2019, **20**:1–14.
59. Felsenstein J: **Maximum-likelihood estimation of evolutionary trees from continuous characters**. *American journal of human genetics* 1973, **25**:471.
60. Revell LJ, Harmon LJ: **Testing quantitative genetic hypotheses about the evolutionary rate matrix for continuous characters**. *Evolutionary Ecology Research* 2008, **10**:311–331.
61. Revell LJ, Harmon LJ, Collar DC: **Phylogenetic signal, evolutionary process, and rate**. *Systematic biology* 2008, **57**:591–601.
62. Hiebert LS, Simpson C, Tiozzo S: **Coloniality, clonality, and modularity in animals: The elephant in the room**. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* 2020,:1–14.
63. Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Guerousov S, Lee LJ, Slobodeniuc V, Kutter C, Watt S, Colak R, others: **The evolutionary landscape of alternative splicing in vertebrate species**. *Science* 2012, **338**:1587–1593.
64. Macrander J, Broe M, Daly M: **Tissue-Specific Venom Composition and Differential Gene Expression in Sea Anemones**. *Genome Biology and Evolution* 2016, **8**:2358–237510.1093/gbe/evw155Available: <http://dx.doi.org/10.1093/gbe/evw155>.
65. Necsulea A, Soumillon M, Warnefors M, Liechti A, Daish T, Zeller U, Baker JC, Grützner F, Kaessmann H: **The evolution of lncRNA repertoires and expression patterns in tetrapods**. *Nature* 2014, **505**:635.
66. Perry GH, Melsted P, Marioni JC, Wang Y, Bainer R, Pickrell JK, Michelini K, Zehr S, Yoder AD, Stephens M, others: **Comparative RNA sequencing reveals substantial genetic variation in endangered primates**. *Genome Research* 2012, **22**:602–610.
67. Levin M, Anavy L, Cole AG, Winter E, Mostov N, Khair S, Senderovich N, Kovalev E, Silver DH, Feder M, Fernandez-Valverde SL, Nakanishi N, Simmons D, Simakov O, Larsson T, Liu S-Y, Jerafi-Vider A, Yaniv K, Ryan JF, Martindale MQ, Rink JC, Arendt D, Degnan SM, Degnan BM, Hashimshony T, Yanai I: **The mid-developmental transition and the evolution of animal body plans**. *Nature* 2016, **531**:637–64110.1038/nature16994.
68. Ma S, Avanesov AS, Porter E, Lee BC, Mariotti M, Zemskaya N, Guigo R, Moskalev AA, Gladyshev VN: **Comparative transcriptomics across 14 Drosophila species reveals signatures of longevity**. *Aging Cell* 2018,:e12740.
69. Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, Storz JF, Antunes A, Greenwold MJ, Meredith RW, others: **Comparative genomics reveals insights into avian genome evolution and adaptation**. *Science* 2014, **346**:1311–1320.
70. Dunn CW, Howison M, Zapata F: **Agalma: an automated phylogenomics workflow**. *BMC Bioinformatics* 2013, **14**:33010.1186/1471-2105-14-330.
71. Guang A, Howison M, Zapata F, Lawrence CE, Dunn C: **Revising transcriptome assemblies with phylogenetic information in Agalma 1.0**. *bioRxiv* 2017, 10.1101/202416Available: <https://www.biorxiv.org/content/early/2017/10/12/202416>.
72. Langmead B, Trapnell C, Pop M, Salzberg SL: **Ultrafast and memory-efficient alignment of short DNA sequences to the human genome**. *Genome Biology* 2009, **10**:R25.

73. Li B, Dewey CN: **RSEM**: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 2011, **12**:32310.1186/1471-2105-12-323.
74. Love MI, Huber W, Anders S: Moderated estimation of fold change and dispersion for RNA-seq data with **DESeq2**. *Genome Biology* 2014, **15**:550.
75. Törönen P, Medlar A, Holm L: **PANNZER2**: a rapid functional annotation web server. *Nucleic Acids Research* 2018, **gky350**10.1093/nar/gky350.
76. Alexa A, Rahnenfuhrer J: *TopGO*: Enrichment analysis for gene ontology. 2016.
77. Young MD, Wakefield MJ, Smyth GK, Oshlack A: Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology* 2010, **11**:R14.
78. Katoh K, Standley DM: **MAFFT multiple sequence alignment software version 7**: Improvements in performance and usability. *Molecular biology and evolution* 2013, **30**:772–780.
79. Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ: **IQ-tree**: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution* 2015, **32**:268–274.
80. Paradis E, Claude J, Strimmer K: **ape**: analyses of phylogenetics and evolution in R language. *Bioinformatics* 2004, **20**:289–290.
81. Church SH, Ryan JF, Dunn CW: Automation and Evaluation of the SOWH Test with **SOWHAT**. *Systematic Biology* 2015, **64**:1048–105810.1093/sysbio/syv055Available: [10.1093/sysbio/syv055](https://doi.org/10.1093/sysbio/syv055).
82. Goolsby EW, Bruggeman J, Ané C: **Rphylopars**: fast multivariate phylogenetic comparative methods for missing data and within-species variation. *Methods in Ecology and Evolution* 2017, **8**:22–27.
83. Harmon LJ, Weir JT, Brock CD, Glor RE, Challenger W: **GEIGER**: investigating evolutionary radiations. *Bioinformatics* 2007, **24**:129–131.
84. Revell LJ: **phytools**: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 2012, **3**:217–22310.1111/j.2041-210X.2011.00169.xAvailable: <http://dx.doi.org/10.1111/j.2041-210X.2011.00169.x>.
85. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y: **ggtree**: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* 2017, **8**:28–36.
86. Li B, Ruotti V, Stewart RM, Thomson JA, Dewey CN: **RNA-Seq gene expression estimation with read mapping uncertainty**. *Bioinformatics* 2009, **26**:493–500.